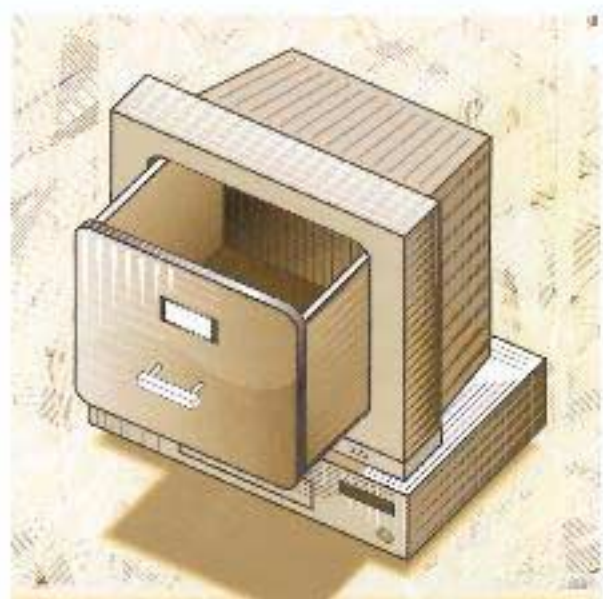





# Sluttrapport

1992-1997



 dokumentasjons  
prosjektet



# Innhold

<i>Om rapporten</i>	3
<i>Dokumentasjonsprosjektet - en oppsummering</i>	5
Bakgrunn	5
Mål og resultater - en kort vurdering	6
Prosjektledelse	9
Organisering og gjennomføring av registreringsarbeidet	10
Dokumentasjonsprosjektets etterorganisasjon	11
Nasjonal styringsgruppe og sentral prosjektstab	14
<i>Prosjektøkonomien</i>	15
Finansiering	15
Oversikt utgifter og bevilgninger 1992 - 1997	15
<i>IT-strategi</i>	17
Den grunnleggende IT-strategien i Dokumentasjonsprosjektet	17
Arbeidet med de ulike materialtypene	18
Det språkorienterte arbeidet	19
Museumsarbeidet	27
Hva kjennetegner Dokumentasjonsprosjektets databaser?	32
Når og hvor?	35
Oppsummering	36
<i>Konverteringsorganisasjonen</i>	39
Fagenhetens rolle	39
Prosjektets fagfolk	40
Tempo og kvalitet	41
Koordinering	42
<i>Innskrivingsmiljøene</i>	45
Registreringsgruppene	45
Registreringssentralene	46
Oversikt	47
<i>Arkivene</i>	51
Arkiver og materiale som er konvertert i Dokumentasjonsprosjektet	51
Arkeologi	53
Runologi	72
Numismatikk	74
Kulturhistorie	76
Historie	83
Leksikografi	89
Navnegransking	100
Folkemusikk og - dans	109
Etnografi	114
<i>Prosjektdirektiv</i>	115
Bakgrunn	115
Målstruktur	117
Oppgavebeskrivelse	118
Retningslinjer for prosjektarbeidet	119
Omfang og tidsrammer	122
Økonomiske rammer og finansiering	122
Planlegging/rapportering/oppfølging	123
Kvalitetssikring	123
Endringshåndtering	123

**Redaksjon:**  
Christian Emil Ore og  
Nina Kristiansen

Form: Torunn Jansen

Grafisk design omslag:  
Terje Syversen

Trykk: Universitetets  
reprosentral

400x/Oslo 9/98

## Om rapporten

Kapittelet *Oppsummering* går gjennom bakgrunnen for at Dokumentasjonsprosjektet ble opprettet, gir en gjennomgang av prosjektets målstruktur, måloppnåelse og oppgavebeskrivelse. Det gis i tillegg en oversikt over hvem som har vært med i prosjektets nasjonale styringsgruppe og prosjektets sentrale stab. Det skisseres også hvordan etterorganisasjonen vil arbeide videre med materialet som er konvertert i prosjektperioden.

Kapittelet *Prosjektøkonomien* viser regnskapet for Dokumentasjonsprosjektet i 1997 og en oversikt over utgifter og bevilgninger for årene 1992 - 1997.

Kapittelet *IT-strategi* har en gjennomgang av den grunnleggende strategien som er valgt i prosjektet, hvordan det er arbeidet med de ulike materialtypene, og hvordan databasene gir nye brukermuligheter. For en grundig gjennomgang av de nye brukermulighetene for arkivene og samlingene som har vært med i Dokumentasjonsprosjektet, viser vi også til boka *Fra skuff til skjerm*, Aukrust og Hodne (red), Universitets-forlaget 1998.

Under *Konverteringsorganisasjonen* er prosjektorganisasjonen beskrevet, hvem de ulike aktørene i prosjektet er, og hva som har vært deres arbeidsoppgaver.

Kapittelet *Innskrivingsmiljøene* går gjennom hvordan konverteringen har blitt gjennomført ved bruk av ulike sysselsettingstiltak. Denne rapporten fokuserer på prosjektets primæraktivitet - konverteringen av samlingene - og vil derfor ikke rapportere i detalj virksomhetene i registreringsgruppene og -sentralene. Her viser vi til års- og sluttrapportene for de ulike registreringsenhetene og til de ulike evalueringene som er gjort i prosjektperioden.

Under *Arkiver* er det foretatt en gjennomgang av alle arkivene, samlingene og verkene som er konvertert i Dokumentasjonsprosjektet. Dokumentasjonsprosjektet har vært et pionerprosjekt. Vi har ikke kunnet støtte oss på andres erfaringer, fordi det ikke tidligere har vært foretatt konvertering på mange av fagområdene. Vi har derfor valgt å gå grundig igjennom materialet, slik at andre som står overfor lignende oppgaver, skal kunne finne matnyttige tips og forslag. Dette kapitlet er basert på innleverte sluttrapporter fra de delprosjektansvarlige og fagkonsulentene.

Nina Kristiansen og Christian-Emil Ore har skrevet og redigert denne sluttrapporten.

Følgende personer har bidratt til sluttrapporten, gjennom årsrapporter, artikler i boka *Fra skuff til skjerm*, Aukrust og Hodne (red), Universitetsforlaget 1998 og andre bidrag.

Jennifer Akselsen	Bodil Kjenstad
Egil Bakka	James Knirk
Alfhild Birkeland	Knut Lunde
Arnold Dalen	Mona Mortensen
Øyvind Eide	Kari Margrete Okstad
Bjørn Eithun	Oddrun Rangsæter
Mette Gismerøy Ekker	Ålov Runde
Asbjørn Engevik jr	Tom Schmidt
Anne Engø	Åge Skjelborg
Bodil Erichsen	Kolbjørn Skaare
Velle Espeland	Sølvi Sogner
Siv-Bente Grongstad	Bergljot Solberg
Geir Grønnesby	Liv Stenersen
Svein Gullbekk	Dikka Storm
Dag Gundersen	Tor Sveum
Sveinulf Hegstad	Karl Johan Sæth
Bottolv Helleland	Eira Söderholm
Bente Hjellum	Siri Sørland
Bjarne Hodne	Toini Thommessen
Jon Holmen	Lars Jørgen Tvedt
Morten Holtet	Espen Uleberg
Tora Hultgren	Mette Vaag
Gitte Høy-Petersen	Lars Vikør
Sonja Innselset	Solveig Wikstrøm
Tor-Erk Jenstad	Einar Østmo
Birgit Hertzberg Johnsen	Lena Aarko

# Dokumentasjonsprosjektet - en oppsummering

Denne oppsummeringen inneholder ikke en gjennomgang av enkeltresultatene innefor hvert fagområde. En slik gjennomgang finnes senere i rapporten. I dette kapitlet prøver vi å sette resultatene inn i en større sammenheng. Vi gir en kort gjennomgang av bakgrunnen for prosjektet og prosjektets målstruktur og generelle resultater. I gjennomgangen av mål og resultater er det lagt vekt på å redegjøre for de nye problemstillinger gjennomføringen av prosjektet reiser for de deltakende avdelingene og institusjonene. Videre er det kort gjort rede for prosjektstyring og finansiering. Kapitlet avsluttes med en beskrivelse av etterorganisasjonen som er etablert for å sikre videre drift og utvikling av databasene.

## 1 Bakgrunn

Oppbyggingen av arkiver og registre ved de ulike samlingsavdelingene og museene på de fire universitetene har foregått over lang tid, for enkelte av museenes vedkommende i rundt 200 år. I årenes løp har det blitt samlet til dels meget omfattende manuelle registre og arkiver. Først rundt 1980 begynte enkelte fagavdelinger å ta i bruk datamaskiner i arkivarbeidet. Den daværende teknologiske virkelighet gjorde at man forestilte seg felles, sentraliserte databaser som kunne betjene flere fagmiljøer på ulike steder i landet. Disse ideene ble av forskjellige årsaker ikke realisert. Dels kunne dette skyldes datidens lite fleksible og svært kostbare tekniske løsninger, dels den totale mangel på tilgjengelig (edb-) personell med tilstrekkelig kompetanse.

Innføringen av små datamaskiner (PCer) fra andre halvpart av 1980-årene betydde både en allminneliggjøring av datateknologien, og oppmerksomheten omkring edb økte. Det ble derfor vanligere at en ved samlingsavdelingene prøvde å bygge opp edbbaserte registre. Utviklingsarbeidet ble på denne tiden gjerne gjort av entusiastiske enkeltpersoner og krevde ofte en uforholdsmessig stor innsats fra disse. Det finnes flere eksempler på at denne enkeltinnsatsen har resultert i varige databaser av høy kvalitet. Men det finnes dessverre også svært mange eksempler på at man etter en tid har gitt opp å bruke dem, at de aldri har blitt vedlikeholdt eller at de har forsvunnet etter at opphavspersonen har sluttet eller skiftet stilling. Om introduksjon av datateknologi skal være vellykket, må de nye systemene integreres i organisasjonen. Denne prosessen krever ofte endret informasjonsflyt og endrede arbeidsoppgaver for enkelte ansatte.

De første planene for Dokumentasjonsprosjektet ble lansert i 1989, og tanken var å gjøre et krafttak for å gi samlingsavdelingene anledning til å ta i bruk datastøttede metoder. De store papirarkivene hadde på sett og vis blitt sin egen fiende idet enhver form for systematisk reorganisering av materialet var blitt så kostbart at det i seg selv krevde ekstra bevilgninger. Dette gjorde og gjør de manuelle arkivene svært lite anvendelige til forskning.

I planleggingsfasen i 1989-90 så man også en økende tendens til at samfunnet utenfor institusjonene ville vite hva som befant seg i samlingsavdelingene og museene og ønsket tilgang til denne informasjonen. Et ønske om en slik demokratisering av tilgangen til kulturelt kildemateriale henger sammen med den internasjonale trenden om rett til tilgang til informasjon om egen kultur. Et slikt krav ble ytterligere forsterket gjennom forventningene til datateknologien som informasjonsformidler. Internasjonalt har blant andre UNESCO grepet fatt i mulighetene av å bruke datateknologi for å fremme denne tilgangen. I Norge har saken blitt satt i fokus på 90-tallet gjennom blant annet Velure-utvalgets plan om revitalisering av norske museer (REVITA) og gjennom planene om et Kulturnett-Norge.

Det historisk-filosfiske fakultet ved Universitetet i Oslo gjennomføre i 1990 et forprosjekt for å få oversikt over omfang og behov og mulige finansieringsmåter for et større dokumentasjonsprosjekt for å overføre arkiver og registre til databaser.

## **2 Mål og resultater - en kort vurdering**

Dokumentasjonsprosjektets hovedmål ble i 1991 formulert som "å gi en samlet og rasjonell tilgang til informasjon om språk og kultur". Tanken var med utgangspunkt i kulturhistoriske og språklige samlinger ved norske universiteter å gjøre kulturrelatert informasjon tilgjengelig for forskere, forvaltning og allmennhet i en form som ville gjøre det mulig å utnytte informasjon fra ulike registre samtidig. Man ville legge vekt på en samlet fremgangsmåte innen de ulike fagfeltene og institusjonene slik at man unngikk å ende opp med et sett av ulike tekniske og faglige løsninger: Det ble også sett som viktig å velge en mest mulig leverandøruavhengig måte å lagre dataene på, og dataløsningene skulle integreres i fagavdelingenes daglige virke. På bakgrunn av hovedmålet ble det formulert fire delmål:

- 1 Dataregistrering av arkiver og samlinger
- 2 Rasjonelle rutiner for samlinger og arkiver
- 3 Optimal tilgjengelighet for forskere, forvaltning og allmennhet
- 4 Samordnet datarepresentasjon og grensesnitt

### **2.1 Delmål 1: Dataregistrering av arkiver og samlinger**

Dokumentasjonsprosjektet var som nevnt ment som et krafttak for å bringe samlingsavdelingenes registre og arkiver over til datastøttede medier. Delmål 1 har derfor hatt aller høyeste prioritet. En tilnærmet komplett overføring til data av materiale ved de humanistiske samlingsavdelingene ved universitetene (biblioteker unntatt) ville kreve omlag 2.000 årsverk. Ut fra praktiske hensyn ble det satt et øvre tak på om lag 600 årsverk sysselsatte og 120 årsverk faglærte. På grunnlag av utredninger foretatt ved de fire universitetene i 1990 - 93, avgjorde prosjektets styringsgruppe hvilke registre som skulle inngå i prosjektet samt tidspunkt for deltakelse. Den endelige ressursfordelingen ble foretatt tidlig i 1994 (listen på side 10 viser hvilke arkiv og registre som har blitt konvertert). Kriteriene for å inkludere registre og arkiv var:

- Registret eller arkivet har materiale av nasjonal interesse
- Det er dokumentert behov for materialet av flere brukere
- Det er materiale som støtter opp om og utfyller øvrig materiale i prosjektet
- Materialet er unikt
- Det er innsatsvilje i miljøene, uttrykt ved prioriteringer innenfor eksisterende rammer og forarbeid utført ved egne midler
- Det er gjennomførbart å inkludere registret i prosjektet

Utvalget av arkiver ble delvis bestemt av det siste punktet. I 1990-93 var det mulig, men for kostbart, å lage digitale foto-, film- og lydsamlinger. Det ble derfor satset på konvertering av tekstlig materiale som krevde liten lagringsplass og bygde på velutviklede metoder. Selve konverteringsarbeidet av tekst er arbeidsintensivt, men kan utføres av ufaglærte. I dag er situasjonen svært annerledes i det lagringsplass ikke



lenger er noe problem og siden digitalisering av foto, produksjon av elektroniske faksimiler av arkivkort og boksider er billigere enn optisk lesing eller avskrift av tekst med tilhørende koding. Dersom Dokumentasjonsprosjektet hadde vært planlagt i dag, ville dermed økonomiske forhold muligens ha gjort prosjektet mer orientert mot oppbygging av bilde og faksimiledatabaser enn mot elektronisk tekst. Fra et faglig synspunkt ville en slik dreining vært svært uheldig, selv om det presentasjonsteknisk sett kunne være gunstig. For mange av fagfeltene er teksten det viktigste studieobjektet. Faksimiler ville for disse fagene være like "dødt" som det originale materialet.

Som man kunne forvente, gikk noe av konverteringsarbeidet raskere, mens annet gikk langsommere enn antatt. For museene er det sett under ett konvertert adskillig mer enn det vi i utgangspunktet hadde planlagt. Vi valgte likevel å holde oss strengt til den opprinnelige ressursfordelingen av faglærte årsverk og heller foreta noe justering av materialmengden, konverteringsmetoden eller etterbehandlingsnivået. Dokumentasjonsprosjektet har som oversiktene senere i denne rapporten viser, oppfylt de konverteringsmål som ble satt ved prosjektets begynnelse.

## ***2.2 Delmål 2: Rasjonelle rutiner for samlinger og arkiver***

Den store datakonverteringen betyr at en rekke samlingsavdelinger nå har hoveddelen av sine registre på elektronisk form. Dersom konverteringen skal ha noen varig verdi, må avdelingene justere informasjonsflyten og arbeidsinnholdet i enkelte stillinger. Det må sikres at ny informasjon kommer på en form som gjøre det mulig å importere den inn i de nye systemene. Begge deler forutsetter IT-kompetanse hos de arkivansvarlige ved samlingsavdelingene.

Dokumentasjonsprosjektet har verken hatt som oppgave eller hatt myndighet til å gjennomføre noen form for aktiv omstrukturering av samlingsavdelingene. Prosjektets etterorganisasjon vil virke som konsulenter i en slik prosess, men det må understrekes at ansvaret for en slik omstrukturering ligger hos de enkelte institusjonene og grunnenhetene. Noen enheter har foretatt stillingsendringer slik at man har fått ansatte med databasevedlikehold og -administrasjon som hovedoppgave. I andre enheter vil gamle og nye arbeidsmetoder og rutiner eksistere side om side i en tid fremover.

I en god del tilfeller har deler av arkiver blitt kopiert fra en institusjon til en annen. Fylkeskommunene på Østlandet har for eksempel kopiert "sine" deler av Topografisk arkiv ved Oldsaksamlingen. Det er heller ikke uvanlig at fagavdelinger har bygd opp bearbejdede registre på grunnlag av materiale som eies av andre. Både gammelnorskregisteret og navneregistrene er basert på materiale fra Riksarkivet og andre arkivinstitusjoner. Folkeminnemateriale er basert på manuskripter fra en lang rekke museer og arkiver. Da arkivene og registrene var manuelle, ble faglig vedlikehold utført av den avdelingen der de befant seg, uansett om det var eget materiale eller materiale kopiert fra en annen institusjon. Introduksjon av elektroniske arkiver bør bidra til en nasjonal rasjonalisering med hensyn til vedlikehold.

Et viktig spørsmål er hvem som har ansvaret for oppdatering, og hvem som skal dele kostnadene. Det er åpenbart at dette bare kan sikres gjennom et forpliktende samarbeid mellom de aktuelle institusjonene. Innad i universitetssektoren vil prosjektets etterorganisasjon ha som en av sine oppgaver å være koordinerende instans for dette. Eksternt må det opprettes avtaler med de ulike aktørene. Ved Dokumentasjonsprosjektets avslutning foregikk det forhandlinger mellom prosjektet (representert gjennom Universitet i Oslo) og Riksantikvaren om et forpliktende samarbeid om fornminnedata. Et tilsvarende samarbeid mellom alle eller enkelte av forvaltnings-

museene og fylkeskommunene om de topografiske arkivene er også tenkelig. Men man støter fort på en rekke andre uavklarte og til dels problematiske forhold mellom ulike deler av forvaltningen og universitetsmuseene. Igjen ser man at ved at informasjonsteknologien gjør det enkelt og raskt både å flytte og å tilgjengeliggjøre informasjon og å synliggjøre eldre og til dels kunstige informasjons-barrierer i samfunnet.

### **2.3 Delmål 3: Optimal tilgjengelighet for forskere, forvaltning og allmennhet**

Dokumentasjonsprosjektet valgte i 1991 å lagre og kode dataene i et smidig og veldefinert format (SGML) som gjør det enkelt å benytte dataene i ulike programprodukter. Dette gjorde det lite kostbart å publisere datamaterialet fra prosjektet på Internett. Størstedelen av materialet fra prosjektet er enten tilgjengelig på nettet eller vil bli det i løpet av 1998. Prosjektet har lagt vekt på å få materialet tilgjengelig og redusert noe på presentasjonsform og sammenstillingsmuligheter. Men sidene er like fullt godt besøkt, omlag 3.000 databasesøk og 5.000 nedlastinger av filer i døgnet pr. 20. august 1998. Sidene brukes også mye av folk utenfor Norge.

Formidlingen gjennom WWW (World Wide Web) er mest rettet mot allmennheten. Det finnes i tillegg nettbaserte databasegrensesnitt som er utelukkende til bruk for forskere og ansatte i samlingsavdelingene. Disse gir også tilgang til interne data som for eksempel konserveringstilstand og hylleplassering for gjenstandsmaterialet og i museene. Det en finner åpent under prosjektets hjemmesider vil derfor ikke være alt materialet som har blitt tilrettelagt av Dokumentasjonsprosjektet.

Hva som er optimal tilgang til materialet, vil endres over tid. Presentasjon av data som en elektronisk kopi av den trykte kilden er allerede noe foreldet. Introduksjonen av informasjonsteknologi i de tradisjonelle samlingsfagene har endret og vil forsette å endre fagfolkernes måte å tenke på sitt grunnlagsmateriale. Det vil derfor kreves en kontinuerlig innsats ikke bare av formidlingstekniske kyndige personer, men også fra fagfolkene selv. Gjennom Dokumentasjonsprosjektet og dets etterorganisasjon blir det laget et informasjonssystem for dataene. Dette er imidlertid ikke en vitenskapelig publikasjon av dataene, men et grunnlag for å kunne bruke dataene grunnlag for vitenskapelig virksomhet. Et optimalt utbytte av materialet vil en ikke kunne få uten en grunnleggende forståelse for IT-baserte metoder ens eget fag samt en viss sans for det som på engelsk kalles "data mining" - kunsten å grave frem verdifull informasjon i store datamengder. Boka *Fra skuff til skjerm* (Aukrust og Hodne red. 1998) gir en gjennomgang av hvordan fagmiljøene i dag ser på mulighetene for bruk og presentasjon av Dokumentasjonsprosjektets materiale.

I "optimal tilgjengelighet" må det også være en vurdering av hva som skal eller kan tilgjengeliggjøres ut fra klausuler, etiske forhold, lovverk og økonomiske interesser. Deler av Dokumentasjonsprosjektets materiale er klausulbelagt og vil bare være tilgjengelig etter søknad. En del bildemateriale er i prinsippet fritt, men vil også bare være tilgjengelig etter søknad ut fra etiske forhold. Dette gjelder særlig det samisk-etnografiske materialet. Opphavsrett og personvern er regulert gjennom lovgivningen.

Økonomiske forhold er mer uklare. Vårt inntrykk gjennom prosjektarbeidet er at universitetene har et uklart forhold til kommersialisering av egen virksomhet og av hvem som skal ha eventuelle inntekter av virksomhet utført av ansatte ved institusjonen. Dette har gjort det vanskelig for prosjektledelsen å kunne gi en rimelig fortolkning av dette delmålet. Det har vært laget en utredning ved Universitet i Bergen angående adgang til museumsdata, men denne har ikke bidratt med forslag ut over det å

oppretholde det uklare status quo. En del publiseringsvalg er foretatt etter skjønn etter konsultasjoner med fagavdelingene. Men mange av fagavdelingene har ikke hatt tilstrekkelig kunnskap på dette feltet.

#### **2.4 Delmål 4: Samordnet datarepresentasjon og grensesnitt**

Dataanalysen var en av forutsetningene for en fornuftig utvikling av arkivene og registrene. Det fantes ikke nødvendig ekspertise i de enkelte fagmiljøene til dette. Analysen ble derfor samordnet sentralt blant annet for å holde oversikten over felles dataelementer. Prosjektets databaser inneholder data fra en rekke forskjellige fagfelt. For å muliggjøre en kobling av dataene på tvers av faggrensene ble det utarbeidet standarder for representasjon av felles datatyper. Disse standardene er, så langt det er mulig, i samsvar med eksisterende nasjonale og internasjonale standarder.

Det er blitt lagt stor vekt på å kode og lagre dataene i et format som er mest mulig uavhengig av enkeltleverandører og at formatet er enkelt og veldefinert slik at det skal være enkelt å forstå. Det skal også være lite arbeidskrevende å benytte dataene i nye programsystemer.

Utvikling av søkesystemer og presentasjon av dataene ble bevisst utsatt til dataregistreringen nærmet seg slutten. Man har på denne måten unngått den type feilsatsninger i utvikling av programsystemer og for tidlige investeringer i utstyr som har forekommet i en lang rekke offentlige og private informasjonsteknologiprojekter. Om man skal trekke frem en ulempe ved den valgte prosjektstrategien, måtte det være at mangelen på demonstrerbare resultater kan ha gjort det vanskeligere å synliggjøre oppfylte milestems mål overfor bevilgende myndigheter i prosjektperioden.

Riktigheten i prosjektstrategien har vist seg gjennom introduksjonen av Internett og den såkalte verdensveven (Word Wide Web). De teknologiske utviklingstrendene man så på slutten av 1980-tallet, utviklet seg mye hurtigere enn man hadde forestilt seg. Allerede i 1994 forelå den nødvendige nett-teknologi. I løpet av et par år er denne nett-teknologien blitt standard hyllevare og er i dag så allminnliggjort at den brukes så vel i klasseværelset som på ungdomsrommet. Man må imidlertid være klar over at konstruksjon av komplekse systemer fremdeles ikke er noen triviell oppgave, og at det her er mulig å feile med svært store økonomiske konsekvenser.

### **3 Prosjektledelse**

I første halvår av 1990 tok daværende prodekanus Bjarne Hodne ved Det historisk-filosofiske fakultet ved Universitetet i Oslo i samarbeid med professor Herman Jervell initiativet til et forprosjekt for å kartlegge fakultetets egne arkiver og samlinger med henblikk på en mer samordnet og massiv innføring av datateknologi.

Gjennom et forprosjekt våren 1990, ledet av førsteamanuensis Christian-Emil Ore, ble det fremskaffet en samlet oversikt over registre og arkiver ved fakultetets tradisjonelle samlingsavdelinger. Forprosjektet ble avsluttet sommeren 1990. Prosjektets referansegruppe med representanter for de ulike avdelingene besluttet at arbeidet skulle fortsette etter de linjer som ble trukket opp i forprosjektrapporten. Det forberedende arbeidet med hovedprosjektet ble tatt opp igjen høsten 1990. I februar 1991 vedtok så fakultetsrådet ved HF-fakultet ved Universitetet i Oslo at man skulle gå i gang med et seksårig dokumentasjonsprosjekt.

I samråd med Kirke-, utdannings- og forskningsdepartementet ble det besluttet å gjøre prosjektet landsomfattende slik at en unngikk parallelle og konkurrerende prosjekter. Det nasjonale dokumentasjonsprosjektet for perioden 1992-1997 ble vedtatt på Det nasjonale dekanmøtet for HF-fakultetene i november 1991. Våren 1992 ble det så oppnevnt en styringsgruppe for prosjektet med en representant fra hver av universitetene, samt en representant for universitetsdirektøren ved Universitetet i Oslo. Medlemmene i styringsgruppen har hatt lokale arbeidsgrupper tilknyttet seg gjennom hele eller deler av prosjektperioden. Dette har vist seg å være hensiktsmessig, særlig i startfasen da utredning av behov og ressursfordelingen ble foretatt.

Prosjektet fikk sin endelige form gjennom et prosjektdirektiv som ble utarbeidet i løpet av 1992. Den interne ressursfordelingen ble foretatt i første halvdel av 1994. I perioden 1993-1997 foregikk arbeidet i henhold til prosjektdirektivet og ressursfordelingen fra 1994. I ettertid har det vist seg svært gunstig idet en har unngått tilbakevendende ressursdebatter og har kunnet foreta en langsiktig planlegging.

#### **4 Organisering og gjennomføring av registreringsarbeidet**

Som nevnt hadde prosjektet en ramme på om lag 600 årsverk sysselsatte og 120 årsverk faglærte. Den endelige ressursfordelingen ble foretatt tidlig i 1994. Som man kunne forvente gikk noe av konverteringsarbeidet raskere, noe langsommere enn antatt. Man valgte likevel å strengt holde seg til den opprinnelige ressursfordelingen av faglærte årsverk. Disse ble oppfattet som den mest "kostbare" ressursen av fagavdelingen. Når det gjaldt ufaglærte årsverk, det vil si innskrivere på sysselsettingstiltak, ble fordelingsnøkkelen fra 1994 justert etter behov.

Prosjektet hadde ansvar for at innskrivingsmiljøene ble etablert og for koordineringen av dem. Det ble etablert i alt 16 mindre registreringsgrupper med medarbeidere på ulike former for sysselsettingstiltak. De lokale arbeidskontorene del- eller helfinansierte lønnskostnadene, mens prosjektet stod for administrative kostnader. Prosjektet inngikk et samarbeid med folkeuniversitetene og andre lokale aktører om administrasjon av gruppene.

Prosjektet "Omstilling gjennom målrettet utdanning" førte til etableringen av fire registreringssentraler i omstillingskommuner med behov for nye tiltak som kunne sikre arbeidsledige formelle utdanningstilbud og relevant arbeidstrening. Sentralene var et samarbeid mellom lokale myndigheter og næringsliv i de aktuelle kommunene, fylkesarbeidskontorene i Nordland og Finnmark, Næringsdepartementet, Kommunal- og arbeidsdepartementet, SND, Statens datasentral og universitetene. Departementene, SND, fylkene og kommunene stod for finansiering av utstyr og lokaler, mens arbeidsmarkedsetaten finansierte driften og kursvirksomheten. Utdanningstilbudet som inngikk i innskrivingspakkene, samt teknisk og administrativ drift av sentralene, ble satt bort på oppdrag til Kurssekretariatet, UiO, som igjen hadde lokale og sentrale samarbeidspartnere.

Dokumentasjonsprosjektet hadde ansvar for 1) sammenkobling mellom delprosjekter og innskrivingsmiljøer, 2) informasjonsvirksomhet 3) organisering og oppfølging av innskrivingsmiljøene og 4) fortløpende evaluering av den samlede innskrivingsaktiviteten ut fra kvalitet, effektivitet og fremdrift.

Selve driften av databasene var ikke en del av prosjektet og var ment å forankres innenfor de enkelte organisasjonene på permanent basis. Ved prosjektets slutt finnes det imidlertid ikke tilstrekkelig kompetanse i fagmiljøene til dette. Dokumentasjons-

prosjektets etterorganisasjon er derfor etablert for perioden 1998 - 2000 for å sikre dataene og bruken av disse og for å bistå med å etablere en permanent driftsorganisasjon. Etterorganisasjonen oppgaver er beskrevet nærmere nedenfor.

## **6 Dokumentasjonsprosjektets etterorganisasjon**

### **6.1 Bakgrunn**

I 1995 ble det startet planleggingsarbeid med mål å videreføre Dokumentasjonsprosjektet i perioden 1998 - 2001. Omfanget av denne fase 2 skulle være om lag den samme som den nåværende fasen 1992 - 1997. På statsbudsjettet for 1996 ble det bevilget kr 800.000 til en oppstart av fase 2. Nytt konverteringsarbeid startet opp, og planleggingsarbeidet fortsatte i 1996. Fase 2 kom imidlertid ikke med i statsbudsjettet for 1997. UiO sentralt fikk derfor tilbud om å kunne benytte prosjektorganisasjonen i et eventuelt museumsprosjekt som skulle omfatte alle universitetsmuseene i Norge. Dette ble sett på som meget interessant, og det ble utarbeidet en prosjektplan og en søknad som ble oversendt KUF i juni 1997. Driften og videreutviklingen av Dokumentasjonsprosjektets databaser er tenkt gjort i nær tilknytning til dette museumsprosjektet. Men de direkte kostnadene ved å sikre drift og videreutvikling av databasene etter Dokumentasjonsprosjektet er holdt utenfor museumsprosjektet.

### **6.2 Rammer og mål**

Rammen for etterorganisasjonen er en førsteamanuensisstilling (leder) og to overingeniører (IT) samt kr 340.000 til utstyr, drift og reiser. Det er på alle mulige måter knapt tatt i betraktning at Dokumentasjonsprosjekt hadde 15 delprosjekter fordelt på fire universiteter og er ansvarlig for utvikling av museumsprogramvare på de kulturhistoriske universitetsmuseene. Aktivitet ut over denne rammen må finansieres på instituttnivå og museumsnivå.

Etterorganisasjonen har ennå ikke noe formelt mandat. I forbindelse med finansieringsarbeidet tidligere i 1997 ble det klart at det er fem hovedmål som må nåes dersom man skal sikre de store investeringene som er foretatt i Dokumentasjonsprosjektet:

*Delmål 1: Ferdigstille brukergrensesnitt for databasene slik at de kan inkorporeres i det daglige arbeidet ved de ulike avdelingene.*

Dokumentasjonsprosjektet hadde som primærmål å bringe papirarkiver over på elektronisk form. Dette er gjort i henhold til planene for alle delprosjektene. Det er laget søkesystem med tilhørende grensesnitt mot dataene. For museene er det også laget enkeltstående systemer for oppdatering av databasene. Slike er ikke laget for leksikografi, navnegransking og folkloristikk. Det er utarbeidet planer for hvordan alle datasamlingene skal kunne vedlikeholdes av fagmiljøene. Men disse er altså ikke implementert i de enkelte organisasjonene. Dokumentasjonsprosjektets delmål "Rasjonelle rutiner for samlinger og arkiver" er dermed ikke fullstendig oppnådd. Den viktigste arbeidsoppgaven for etterorganisasjonen er derfor i samarbeid med de enkelte fagavdelingene å lage helhetlige databasesystem som skal erstatte dagens delsystemer og manuelle rutiner. Dette arbeidet kan bare bli vellykket dersom fagavdelingene forplikter seg til å delta i den nødvendige omstruktureringen av det daglige samlingsarbeidet. Her har museene og fakultetene en viktig rolle.

*Delmål 2: Lage opplegg for og administrere opplæringen i bruk av databasene ved de enkelte avdelingene.*

Etterorganisasjonen kan i samarbeid med det enkelte fagmiljø lage kursopplegg for bruk av databasene. Slike kursopplegg kan spenne fra rent praktisk instruksjon i bruk av skjerm bilde til undervisning i IT-metode i det enkelte faget. Etterorganisasjonen har kompetanse til å lage og administrere opplæringen, men ikke tilstrekkelig ressurser til å gjennomføre det i noen stor skala. Til dette trengs tilskudd fra institusjonene.

*Delmål 3: Lage eller tilpasse IT-standarder for de ulike fagområdene på nasjonal basis.*

I Dokumentasjonsprosjektet er store datamengder konvertert til elektronisk form etter samme mal. Man er derfor i den enestående situasjon at man innenfor de enkelte fagfelt har en nasjonal de facto dataformatstandard. Disse standardene er basert på tilsvarende internasjonale anbefalinger. Standardene brukt i Dokumentasjonsprosjektet må antakeligvis justeres noe for å kunne brukes i fagavdelingenes daglige arbeid. Det er også viktig at standardene videreutvikles med fagene slik at de opprettholdes som standarder på nasjonalt nivå. For å åpne dette må det etableres landsdekkende kontaktgrupper innenfor de enkelte fagene, men også en tverrfaglig koordinerende gruppe for ivaretagelse av standarder som griper over flere fag (f.eks foto, geografiske data og tidsbestemmelse).

*Delmål 4: Videreutvikle databasene og IT-metoder knyttet til de ulike fagområdene.*

Databasene etter Dokumentasjonsprosjektet er forskningsdatabaser og ikke administrative databaser. For å få full nytte av dem vil det derfor være nødvendig å videreutvikle dem i takt med behovene hos forskerne som bruker dem. Stikkord er nye søkemuligheter, nye datafelt og forbedrede grensesnitt for søk og formidling. Videreutvikling av IT-metodene vil innenfor de enkelte fagene være knyttet til bruk av innsamlede data, men også til feltarbeid og dataregistrering. Med bruk menes dataanalyse i forbindelse med forskningsprosjekter, bruk av (bearbejdede) data i for eksempel språkteknologiske verktøy og bruk av dataene i undervisning og formidling. Dersom en skal kunne benytte datamaskinens kraft både for dataanalyse og for formidling og presentasjon, stilles det strengere krav til de innsamlede (referanse)dataene. Det er svært viktig å utvikle IT-baserte metoder for innsamling av (referanse)data slik at feltarbeidet kan effektiviseres og slik at man for all del unngår den opphopningen av de til dels usystematiske papirsamlinger man har hatt til nå. Dette punktet henger således sammen med punktene 1 - 3 ovenfor.

*Delmål 5: Utrede og etablere en permanent drift av databasene.*

Dokumentasjonsprosjektets oppgave var i henhold til prosjektdirektivet å etablere konvertere data og etablere databaser, men ikke forestå noen drift av disse. Etterorganisasjonen vil drive databasene inntil en permanent løsning er funnet. Ved årsskiftet 1997/1998 finnes det ingen andre alternativer. Etterorganisasjonen skal imidlertid raskest mulig etablere andre løsninger. Databasene og den tilhørende driften kan enten splittes opp på de enkelte universitetene eller drives som en sentral database. Den førstnevnte løsningen vil kunne gi den enkelte institusjonen følelse av bedre kontroll over egne data, men vil være noe mer ressurskrevende å administrere dersom en ønsker at databasene også i framtiden skal fremstå som en tverrfaglig nasjonal database.

### **6.3 Lokale forankringspunkter for etterorganisasjonen**

Etterorganisasjonen skal søke samarbeid med eksisterende miljøer ved de ulike universitetene slik at den fremtidige drift og videreutvikling av databasene kan bli mest mulig lokalt forankret. For universitetene i Bergen og Tromsø er henholdsvis Humanistisk informasjonsteknologi-senter (HIT) og Registreringsentralen for historiske data (RHD) naturlige samarbeidspartnere. I Oslo er miljøet ved Seksjon for humanistisk informatikk og Tekstlaboratoriet ved Institutt for lingvistiske fag et naturlig forankringspunkt. I Trondheim finnes ingen tilsvarende enheter i de humanistiske miljøene.

Ved alle de fire universitetene gjelder det at kompetansen innenfor historisk og språklig informatikk er relativt godt utbygd. Innenfor språkteknologi er det et godt utbygd samarbeid gjennom Norsk infrastruktur for språkteknologi (NIFST). Dokumentasjonsprosjektets etterorganisasjon deltar aktivt i dette samarbeidet.

Feltet museumsinformatikk er ikke så godt utbygd. I tillegg til kompetansen som er bygd opp rundt Dokumentasjonsprosjektets delprosjekter, finnes det kompetanse ved HIT-sentret i Bergen og ved Vitenskapsmuseet i Trondheim. En viktig oppgave for etterorganisasjonen er derfor å bidra til at det etableres kompetanse innenfor museumsinformatikk ved alle universitetene.

### **6.4 Eksternt samarbeid**

Etterorganisasjonen vil søke kontakt med eksterne samarbeidspartnere og knytte til seg eksterntfinansierte prosjekter. Med eksternt menes her institusjoner og prosjekter utenfor universitetssektoren i Norge. I løpet av høsten 1997 og våren 1998 er det etablert kontakter og innledet formelt samarbeid med en rekke eksterne institusjoner. I listen under er det i tillegg til tema og eksternt institusjon ført opp de faste universitetsavdelingene som deltar i samarbeidet.

- Riksarkivet og INL (Gammelnorsk ordboksverk): samarbeid om kombidatabase for middelalderdiplomer.
- Norsk språkrådet og INL (seksjon for leksikografi og målføregransking): samarbeid om leksikalske databaser og ordregistrant for norsk.
- Ivar Aasen-stiftinga, INL (seksjon for leksikografi og målføregransking): samarbeid om databaser over norske dialekter og nynorsk skriftspråk.
- Ortnamninstutet i Uppsala og INL (seksjon for navnegransking): samarbeid om videreutvikling og drift av skandinavisk stedsnavnsdatabase.
- Riksantikvaren og forvaltningsmuseene: samarbeid om sammenslåing og videreutvikling av det nasjonale fornminneregisteret.
- Nationalmuseet i København, Nordiska Museet i Stockholm, Båhuslän länsmuseum, Norsk Folkemuseum og Dokumentasjonsprosjektet gjennomfører et forprosjekt for et fellesnordisk søkesystem for kulturhistoriske museumsdatabaser.
- Primusprosjektet (Norsk Folkemuseum, Maihaugen, Norsk teknisk museum, Telemuseet): samarbeid om design av databaser for større norske museer.

## 7 Nasjonal styringsgruppe og sentral prosjektstab

### Følgende personer har vært medlemmer av Den nasjonale styringsgruppen:

Leder:	Tellef Kvifte, UiO	(1992)
	Knut Aukrust, UiO	(1993 - 94)
	Bjarne Hodne, UiO	(1995 - 97)
Medlemmer:	Thor Bjørn Arlov, NTNU	(1992 - 97)
	Arne Svindland, UiB	(1992 - 94)
	Christopher Prescott, UiB	(1994)
	Karen Marie Heimvik, UiB	(1995)
	Svein Indrelid, UiB	(1996 - 97)
	Gunnar Thorvaldsen, UiTø	(1992 - 97)
	Dikka Storm, UiTø	(1996 - 97)
Sekretær:	Knut Aukrust	(1995 - 97)

Svend Erik Svendsen/Per Heitmann har vært representant og vararepresentant for Unversitetsdirektøren i Universitetet i Oslo.

### Følgende personer har vært ansatt i Dokumentasjonsprosjektets sentrale prosjektorganisasjon:

#### *Prosjektets daglige ledelse:*

Knut Aukrust	administrativ prosjektleder	(1992)
Christian-Emil Ore	faglig leder	(fra 1991)
	prosjektkoordinator	(1993 - 97)
Svein Bjørkås	registreringsleder	(1992)
Nina Kristiansen	registreringsleder	(1992 - 97)
Jennifer Akselsen	driftskoordinator, kursansvarlig, ansatt i Kurssekretariatet, UiO	(92-97)

#### *Administrativ prosjektstab:*

Oddrun Rangsæter	personalkonsulent	(1995 - 97)
Else Fjeld	konsulent, tiltakssiden	(1995 - 97)
Christine Cellucci	prosjektsekretær	(1993 - 94)
Kaja Videng	prosjektsekretær	(1994 - 95)
Elin Rognerud	prosjektsekretær	(1995 - 97)

#### *IT-faglig stab:*

Sølvi Andersen	numismatikk, nyere kulturhistorie, arkeologi	(1991 - 96)
Jon Holmen	koordineringsansvar for museums- arbeidet arkeologi	(1992 - 97)
Silvester Pozarnik	leksikografi	(1992 - 95)
Lars Jørgen Tvedt	navnegransking og leksikografi	(1994 - 97)
Gunn Fossdal	museumssamlingene	(1995 - 97)
Jarle Stabell	museumssamlingene	(1995 - 97)



# Prosjektøkonomien

## Finansiering

Prosjektet ble i 1991 igangsatt med midler fra Det historisk-filosofiske fakultet i Oslo. I 1992, 1993 og 1994 ble det fra Kirke-, forskning,- og undervisningsdepartementet (KUF) avsatt henholdsvis 4 millioner, 4 millioner og 5,3 millioner kroner som øremerkede midler på statsbudsjettene og 2 millioner kroner pr. år fra universitetenes side. For å få gjennomført den store konverteringen var dette langt fra nok, og man hadde allerede i 1991 begynt arbeidet med alternative finansieringsmuligheter. I samarbeid med Næringsdepartementet, SND, Arbeidsdirektoratet samt lokale myndigheter og arbeidskontor ble det etablerte en rekke opplæringstiltak for arbeidsledige i Nord-Norge. I perioden 1993-1994 ble aktivitetene ved disse sentralene inklusive faglærte årsverk til kvalitetssikring og oppfølging, fullfinansiert av Arbeidsmarkedsetaten. Fra 1995 ble finansieringen av kvalitetssikring og oppfølging overtatt av KUF som økte rammebevilgningen til henholdsvis 13 millioner, 13 millioner og 10 millioner for årene 1995, 1996 og 1997. Ved siden av registreringsentralene har en stor del av konverteringsarbeidet vært finansiert gjennom bruk av ulike andre sysselsettingstiltak (f.eks KAJA).

Gjennom å finansiere konverteringen på denne måten fikk prosjektet en viktig tilleggsoppgave: å drive gode, kompetansehevende sysselsettingstiltak. To målsettinger skulle nås samtidig: å konvertere arkivene på en effektiv måte med høy kvalitet, samt å gi deltakerne en meningsfull tiltaksplass som gav økte sjanser for ordinært arbeid. Målinger og evalueringene av Dokumentasjonsprosjektet som sysselsettingstiltak viser at resultatene er svært gode når det gjelder deltakernes overgang til ordinært arbeid eller studier.

Det er altså mulig å skaffe finansiering og å gjennomføre et prosjekt av Dokumentasjonsprosjektets størrelse dersom en velger å tenke utenfor de tradisjonelle universitets-rammene og å benytte universitetenes brede kompetanse på utradisjonelle måter.

## Oversikt utgifter og bevilgninger 1992 - 1997

<i>Utgifter:</i>		%
Prosjektadministrasjon og konverteringsstøtte	13.500.000	6
Systemutvikling, innskrivingsprogrammer, drift	10.000.000	4
Kvalitetssikring, virksomheten i delprosjektene	41.000.000	17
Registreringsarbeidet	26.000.000	11
Utdannings- og kursvirksomhet arbeidssøkende	51.000.000	21
Sum prosjektutgifter	141.500.000	
Lønn og ytelser arbeidssøkende	99.000.000	41
Totale utgifter:	240.500.000	100

**Bevilgninger:**

Kirke- undervisnings og forskningsdepartementet	46.500.000	19
Universitetene	15.500.000	6
Nærings- og energidepartementet	6.500.000	3
Kommunal- og arbeidsdepartementet	3.600.000	1
Andre bidragsytere	2.600.000	1
Arbeidsmarkedsetaten	165.800.000	69
Totale bevilgninger:	240.500.000	100

**Kommentar til oversikten:**

Regnskapene for Dokumentasjonsprosjektet avsluttes ikke før i 1998, med utbetaling av feriepenger og andre restkostnader og de siste overføringene fra Arbeidsmarkeds-etaten. Endelige tall vil derfor ikke være klare før høsten 1998.

Administrasjon, drift og kursvirksomheten ved registreringsentralene i Nord-Norge finansieres gjennom direkte bevilgninger av Arbeidsmarkeds-etaten og er regnskapsført i prosjektet. Arbeidsmarkeds-etaten står også for stønadene til medarbeiderne på sentralene. Disse går direkte til den enkelte medarbeider og er ikke regnskapsført av prosjektet. Tallene for disse ytelsene er derfor beregnet etter gjeldende satser for de enkelte år.

Lønn til medarbeiderne i registreringsgruppene finansieres delvis av Arbeidsmarkeds-etaten, mens administrasjon av gruppene og deler av lønnskostnadene finansieres via prosjektet. Ved noen av tiltaksformene overføres lønnsmidlene til arbeidsgiver som står for utbetalingene, mens stønadene ved andre tiltaksformer går direkte fra arbeidskontorene til den enkelte tiltaksdeltaker. Tallene for disse siste ytelsene er også beregnet etter gjeldende satser.

## IT-strategi

Kapittelet er basert på Christian-Emil Ores artikkel: Hvordan lage databaser for språk og kulturfag i *Fra skuff til skjerm*, Aukrust og Hodne (red), Universitetsforlaget 1998.

En elektronisk tilgjengeliggjøring av kildemateriale fjerner behovet for nærhet til de fysiske kildene for store deler av forskersamfunnet og for allmennheten. Informasjonsteknologi vil derfor kunne bidra til en "demokratisering" av kilder i form av mer allmenn tilgang. Det er viktig å understreke at hensyn til personvern, hensyn til etikk og kommersielle hensyn legger begrensninger på hvor åpen tilgangen til materialet kan være. I dette kapittelet har jeg imidlertid sett bort fra den typen forhold. Når jeg beskriver systemet som et bibliotek hvor man kan hente ut informasjon til eget bruk, er det derfor forutsatt at det eksisterer sikkerhetsmekanismer som hindrer uautorisert bruk av sensitivt materiale.

### Den grunnleggende IT-strategien i Dokumentasjonsprosjektet

Formålet med Dokumentasjonsprosjektet har vært å omstille samlingsavdelingenes behandling av informasjon til moderne datateknikk og dermed effektivisere det interne samlingsarbeidet, eksternt samarbeid og utveksling av eksisterende informasjon samt innhenting av ny informasjon.

Et slik omstillingsprosjekt består av to hoveddeler: å utarbeide elektroniske informasjonssystemer for de deltagende avdelingene, og å konvertere den eksisterende papirbaserte informasjonen til elektronisk form. Det har i tilsvarende prosjekter ikke vært uvanlig først å anskaffe utstyr og å utvikle de elektroniske informasjonssystemene for så å skulle legge den papirbaserte informasjonen i systemene. Fordelen med denne strategien er at man kan begynne å bruke de konverterte dataene etter som de legges inn. Ved et konverteringsarbeid av noen varighet risikerer man imidlertid at de valgte løsningene er foreldede når datainnlastingen er ferdig, eller - enda verre - at de i sin tid ambisiøse egenproduserte løsningene finnes tilgjengelig som hylleware på markedet. I det siste tilfellet sitter man med et i og for seg godt system, men også med et meget kostbart vedlikeholdsansvar siden det er svært få å dele vedlikeholdskostnadene med. Man kan også risikere at det lagringsformatet man lagde for sine egne løsninger, gjør det vanskelig å konvertere dataene til andre, mer standard, formater. Det finnes dessverre mange eksempler på både store og små omstillingsprosjekter hvor det har gått til dels svært galt nettopp på grunn av en dårlig koordinering av konverteringsarbeid, maskininnkjøp og egenproduksjon av programvare med de generelle utviklingstrendene innenfor IT-sektoren.

Da Dokumentasjonsprosjektet ble planlagt i 1990-1991, satte vi oss som mål å gjøre de konverterte dataene tilgjengelig for et bredest mulig publikum. Vi så for oss distribusjon via disketter, via CD-ROM, men helst via datanett. I 1990 var det klart at det norske universitetsnettet kunne brukes til en slik distribusjon for i alle fall høyskolene og universitetene. Det lå også klare tegn i tiden på at de store programvareleverandørene ville satse på nettbaserte løsninger. For å unngå å havne i den fallgruben som er beskrevet over, valgte vi på bakgrunn av den forventede utviklingen å holde produksjon av egen programvare på et absolutt minimum. I stedet ble prosjektets virksomhet sentrert rundt konvertering av data. Riktigheten av prosjektets IT-strategi har vist seg i den utviklingen Internett har hatt siden 1994. Utviklingen har faktisk gått fortere enn vi våget å tro i 1990. Om prosjektet hadde utviklet egne løsninger i 1992-1993, ville disse i dag ha vært akterutseilt av den massive innsatsen som har blitt gjort internasjonalt for å utvikle programvare for nettbaserte systemer.

Vi har fra begynnelsen av tilstrebet å konvertere den papirbaserte informasjonen til et mest mulig plattformuavhengig format. Prosjektets strategi har vært å legge stor vekt på å lage formelle datamodeller for informasjonen i arkivene i henhold til en grundig utredning av brukernes nåværende behov, men også ut fra en gjennomgående analyse av brukspotensialet i materialet. Informasjonen lagres stor sett som rene tekstfiler. Disse inneholder koder i henhold til datamodellene. I praksis brukes industristandarden SGML (Standard Generalized Markup Language) Goldfarb 1991 som verktøy. Kodingen er gjort i henhold til retningslinjene som er trukket opp av det internasjonale samarbeidsprosjektet TEI (Text Encoding Initiative) Sperberg-McQueen og Burnard 1994. Det er klart at selv denne metoden også kan bli foreldet, men informasjonen ligger lagret som ren tekst med en veldefinert beskrivelse av kodingen i klar tekst. Den kan inspiseres og leses ved hjelp av det enkleste tekstbehandlingsprogram. Informasjonen kan også omformes til å kunne lastes inn i ulike databaser og søkesystemer.

Dokumentasjonsprosjektets noe trygge og reserverte "la teknologiutviklingen arbeide for oss mens vi konverterer data"-holdning har gjort at resultatene fra prosjektet var lite synlige før i slutten av prosjektperioden. Dette har nok strukket tålmodigheten hos våre finansieringskilder, men sett under ett har det vært en riktig prioritering av ressursene. Når det gjelder metoder for å kode og modellere dataene har vi valgt teknikker som gjør oss uavhengige av program- og maskinvareprodusentenes krumspring. Vi har likevel prøvd å satse på gjennomarbeidede teknikker som synes å ha framtiden for seg. Metoden vi i 1991 valgte for å kode og lagre våre tekstlige data, er siden den gang blitt den normale måten å produsere elektronisk tekst på. Den benyttes blant annet for å produsere tekst som ligger i verdensveven (World Wide Web).

I 1990 var det svært kostbart å lagre både bilder og lyd i digital form. Vi satset derfor utelukkende på tekstkonvertering. I de senere årene har det blitt ganske billig å fremstille og lagre digitale bilder og lyd i datasystemer. Vi har derfor introdusert faksimiler av originaldokumentene i mange av databasene samt bygd opp en del digitale fotosamlinger. I en eventuell videreføring av prosjektet vil det være naturlig å foreta en storstilt konvertering av bildematerialet i museene.

I prosjektets siste år har det blitt fokusert mer på å utvikle og tilrettelegge verktøy for bruk av dataene og å "pakke" dem inn i et brukervennlig system. Det vil likevel være en løpende utvikling av systemet i årene som kommer. Dels skyldes dette IT-utviklingen, dels at det ikke er mulig å forutsi alle fremtidige behov forskere vil ha.

## **Arbeidet med de ulike materialtypene**

Dokumentasjonsprosjektet kan grovt deles inn i en museumsdel og en språklig orientert del. I den museumsrettede delen av prosjektet arbeider vi med å bygge opp forskningsdatabaser for de arkeologiske museene i Bergen, Oslo og Tromsø. I Tromsø arbeider vi også med nyere kulturhistorisk materiale. Den språklige delen består i hovedsak av tilrettelegging av bakgrunnsmateriale for ordboksavdelingene (bokmål, gammelnorsk og nynorsk) ved Avdeling for leksikografi i Oslo, for Trønderordboka ved Norges teknisk-naturvitenskapelige universitet (NTNU) og for navnegranskingsmiljøene i Oslo og Tromsø. I tillegg kommer arbeidet med tingbøker fra 1600-tallet (Tingbokprosjektet), reinbeitekommissjonens rapport fra 1920-tallet, balladeoppskriftene fra forrige århundre og minneoppgavene fra 60-, 70- og 80-tallet. Disse delprosjektene faller også inn under det man kan kalle språklig orientert materiale, selv om de primære brukerne er mest interessert i innholdet og ikke i hvordan den språklige ytringen er konstruert. For å gjøre bildet komplett må vi også nevne arbeidet med

registrene til Rådet for folkemusikk og folkedans og arbeidet med en nordnorsk bibliografi. Disse delprosjektene kan ikke plasseres verken som språkorienterte prosjekter eller som museumsprosjekter.

En annen måte å dele inn de ulike samlingsavdelingene der Dokumentasjonsprosjektet har vært engasjert på, er å se om samlingene består av fysiske gjenstander. De arkeologiske museene har åpenbart slike fysiske gjenstander, og en vesentlig del av dataene i deres arkiver og kataloger beskriver objektene selv og omstendigheter rundt funnene av dem. For denne typen samlingsavdelinger vil gjenstandene ha en svært viktig plass. Databasene er også orientert rundt gjenstandene. Men i selve konverteringen av dataene og i konstruksjonen av databasene har vi tatt i bruk teknikker som vanligvis er brukt i forbindelse med språklige data. Jeg vil her nevne tekstkoding, fritekstsøk og bruk av konkordanser som fremvisningsverktøy. Hvordan dette er gjort, er av vesentlig betydning for å skjønne mulighetene og begrensningene i databasene, og er forklart inngående i avsnittet "Museumsarbeid".

## **Det språkorienterte arbeidet**

Samlinger uten fysiske gjenstander kan virke selvmotsigende. Arkivskuffene ved Avdeling for leksikografi er fylt med over 9 millioner kort og er definitivt en fysisk realitet. Men det som står skrevet på kortene, er bare opplysninger om språket selv. Språk observeres vanligvis i form av lyd, skrift eller bilde. For de aller fleste formål innenfor språklig forskning regnes det som fullgodt å studere de språklige ytringene gjennom representasjoner som skrift i ulike medier, lydopptak eller film og bilder. Felles for disse representasjonsmåtene er at datamaskinene kan brukes som representasjonsmedium. De språklige samlingene kan derfor lagres i en datamaskin. Man er altså uavhengig av de språklige samlingsavdelingenes originalmateriale. Det har i løpet av de siste 30-40 årene blitt utviklet en rekke datamaskinelle versjoner av tradisjonelle metoder samt nye datamaskinelle metoder for datafangst og analyse innenfor de ulike grenene av språkforskningen. Stikkord er her optisk lesing av tekst (OCR), elektroniske tekstkorpus, programmer for hel- og halvautomatisk lemmatisering (bøyningsanalyse og påføring av grunnord) av løpende tekst, ulike konkordansverktøy og statistikkpakker. I prinsippet trenger denne type samlingsarbeid derfor bare kraftige datamaskiner, digitalt lyd- og bildestudio samt digitaliseringsutstyr for kartdata.

Hva så med samlingene som har originaloppskrifter fra forrige århundre eller tingbøker fra 1600-tallet, for ikke å snakke om middelalderdiplomene? Disse kan ikke legges inn i en datamaskin. Her kan vi bemerke at samlingsavdelingene ved universitetene i liten grad har denne typen materiale. Der man gjør det, er det en liten del av aktiviteten ved avdelingen som er knyttet til de fysiske objektene. For språkforskere er de likevel ikke de primære studieobjektene. Det primære målet for de språklige orienterte forskerne ved HF-fakultetene er ikke å forske i gjenstandenes fysiske eller kjemiske beskaffenhet, men å kopiere og analysere den informasjonen som gjenstandene formidler, altså en av de mest tradisjonelle humanistiske oppgavene: publisering gjennom tekstkritiske utgaver og faksimileutgaver (mikrofilm eller trykte bilder).

Innenfor de humanistiske fagene synes det å være en begynnende trend å gi ut selve kildematerialet i elektronisk form utstyrt med søkeverktøy og/eller et elektronisk note- eller kommentarapparat. En slik publikasjon er på mange måter svært lik en mikrofilmversjon av et arkiv eller bibliotek, men synes likevel å være nærmere intensjonen bak tekstkritiske utgaver. Denne publiseringsmetoden er også et svært godt alternativ til det tradisjonelle filologiske arbeidet der kildematerialet kun refereres gjennom noteapparatet. Ved å legge selve bakgrunns materialet sammen med det

vitenskapelige arbeidet, vil forholdet mellom kildematerialet og den vitenskapelige artikkelen endres. Som George Landow påpeker, får fotnoten eller henvisningen en annen rolle. De er ikke lenger underordnede tekstfragmenter, men pekere inn i andre komplette tekster Delary og Landow 1991. I stedet for å stå som en egen overordnet enhet vil artikkelen på mange måter kunne bli et kommentarverk som eksisterer parallelt med sine kilder. Artikkelen vil således bevege seg i retning av en spesialisert kommentert tekstutgave.

Muligheten for å publisere hele arkiver med et noteapparat vil på den annen side kunne viske ut forskjellen mellom arkiv eller bibliotek og de filologiske tekstkritiske samlingsutgavene. Dette vil stoppe utgivelsen av de store trykte kommenterte tekstutgavene. Av økonomiske årsaker, men også av praktiske årsaker, vil de bli erstattet av utgivelser av arkiver der grunnmaterialet vil være mer eller mindre bearbejdede rådata i form av avskrifter eller faksimiler supplert med ulike fortolkede utgaver av de samme tekstene. Et eksempel er utgivelsen av islandske ættesagaer på CD-ROM Myklebust 1996 eller CD-ROM-en med Dag Solstads romaner Solstad 1996. Den førstnevnte er riktignok en ren tekstsamling. Solstad-CD'en har i tillegg søkemuligheter og litt lyd. Beveger i oss mer i retning av det filologiske har Samlaget en CD-ROM med stoff av og rundt Ivar Aasen. Det er egentlig ingen kommentert utgave av Ivar Aasens verker, men en blanding av en elektronisk variant av en gjennomillustrert bok om Ivar Aasen med lite tekst og en samling av de viktigste verkene Aasen skrev. Men kommenterte verk eksisterer allerede. Det finnes for eksempel tre ulike Shakespeare-utgivelser på CD-ROM. Det virkelige filologiske flaggskipet for elektroniske kommenterte tekstutgaver er "The Wife of Bath's Prologue" publisert som CD-ROM gjennom "The Canterbury Tales Project" Robinson 1996. Her er om lag 40 ulike manuskripter lemmatisert (det vil si at hvert ord er gitt en grammatisk analyse), og alle manuskriptene er sammenstilt med et felles kommentarapparat og søkesystem.

Den lille gjennomgangen over viser at elektronisk publisering gjør det vanskeligere å holde orden på hvordan man skal typebestemme innholdet i en publikasjon. Har vi å gjøre med et bibliotek, en dokumentsamling, et arkiv eller er det et nyskrevet verk? Elektronisk publisering kan også løse opp sjangerne. Det er ikke alltid så lett å si om innholdet på en CD-ROM er litteratur eller film. CD-ROM-en *Sofies verden* Gaarder 1997 er ikke en CD-versjon av Jostein Gaarders bok, slik Solstad-CD-en er en samleversjon av romanene til Solstad. Den er nærmere en filmatisering, men siden den er bygd opp som et dataspill, kan den vel heller kalles en ludifisering av romanen.

Dokumentasjonsprosjektets oppgave har vært å bringe samlingsavdelingene et steg på veien mot den digitale hverdagen. Vår oppgave har sånn sett vært å publisere informasjonen i papirsamlinger elektronisk, slik at papiroriginalene blir overflødige. En slik publisering må være vitenskapelig i den forstand at brukerne kan stole på at det vi har lagt inn i datamaskinen, gjengir informasjonen i originalmaterialet korrekt. Gjennom dette arbeidet har vi møtt de samme utfordringene og metodologiske problemene som vi finner i den tradisjonelle humanistiske kildepublikeringen. Vår "publisering" er mer som Solstad-CD-en, riktignok med et svært mye bedre søkeapparat (se også avsnittet "Hva særmerker Dokumentasjonsprosjektets databaser"). Vi har ikke hatt som oppgave å lage en spillversjon av norsk språkhistorie, selv om det kunne vært interessant. Databasene er derimot laget slik at det skal være lett å ta ut deler av dataene for å lage mer publikumsorienterte publikasjoner, gjerne spill. Systemet er laget slik at det skal også være mulig å koble disse publikasjonene sammen med de originale databasene. For et eksempel på slike publikasjoner henviser jeg til de to utstillingene "Norsk mynt i 1000 år" og "Folkeviserautomat". Den ene er

laget til Myntkabinettets jubileumsutstilling av samme navn og den andre er laget som en frittstående stand til vandreutstillingen "Kalmarunionen 600 år".

Erfaringene viser at det å lage gode utstillinger eller andre publikumsvennlige applikasjoner er noe langt mer enn bare å gjøre opprinnelig trykte tekster tilgjengelig elektronisk. Men selv det å gjøre ren tekst tilgjengelig med henblikk på anvendelighet i ulike applikasjoner er noe som må gjøres med omtanke. Stikkordet her er koding eller merking av teksten.

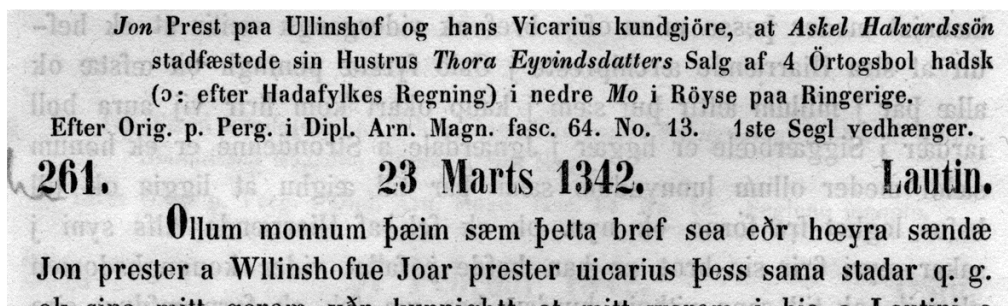
### ***Tekstkoding***

I tillegg til de rene kortarkivene og kortkatalogene er det i Dokumentasjonsprosjektet bygd opp et lite elektronisk bibliotek tilsvarende om lag 150.000 trykksider eller 500 middelstykke bind. Denne tekstsamlingen består av skjønnlitteratur, poesi, drama og sakprosa fra 1100-tallet til i dag. Tekstene er ment dels som erstatning for leksikografiske seddelarkiver, dels som selvstendige kilder for historikere, navnegranskere og språkforskere. I tillegg til dette filologisk-historiske materialet består tekstsamlingen av om lag 20.000 sider arkeologiske gjenstandsbeskrivelser. Dette tekstmaterialet er i utgangspunktet skrevet av fra originalene eller skannet inn ved hjelp av såkalte optiske leseprogrammer (OCR). Det ble så foretatt en grundig korrekturlesing mot originalmaterialet. En slik elektronisk råtekst kan være bra til en lang rekke formål. Men den er ikke tilstrekkelig dersom man vil bruke tekstene i avanserte analyse- og søkesystemer eller lage mer avanserte elektroniske publikasjoner med krysskoblinger og elektroniske noter. Til denne bruken må tekstenes struktur og andre relevante ting markeres i teksten på en måte som er forståelig for de programmene som skal brukes. For å møte denne typen behov og en lang rekke beslektede behov er det i løpet av 1980-årene utviklet en internasjonal standard for tekstkoding kalt Standard Generalized Markup Language (SGML, se Goldfarb 1991). SGML blir brukt i stadig større omfang i forlags- og publiseringsvirksomhet. HTML (Hyper Text Markup Language som brukes til å kode tekstene i verdensveven (World Wide Web eller WWW), er - noe unøyaktig sagt - en enkel form for SGML. Selve SGML-standardten er svært omfattende og temmelig tung å sette seg inn i. Men siden alle tekstene i Dokumentasjonsprosjektet er kodet ved hjelp av SGML, vil jeg bruke noe plass på å gi en forståelse av hva dette går ut på.

Når vi leser en roman, et dikt eller en artikkel i et fagblad, tenker vi kanskje ikke så mye over hvordan en tekstsider ser ut. Vi er mest interessert i å få med oss det som står på siden, eller å finne frem i den. Studerer vi teksten litt nøyere, vil vi finne ut at romanen er delt inn i kapitler, som igjen er delt inn i avsnitt. Kapitlene har kanskje overskrift. Diktet vi leser, består av en eller flere strofer og står gjerne trykt i en diktsamling. Dersom diktsamlingen inneholder dikt av flere forfattere, kan diktene ha angitt forfatterens navn i tillegg til tittelen på selve diktet. Fagartikkelen består gjerne av deler med enkeltavsnitt og har illustrasjoner med illustrasjonstekst. I alle disse trykksakene har utgiverne prøvd å bruke forskjellige skriftstørrelser, kursiv, halvfet og andre typografiske virkemidler for at leseren skal kunne forstå hva som er overskrift, hva som er bildetekst, hva som er nytt avsnitt, hva som er replikker i et skuespill osv.

Når en tekst skal lagres i en datamaskin, er man som regel påpasselig med å kode inn informasjon om skrifttype, kursiv og halvfet, slik at teksten ser riktig ut på skjermen eller når den trykkes. Vanligvis gjøres denne markeringen ved at forfatteren eller inntasteren reproducerer teksten i et tekstbehandlings- eller setteprogram ("desktop publishing"). Dette kan være praktisk for den ene gangen teksten skal trykkes, men gjør gjenbruk av teksten vanskelig. Vi har alle erfart hvor vanskelig det er å flytte en tekst

fra ett tekstbehandlingsprogram til et annet. De ulike programmene griser dessuten til teksten med kryptiske tegn som gjør det umulig å bruke den elektroniske teksten i tekstanalyse- eller søkeprogrammer uten å fjerne all typografisk informasjon og lagre teksten som "ren tekst". En slik ren tekst er imidlertid ikke så gunstig siden humanister gjerne vil kunne vite i hvilken del av en tekst ord og fraser forekommer. På den annen side er ikke de typografiske virkemidlene tilstrekkelige til å gi en entydig markering eller koding av de enkelte delene av en tekst. Kursivering kan for eksempel brukes både for å utheve tekst og markere et sitat. SGML er en standard for å kunne kode slike innholdsmessige og strukturelle trekk ved en tekst. Som et eksempel på bruk av SGML har jeg valgt å bruke vårt arbeid med *Diplomatarium Norvegicum* (*Diplomatarium Norvegicum* 1849-1990). *Diplomatarium Norvegicum* er en samleutgave av avskrifter av såkalte diplomer (rettsdokumenter og offisielle brev) fra norsk middelalder. Første del av verket kom i 1847, og man arbeider for tiden med bind 23.



Figur 1 Utsnitt av trykt diplom i *Diplomatarium Norvegicum*

Figur 1 viser et lite utsnitt av en side. For hvert diplom er det et sammendrag, kildeopplysninger, et løpenummer samt utstederdato og sted. Sammendraget er rykket inn og satt med liten skrift. Egennavn er satt i kursiv. Utstedersted og dato er satt i større skrift med halvfet, mens selve brevtteksten er satt med vanlig skrift. Utgiverne har altså brukt typografien bevisst for å veilede brukeren.

```
<BREV ID="03734">
<INGRESS><K>Jon</K> Prest paa Ullinshof og hans Vicarius kundgjöre, at <K>Askel
Halvardssön</K><LINJE>stadfæstede sin Hustrus <K>Thora Eyvindsdatters</K> Salg af 4 Örtogsbol
hadsk
<LINJE>(dvs: efter Hadafylkes Regning) i nedre Mo i Röyse paa Ringerige.</INGRESS>
<KILDE>Efter Orig. p. Perg. i Dipl. Arn. Magn. fasc. 64. No. 13. 1ste Segl vedhænger.</KILDE>
<NR>261.</NR><DATO AAR="1342">3 Marts 1342.</DATO><STED>Lautin.</STED>
<TEKST><SPRAAKFORM>gammelnorsk</SPRAAKFORM>Ollum monnum /th/æim sœm /th/etta
bref sea e/edd/r h/oe/yra sændæ<LINJE>Jon prester a Wllinshofue Joar prester uicarius /th/ess sama
stadar q. g.<LINJE>ok sina mitt gerom y/edd/r kunnighitt at mitt warom j hia a Lautini
a<LINJE>pallmsunnu æptæn a /th/ridiæ are ok tythugta rikis Maghnusar kononghs<LINJE> ok soom er
/th/au helldo handum soman Askiaell Hallwardz sson ok /Th/oræ<LINJE>Æiwindar dotter huspr/oe/yiæ
hans sam/th/yktti /th/a Askiaell fyrnæmder salu<LINJE>/th/a er /Th/oræ hafde sælltt j nedræ gardenom a
Moe iii /oe/rttoghær boll<LINJE>hatdz er lighær a R/oe/yisi a Ringhærriki ok till sanyndæ sættum
mitt<LINJE>okor jnsigli firir /th/etta bref er gort war a dægghi ok are sœm
fyr<LINJE>sæghir.</TEKST></BREV>
```

Figur 2 SGML-kodet versjon av de trykte diplommet

Figur 2 viser teksten i diplom nummer 261 etter at den er skrevet inn og kodet med SGML. De enkelte elementene i brevet er markert med koder på formen "<. .>" og



"</ . .>". Disse kalles tagger. Selve kodeprosessen kalles tagging. Den innholdsmessige informasjonen blir tatt vare på ved at taggene markerer hva som er sammendrag, kilde, utstederdato, sted og brevtekst. Skriftbildet kan reproduseres på grunnlag av kodingen ved at setteprogrammet bruker ulike stiler i henhold til taggene. Teksten mellom "<INGRESS>" og "</INGRESS>" skal for eksempel settes innrykket og med liten skrift. Taggingen gjør at det er lettere å søke i teksten. For å finne brev i Diplomatarium Norvegicum utstedt i år 1342 søker vi etter alle forekomster av "1342" omsluttet av "<DATO>" og "</DATO>". De fleste moderne tekstsøkesystem har mulighet for å formulere denne typen søk. Den elektroniske utgaven av Diplomatarium Norvegicum som ligger tilgjengelig på nettet, er basert på et slikt tekstsøkesystem. Det er taggene i teksten som gjør det mulig å finne frem til for eksempel alle diplomer skrevet på Hamar mellom 1340 og 1400. Hvis man hadde brukt den trykte originalen, hadde dette vært et omfattende arbeid. Siden hvert bind spenner over hele perioden fra 1100 til 1570, ville man måtte lete i alle de 22 bindene for å finne diplomene.

For å få full nytte av en kodet tekst er det nødvendig at taggene er plassert etter bestemte regler. I Diplomatariet er det for eksempel bare er utstederdatoen som er merket med "<DATO>" og "</DATO>". "<DATO>"-taggen forekommer dermed ikke innenfor andre tagger enn selve "<BREV>"-taggen. Dette gjør det enklere å formulere riktige søk. Dersom alle datoangivelsene var merket med "<DATO>"-taggen med datoer og årstall i sammendragene med "<DATO>"-taggen, måtte vi ta hensyn til det når vi søker i teksten og når tekstbildet skal gjenskapes. Søket etter diplomer fra året 1245 måtte da formuleres omtrent som "finn alle forekomster av 1245 omsluttet av "<DATO>" og "</DATO>", men ikke innenfor "<INGRESS>" og "</INGRESS>" for å utelate diplomer der årstallet 1245 bare er nevnt i sammendraget og ikke i utstederdatoen.

```

<!ELEMENT DIPLOMATARIUM - - (bind+)>
<!ELEMENT BIND - - (FORORD?,BREV+, SAKSREGISTER?, NAVNEREGISTER?,RETTELSER?)
+ (K|MERK|FMARK|FNOTE|SIDE)>
<!ELEMENT BREV - O (SPRAAKFORM?), INGRESS, KILDE, NR, DATO, TEKST,
TILLEGG*)>
<!ELEMENT DATO - O (#PCDATA)>
<!ELEMENT INGRESS - O (#PCDATA)>
<!ELEMENT K - - (#PCDATA)>
<!ELEMENT LINJE - O EMPTY>
<!ELEMENT KILDE - O (#PCDATA)>
<!ELEMENT NR - O (#PCDATA)>
<!ELEMENT SPRAAKFORM- O (#PCDATA)>
<!ELEMENT STED - O (#PCDATA)>
<!ELEMENT TEKST - O (#PCDATA)>
<!ELEMENT TILLEGG - O (#PCDATA)>
<!ATTLIST BREV ID cdata implied>
<!ATTLIST BIND NR cdata implied>
<!ATTLIST DATO AAR cdata implied>

```

Figur 3 Utsnitt av DTD for Diplomatarium Norvegicum

I SGML beskrives reglene for hvordan tagger kan settes i en tekst i den såkalte "Document Type Defintion" (DTD). Dette er en helt sentral del av SGML. Alle tekster som er kodet i henhold til SGML-standard, må være utstyrt med en slik DTD. I figur 3 ser vi en del av DTD-en for Diplomatarium Norvegicum. Linjene som begynner med "<! ELEMENT" definerer hva som kan forekomme innenfor en tagg. Navnet på taggen står til venstre, mens strukturen er angitt til høyre. Hele Diplomatariet ("DIPLOMATARIUM") består av en rekke bind ("BIND+"). Tegnet '+' angir at det kan

komme ett eller flere bind. Hvert bind består av et forord og en rekke brev fulgt av saksregister, navnerregister og rettelsler. Et brev består av sammendrag, kilde, nummer, dato, sted, tekst og eventuelle tillegg. Dette stemmer med den taggedede teksten i figur 1. Linjene som begynner med "<! ATTLIST", forteller hvilke attributter en tagg kan ha. Taggen "BREV" har attributtet "ID" slik det er vist i figur 2 ("<BREV ID="03734">"). DTD-en for Diplomatarium Norvegicum er en enkel DTD. DTD-ene blir imidlertid fort svært kompliserte, og vi kan trøste oss med at de mest er ment for datamaskiner og ikke for mennesker. Tekstkoding er imidlertid nyttig og kan sies å være en elektronisk variant av en sentral del av filologers tradisjonelle arbeid, nemlig å påføre noter og kommentarer i eksisterende tekster. Tross en del svakheter er SGML godt egnet til å kode tekster med den typen informasjon som humanister er interessert i. Siden 1987 har det derfor pågått et internasjonalt samarbeid blant humanister for å utvikle DTD-er for alle tenkelige tekster som brukes innenfor humaniora. Samarbeidet lyder navnet "The Text Encoding Initiative (TEI)" og har resultert i et svært omfattende sett med DTD-er og tilhørende anbefalinger for koding av tekster (Sperberg-McQueen og Burnard 1994).

Dokumentasjonsprosjektet har brukt SGML siden 1991. Alt det tekstlige materialet ligger som sagt kodet i SGML. Siden vi startet såpass tidlig, og siden vi har brukt ufaglært arbeidskraft til det meste del av kodingen, var vi nødt til å utvikle våre egne dokumentbeskrivelser (DTD-er). For eksempel er alle våre taggnavn norske og ikke engelske. Men vi har hele tiden skjelt til det pågående arbeidet i TEI, så det er fullt mulig maskinelt å omforme de norsktagede dokumentene slik at de er i henhold til TEIs anbefalinger også med hensyn til taggnavn.

Selv om alle tekstene våre er kodet i SGML, er finhetsgraden svært varierende. For en del av tekstene er bare strukturen markert, altså slike ting som side, avsnitt, linje og kapittel. Dette gjelder stort sett alle de nynorske tekstene. Bokmålstekstene er mer grundig tagget. Her er det lagt stor vekt på å få markert sjanger, overskrifter, alle elementene i dramatikk (replikker, talere, sceneanvisninger med mer) og i poesi. Jeg anbefaler alle som har tenkt å benytte tekstene i Dokumentasjonsprosjektets tekstsamling, å sette seg inn i tekstkoding. Da vil man kunne utnytte tekstene bedre og kunne legge til sin egen koding. Tekstkoding er et sentralt verktøy for en moderne edisjonsfilolog.

### ***Konvertering av språklige kortarkiver***

Den opprinnelige målsettingen for delprosjektene ved Avdeling for leksikografi var "å gjøre seddelsamlingene tilgjengelig på elektronisk form". Til sammen er det om lag 9 millioner sedler i avdelingens samlinger. En typisk ordseddel inneholder et ord i grunnform, en liten tekstbit som gir et eksempel på en bruk av ordet, og opplysninger om grammatikk, uttale og eventuelle andre forhold rundt dette eksempelet. I de siste 130 årene har ordsedler vært den tradisjonelle systematiske metoden for å samle belegg og opplysninger om bruk av ord for redigeringen av ordbøker. I de siste tiårene har imidlertid datateknikken muliggjort en mye mer effektiv oppbygning av den informasjonen som en seddelsamling representerer. Optisk lesning av tekst (OCR) og datastøttede hjelpemidler for (halv-)automatisk markering av grammatisk informasjon til ord i løpende tekst brukes nå til å bygge opp elektroniske tekstsamlinger. Konkordansprogrammer og ulike søke- og analyseprogrammer for tekst brukes til å hente ut informasjon fra disse tekstsamlingene. På denne bakgrunn var det derfor riktig å vurdere om det var hensiktsmessig å skrive av alle sedlene og lage en database over disse avskriftene.

I Bokmålsavdelingen er den alt overveiende delen av ordsedlene laget på grunnlag av hele forfatterskap (Wergeland, Bjørnson, m.m.). På disse sedlene står det en liten tekstbit med det interessante ordet understreket, grunnformen av ordet, ordklassen og opplysning om hvor tekstbiten er hentet fra. Ser vi bort fra ordklassen og grunnformen, er dette akkurat den informasjonen vi kan få ut av en såkalt KWIC-konkordans (Key Word In Context). Slike konkordanser kan lages maskinelt ut fra en elektronisk tekst. En elektronisk tekst åpner også for mange interessante anvendelser. Vi valgte derfor å bygge opp en elektronisk tekstsamling som en erstatning for 95 % av sedlene. Tekstene er SGML-tagget, slik det er beskrevet tidligere i dette kapitlet. Den eneste seddelsamlingen som er skrevet inn, er det såkalte nyordsmaterialet (200.000 sedler) for perioden før 1976. Dette materialet supplerer det allerede eksisterende elektroniske nyordsmaterialet for perioden etter 1976, hvor det er påført ekstra informasjon til hvert enkelt ord.

Gammelnorskavdelingens seddelsamling består også av sedler av KWIC-konkordans typen. Men sedlene er så systematisk bygd opp at vi her har valgt å bruke informasjonen på dem til å gjenoppbygge de originale sagaene som elektroniske tekster der hver ordform har fått markert grunnord, ordklasse og aktuell bøyingsform. Seddelarkivet er dermed blitt til det man kaller et tagget tekstkorpus. For dem som ønsker å se informasjonen som tradisjonelle sedler, er dette mulig, men det er også mulig å søke på fraser der ordene skal høre til gitte ordklasser eller være i bestemte former.

Seddelsamlingen til Norsk Ordbok skiller seg ut fra de to andre ved at det er bygd opp over en lang periode (om lag 60 år) av mange hundre frivillige uten spesiell leksikografisk utdannelse. Dette har resultert i en heterogen samling som består både av enkle sedler som for bokmål, og av sedler med mye ekstra informasjon om blant annet bruk og uttale. Den opprinnelige planen for konverteringen var at de enkle sedlene skulle frasorteres og erstattes av elektroniske tekster. Resten av sedlene skulle skrives inn og SGML-kodes. Etter at om lag 10% av sedlene var behandlet, viste det seg at det i praksis var vanskelig å sortere samlingen på en effektiv måte. Det til dels uryddige oppsettet av informasjonen på sedlene samt mye håndskrift gjorde at konverteringsarbeidet gikk for langsomt. Det ble også klart at dersom man skulle kunne stole på den elektroniske versjonen, måtte en faksimile være tilgjengelig i databasen. Vi valgte derfor å gå bort fra sorteringen og avskriften av sedlene. I stedet har vi laget en faksimiledatabase over samtlige 3 millioner nynorsksedler. Denne samlingen av faksimiler har oppslagsord, ordklasse og uttømmende kildeopplysninger som søkenøkler. Man mister på denne måten mulighetene til å søke i den løpende teksten på sedlene, men har fremdeles muligheten til å finne sedler etter grunnord, ordklasse, sted i landet, kildetype og hvem som har skrevet seddelen. Se Ore 1996 for en inngående diskusjon av fordeler og ulemper ved denne løsningen.

### ***Dokumentasjonsprosjektet og leksikalske databaser***

Med leksikalske databaser menes databaser som inneholder informasjon om ord så som deres oppbygning og bøyning (morfologi), grammatisk funksjon, mening, relativ frekvens og så videre. En leksikalsk database er altså datateknikkens svar på en ordbok. Den skiller seg fra en tradisjonell ordbok også ved at det ikke er meningen at all informasjonen skal leses av mennesker. I mange tilfeller kan leksikalske databaser inneholde informasjonen kodet til bruk i ulike språkteknologiske verktøy så som morfologiske analysatorer, syntaksanalysatorer og oversettelsesstøttesystemer. Men en leksikalsk database vil typisk kunne inneholde teksten fra en eller flere tradisjonelle ordbøker som hjelp og supplement til menneskene som bruker basen. Dokumentasjonsprosjektet har laget orddatabaser (leksikalske databaser) for bokmål,

nynorsk og gammelnorsk. Nynorskdatabasen er mest mangeartede og er her brukt som eksempel, se figur 4.



Figur 4 En skisse av den leksikalske databasen. De deler som er digitalisert eller planlagt digitalisert gjennom Dokumentasjonsprosjektet eller andre er satt i halvfet.

Nynorskdatabasen inneholder informasjon fra mange kilder, seddelarkiv, ordbøker og ordsamlinger. Blant ordbøkene kan vi nevne Aasens Norsk Ordbog Aasen 1872, Nynorskordboka Hovdenak et al. 1994 og Grunnmanuskriptet Skard et al. 1932, som er et første utkast til Norsk Ordbok fra 1930-tallet. I tillegg finner vi 20-30 ordsamlinger fra 1600-, 1700- og 1800-tallet. I tillegg til dette historiske materialet inneholder orddatabasen også bøyningsinformasjon og annen grammatisk informasjon hentet fra IBM Eng 91 og fra arbeidet med "Komputasjonelt leksikon (NorKompLeks)" ved Lingvistisk institutt, NTNU (Nordgaard 1995). Dokumentasjonsprosjektets nynorske ordbaser vil dermed være en blanding av en leksikalsk database i ordets mer tekniske betydning og en samling av elektroniske versjoner av ordbøker, ordsamlinger og annet leksikografisk bakgrunnsmateriale. Figur 4 gir en skjematisk oversikt over den nynorske leksikalske databasen slik den er, og slik den kan bli i framtiden. Den leksikalske databasen for bokmål er satt sammen på en tilsvarende måte. Gammelnorskdatabasen er foreløpig bare basert på seddelarkivene.

Et viktig punkt, som krever norsk-filologisk og helst leksikografisk ekspertise, er å etablere en kobling av grunnord mellom de ulike delbasene innenfor hvert språk, men også på tvers av språkgrensene. Den interne koblingen kompliseres av alle de ulike rettskrivingsstandardene. De eldre ordsamlingene og ordbøkene følger sin egen rettskriving. Nynorsksamlingen og grunnmanuskriptet følger ideelt sett 1938-rettskrivingen, mens det nyere materialet følger moderne rettskriving.

Sammenkoblingsarbeidet krever altså at alle ordartiklene i de ulike samlingene og verkene i tillegg til det originale oppslagsordet får påført en variant av dette som følger 1938-retskrivningen. Denne oppgaven er utenfor Dokumentasjonsprosjektets ansvarsområde, men er helt nødvendig for å få full utnyttelse av basen. Så lenge denne normaliseringen ikke er foretatt, må brukeren bruke sin egen kreative språksans for å formulere søk som får tak i oppnåelig informasjon.

## **Museumsarbeidet**

Dokumentasjonsprosjektet har delprosjekter ved Oldsaksamlingen, Universitetets Myntkabinett, Bergen Museum (Historisk Museum), ved arkeologisk avdeling ved Vitenskapsmuseet i Trondheim og ved avdelingene for arkeologi, nyere kulturhistorie og samisk etnografi ved Tromsø Museum. Det er tatt utgangspunkt i eksisterende trykte tilvektskataloger (se f.eks Universitetets Oldsaksamlings Årbok 1958-1959, utkast til tilvektskataloger, Per Fetts "Førhistoriske minne" (se f.eks Fett 1960), som er en katalogserie over fornminner og gjenstander funnet i de tre vestlandsfylkene, samt en del kortkataloger (noe over 300.000 kort). Hovedmålet for disse delprosjektene har vært å etablere databaser over gjenstandssamlingene. I tillegg til det rent gjenstandsorienterte materialet er det laget en elektronisk faksimileversjon av det såkalte topografiske arkivet ved Historisk museum i Bergen. Dette er et brev- og saksarkiv ordnet etter gårds- og bruksnummer. Det er også laget fotodatabaser av utvalgte deler (40.000 bilder) av fotosamlingene til avdelingene for nyere kulturhistorie og samisk etnografi ved Tromsø Museum. Disse ekstra delprosjektene har vært "pilotprosjekter" i full skala med sikte på å få erfaring i hvordan de gjenstandsorienterte databasene kan utvides til et komplett informasjonssystem for museene. Dokumentasjonsprosjektets museumsdatabaser fremstår derfor i dag som svært gjenstandsfikserte, og bare for noen museer inneholder de annen informasjon.

Et komplett informasjonssystem for arkeologi bør imidlertid også inneholde registre over registrerte og slettede fornminner (gravhauger, boplasser o.l). Arbeidet med å etablere den typen registre har pågått i mange år (på Vestlandet siden før den andre verdenskrig). Dokumentasjonsprosjektet tok i 1991 over driften av databasen med de såkalte ØK-registreringene (Økonomisk kartverk) av faste fornminner på Østlandet. Dette og tilsvarende arkiver er i henhold til lov om kulturminner helt sentrale i forbindelse med all arealplanlegging i landet. Ansvar for ØK-registrene ble i 1993 overført til Riksantikvaren. Databasene inngår imidlertid i det felles arkeologiske gjenstands- og fornminnesystemet. Et slikt register dekker blant annet fylkeskommunenes behov i forbindelse med den lovpålagte plikten til å utrede mulige fornminner i forbindelse med arealplanlegging ved alle utbyggingsprosjekter. Et felles fornminneregister og gjenstandsregister har faktisk vært planlagt siden 30-årene, men har av ulike årsaker ikke latt seg gjennomføre i praksis. For universitetsmuseene garanterer registeret en felles tilgang til gjenstands- og fornminnedata som er av stor betydning i forskning og formidling. Deler av materialet er av sikkerhetsmessige og kommersielle årsaker passordbeskyttet.

### ***Metoder og gjennomføring av arbeidet***

Det arkeologiske kildematerialet består av arkivkortsamlinger, brev- og saksarkiv (kalt Topografisk arkiv etter måten det er ordnet på), trykte og håndskrevne tilvektstprotokoller samt fotografier. De arkeologiske kortsamlingene er enklere å konvertere enn de språklige kortsamlingene siden de ikke inneholder lydskrift og det sjelden er brukt annet enn det vanlige norske alfabetet. Vi har skrevet inn kortene og bygd opp en relasjonsdatabase (se under). Et Topografisk arkiv inneholder alt fra svært

store tegninger og kalkeringer til små brevark. Erfaringene fra arbeidet med det tidligere omtalte nynorskarkivet, tilsa at det eneste forsvarlige var å lage et faksimilearkiv. Alle dokumentene i arkivet er derfor blitt skannet og lagret som elektroniske bilder. For å kunne finne frem enkeltdokumenter er det laget en katalog-database der det for hvert bilde er lagret de mest sentrale opplysningene som gårds- og bruksnummer, adressat, avsender, type dokument m.m. For brukeren virker det hele som et elektronisk dokumentarkiv der man kan hente frem faksimiler av dokumentene i det originale arkivet. Men det er altså bare mulig å søke etter visse nøkler og ikke i selve dokumentteksten. Det fotografiske materialet er behandlet på en tilsvarende måte. En database med elektroniske fotografier er jo også en faksimiledatabase.

De arkeologiske museene i Norge skiller seg fra de fleste andre slike museer ved at det med jevne mellomrom utgis såkalte tilvekstkataloger med opplysninger om nye gjenstander ved museene. Denne praksisen har vært fulgt i over 140 år. De trykte katalogene supplert med upubliserte håndskrevne varianter gir til sammen det mest komplette bildet av gjenstandssamlingene. På den annen side er katalogene blitt til over nesten 200 år. I løpet av denne tiden har det skjedd en svært stor endring i arkeologisk nomenklatur og gjenstandsbeskrivelse. Espen Uleberg gir interessante eksempler på teoriavhengig bruk av gjenstandsbetegnelser i Holmen og Uleberg 1996. I tillegg har skriftspråket i Norge endret seg fra dansk i 1829 til moderne bokmål og nynorsk i vår tid med en lang rekke mellomstadier. Det er på denne bakgrunn ikke å undres over at fagmiljøet ved Oldsaksamlingen i 1990 mente at den beste måten å bygge opp en gjenstandsbase på bakgrunn av tilvekstkatalogene på, var å la en eller flere kvalifiserte personer skrive inn opplysningene i materialet i et ferdiglaget databaseskjema, slik det er blitt gjort med arkivkortsamlingene Ore 1991. Dokumentasjonsprosjektet var imidlertid basert på størst mulig bruk av ufaglærte engasjert gjennom ulike former for sysselsettingstiltak. Det var derfor uaktuelt å benytte denne metoden. Vi valgte i stedet å benytte den samme metoden som for det andre tekstlige materialet i prosjektet (litteratur, Diplomatarium Norvegicum m.m), altså konvertering til SGML-tagget tekst. Dette skiller seg ut fra hva som har vært vanlig i museumsbransjen. I de arkeologiske katalogene har det vært viktig å markere selve innholdet slik at informasjonen kan hentes ut til en relasjonsdatabase. Taggingen er dermed noe annerledes en det vi har sett eksempler på så langt i dette kapitlet. En slik innholdstaggning er viktig også for historikere og språkforskere. Jeg har derfor valgt å bruke noe ekstra plass på arbeidet med de arkeologiske katalogene.

### ***Bruk av SGML i konvertering av gjenstands- og fornminnekatalogene***

Arbeidet med å konvertere arkeologisk tekstmateriale startet i august 1992 da det ble etablert et delprosjekt ved Historisk museum i Bergen. For å komme i gang ble det lagt vekt på å finne et systematisk materiale som kunne egne seg for konvertering av ufaglært arbeidskraft. Valget falt på Per Fetts forbillidige serie "Førhistoriske minne på Vestlandet" (se f.eks. Fett 1960 som er en oversikt over gjenstander og fornminner i fylkene Hordaland, Sogn og Fjordane samt Sunnmøre). Oversikten er ordnet etter matrikkelen og ajourført opp til 1976 da Per Fett gikk av med pensjon.

I korthet foregikk arbeidet ved at heftene ble lest optisk (OCR) og korrekturlest. Deretter ble teksten kodet ved at gjenstands- og fornminne-beskrivelser ble markert i teksten ved hjelp av SGML-koding (se under). Kodeprosessen var i utgangspunktet ingen suksess, og det krevdes en del dyrekjøpt erfaring for å få den til å gå glatt. Men arbeidet med Fett-heftene skaffet oss den nødvendige erfaringen. Arbeidet med tilvekstkatalogene ved alle de fire universitetsmuseene fulgt den malen som ble satt opp gjennom arbeidet med Fett-heftene.

Den vanlige databasetankegangen i norske museer har vært å definere felter enten som et feltbibliotek eller som ferdige feltskjemaer. Filosofien er altså å overføre informasjon fra eksisterende kortkataloger til et ferdiglaget skjema. Vår metode er i utgangspunktet en på sett og vis omvendt prosess. En SGML-koding av en tekst er å projisere et skjema inn i en løpende tekst. Vi skal prøve å klargjøre dette nedenfor.

**ORIGINALTEKST**

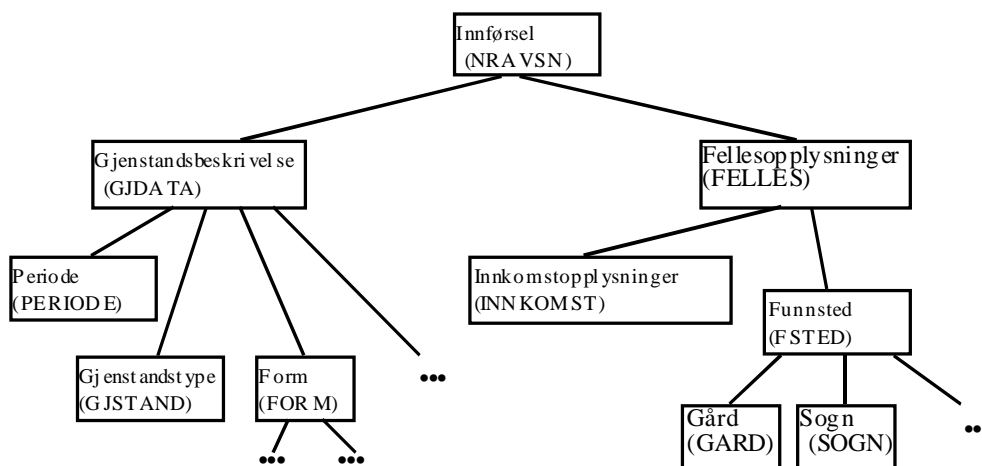
30208. Pilespiss av jern fra vikingtiden, som R. 539 , men med avfasettering av bladets midtparti istedenfor midtribbe, jvf. 537. Ganske bra bevart. St. lengde 13,8 cm, bladets st. bredde 2,1 cm.F. ca. 100 m ovenfor Espa st. "på en plass under Råvangen" (g.nr. 179, b.nr. 2), Tangen s., Stange pgd., Hedmark. Gave. Innsendt av Odd Stensrud, Espa st.

**MERKET TEKST:**

```
<NRAVSN><MUSEUMSNR NRID="30208">30208.<MUSEUMSNR><GJDATA><GJSTAND>
Pilespiss </GJSTAND> av <MAT>jern</MAT> fra <PERIODE>vikingtiden</PERIODE>, <FORM>
som<LITTRREF>R.<FIGNR> 539</FIGNR> </LITTRREF>, men med avfasettering av bladets midtparti
istedenfor midtribbe, jvf. <SENR>537</SENR></FORM>. Ganske bra bevart. <MAAL>St. lengde 13,8
cm</MAAL>, <MAAL>bladets st. bredde 2,1 cm.</MAAL></GJDATA>
<FELLES>F.<FSTED>ca. 100 m ovenfor Espa st. "på en plass under <GARD><GNAVN>Råvangen </
GNAVN>"(g.nr.<GNR> 179</GNR>, b.nr.<BNR> 2</BNR>)</GARD>, <SOGN>Tangen</SOGN>
s.,<PGJ>Stange </PGJ>pgd., <FYLKE>Hedmark</FYLKE></FSTED>. <INNKOMST>Gave. Innsendt
av <INNKVED>Odd Stensrud, Espa st</INNKVED></INNKOMST>
</FELLES> </NRAVSN>
```

Figur 5 Eksempel på kodet tekst

Figur 5 viser den originale teksten i en innførsel fra Oldsaksamlingens tilvekstkatalog og den tilsvarende innførselen ferdig kodet. Selve innførselen er blitt omsluttet av et taggpar 'NRAVSN'. Vi kan også legge merke til at inne i innførselen er "pilespiss av jern" blitt kodet '<GJDATA><GJSTAND>Pilespiss</GJSTAND> av <MAT>jern</MAT></GJDATA>'. Dette viser hvorledes kodingen kan brukes til å gruppere informasjonen. Tilsvarende koding finner vi for stedsangivelsen. Det at taggene står inni hverandre slik, gjør at de danner et hierarki slik som det er vist i figuren under. Med en teknisk term kan man si at en SGML-kodet tekst sammen med et fritekstsøkesystem utgjør det som kalles en hierarkisk database.



Figur 6 Taggehierarkiet i tilvekstkatalogen (noe forenklet)

Når en skal kode tekst på denne måten, må det lages en "Document Type Definition" (DTD), som er en grammatikk for eller formell beskrivelse av den tillatte kodestrukturen. Tilvekstkatalogene til Oldsaksamlingen skiller seg fra Per Fetts protokoller ved det at de er blitt til gjennom lang tid og er skrevet av mange personer. Det viste seg derfor vanskelig å lage en enkelt grammatikk som skulle passe til strukturen i alle mulige tilvekstkataloger, binde relevante opplysninger sammen (som vist over) og være anvendelig for ufaglært arbeidskraft. Vår erfaring er at det sistnevnte punktet er viktigst. Det løser på mange måter det førstnevnte idet dokumentdefinisjonen må tillate en temmelig fri bruk av taggene. Denne friheten til å sette tagger nær sagt hvor man vil, kan av og til føre til at informasjon som egentlig hører sammen, havner i ulike "tagparenteser". Dette vil lett kunne skje i innførsler med mange gjenstander fra ulike perioder. Da er ofte gjenstandenes form og utseende beskrevet ett sted og periodene et annet. Med vår SGML-koding kan det da kunne bli umulig å koble riktig gjenstand til riktig periode. Dette er for så vidt ikke en svakhet ved SGML. Det er mulig å foreta denne koblingen ved hjelp av såkalte "id" og "idref", men det ville kreve uforholdsmessig mye arbeid i forhold til resultatet. I vårt system får derfor alle gjenstandene i en slik innførsel alle de aktuelle periodene knyttet til seg. Dette vil resultere i en del ekstra treff ved søking. Men det skal understrekes at problemet er lite, siden tilvekstkataloger som oftest er ført svært systematisk.

### ***Hva oppnår man med en SGML-tagget tekst?***

Den korte gjennomgangen over viser at en SGML-koding trer tabeller over teksten. Men som vi så, er det ikke all informasjon som hører sammen, som blir rubrisert sammen. Dette viser seg å ha lite å si for de fleste praktiske formål. Arkeologiske tilvekstprotokoller representerer en ganske fastlagt genre, hvor formålet er å uttrykke seg klart. Det er derfor mulig å foreta en maskinell tolkning av de taggede tilvekstkatalogene og med stor grad av nøyaktighet gruppere informasjonen riktig. En slik tolkning brukes i det fritekstsystemet vi har lagt tilvekstkatalogene inn i. For den jevne bruker er det vanskelig å se forskjell på dette systemet og en relasjonsdatabase.

Det er imidlertid ett problem som nå melder seg med full tyngde. Våre registratorer og fagkonsulenter har fått beskjed om å gjenskape de originale tilvekstkatalogene i elektronisk form og å markere de enkelte informasjonskategoriene som forklart ovenfor. Det foretas ikke rettelser i den originale teksten. Vi kan dermed hente ut komplette lister over termene som har vært brukt i tilvekstkatalogene ved de norske arkeologiske museene de siste 200 årene. På den annen side kan man ikke få fullgode søk dersom man ikke har oversikt over alle disse termene. For å få tilslag på alle slags spenner må man blant annet søke på "fibula", "spænde", "spende" og "spenne" samt alle slags bøyde former av disse variantene. Blant materialtermer for den samme typen gjenstand finner vi "bronze", "bronce" og "bronse".

Denne mangelen på standardiserte termer er blitt kritisert. Det er implisitt blitt hevdet at vi egentlig ikke har laget noe felles gjenstandssystem for museene. Som et svar på denne kritikken vil jeg påpeke to forhold. For det første har vi nå muligheten til å ta ut komplette lister over termene og dermed foreta en fullstendig analyse av termbruken. En slik analyse er i seg selv av forskningshistorisk interesse, og er etter min mening helt nødvendig for en fornuftig standardisering av termene. Det andre momentet er at dersom en foretok en standardisering av betegnelsene i tilvekstmaterialet ved hjelp av f.eks. klassifiseringssystemet Outline, vil dette representere en rekatalogisering av materialet.



### *Et kombinert system for arkeologi*

Hovedmålet for museumsarbeidet i Dokumentasjonsprosjektet var å skape det elektroniske datagrunnlaget for å kunne bygge opp komplette informasjonssystemer for de arkeologiske universitetsmuseene. Slike informasjonssystemer vil omfatte alt fra utgravning via katalogisering og konservering til utstillingsarbeid. De nåværende databasene er ikke så omfattende. Men de inneholder svært mye av den (arkeologiske) gjenstandsorienterte informasjonen i museene, og er godt egnet til å få svar på spørsmål om gjenstandsmaterialet og om hvor og når det er funnet. Koblet sammen med Dokumentasjonsprosjektets andre databaser, særlig steds- og navnedatabasene samt Diplomatarium Norvegicum, gir basene helt nye muligheter til å sammenstille materiale fra historisk tid.

Det har blitt lagt ned et stort arbeid i å analysere virksomheten ved museene med henblikk på å konstruere en generell datamodell for informasjonsflyten i museene. Sentralt i modellen står et observasjons- eller hendelsesbegrep. Dette er en generalisering av hendelsesbegrepet som er brukt av Dokumentationsavdelingen ved Nationalmuseet i København Røll 1992 i dens arbeid med å lage arkeologiske gjenstandsdata-baser. Dette hendelsesbegrepet har igjen sitt utspring i det datalingvistene kaller "event semantics". De danske arkeologene bruker hendelsesbegrepet for å kunne modellere en arkeologisk gjenstands historie fra den ble skapt. Vi har generalisert begrepet og bruker det mest for å kunne fortelle hvilke hendelser som har produsert de dataene som ligger i databasene. Noen eksempler på slike hendelser er en arkeologisk utgravning, en gjenstandskonservering eller en katalogisering av gjenstander ved et museum.

Tilvekstkatalogene kan sies å være en slags første katalogisering av gjenstandsmaterialet i museene. Ved hjelp av taggene i teksten er data om de enkelte museumsnumrene (gjenstandene) blitt lastet over i en relasjonsdatabase. Resultatet er en database med visse unøyaktigheter på grunn av taggingen (se ovenfor ) og med delvis foreldet bruk av fagtermer. Til tross for slike mangler representerer denne databasen en komplett katalogisering av gjenstandsmaterialet, og den vil bli stående inntil det blir foretatt en magasinrevisjon.

For noen av museenes vedkommende er det allerede utført magasinrevisjoner (nykatalogisering) for deler av gjenstandssamlingen. Resultatet ligger i noen av de arkivkortsamlingene som er skrevet inn. I disse tilfellene er informasjonen fra revisjonene også lagt inn i databasene. Det er viktig å legge merke til at den gamle informasjonen (i dette tilfellet fra tilvekstkatalogen) ikke blir fjernet. Dette vil også være tilfellet når informasjonen fra fremtidige magasinrevisjoner og annen rekatalogisering legges inn i databasen. Det er mulig å hente frem all informasjon som ligger lagret om gjenstandene, og ikke bare den aller nyeste. På denne måten sikres faghistorien, slik denne kommer til uttrykk i museumsdatabasene.

De taggedede tilvekst- og fornminnekatalogene er ikke lagt inn i denne relasjonsdatabasen. De er ferdige publikasjoner og er lagret separat i et friteksts-system. Men ved å bruke museumsnumrene som søkenøkler (identifikatorer) er det en toveis forbindelse mellom dette friteksts-systemet og relasjonsdatabasene, og man kan hoppe fra det ene til det andre ved et museklikk. Tilsvarende er det en kobling mellom det digitaliserte topografiske arkivet og de andre databasene. Grovt sett kan vi si at datasystemet består av en foto- og faksimiledatabase, en fritekstdatabase og en vanlig relasjonsdatabase.

## **Hva kjennetegner Dokumentasjonsprosjektets databaser?**

Dokumentasjonsprosjektet har arbeidet med materiale av svært ulik karakter. Det omfatter arkiver, kataloger, tekstsamlinger og litterære verk. De fleste database-systemer er det man kan kalle kataloger. De inneholder referanseopplysninger om objekter som kan observeres. Dokumentasjonsprosjektets databaser er noe annerledes, idet de til dels inneholder objektene selv eller replikaer av disse. Man kan bruke flere metaforer for å forklare hvordan innholdet i databasene skal forstås.

### ***Et bibliotek***

Det enkleste mentale bildet er biblioteket. Dette bildet passer godt for de store mengdene litterære verker som er konvertert til elektronisk tekst. Men det kan kanskje synes litt fjernt å bruke en slik metafor for å forstå den elektroniske versjonen av det topografiske arkivet ved Historiske Museum i Bergen, som er et brev- og rapportarkiv sortert etter gårds- og bruksnummer. Det er mulig å forestille seg at et slikt arkiv en gang kunne ha blitt publisert i en faksimileutgave trykt på papir. Topografisk arkiv ville riktignok ha blitt en samling på 150 bind à 1.000 sider og ruvet godt i bokhyllen. Dokumentasjonsprosjektet har laget en faksimileutgave av dette arkivet. Den er riktignok elektronisk og vil bare beslaglegge 15 hyllecentimeter om den ble publisert på CD-ROM. Materialet skal derfor på det mest generelle nivået kunne sees som et bibliotek med en rekke verker, kataloger og faksimileutgaver. I papirutgave ville dette ha bestått av i underkant av tre tusen tykke bind. Denne biblioteksmetaforen ligger til grunn for måten dataene er blitt konvertert til elektronisk form på.

Noen av tekstene er fra forfatterens side ment som oversikter og kataloger. Andre er forfattet for å bekrefte avtaler, dommer, utnevelser med mer. De litterære tekstene kan være forfattet med helt andre hensikter. For alle tekstene gjelder det at de er inkludert i systemet fordi de brukes av humanistiske forskere som informasjonskilde. Den informasjonen forskerne er interessert i, er noen ganger den som forfatteren ønsket å formidle, men ofte ikke. Et eksempel på det siste er språkforskernes bruk av middelalderbrev som kilde for informasjon om gammelnorsk syntaks. Siden vi som laget systemet ikke kan forutsi alle forskeres behov og analyseønsker, er det laget som et bibliotek der det er mulig for en forsker å hente frem utvalgte deler av materialet til eget bruk.

### ***Elektroniske bøker***

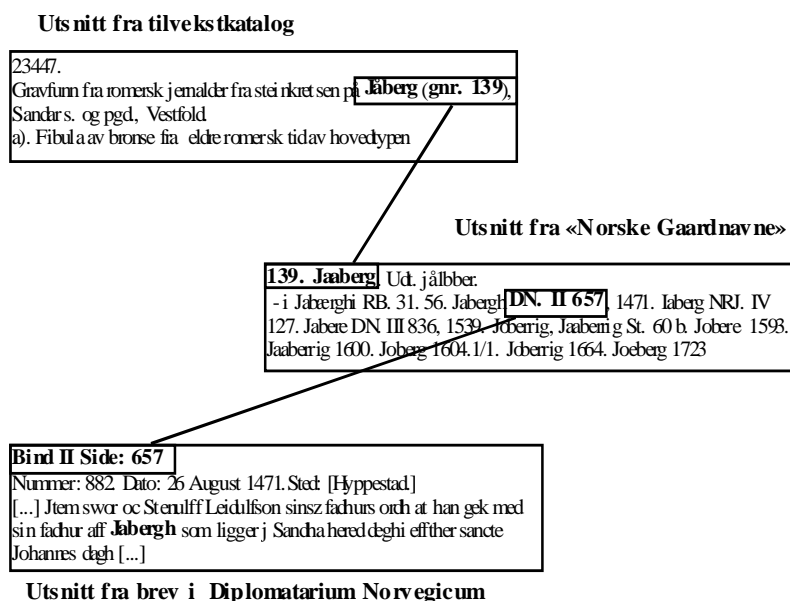
Det neste skrittet inn i materialet er å åpne de enkelte verkene i dette elektroniske biblioteket. Med dette mener jeg at systemet er laget slik at de enkelte verkene i biblioteket er utstyrt med søkemuligheter av ulike slag. Faksimileutgaven av det topografiske arkivet består av en arkivindeks eller katalog der inngangene er ulike arkivopplysninger så som sender og mottager av brev, dato for utstedelse og dato for innlemmelse i arkivet. Tilsvarende er det for de andre faksimilesamlingene som nynorsksamlingen og deler av navnematerialet. For de andre verkene vil vi også få tilgang til søkesider som svarer til registre og katalogbind i papirbibliotekene. Ved å bruke disse søkemulighetene kan brukeren altså få hentet frem utvalgte deler av de enkelte verkene. Så langt er altså databasene en elektronisk utgivelse av de originale katalogene, samlingene og verkene utstyrt med svært omfattende registre og indekser. Det er ikke tilført ny informasjon. Databasene gir bare en helt annen mulighet for tilgang til den opprinnelige informasjonen og åpner for anvendelse av kvantitative metoder.

### Et tverrvitenskapelig oppslagsverk

En vesentlig hensikt med Dokumentasjonsprosjektet er muligheten for tverrfaglige søk. Men hva innebærer dette? Vindusgrensesnittet i moderne operativsystemer gjør det mulig å ha flere databaser åpne samtidig og sammenlikne resultatene. Skal man kunne snakke om et tverrfaglig system, må det imidlertid være noe mer. Våre databaser har to muligheter for tverrfaglige søk. Den første kan vi kalle systemlagde henvisninger eller hypertextforbindelser. Den andre kan vi kalle brukerstyrte "en bloc"-søk. For å forklare disse begrepene og hvordan de virker i våre databaser, skal jeg bruke *Diplomatarium Norvegicum*, tilvekstkatalogene ved de arkeologiske museene og Oluf Ryghs *Norske Gaardnavne* Rygh 1887-1936. De to første er beskrevet tidligere. *Norske Gaardnavne* trenger en kort introduksjon.

*Norske Gaardnavne* er gjennomgang av navnene på 45.000 norske matrikelgårder trykt i 17 bind. For hvert navn er det gitt opplysninger om uttale, etymologi og en lang rekke kildehenvisninger til eldre kilder, for eksempel *Diplomatarium Norvegicum*, der navnet forekommer. Verket er basert på Oluf Ryghs arbeid i forbindelse med den nye eiendomsoversikten (matrikkelen) som ble lagd for Norge i 1880-årene, men var ferdig først lenge etter Ryghs død. *Norske Gaardnavne* er ordnet etter matrikkelen, det vil si etter fylke (amt), herred og gårds- og bruksnummer.

La oss først se hva vi mener med systemlagde forbindelser. Sett at en bruker av våre databaser har gjort et søk i tilveksten og fått frem innførselen om et gravfunn gjort på Jåberg i Vestfold (se figur 7). Det er da mulig å klikke på gårdsnavnet for å få opp flere opplysninger om gården, blant annet om den er nevnt i *Norske Gaardnavne*. Når artikkelen i *Norske Gaardnavne* kommer frem på skjermen, føles det naturlig å kunne klikke på en kildehenvisning i *Norske Gaardnavne* og hente frem diplomtet det vises til. I begge tilfeller gjør vi noe som kan sies å være i samsvar med de originale tekstene. De elektroniske henvisningene, eller hyperlenkene, som de også kalles, er laget av systemet på bakgrunn av SGML-kodingen. Bak kulissene blir det foretatt et fritekstsøk i henholdsvis den elektroniske *Norske Gaardnavne* og det elektroniske *Diplomatarium Norvegicum*.



Figur 7 Systemgenererte henvisninger mellom ulike verk

Sammenstillingen av de tre verkene i ett søkesystem effektiviserer slike parallelle søk, slik at det er mulig å formulere enbloksøk av typen "For alle gårder i Vestfold med gravfunn finn de middelalderdiplomene der gården er nevnt ved å bruke Oluf Ryghs *Norske Gaardnavne*."

Dersom dette søket skulle formuleres ved separate søk, måtte vi først finne alle gravfunn i Vestfold og lage en liste over gårdsnavnene der disse var oppgitt. Deretter måtte vi lete frem de riktige artiklene i *Norske Gaardnavne* og lage en liste over alle henvisningene til Diplomatariet. Til sist måtte vi søke frem de aktuelle diplomtekstene. Det samlede systemet automatiserer denne koblingen og bruker de tre verkene som én database.

Man kunne kanskje i effektiviseringens navn bemerke at det da burde holde å formulere søket som "For alle gårder i Vestfold med gravfunn finn de middelalderdiplomene der gården er nevnt". Formuleringen "ved å bruke Oluf Ryghs *Norske Gaardnavne*" kan synes pedantisk. Men fra et kildekritisk synspunkt er dette absolutt nødvendig. Brukeren må kunne få vite eller bestemme hvor påstanden om at det er en kobling fra en moderne gård i Vestfold til et brev skrevet i 1471. Dette er særlig viktig fordi et verk som *Norske Gaardnavne* er av varierende kvalitet og verken er absolutt korrekt eller fullstendig.

Av samme årsak skal man være forsiktig med å snu henvisninger. Det er nemlig teknisk sett lett å vende forbindelsene mellom de tre verkene, slik at vi for et gitt diplom kan få alle innførsler i *Norske Gaardnavne* som har referanse til diplom, og alle innførsler i tilvektskatalogene. Forbindelsen mellom Diplomatarium Norvegicum og tilvektskatalogene gjennom *Norske Gaardnavne* går tilsynelatende begge veier. Men i det nevnte eksempelet er det intet i den originale teksten i Diplomatarium Norvegicum som gir oss et hint om at vi kan gå til *Norske Gaardnavne* og videre til tilvektskatalogene. Så her føyer systemet til alternativet "Vis innførsler i *Norske Gaardnavne* som henviser til dette diplom". Ville det ikke også være kjekt å ha alternativet "Vis alle gjenstander funnet på stedene som er omtalt i dette diplom"? Svaret er selvfølgelig ja. Men hvordan skal vi finne disse gjenstandene? Vi kan selvfølgelig la systemet bruke *Norske Gaardnavne* som forbindelse til tilvektskatalogen, men da må dette fremgå.

Mye av det som inngår i våre databaser, er altså i utgangspunktet laget som kataloger eller systematiske sammenstillinger. Når disse tekstene gjøres til søkbare elektroniske bøker, lager vi i praksis en modellverden som bygger på det intellektuelle arbeidet til forfatterne av disse tekstene. Påliteligheten av informasjonen man kan få ut av denne modellverdenen, er avhengig av forfatterens arbeid. Så lenge vi følger eksplisitte forbindelser i de enkelte tekstene er det faglig sett uproblematisk å koble flere av disse modellverdenene sammen. Men når vi skjuler forbindelsen eller definerer nye koblinger ved å ta en skjult snarvei, kan resultatet bli et system der det er mulig å trekke slutninger som ikke kan begrunnes i de originale kildene. Det er derfor viktig at brukeren har en forståelse av hva som foregår bak kulissene - altså hvilken informasjon som er trukket direkte ut av originalmaterialet, og hvilken informasjon som er lagt til gjennom arbeidet med å lage databasene. I siste instans er dette et spørsmål om tillit. Men vi har også bestrebet oss på å lage systemet etterprøvbart. Dersom systemet alltid skjulte denne typen fakta for brukerne, ville systemet ikke lenger fylle sin vitenskapelige hensikt.

## Når og hvor?

Dette kapittelet avsluttes med en diskusjon av hvordan vi forholder oss til kjernedataene tid og sted. Om ikke noe annet binder ulike fagområder sammen, vil alltid tid og sted være koblingspunkter. Men ved nærmere ettersyn viser det seg at tids- og stedsangivelser ikke er uproblematiske. Begge deler har en tendens til å bli uttrykt på en lite presis og ofte relativ måte. Diskusjonen nedenfor beskriver hvordan vi har løst dette i våre systemer, og formulerer noen tanker om hvordan dette kan gjøres i framtiden.

Av de to kategoriene er angivelse av tid og dato det som er best tatt hånd om av moderne databasesystemer. De fleste slike systemer kommer med et innebygd system for å regne ut antall dager mellom to gitte datoer o.l. For historisk orienterte databaser må man imidlertid ta hensyn til de ulike kalenderne som har vært i bruk i det kristne Europa, nemlig den julianske og den gregorianske. Den romersk-katolske kirken gikk over fra den julianske til den gregorianske på slutten av 1500-tallet, det protestantiske Danmark-Norge holdt på den julianske til år 1700, mens den gresk-ortodokse kirke fremdeles holder på den julianske kalenderen. Det yngste diplommet i *Diplomatarium Norvegicum* som er utstedt i Roma i 1570-årene er antageligvis datert etter den gregoriansk kalenderen. Det betyr lite i dette konkrete tilfellet. Men i en database over historiske kildekrifter må man være klar over forskjellene dersom man ønsker nøyaktige søk. Den databasetekniske løsningen på dette problemet er å legge inn tileggsinformasjon om hvilken kalender dokumentet følger. Det er foreløpig ikke gjort i våre databaser.

Et annet problem er upresis datering av typen "1349" eller "sommeren 1430". Vår løsning er å la et tidspunkt være et tidsrom. Tidsangivelsen "1349" er dermed perioden 1.1-31.12.1349. Tilsvarende er "april 1540" være perioden 1.4-30.4-1540. Dateringer av typen "sommeren 1430" kan uttrykkes på tilsvarende måte ved at man definerer hva som menes med "sommer". Dette blir noe tilfeldig, idet både hele sommerhalvåret (fra første sommerdag til første vinterdag på primstaven) og juni-august kan være fornuftige tolkninger. Ønsker man å være på den trygge siden bør man velge det største tidsrommet. Dette vil kunne medføre at en del ikke overlappende hendelser tilsynelatende har skjedd samtidig. Slike feilslutninger kan løses ved at man markerer nøyaktighetsgraden for hver enkel datering. Et spørsmål om to dokumenter er utstedt i samme måned, vil da kunne gi svaret "Dette er ukjent, men de er utstedt i samme år". Dette svarer til hva en i fysikken kaller antall gjeldende desimaler eller nøyaktigheten ved ulike målinger. Den minst nøyaktige målingen bestemmer nøyaktighetsgraden for hele settet.

Når (historisk) stedsinformasjon skal lagres systematisk, kan man enten bare lagre stedsnavnene slik de forekommer i teksten, eller koble dem til et forhåndslagd hierarki. Et eksempel på det siste er å kode alle stedsnavnene med de moderne kommune- og fylkesnumrene. Et sted i Aker kommune (slått sammen med Oslo i 1948) får dermed kommunekode 0301 som er Oslos kommunenummer. Denne metoden er mye brukt med den variasjon at fagmiljøene har valgt en noe eldre inndeling, eksempelvis fra 1946 (leksikografi) eller fra rundt 1900 (navnegransking). Arkeologene på Vestlandet har valgt å holde seg til den til enhver tid gjeldende kommuneinndeling, mens arkeologene på Østlandet bruker Kirkens inndeling i sogn og prestegjeld, siden denne er mer stabil. Vi har altså tilsvarende problemer i de ulike administrative inndelingene av landet som vi har med ulike tidsregninger og kalendere. I tillegg til de administrative hierarkiene kommer betegnelser på enkeltstående områder (Viken, Hardangervidda, Oslofjorden) og naturformasjoner som fjell, åser, teiger, bekker. Den fysiske

utstrekningen til slike områder og formasjoner er som oftest noe uklar. Det samme gjelder for øvrig eldre også administrative enheter. Det er ikke alltid så greit å finne de nøyaktige grensene til et fogderi på 1700-tallet.

I det materialet Dokumentasjonsprosjektet har arbeidet med, brukes det en lang rekke moderne og historiske områdeinndelinger og betegnelser. Der vi har hatt kapasitet til det, er slike betegnelser merket i teksten. Men det er ikke påført noen moderne referansekoder av den typen som er nevnt ovenfor. I stedet har vi utviklet et rammeverk for å lagre stedsinformasjon. I dette rammeverket opererer vi grovt sett med to begreper, "administrativ enhet" og "sted". De administrative enhetene er koblet sammen i administrative hierarkier, f.eks. fylke-kommune-hierarkiet. For å kunne holde orden på hvor informasjonen om slike koblinger kommer fra, bruker vi en variant av hendelsesbegrepet i de arkeologiske databasene. I stedsdatabasen snakker vi om en observasjon. En slik observasjon er på sett og vis en kildeangivelse. Når gårder, herreder med mer i Oluf Ryghs *Norske Gaardnavne* lastes inn i stedsdatabasen, blir det koblet en merkelapp eller observasjon til hver enhet som forteller at denne enheten er laget på grunnlag av informasjonen i *Norske Gaardnavne*. I tillegg til *Norske Gaardnavne* har vi også lagt inn skattematrikkelen (gårds- og bruksoversikten) fra 1950. Gårdene i denne ligger lagret som enheter med merkelappen "hentet fra skattematrikkelen i 1950". I utgangspunktet er det ikke laget noen kobling mellom *Norske Gaardnavne*-hierarkiet (amt-herred-sogn-gård) fra 1880-årene og hierarkiet i 1950-matrikkelen (fylke-kommune-gård). Men for at stedsdatabasen skal bli brukbar, bør det gjøres en slik kobling eller lenking. Alle disse lenkene må da få en merkelapp som sier hvem eller hva som påstår sammenhengen mellom enhetene (dvs. gårdene). Matrikler og folketellinger gir oss hele hierarkier på bestemte tidspunkt, mens lenkingen gir den tidsmessige forbindelsen. Statens kartverk bruker en tilsvarende modell for helt moderne tid. Men der er man mer fokusert på at en administrativ enhet eksisterer. Vi forholder oss bare til det vi kan finne om enheten i skriftlige kilder. Rammeverket for stedsdata inneholder også enkeltsteder av typen "Hardangervidda" og koblinger til dagens kartkoordinater der stedets eller enhetens utstrekning er kjent.

I årene som kommer, vil antageligvis informasjonsteknologien utvikle seg videre, slik at verken lagringskapasiteten eller overføringshastigheten vil være noen begrensning for den typen data vi finner i humanistiske arkiver og samlinger. Det som vil forsette å være et til dels utfordrende problem, er å finne frem i datamassen og å koble de dataene som hører sammen. Det finnes i den forbindelse innenfor all oppbygging av elektroniske tjenester for (historiske orientert) humaniora et behov som er et system for historisk orientert stedsinformasjon slik det er beskrevet over. Et slikt rammeverk vil aldri bli ferdig utfylt. Men det er lagd slik at ulike forskere kan bidra med deler og dermed gjøre det gradvis bedre. Systemet for historisk stedsinformasjon kan altså langt på vei bygges opp innenfor de humanistiske fagmiljøene selv. Et slikt system er egentlig et elektronisk historisk atlas, og vil binde sammen de ulike dataene fra Dokumentasjonsprosjektet, IT-prosjekter innenfor kulturminnevernområdet og andre historisk orienterte humanistiske IT-prosjekter. For å få registrert og presentert data på en fornuftig måte trengs det også gode elektroniske kart. Disse må leveres av profesjonelle til en overkommelig pris, noe som dessverre ikke er tilfellet i dag.

## **Oppsummering**

Denne artikkelen har vært en rask gjennomgang av IT-strategien og metodene i Dokumentasjonsprosjektet. Den er altfor kort til å kunne fungere som en lærebok i humanistisk IT-metode. Men det denne gjennomgangen viser, er at de ulike humanistiske fagfeltene har svært sammenfallende behov innenfor IT-baserte metoder.

Ulike metoder for elektronisk tekstbehandling (koding, analyse og informasjons-gjenfinning i fritekst) er sentrale i samtlige fagfelt. Dette avspeiler de humanistiske fagenes sterke tilknytning til tekststudier. Noe som kanskje ikke kommer så klart frem i Dokumentasjonsprosjektets arbeid, er det felles behovet de humanistiske fagene har for elektronisk bildebehandling. Datamaskinenes evne til å lagre, analysere og bearbeide bilder på ulike måter er allerede svært nyttig i alt fra dokumentstudier i gammelnorsk filologi til stilstudier i kunsthistorie. Elektronisk lagring av lyd er heller ikke drøftet i dette kapittelet. Dette skyldes dels at digital lagring allerede er den alminnelige måten for å lagre lyd på, dels at Dokumentasjonsprosjektet har vært orientert mot tekst- og bildemateriale. De språkavdelingene som har deltatt i Dokumentasjonsprosjektet, har riktignok store mengder lydbånd, men den tradisjonelle filologiske presentasjonsformen for talemateriale er skrift. Det er å håpe at dette store lyd materialet, sammen med lyd materialet i (folke)musikkavdelingene, blir digitalisert, slik at også det kan bli tilgjengelig for ulike datamaskinelle analyse- og presentasjonsverktøy.

For å kunne anvende datamaskinene er det viktig å ha store mengder tekst og bilder elektronisk tilgjengelig. Dokumentasjonsprosjektet har hatt som primært mål å konvertere papirsamlingene til elektronisk form. Vi har i dette kapittelet sett hvordan dette målet er nådd ved at alt materialet er konvertert til kodet tekst eller faksimiler. Det ligger dermed klart for forskere som disponerer IT-kompetanse, og som ønsker å bruke materialet i det analyseverktøyet de foretrekker. Men for de fleste humanistiske forskere er det trolig ønskelig å få dataene presentert gjennom et databasesystem. Akkurat hvordan dette systemet ser ut, har jeg viet liten plass. Årsaken til det er at hva brukeren ser på skjermen, vil ganske sikkert endre seg etter hvert som man vinner erfaring med systemene, og etter hvert som den tekniske utviklingen fortsetter. I avsnittet "Hva særtegnar Dokumentasjonsprosjektets databaser?" beskrev jeg hvordan dataene kunne sees på ulike måter: både som åpne bibliotek og som lukkede systemer. De leksikalske dataene og særlig gjenstandsdataene vil primært ligge i slike "lukkede systemer". Dataene ligger lagret i en relasjonsdatabase som blir oppdatert fortløpende. Men det er selvfølgelig mulig å ta ut deler av dem til bruk i andre systemer. For de andre fagfeltene og materialtypene er situasjonen omvendt. Dataene ligger primært i statiske tekst- og bildesamlinger (bibliotek) og sekundært i et søke- og presentasjonsverktøy. Jeg tror at en slik blanding av ferdige databaser, spesiallagde elektroniske utstillinger og publikasjoner samt elektroniske samlinger av bilder, lyd og tekst vil være kjernen i et fremtidig elektronisk universitetsbibliotek.





# Konverteringsorganisasjonen

Konseptet var på plass fra begynnelsen av i forprosjektet som startet ved Universitetet i Oslo i 1991. De ulike aktørene var:

- fagenheten med ett eller flere arkiv som skulle konverteres til data
- den delprosjektansvarlige, som er en vitenskapelig ansatt i avdelingen med ansvar for det faglige resultatet
- en eller flere fagkonsulenter som ledet konverteringsarbeidet
- korrekturlesere, oftest hovedfagsstudenter eller tidligere innskrivere
- programmererne, som i dialog med fagkonsulentene og fagenhetene bestemte it-løsningene
- innskriverne organisert i grupper på 5-10 personer eller på sentraler med 35-75 personer
- lokaladministrasjonen, som sørget for driften av gruppa eller sentralen
- produksjonskonsulenten i sentralene som var delprosjektets forlengede arm
- sentraladministrasjonen i prosjektet som var konverteringsstøtte for de ulike delene av organisasjonen
- prosjektledelsen, med daglig ansvar for gjennomføringen av prosjektet
- Nasjonal styringsgruppe, med strategisk ansvar for gjennomføringen av prosjektet.

## Fagenhetens rolle

Dokumentasjonsprosjektet var delt opp i femten delprosjekter. Arkivene på en avdeling eller ett institutt utgjorde ett delprosjekt. Fagenheten utpekte en vitenskapelig ansatt som delprosjektansvarlig. Disse femten delprosjektansvarlige har utøvd sin rolle i prosjektet på femten ulike måter, og først sent i prosjektperioden utarbeidet Nasjonal styringsgruppe en beskrivelse av arbeidsoppgaver og ansvarsområder. Den delprosjektansvarliges fremste oppgave var å ha det emnespesifikke fagansvaret, det vil si å borge for at det konverterte materialet og løsningene som er valgt, holder faglig mål, og at det arbeidet som ble utført i delprosjektet, var det fagenheten ønsket og hadde bruk for. De delprosjektansvarliges engasjement spente over en skala fra full deltakelse - kontakt med innskrivingsenhetene gjennom besøk og telefoner og daglige samtaler med fagkonsulentene - til distansert rådgivning - der de var tilgjengelige for faglige spørsmål, men ellers delegerte ansvar til fagkonsulentene. Prosjektet trakk de delprosjektansvarlige med i rekrutteringen av fagkonsulenter og i økonomistyringen av delprosjektet. De var også med på de halvårige framdriftsmøtene og skrev eller undertegnet delprosjektets årsrapporter. Den delprosjektansvarlige var også fagkonsulentenes "fadder". Fagkonsulentene hadde sin arbeidsplass i fagenheten, og det var ofte den delprosjektansvarlige som drog dem med i fagmiljøet og sørget for at de fikk mulighet til å delta i fagenhetens aktiviteter og diskusjoner.

Den delprosjektansvarlige er også Dokumentasjonsprosjektets talskvinne eller -mann i fagenheten. Graden av deltakelse fra de delprosjektansvarliges side samsvarer ikke nødvendigvis med kvaliteten på det ferdige produktet, men de delprosjektansvarliges

engasjement vil nok være viktig når det gjelder lanseringen og bruken av det ferdige informasjonssystemet i fagenheten. Det er den delprosjektansvarlige som vil kunne fremme bruken av databasene og være drivkraft for at rutiner og arbeidsmetoder legges om i tråd med intensjonene i Dokumentasjonsprosjektet.

## Prosjektets fagfolk

Fagkonsulentene var Dokumentasjonsprosjektets representanter i fagenheten og fagenhetens representanter i prosjektet. De var ansatt i prosjektet, men hadde sitt arbeidssted i fagenheten. På det meste, i 1995, var det 35 årsverk satt av til kvalitetssikring. Delprosjektene fikk tildelt fagkonsulentårsverk ut fra størrelse og omfang på det papirbaserte arkivet, og antallet varierte fra én til ti stillinger pr. år. Fagkonsulentene hadde som regel hovedfag i faget arkivet omhandlet, var relativt eller helt nyutdannet og hadde ofte et ønske om en videre faglig karriere med doktorgradsstipend som nærmeste mål. Dokumentasjonsprosjektet ble en måte å fortsette i universitetssystemet og fagmiljøet etter endt hovedfag.

Fagkonsulentenes stillingsbeskrivelse omfattet utredning og analyse av arkivene ut fra faglige og datatekniske kriterier; utarbeidelse og vedlikehold av innskrivingsopplegget og taggesystemet, kopiering av originalmaterialet, opplæring, organisering og veiledning av innskriverne i skriverrutiner, tagging, skanning og korrekturlesing og administrasjon av time- og ekstrahjelpere. Dette er varierte oppgaver som innebærer ulike utfordringer, men det meste av fagkonsulentenes tid ble brukt til korrekturlesing og kvalitetssikring av det konverterte materialet.

Ideelt sett skulle alle arkivene vært utredet nøye og hatt en grundig faglig gjennomgang før oppstart av konverteringen. Men særlig for de første delprosjektene ble utredningstiden kort, fordi arkivene var så omfattende at en helhetlig analyse ville ta for mye av de tilmålte ressursene, og fordi innskriverne satt ved maskinene og ventet på arbeidsoppgavene. Ulempene med dette var at man kunne få overraskelser som den innledende stikkprøvegjennomgangen eller prøveutvalget ikke hadde dekket, mht klassifiseringer, terminologi eller mengder av et bestemt innhold. Innskrivingsprosedyrene måtte derfor på flere delprosjekt endres underveis. Fordelene med en rask oppstart var at konverteringsarbeidet faktisk kom i gang. En grundigere faglig utredning ville for de fleste arkivene ha tatt for lang tid og blitt for dyrt.

Underveis i delprosjektene kom det ofte ønske fra avdelingen eller instituttet til fagkonsulenten om å foreta endringer i eller forbedringer av arkivet. Man så muligheten for samtidig med konverteringen å komplettere arkivet eller sjekke det opp mot gjenstandene det omhandlet, for å luke ut unøyaktigheter og mangler. Dokumentasjonsprosjektets oppgave var imidlertid å konvertere det papirbaserte arkivet slik det stod. Magasinrevisjoner og innlegging av nytt materiale inngår i fagenhetenes ordinære oppgaver og skulle ikke tas i prosjektets dugnadsløft med å edb-føre eksisterende papirarkiver. De databaserte arkivene har derfor ikke høyere kvalitet enn de papirbaserte, men gir nye muligheter for arbeidet med materialet. Noen av fagenhetene valgte å foreta revisjonene med egne midler parallelt med konverteringen, mens andre vil arbeide videre med de ferdige dataarkivene. Dataverktøyet vil gjøre arbeidet med magasinrevisjon og oppdatering langt enklere.

Fagkonsulenten hadde altså ansvar for sin(e) konverteringsenhet(er) gjennom å gi opplæring i materialet, sette standarder for innskrivingen og å gi veiledning og tilbakemelding på arbeidet. Prosjektet la vekt på at det var rom for at den enkelte fagkonsulent skulle kunne utøve lederfunksjonen på sin måte, fordi noe av nøkkelen til

kvalitetsmessig forsvarlig materiale nettopp lå i forholdet mellom fagkonsulenten og innskriverne, selv om det etter hvert ble utarbeidet detaljerte retningslinjer og stillingsbeskrivelser for fagkonsulentene.

For nyutdannede akademikere med et langt opphold i universitetsverdenen bak seg var møtet med "folket" ute i innskrivingsenheten ofte en utfordring. Innskriverne var spente på møtet med en akademiker, og akademikeren var ikke mindre spent på sin rolle. I prosjektet var man opptatt av at innskriverne skulle få innsikt i faget som arkivet omhandlet, og fagkonsulentene gav derfor innskriverne en innføring i det. Dette kunne noen ganger tas for langt, ved at innskriverne fikk lange, tunge foredrag om pilespisser eller bøyningformer, men som oftest fungerte det som inspirasjon og motivasjon. Det var meget vanlig at innskriverne fikk både interesse for faget og lojalitet til arkivet.

Et godt resultat var avhengig av at forholdet mellom oppdragsgiveren, representert ved fagkonsulenten, og innskriverne fungerte bra. Innskriverne satt langt borte fra oppdragsgiveren, og konflikter kunne vokse seg store under manglende eller dårlig kontakt. Noen av fagkonsulentene hadde et tett forhold til innskriverne og fulgte dem opp med daglige telefoner og hyppige reiser, mens andre kunne la gruppene arbeide mer selvstendig og hadde sjeldnere kontakt. Dette var avhengig av vanskelighetsgraden på materialet og hvor i innskrivingsprosessen det befant seg.

Til å begynne med ble fagkonsulentene rekruttert etter utelukkende faglige kriterier. Karakteren på hovedfaget og faglig innsikt i samlingsen og arkivet avgjorde ansettelsen. Etter hvert ble det lagt mer vekt på kunnskap om dataverktøy og lederegenskaper. Man så at opparbeidet konverteringskompetanse kunne være like viktig som en god akademisk innsats. Det at man la så stor vekt på faglige prestasjoner i prosjektets første år, var imidlertid nødvendig for å oppnå tillit i fagenheten. Dokumentasjonsprosjektet var kjent som et "dataprojekt" og møtte skepsis og tvil hos fagenhetens vitenskapelig ansatte med varierende innsikt i moderne it-baserte metoder. Det var ikke uvanlig at man møtte holdninger hos enkelte som gikk ut på at databaser aldri ville kunne erstatte de gamle arkivkortene. Det at prosjektets medarbeidere hadde god faglig innsikt og gode faglige kvalifikasjoner ble derfor betryggende, og diskusjonen kunne foregå innenfor fagets rammer, og ikke mellom datafolket og fagfolket med de kommunikasjonsproblemene det ville medføre. Etter hvert som prosjektet ble kjent i fagenheten og det ble klart at arbeidet skulle utføres på deres premisser, ble det mulig å også vektlegge andre ferdigheter og egenskaper hos fagkonsulentene, selv om faglig kompetanse fortsatte å være det viktigste rekrutteringskriteriet.

I tillegg til fagkonsulentene kunne delprosjektet ha en eller flere korrekturlesere knyttet til seg. Disse var ofte hovedfagsstudenter eller tidligere innskrivere som arbeidet på timebasis med ren korrekturlesing. En bieffekt av bruken av studenter var at disse fikk innsikt i bruk av dataverktøy på det aktuelle fagområdet. Prosjektets ansatte har deltatt på en rekke seminarer og konferanser i inn- og utland, ofte med presentasjon av prosjektet og demonstrasjon av databasene.

## **Tempo og kvalitet**

Fra begynnelsen av ble det lagt vekt på at innskriverne skulle få tilbakemeldinger på sitt arbeid for å sikre kvaliteten. Denne tilbakemeldingsaktiviteten ble vektlagt mer og mer i løpet av prosjektiden, fra det første årets telefoniske "dette var bra" til detaljerte statistikker over enkeltfiler med oversikt over antall feil og skriftlig påpeking av feiltyper. Vi så at jo grundigere og mer konkrete kvalitetsrapportene var, jo bedre ble også kvaliteten. Det ble også stilt kvalitetskrav til innskrivingsenheten avhengig av

materialets art og fagets krav til nøyaktig tekstgjengivelse. Kravene kunne variere fra ikke mer enn én feil på ti utskriftssider til ikke mer enn én feil pr. side. En feil kunne være punktum i stedet for komma, stor bokstav i stedet for liten, feilkoder eller feil tolkning av teksten. Leverte innskriverne datafiler med flere feil enn kravet, fikk de det tilbake for ny korrektur. Dette var nødvendig for å sikre et materiale med høyt kvalitetsnivå, men også fordi det ville være for tidkrevende for fagkonsulentene å rette opp store mengder materiale med mye feil.

Tilbakemeldingen fikk også en positiv effekt på kvaliteten på sysselsettingstiltaket. Mange av innskriverne var langtidsledige med manglende tro på seg selv som arbeidstakere. Tap av arbeid med påfølgende ledighet skaper usikkerhet og lav selvtillit. Konkrete tilbakemeldinger om kvaliteten på arbeidet, i positive rammer, var et bidrag til å gjenopprette selvtilliten. Det er sjelden ellers i arbeidslivet at en får så rask og konkret respons på arbeidet en utfører.

Mens kvalitetskravene som regel ble satt rett etter innskrivingsoppstarten på hvert enkelt materiale, var praksisen mer ulik når det gjaldt kravene til tempo. Det ble tidlig klart at vi ikke bare kunne konvertere i vei og håpe på at vi kom tidsnok i mål. Vi måtte holde nøye oversikt med framdriften på registreringsarbeidet for å justere ressursbruken og metodene. Delprosjektene gav halvårige framdriftsrapporter, som ble fulgt opp med møter og eventuell justering av kursen. Gruppene og sentralene rapporterte inn sitt arbeid i ukentlige eller månedlige produksjonsoversikter. Mens kvalitetsrapportering ble ønsket velkommen av både fagkonsulenter og innskrivere, var skepsisen til temporapportering større. Kvalitetskravene kunne være meget strenge, men dette ble oppfattet som forståelig og ikke utilbørlig press. Når farten på konverteringsarbeidet ble tatt opp til diskusjon, var argumentene mot tempokrav mange. Det viktigste var at press på tempoet ville føre til dårligere kvalitet. Det var også svært vanskelig å sette tempokrav, simpelthen fordi vi ikke hadde noe sammenligningsgrunnlag. Hvor fort er det rimelig å skrive inn tingbøker fra 1600-tallet eller dialektkort fra 1920? Det var heller ikke alltid mulig å sammenligne på tvers av delprosjektene.

Allikevel ble det gitt detaljerte produksjonsrapporter, som registreringsledelsen brukte til å beregne framdriften, avgjøre ressursbehovet og som mål på tilstanden i konverteringsenheten. Gikk tempoet ned over tid, var det et varsel om at noe var galt. Det kunne skyldes manglende opplæring i materialet eller dataverktøyet, manglende motivasjon eller konflikt innad i konverteringsenheten eller mellom innskriverne og fagkonsulentene - og det ble satt i verk tiltak for å løse problemene avhengig av hva årsaken var.

Etter hvert fikk prosjektet nok erfaring til å kunne bedømme hva som var en rimelig innskrivingsfart på materialet, og det ble satt mål for hvor mye hvert dagsverk skulle produsere av sider eller kort. Disse målene ble ikke alltid kommunisert ut til innskrivingsstedet, men ofte fikk sentraler og grupper beskjed om at totalproduksjonen burde være av en bestemt størrelse.

## **Koordinering**

Prosjektets sentrale administrasjon hadde ansvar for konverteringen og koordinerte dette arbeidet i prosjektet. Administrasjonen bestod fra 1995 av fire personer som hadde ulike oppgaver: personalarbeid, produksjonsrapportering og registreringsledelse. Prosjektet bestod av 15 delprosjekter og 15 registreringsenheter fordelt på 20 ulike steder i Norge. Å holde prosjektet samlet med felles mål og kunnskap var en stor

utfordring. De ulike medarbeidernes følelse av prosjektidentitet varierte sterkt. Noen følte seg først og fremst som tilhørende "gruppa i Larvik" eller "sentralen i Mo" eller som arkeologer eller folklorister, mens tilhørigheten til det nasjonale prosjektet var mindre. Andre igjen var først og fremst "dokpro-ere".

Mye av sentraladministrasjonens arbeid bestod i å formidle erfaringer utover landet og sørge for at gode ideer fikk rask spredning. Kommunikasjonen i prosjektet gikk via telefon, faks og elektronisk post, men vi innså tidlig at en viss grad av personlig kontakt var nødvendig for å få gjort jobben. Reisebudsjettene vokste derfor år for år.

Dokumentasjonsprosjektet kan også ses på som en moderne arbeidsplass hvor datateknologien gjorde geografisk samlokalisering unødvendig. Prosjektet hadde i løpet av de syv årene kun én felles samling for de ansatte. En svakhet med dette var at det kunne være vanskelig for de ansatte i delprosjektene og innskrivingsenhetene å få oversikt over helheten i prosjektet. Men på grunn av tverrfagligheten, organiseringen i delprosjekter og spredte innskrivingsenheter var behovet for en helhetlig prosjektfølelse av varierende betydning. Dokumentasjonsprosjektet betydde forskjellige ting for de ulike medarbeiderne: en jobb etter år med arbeidsledighet, dataopplæring eller relevant praksis etter datakurs, en mulighet for utdanning på hjemstedet, en jobb hvor en kan få brukt faglig kompetanse, en nødvendig oppgradering av et tungrodd arkivsystem, en mulighet til å gjøre et spennende faglig materiale mer tilgjengelig eller et unikt tverrfaglig pionerprosjekt med henhold til edb-baserte metoder innenfor humaniora.

204 personer har vært ansatt i Dokumentasjonsprosjektet i årene 1991 - 1997, fordelt på følgende kategorier:

Prosjektets daglige ledelse:	5
IT-faglig stab:	9
Sentral prosjektstab:	9
Fagkonsulenter:	63
Korrekturlesere:	94
Stab registreringsentralene:	24



## Innskrivingsmiljøene

Denne rapporten fokuserer på Dokumentasjonsprosjektets primæraktivitetet - konverteringen av arkiver og samlinger. Men prosjektet har også vært et stort sysselsettings-tiltak for mer enn ett tusen arbeidsledige over hele landet. Vi vil imidlertid ikke gå detaljert gjennom hvordan Dokumentasjonsprosjektet har fungert som sysselsettings-tiltak. For mer informasjon om dette perspektivet på prosjektet viser vi til års- og sluttrapportene for de ulike registreringsenhetene, evalueringene av gruppene som sysselsettningstiltak gjennomført i 1995 og 1997 og de ulike evalueringene som har blitt gjennomført lokalt ved de fire registreringsentralene.

Selve inntastingen eller innskanningen av arkivene har vært den mest arbeids- og kostnadskrevende delen av prosjektet. Dette ble gjort mulig ved at Dokumentasjonsprosjektet engasjerte seg i ulike former for sysselsettningstiltak.

I 1991 startet konverteringsarbeidet opp, og små grupper bestående av 5-10 arbeidsledige ble opprettet først i Hokksund og Sigdal, deretter videre utover i prosjektperioden i Drammen, Røyken, Lier, Larvik, Sandefjord, Tønsberg, Horten, Trondheim, Oppdal, Bergen, Oslo og. Fra 1992 og utover ble registreringsentralene startet i Mo i Rana, Indre Salten, Narvik og Sør-Varanger. På det største bestod konverteringsorganisasjonen av 240 personer. Fagkonsulenten reiste til sin gruppe eller sentral, gav innføring i arkivet og fortalte om bruken av det, formålet med det og fordelene med digitalisert versjon. Deretter startet opplæringen i å skrive inn eller skanne inn materialet. Innskriverne fikk standarder for hvordan teksten skulle behandles og materialet organiseres.

## Registreringsgruppene

Gruppene bestod av langtidsledige, yrkeshemmede og folk på attføring. Prosjektet brukte tiltakstypene Arbeid for trygd, SKAP, KAJA, sysselsetting i offentlig sektor og hospitant- og praksisplasser. De fleste innskriverne var i prosjektet i 10 måneder. Etter hvert ble kategorien "voksne kvinner uten dataerfaring" den typiske innskriveren. Det lokale arbeidskontoret rekrutterte folk til gruppene, som ble administrert av det lokale Folkeuniversitetet eller andre egnede aktører, på oppdrag fra Dokumentasjonsprosjektet. Hver gruppe hadde en gruppeleder, som kunne være opptil to år i prosjektet og var kontaktperson overfor delprosjektet. Fordi gruppene var små, oppsto det ofte nære relasjoner mellom fagkonsulenten og innskriverne. Idealismen og pionerfølelsen som preget fagkonsulentene, smittet ofte over på innskriverne. Om ikke alle innskriverne var like engasjerte, var lojaliteten til prosjektet stor. En undersøkelse som ble foretatt i 1995 over alle de ledige som hadde vært i innom gruppene til da, viste at interessen for materialet og registreringsarbeidet var stor. Tidligere innskrivere skrev f.eks. at "jeg ble interessert i gammel litteratur og begynte å lese på egenhånd", og at "det var spennende å lære om tidligere tider og arkeologi". Undersøkelsen viste også at gruppene var vellykkede som sysselsettningstiltak. 64% av innskriverne hadde fått arbeid eller var i utdanning. Når de tidligere medarbeiderne skulle skrive med egne ord hva som hadde vært best med tiltaket, var det imidlertid gruppa selv som oftest ble trukket fram i uttalelser som: "Jeg fikk ny selvtillit og ble inspirert til å søke jobber", "den beste arbeidsplassen jeg har hatt" og "miljøet i gruppa var strålende, og vi holder fortsatt kontakten".

Selv om gruppene ble ledet av fagkonsulentene og detaljstyrt på utførelsen av arbeidet ned til kommabruk og antall mellomrom mellom ordene, er det tydelig at de følte en stor grad av selvstendighet. Dette skyldes nok i hovedsak at gruppene satt langt fra fagkonsulenten, og at lokaladministrasjonen ikke hadde innsikt i selve arbeidet, slik at det daglige arbeidet ble organisert av gruppa selv.

## Registreringsentralene

I 1992 ble konseptet med konverterings- og Edb-praksis som sysselsettingstiltak videreutviklet og Universitetenes registreringssentral i Mo i Rana startet opp. Deretter åpnet sentraler i Indre Salten, Narvik og Sør-Varanger. Sentralene var et samarbeid mellom lokale myndigheter og næringsliv i de aktuelle kommunene, fylkesarbeidskontorene i Nordland og Finnmark, Næringsdepartementet, Kommunal- og arbeidsdepartementet, SND, Statens datasentral og universitetene. Departementene, SND, fylkene og kommunene stod for finansiering av utstyr og lokaler, mens arbeidsmarkedsetaten finansierte driften og kursvirksomheten.

Sentralene ble opprettet i omstillingskommuner hvor det var behov for nye tiltak som kunne sikre arbeidsledige formelle utdanningstilbud og relevant arbeidstrening, og utgjorde prosjektet "Omstilling gjennom målrettet utdanning". Sentralene hadde fra 35 til 75 personer som arbeidet 50% av tiden med konvertering av arkivmateriale, mens resten av arbeidsuken ble brukt til studier.

Hver sentral hadde sin egen administrative stab som var ansatt i eller leid inn av Dokumentasjonsprosjektet, og som bestod av en daglig leder, en produksjonskonsulent og en sekretær. Staben hadde ansvar for rekruttering, personforvaltning, gjennomføringen av konverteringsarbeidet, kursadministrasjon og formidling av arbeidsledige ut i jobbmarkedet. Staben ble også en forlenget arm av de lokale arbeidskontorene.

Kurssekretariatet, Studieavdelingen, Universitetet i Oslo hadde ansvaret for kvalitets-sikringen av utdanningsdelen og målrettet studieveiledning samt administrativ og teknisk drift av sentralene.

Dokumentasjonsprosjektet hadde ansvar for sammenkobling mellom delprosjekter og innskrivingsmiljøer, informasjonsvirksomhet, organisering og oppfølging av innskrivingsmiljøene og fortløpende evaluering av den samlede innskrivingsaktiviteten ut fra kvalitet, effektivitet og fremdrift.

De arbeidsledige deltok på universitets- og høyskolestudium eller kurs ved videregående skoler og oppnådde formell kompetanse. Sentralene benyttet kurstilbudene som finnes innenfor Norgesnett. I prosjektperioden ble det ved sentralene avholdt mer enn 220 kursmoduler, som også var tilgjengelige for ansatte i det lokale næringslivet og offentlige etater. Det viste seg raskt at de mest givende kursene for innskriverne var de som var relatert til konverteringsoppgavene, f.eks. gav studier som nordisk grunnfag innsikt som var nyttig i registreringen av nynorskarkivet og kurs i edb organisert over Internett gav kunnskaper om databaser og edb-verktøyet de brukte i arbeidet.

For kommunene har "Omstilling gjennom målrettet utdanning" blitt et viktig omstillingsverktøy. Lokale utviklingsselskap brukte tiltaket til å sikre kompetanseheving i næringslivet, etablere spinn-off-virksomheter og å skape dialog mellom de lokale utdanningsaktørene. Universitetene har gjennom prosjektet fått status som kompetanseleverandør i omstillingsdistriktene.

For universitetene og høyskolene som var involvert i kursdelen, har sentralene vært en mulighet til å utprøve fjernundervisning og desentraliserte kurstilbud innenfor etter- og videreutdanning, gjennom utsendte foredragsholdere, kurs via Internett, veiledning over elektronisk post eller telefon og brevundervisning. De lokale videregående



skolene og deres ressursentre tilpasset undervisningsmoduler til sentralenes tidsskjema. På sentralene ble det utprøvd nye metoder innenfor voksendidaktikk, med tilrettelagt undervisning for voksne elever med ulik livserfaring og metoder for opplæring av personer med lese- og skrivevansker.

For Dokumentasjonsprosjektet var sentralene en mulighet til å konvertere arkivmateriale i stor skala. Ledelsen ved sentralene prøvde ut ulike måter å organisere konverteringsarbeidet på og utviklet rutiner og metoder som var nyttige og ressursbesparende. Alle sentralene hadde en ansatt med spesielt ansvar for konverteringsarbeidet, produksjonskonsulenten, som fungerte som delprosjektens lokale representant. Modellen fra gruppene med fagkonsulenten med direkte arbeidsledelse ble oppgitt til fordel for en modell med mer lokal styring. Sentralene fikk et konverteringsoppdrag med måltall på kvalitet og tempo og utførte oppdraget selvstendig, men med fagkonsulentenes opplæringsstøtte, faglig veiledning og tilbakemelding på det utførte arbeidet.

Cirka 60% av innskriverne ved sentralene gikk ut i jobb, og 15% gikk over i ordinært studium. Det arbeides også lokalt med å etablere varige arbeidsplasser med bakgrunn i sentralenes virksomhet.

## Oversikt

### *Innskrivingskapasiteten:*

1992	47 årsverk	gruppene: 38 årsverk	sentralene: 9 årsverk
1993	104 årsverk	gruppene: 48 årsverk	sentralene: 56 årsverk
1994	132 årsverk	gruppene: 55 årsverk	sentralene: 77 årsverk
1995	141 årsverk	gruppene: 60 årsverk	sentralene: 81 årsverk
1996	143 årsverk	gruppene: 62 årsverk	sentralene: 81 årsverk
1997	100 årsverk	gruppene: 40 årsverk	sentralene: 60 årsverk

På sentralene representerer hvert årsverk 2 personer, da medarbeiderne registrerer 50% av tiden de er i prosjektet.

### *Registreringsgrupper*

<i>Sted</i>	<i>Periode</i>	<i>Materiale</i>	<i>Antall</i>	<i>Lokaladministrasjon</i>
Hokksund	91 - 93	gammalnorsk, nynorsk	5	Folkeuniversitetet i Hokksund
Sigdal	91 - 93	nynorsk	5	Folkeuniversitetet i Sigdal
Båtsfjord	92 - 93	bokmål		Båtsfjord kommune
Larvik	92 -	mynt, nynorsk, bokmål, etnografi	5 - 6	Folkeuniversitetet i Larvik
Tønsberg	92 - 97	mynt, bokmål, arkeologi, navn, etnografi	3 - 6	Folkeuniversitetet i Tønsberg

Sandefjord	92 -	mynt, arkeologi nynorsk	5 - 7	Folkeuniversitetet i Sandefjord
Horten	92 - 97	navn, runer, ballader, nynorsk	5 - 9	Folkeuniv. Horten og Holmestrand
Lier	92 - 97	nynorsk, navn tingbøker	6	Folkeuniv. Lier, Røyken og Hurum
Drammen	92 -	nynorsk, arkeologi etnografi	5 - 7	Folkeuniversitetet i Drammen
Slemmestad	92 - 97	nynorsk, gammalnorsk	5 - 8	Folkeuniv. Lier, Røyken og Hurum
Oslo	94 - 97	gammalnorsk, ballader	8	Kurssekretariatet, UiO
Oppdal	94 - 96	folkedans	3	Oppdal spele- og danse- lag
Bardu	95 - 97	navn	10	Bardu kommune
Trondheim	96 - 97	folkedans	3	Rådet for folkemusikk og - dans
Bergen	96 -	arkeologi	6	Bergen museum
Nordland	97 -	folkloristikk	8	Registreringsentralen i Mo

### ***Registreringsentraler***

<i>Sted</i>	<i>Periode</i>	<i>Materiale</i>	<i>Antall</i>
Mo i Rana	92 - 96	nynorsk, trønder arkeologi	75 - 90
Indre Salten	93 -	bokmål, navn arkeologi, foto, historie	35 - 50
Sør-Varanger	93 - 97	navn, folkloristikk	32 - 33
Narvik	95 -	folkloristikk, nynorsk	35 - 45

### *Stab Mo i Rana:*

Beate Aspdal	daglig leder	(1992 - 96)
Sissel Holdahl	sekretær	(1993 - 96)
Inger Dalåsen	sekretær	(1995 - 96)
Nina Kristin Olsen	kvalitetssikrer	(1995)

Svein Gullbekk	produksjonskonsulent	(1993)
Svein Sanden	produksjonskonsulent	(1994 - 96)
Per Nonskar	produksjonskonsulent	(1996)

*Stab Indre Salten:*

Jan-Åke Storlund	daglig leder	(1994 - 97)
Kjell Børge Jordbro	produksjonskonsulent	(1994 - 97)
	daglig leder	(1997 - )
Solfrid Eliassen	sekretær	(1994 - )

*Stab Sør-Varanger:*

Unni Hogsnes Tandberg	daglig leder	(1994 - 97)
Randi Simonsen	sekretær	(1994)
Tone Merethe Johansen	sekretær	(1995)
Signe Vibeke Hauge	sekretær	(1995 - 97)
Line Moe	produksjonskonsulent	(1995 - 97)
Unni Naustdal	produksjonskonsulent	(1997)
Lani Wosnitza	edb konsulent	(1997)
Björg Karin Wikan	konsulent	(1997)

*Stab Narvik:*

Gørill Hørrigmoe	daglig leder	(1995 - )
Inger Almli	sekretær	(1995 - )
Karoline H. Mathisen	produksjonskonsulent	(1995 - 96)
Halvard Larssen	produksjonskonsulent	(1996 - )



# Arkivene

## Arkiver og materiale som er konvertert i Dokumentasjonsprosjektet

Førhistoriske minne på Vestlandet	Bergen museum, UiB
Gjenstandsregisteret	Bergen museum, UiB
Tilvekstfortegnelsene 1850 - 1990	Bergen museum, UiB
Topografisk arkiv	Bergen museum, UiB
Oldsaksamlingens hovedkatalog	Oldsaksamlingen, IAKN, UiO
Middelalderkort	Oldsaksamlingen, IAKN, UiO
Fotoarkivets kortarkiv	Oldsaksamlingen, IAKN, UiO
Konserveringsavdelingens arkiv	Oldsaksamlingen, IAKN, UiO
Runer: skjemaer	Runearkivet, IAKN, UiO
Runer: innholdsfortegnelser	Runearkivet, IAKN, UiO
Norges Innskrifter yngre Runer	Runearkivet, IAKN, UiO
Myntarkivet	Myntkabinettet, IAKN, UiO
Utstilling: Norsk mynt i 1000 år	
Fotolister	Arkeologisk avdeling, Tromsø museum, UiTø
	Arkeologisk avdeling, Tromsø museum, UiTø
Tilveksten/funnkatalogene	Vitenskapsmuseet, NTNU
	Fagenhet for samisk-etnografi, Tromsø museum, UiTø
Vitenskapsmuseets tilvekster	Fagenhet for nyere kulturhistorie, Tromsø museum, UiTø
Fotoarkivet	Institutt for kulturstudier, UiO
Fotoarkivet	Institutt for kulturstudier, UiO
	Historisk institutt, UiO
Norske ballader	UB Tromsø, avd. Tromsø museum
Utstilling: Folkevisautomaten	UB Tromsø, avd. Tromsø museum
Minneoppgavene	UB Tromsø, avd. Tromsø museum
Tingbøker	UB Tromsø, avd. Tromsø museum
Reinbeitekommissjonen, bind 2	UB Tromsø, avd. Tromsø museum
Schnitlers protokoller, bd.1	UB Tromsø, avd. Tromsø museum
Qvigstads brev	Seksjon for leksikografi og målføregransking, INL, UiO
Bibliografi - nordnorsk topografi	Seksjon for leksikografi og målføregransking, INL, UiO
Friis' etnografiske kart	Seksjon for leksikografi og målføregransking, INL, UiO
Setelarkivet, Norsk Ordbok	Seksjon for leksikografi og målføregransking, INL, UiO
	Seksjon for leksikografi og målføregransking, INL, UiO
Nynorske tekster	Seksjon for leksikografi og målføregransking, INL, UiO
	Seksjon for leksikografi og målføregransking, INL, UiO
Grunnmanuskriptet	Seksjon for leksikografi og målføregransking, INL, UiO
	Seksjon for leksikografi og målføregransking, INL, UiO
Nyordsarkivet	Seksjon for leksikografi og målføregransking, INL, UiO
	Seksjon for leksikografi og målføregransking, INL, UiO
Litterære tekster - bokmål	Seksjon for leksikografi og målføregransking, INL, UiO
	Seksjon for leksikografi og målføregransking, INL, UiO
Gammalnorsk seddelarkiv	Seksjon for leksikografi og målføregransking, INL, UiO
	Seksjon for leksikografi og målføregransking, INL, UiO
Norske diplomer	Seksjon for leksikografi og målføregransking, INL, UiO
	Seksjon for leksikografi og målføregransking, INL, UiO
Diplomatarium Norvegicum	Seksjon for leksikografi og målføregransking, INL, UiO
	Seksjon for leksikografi og målføregransking, INL, UiO
Trønderordboka	Institutt for nordistikk og litteraturvitenskap, NTNU
	Seksjon for navnegransking, INL, UiO
Etymologiregisteret	Seksjon for navnegransking, INL, UiO
Matrikelutkastet av 1950	Seksjon for navnegransking, INL, UiO

Bustadnavnregisteret	Seksjon for navnegransking, INL, UiO
Norske Gaardenavne	Seksjon for navnegransking, INL, UiO
Talemåls- og navnearkivet	Det humanistiske fakultet, UiTø
Skulebarnoppskrifter/nordnorsk del	Det humanistiske fakultet, UiTø
Informantskjema	Rådet for folkemusikk og -dans, NTNU
Film 8 mm	Rådet for folkemusikk og -dans, NTNU
Film 16 mm	Rådet for folkemusikk og -dans, NTNU
Lydbånd - Rff-serien	Rådet for folkemusikk og -dans, NTNU
Lydbånd - MIT-serien	Rådet for folkemusikk og -dans, NTNU
Hovedkatalogen	Etnografisk museum, UiO

## Arkeologi

### Generelt: Bergen museum, Universitetet i Bergen

Museumsseksjonen ved Arkeologisk institutt er, sammen med de andre museumsseksjonene ved Universitetet i Bergen, en del av Bergen Museum. Bergen Museum v/ Arkeologiske institutt er landsdelsmuseum for Hordaland, Sogn og Fjordane og Sunnmøre og omfatter gjenstandssamlinger fra forhistorisk tid og middelalder. I tillegg finnes en rekke arkiv som dokumenterer faste fornminner og gjenstander fra museumsregionen.

Informasjon fra arkivene og samlingene utgjør selve fundamentet for arkeologisk forskning. Gjennom Dokumentasjonsprosjektet vil store deler av denne informasjonen bli elektronisk tilgjengelig. I Dokumentasjonsprosjektet var det de forhistoriske arkivene som ble konvertert til elektronisk format.

Delprosjektansvarlig: Professor Bergljot Solberg, Bergen Museum

Fagkonsulenter i prosjektperioden: Mona Mortensen, Eva M. Walderhaug, Oddgeir Hoftun, Sonja Innselset og Asbjørn Engvik jr.

Korrekturlesere/andre medarbeidere i delprosjektperioden: Marianne Vedeler Nilsen, Marit Kvitne, Grethe Larsen, Ragnhild Todnem, Eva Moberg, Lene Elisabeth Walle, Anne-Cathrine Johannessen og Anne K. Sæther

IT-faglig ansvar: Jon Holmen

Antall fagkonsulentårsverk i perioden 1992 - 1997: 15,5

Kontaktpersoner i fagenheten fra 1.1.98: fagkonsulentene Sonja Innselset og Asbjørn Engvik jr.

### *Førhistoriske minne på Vestlandet*

#### *Om arkivet:*

Førhistoriske minne på Vestlandet er en registreringsserie utarbeidet av arkeolog og konservator Per Fett, som gir en samlet oversikt over faste fornminner og gjenstander i de 77 kommunene som hører inn under Bergen Museums distrikt.

#### *Status på materialet pr. 31.12.97:*

Total størrelse:	78 hefter, ca 5.000 sider
Konvertert totalt:	78 hefter, ca 5.000 sider
Korrekturlest totalt:	78 hefter, ca 5.000 sider
Totalt ferdigstilt materiale:	78 hefter, ca 5.000 sider

#### *Beskrivelse av arbeidet:*

Heftene er skrevet inn og tagget, korrekturlest i innskrivingsenheten og i delprosjektet. Heftene er parsert og kjørt gjennom Author Editor som sjekker at taggene er plassert riktig i forhold til "grammatikken" som taggesystemet er basert på.

#### *Beskrivelse av digitalt arkiv / database:*

Det konverterte materialet er lagret som SGML-tagget tekst og er gjort tilgjengelig over Internett.

## ***Gjenstandsregisteret***

### *Om arkivet:*

Magasinet for de forhistoriske samlingene inneholder ca 150.000 enkeltgjenstander. Data om selve gjenstanden; mål, tilstand og funnsted, registreres elektronisk og danner utgangspunkt for en gjenstandsdatabase som hovedsakelig skal benyttes internt i magasinet.

Registreringskortene for eldre jernalder:

### *Status pr. 31.12.97:*

Total størrelse på materialet:	8.000 kort
Konvertert totalt:	8.000 kort
Korrekturlest totalt:	8.000 kort
Ferdigstilt totalt:	8.000 kort

Registreringskortene for steinalder og bronsealder

### *Status pr. 31.12.97:*

Størrelse på materialet:	3.000 kort
Konvertert totalt:	3.000 kort
Korrekturlest totalt:	3.000 kort
Ferdigstilt totalt:	3.000 kort

Registreringskortene for yngre jernalder

### *Status pr. 31.12.97:*

Total størrelse på materialet:	5.340 kort
--------------------------------	------------

### *Beskrivelse av arbeidet:*

Registreringskortene for eldre jernalder er skrevet inn av registreringsgruppe og korrekturlest i delprosjektet.

Mesteparten av tilretteleggingen, konverteringen og kvalitetskontrollen på de 11.500 registreringskortene for steinalder og bronsealder ble utført i regi av Bergen Museum, men en restanse på ca 3.000 registreringskort ble skrevet inn av registreringsgruppen og korrekturlest i delprosjektet.

Arbeidet med å sjekke innholdet på kortene opp mot gjenstandene de beskriver og foreta eventuelle nye faglige vurderinger, er påbegynt, men ikke fullført.

5340 registreringskort for yngre jernalder ble konvertert fra et ikke-operativt database-format (Mapper) til museets nåværende brukersystem (HyperCard). Delprosjektet påbegynte kvalitetssikringen av det konverterte materialet. Endelig tilrettelegging er lagt på is, da kvalitetssikringen viser at det mangler en del B-nummer.

### *Bruk av digitalt arkiv/database:*

Kortene er lagt i en gjenstandsdatabase som hovedsakelig benyttes internt i magasinet.



## ***Tilvekstfortegnelse 1850 - 1990***

### *Om arkivet:*

Magasinet for de forhistoriske samlingene inneholder ca 150.000 enkeltgjenstander. Tilvekstfortegnelse inneholder beskrivelser av disse gjenstandene samt deres funnsted og funnomstendigheter. Fortegnelsene, som er ordnet i numerisk rekkefølge, har blitt publisert helt siden andre halvdel av forrige århundre.

### *Status på materialet pr. 31.12.97:*

Total størrelse:	90 dokumenter, ca 4.500 sider
Konvertert totalt:	90 dokumenter, ca 4.500 sider
Korrekturlest:	90 dokumenter, ca 4.500 sider
Totalt ferdigstilt materiale:	90 dokumenter, ca 4.500 sider

### *Beskrivelse av arbeidet:*

Alle tilvekstfortegnelsene er konvertert til elektronisk tekst og tagget. Originalen under innskrivingen har i prinsippet vært den originale, håndskrevne tilvekstfortegnelse (hovedprotokollen). Ettersom innskriverne har skrevet inn teksten etter de publiserte, trykte tilvekstfortegnelsene, er den innskrevne teksten kontrollert mot hovedprotokollen. I tilfeller med divergerende tekst i hovedprotokoll og publisert utgave er hovedprotokollen fulgt. Endringer/tilføyelser skal være merket med "Tilføydelse fra/i hovedprotokoll:....". De taggedde tilvekstfortegnelsene er korrekturlest, rettet og parsert.

### *Beskrivelse av digitalt arkiv/database:*

Det konverterte materialet er lagret som SGML-tagget tekst.

### *Teknisk beskrivelse:*

Alle selvstendige filer skal ha såkalt "statusoversikt" øverst i dokumentet. Ut fra koden satt inn i taggen <STATUS>, ser en hvor langt filen er behandlet. Alle filene er navngitt etter den trykte, publiserte tilvekstfortegnelse årstall, samt en endelse som kan være innskrivers initialer e.l. Noen av tilvekstene er så store at de er delt i flere filer. Disse har et romertall i tillegg til et felles navn. I påvente av ferdigbehandling er filene foreløpig lagret ved Dokumentasjonsprosjektets filtjenere i Oslo.

### *Bruk av digitalt arkiv/database:*

Tilvekstfortegnelsene vil tilgjengeliggjøres i samråd med Bergen museum.

## ***Topografisk arkiv***

### *Om arkivet:*

Det topografiske arkivet inneholder skriftlig dokumentasjon (korrespondanse, rapporter, kart, tegninger, plansaker mm) angående faste og løse fornminner i museumsregionen. Arkivet omfatter 7.650 gårder i de 77 kommunene i museumsdistriktet og inneholder dokumenter helt tilbake til museets stiftelse i 1825 og fram til i dag. Dokumenter blir først skannet inn. Deretter blir dokumentets sted og saksomstendigheter registrert i et elektronisk registreringsskjema. Målet var et informasjonssystem, hovedsakelig til intern bruk, som skal lette tilgangen til det enkelte dokument. Et elektronisk tilgjengelig topografisk arkiv vil også øke mulighetene for søking etter sak, hendelse og sted.

*Status på materialet pr. 31.12.97:*

Total størrelse: trolig omlag	105.000 sider
Totalt skanna materiale:	97.139 sider
Gjenstår å skanne i 1998:	7.861 sider*
Totalt registrert materiale:	91.578 sider
Gjenstår å registrere:	13.422 sider*
Totalt korrekturlest materiale:	21.742 sider
Totalt opprettet materiale:	12.962 sider

\* Delprosjektet fortsatte i 2 måneder i 1998, finansiert av Bergen museum.

*Beskrivelse av arbeidet:*

Prosjektet ble satt i gang i februar 96 med 2 fagkonsulenter og 4 innskrivere; innenfor april var totalt 6 innskrivere engasjert. Ved starten av prosjektet var det mangler når det gjaldt både datamaskiner, nettverk og dataapplikasjon. Fagkonsulentene brukte mesteparten av arbeidstiden på disse områdene i tillegg til å prøve å gi innskriverne en liten innføring i det arbeidet som ventet dem. Skanninga kom i gang i midten av mars. Innskrivarane arbeidet i to grupper som delte arbeidsdagen to økter: skanning og repetisjon av Word/Windows, mest på egen hånd. Fra slutten av april jobbet 2 faste personer i skanninga, og prognosen for arbeidet som gjenstår, er utregnet på bakgrunn av deres produksjon.

Registreringen av dokument startet 22.4.96, med 4 innskrivere. Dataapplikasjonen var da ferdig, men det var behov for en god del justering av både applikasjonen og nettverksopplegget etter hvert som ulike problem meldte seg.

Den ene fagkonsulenten hadde 2 måneders permisjon i perioden april-juni. Derfor ble det i tillegg til løpende arbeidsoppgaver bare tid til å tilrettelegge materialet for skanning og registrering. Dessuten tok det tid å utarbeide en korrekturapplikasjon for databasen.

I midten av oktober 1996 var det såpass klart m.h.t. applikasjoner, opplæring, m.m. at det ble tid til å starte på korrekturlesingen av de registrerte dokumentene. En av fagkonsulentene gikk da over til å lese korrektur på heltid (i tillegg til ulike andre gjøremål). Korrekturlesingen gikk markert raskere enn registreringen, men det ble raskt klart at det var nødvendig med mer korrekturhjelp.

I første halvår 1997 gikk innleggelsen av Topografisk arkiv forholdsvis greit. Det var et par "flaskehalser" som forårsaket mindre forsinkelser. En av disse var at ordningen med innskrivere finansiert gjennom Kaja-midler opphørte. Dette førte til forsinkelser og problemer med å få nye innskrivere når eksisterende Kaja-kontrakter gikk ut. Etter hvert har en imidlertid fått innskrivere gjennom andre arbeidsmarkedstiltak, fortrinnsvis gjennom "sysselsetting i offentlig sektor". I 1997 var det bare tre av seks innskrivere som fortsatt gikk på Kaja-kontrakt.

Disse nye sysselsettingstiltakene fungerte godt. Kontraktene som var på tre måneder om gangen, kunne forlenges flere ganger. Dette gjorde at delprosjektet beholdt de samme innskriverne gjennom hele året, og dermed økte også hastigheten på registrering og skanning gjennom hele 1997.

Maskinene som innskriverne brukte, har etter hvert vist seg å være noe seine i forhold til de oppgavene de skal utføre. Det ble derfor bestilt mer internminne til disse høsten 1997.

Det har vært gjort forsøk på å få tak i et brukbart skanneprogram som også fungerer med gråtoner. Dette har foreløpig ikke lyktes. Et slikt program er viktig for å få skannet dokumenter som vanskelig lar seg skanne i 1-bit format.

En har i lengre tid sett at en ikke ville bli ferdig med hele arkivet innenfor prosjektets tidsramme. Det ble derfor på et tidspunkt bestemt at området Sunnhordland ikke skulle tas med i første omgang. Det viste seg imidlertid etter hvert at både hastigheten på skanning og innskriving økte så pass mye at man gikk i gang med å skanne Sunnhordland også. En ufullstendig versjon av arkivet vil dessuten være av liten vitenskapelig verdi. Pr. 31.12.97 er 75 av 77 kommuner skannet, mens rundt 72 kommuner er ferdig innskrevet.

#### *Arbeidsoperasjoner i delprosjektet*

Siden innskrivingsgruppen og fagkonsulentene er lokalisert i samme by, har det vært nær kontakt mellom enhetene. Arbeidet har vært organisert på følgende måte:

- 1 Fagkonsulentene har hatt ansvar for tilrettelegging og gjennomgang av dokumenter som ligger i Topografisk arkiv. Tilretteleggingen har bestått i sortering, opprydding og stempeling av dokumenter. Dokumenter og saker som hører sammen, har også blitt merket på en slik måte at innskriverne senere kan se at disse hører sammen. Hvert dokument (som kan inneholde mange sider) blir stemplet med et eget nummer som senere også brukes som filnavn på det ferdig skannede og lagrede dokumentet.
- 2 Etter at dokumentene er ferdig tilrettelagt, blir de skannet inn av to innskrivere. Dette arbeidet gjøres i umiddelbar nærhet av selve arkivet for at ikke dokumenter skal forsvinne fra arkivet.
- 3 Når dokumentene er skannet, blir de lagrede filene kopiert ned til en egen NT-server som tilhører delprosjektet.
- 4 Fra denne serveren blir siden aktuelle dokumenter overført via nettverk til de ulike innskriverne som er lokalisert i et annet bygg på universitetsområdet.
- 5 Når innskriverne har fått overført nye dokumenter blir disse først koblet opp mot deres innskriverapplikasjon. Etter at dette er gjort, leser hver innskriver gjennom hvert eneste dokument før opplysninger om disse skrives inn i databasen. Gjennomlesningen av dokumentene gjøres på kopier av originaldokumentene. Kopieringen av dokumentene er gjort av de som skanner inn dokumentene på museet.
- 6 Etter endt arbeidsdag legger hver innskriver over en kopi av den databasen (kommunen) de arbeider med, til prosjektets NT-server som står på museet. Her blir også de ferdige databasene lagret. Men før dette gjøres, skal hver innskriver lese korrektur på alt arbeidet han/hun har gjort.

#### *Beskrivelse av digitalt arkiv*

Da prosjektet startet opp, ble innskrivingsapplikasjonen laget i Access. Etter en stund viste det seg at kapasiteten til programmet var noe i underkant av behovet. Det ble derfor i en periode arbeidet ganske intenst med å lage innskriverversjon i Delphi. Denne ble imidlertid ikke ferdig før ut på sommeren 1997. Da var det så kort tid igjen av prosjektet at en bestemte seg for å vente med å gå over til denne nye applikasjonen. 73 av de totalt 77 kommunene var pr. 31.12.97 ferdig innskrevet. 72 av disse er lagret i Access, mens 1 kommune (Solund) er skrevet inn og lagret i Delphi.

### *Teknisk beskrivelse*

De skannede dokumentene ble lagret med sitt eget dokumentnummer+.tif som filnavn. Eks. 12863.tif. Siden filene er multipage-filer, kan en fil inneholde fra en til flere hundre sider. Det er ikke laget noe system for statusmerking av filene. Filene er i dag lagret på prosjektets egen NT-server. Denne serveren blir det tatt full backup av hver eneste uke, samt backup av oppgraderte filer to ganger i uken.

### *Videre arbeid*

Det har lenge vært klart at en komplett innlegging av Topografisk arkiv ikke vil bli ferdig innenfor prosjektperioden. Det sier seg selv at arkivet ikke er fullgodt før hele området som museet har ansvar for, er lagt inn. Det er også viktig å poengtere at selv om mesteparten nå er skannet og registrert, har det ikke vært tid til korrekturlesing og oppretting av det arbeidet som er gjort. Dette er et arbeid som må gjøres før en elektronisk versjon av arkivet kan tas i bruk i forvaltning og forskning. Foreløpig har Bergen Museum bevilget fire månedsverk til en ferdigstillelse av Topografisk arkiv. Dette betyr at en vil klare å få skannet og registrert det gjenstående materialet, men det vil bare være begrenset tid til korrekturlesning. Når dette korrekturarbeidet er ferdig, vil en ha en råversjon av et elektronisk topografisk arkiv. Foreløpig ligger de ulike kommunene på små, selvstendige databaser, og disse må derfor kobles sammen før det i det hele tatt vil være mulig å gjøre søk i hele arkivet. Det er heller ikke laget noe system for en fortløpende skanning og registrering av dokumenter som tillegges arkivet etter hvert. Med andre ord er det mye arbeid som gjenstår før en digital versjon av Topografisk arkiv kan brukes i forvaltning, undervisning og forskningsøyemed.

### *Bruk av arkivet*

Det har vært stor interesse for bruk av det ferdige digitale arkivet. Med sine mange helt nye søkefunksjoner vil det bare ta sekunder å hente ut informasjon som det tidligere tok uker og dager å finne frem til. Spesielt personale som arbeider med forvaltning, ser frem til å kunne benytte seg av de nye mulighetene, men også forskere og studenter har allerede henvendt seg for å få hjelp til å finne fram til informasjon.

I løpet av 1998 vil en også gå i gang med å tilrettelegge deler av Topografisk arkiv for bruk i grunnskolen. Nasjonalt læremiddelsenter har bevilget 250.000 kroner til utvikling og produksjon av en CD-ROM-plate på bakgrunn av dette datamaterialet som i dag er tilgjengelig.

## **Generelt: Oldsaksamlingen, Universitetet i Oslo**

Universitetets Oldsaksamling ble opprettet i 1829. Oldsaksamlingen var tidligere et eget universitetsinstitutt, men er nå del av Institutt for arkeologi, kunsthistorie og numismatikk. Som et av de fem arkeologiske landsdelsmuseene har Oldsaksamlingen forvaltningsansvar for Østlands- og Sørlandsfylkene; Oslo, Akershus, Østfold, Hedmark, Oppland, Buskerud, Vestfold, Telemark, Aust-Agder og Vest-Agder. Viking-skipshuset på Bygdøy er en del av Oldsaksamlingens museumsvirksomhet.

Delprosjektet ved Oldsaksamlingen begynte i 1993. I tillegg til tilrettelegging av tekst og korrekturarbeid, har fagkonsulentene også arbeidet med edb-basert tilvekstføring, programmert applikasjoner for bruk ved museet, konvertert eksisterende baser fra tidligere prosjekter ved Oldsaksamlingen og arbeidet med geografiske informasjons-systemer (GIS).

Delprosjektansvarlig: Professor Einar Østmo, IAKN, UiO

Fagkonsulent: Espen Uleberg

Korrekturlesere/andre medarbeidere i delprosjektperioden: Laila Buskum, Gro Anita Bårdseth, Hege Roaldset, Lillian B. Trinterud, Margrethe Simonsen, Lill Irene Uleberg, Geir Noddeland, Trude Tønnes, Ellen Anne Pedersen

IT-faglig ansvar: Jon Holmen

Kontaktperson ved fagenheten er fra 1.1.98: Databaseadministrator Espen Uleberg

Antall fagkonsulentårverk i perioden 1993 - 1997: 6,6

### ***Oldsaksamlingens hovedkatalog***

#### *Om materialet:*

Hovedkatalogen (tilvekstene) for Oldsaksamlingen i Oslo er en oversikt over gjenstandsmaterialet som har kommet inn til museet siden slutten av 1820-årene og til i dag. Hovedkatalogen beskriver alle gjenstandene ved museet. For de første årene finnes hovedkatalogen bare som håndskrevne protokoller. Hovedkatalogen er for en del publisert i form av trykte tilvekster, dels som del av andre tidsskrift og dels som egne hefter.

#### *Status på materialet pr. 31.12.97:*

Total størrelse:	5.700 sider
Konvertert totalt:	5.700 sider
Korrekturlest totalt:	5.700 sider
Ferdigstilt totalt:	5.700 sider *

\* Enkelte numre skal legges til; det genereres stadig ny tilvekst, som vil bli tagget opp.

#### *Beskrivelse av arbeidet:*

Innskrivingsarbeidet og tagging av tekst er gjort i registreringsgruppa i Sandefjord. Dette arbeidet ble avsluttet i februar 1997. I utgangspunktet var planen å skrive inn Oldsaksamlingens gjenstandsfortegnelse til og med 1990 ut fra trykt tilvekst og supplere denne med hovedkatalogen. Dette målet er nådd. Det er 5.700 sider som er ferdigstilt på denne måten.

I den grad det har eksistert trykte tilvekster, er teksten hentet fra disse. Der det ikke har vært trykte tilvekster, er opplysningene lagt inn fra den håndskrevne hovedkatalogen. Enkelte katalognumre er ikke detaljert beskrevet i den digitale versjonen på grunn av de trykte tilvekstenes summariske fremstilling. Det blir stadig generert ny tilvekst. Tilrettelegging og formatering av tilvekst som ikke var kommet under Dokumentasjonsprosjektet, vil bli gjort av databaseadministratoren ved IAKN, slik at det nye materialet vil bli koblet sammen med det som er konvertert i prosjektet. Databaseadministratoren vil også arbeide med komplettering av teksten ved å legge til opplysninger fra hovedkatalogen.

Tilretteleggingen har bestått i oppkopiering av nødvendig materiale og utarbeiding av standard for innskrivning. Teksten er blitt

SGML-kodet. I gruppene bestod første del av arbeidet enten i innskanning, OCR-behandling og oppretting av den innskannede teksten eller av innskrivning av tekst som ikke lot seg skanne på grunn av trykkkvalitet eller fordi det var håndskrift. Andre delen av arbeidet bestod av koding av teksten etter et taggeskjema utarbeidet av prosjektet.

*Beskrivelse av digitalt arkiv/database:*

Basen ligger i SGML-tagget tekstformat og er gjort tilgjengelig over Internett. Bruk av basene forutsetter også en del kunnskap om skiftinger i stedsnavn og gjenstandsbetegnelser siden originalteksten ble skrevet. Gjenstandsfortegnelsen er i stadig vekst. Den digitale hovedkatalogen er i hovedsak fullstendig, men det er enkelte numre som skal legges til og dessuten enkelte tilføyelser som er blitt gjort i hovedkatalogen. I tillegg genereres det stadig ny tilvekst som vil bli tagget opp slik at basen vil bli komplett.

*Teknisk beskrivelse:*

Innskrivningen ble gjort i bolker på ca. 20 sider. Filene fikk navn etter systemet uo<tilvekstaar>\_<sidenr\_i\_tilvekst>.lko for første gangs innskrivning, .2ko ved andre korrekturlesing.(uo1939\_230.2ko) Innsendte filer ble satt sammen år for år og fikk navn etter systemet <tilvekstår>.txt. (1939.txt) S sammensatte filer som ikke var fullstendig korrekturlest og opprettet, fikk en "u" i tillegg til årstallet i navnet. (1939u.txt) For filer med enkeltnumre ble det gjennomført at filen fikk navn etter første museumsnummer på fila. Ved taggingen ble de rene tekstfilene lagt på et område hvis navn viste til sted for taggingen og hvilken taggeomgang det var (txtsfj6). Autotaggingen la filene på et område som viste at filene var klare til å sendes (tilsfj6).

De taggete filene er samlet under katalogen taggpars, under hvilken det finnes en katalog som heter "tolket", der gjennomsette filer ligger, og en katalog som heter "klar", der filer som er ferdigtagget og SGML-sjekkert med Author Editor, ligger. Filer eksportert fra AuthorEditor ligger på et område som heter "valid".

*Videre arbeid:*

Det de fleste etterlyser er standardiserte geografiske navn ved søk i gjenstandsbasen. Det vil også være ønskelig å kunne ta ut informasjon i tabellform, og også å kunne ta ut større mengder informasjon. (Det har nå vært vanskelig å hente ut mer enn tre nummer av gangen.) Det bør også bli mulig å bla side for side i den elektroniske versjonen.

En videre bearbeiding av materialet som også innebærer en kvalitetsheving, vil dermed være å gjøre det mulig å søke på dagens geografiske navn og dagens gjenstandsbetegnelser. Dette vil for en del kunne gjøres ved synonymlister, og dels kan en støtte seg på nykatalogiseringer. Det vil også være ønskelig å oppdatere informasjonen med ekstra tilførsler i hovedkatalogen og også legge til annen informasjon som ble utelatt ved trykking av tilvekstteksten.

*Bruk av digitalt arkiv/database*

Hovedkatalogen er gjort tilgjengelig for fritekstsøk for studenter, forskere og allmennheten. De digitale arkivene er for en del tatt i bruk. Nettversjonen av hovedkatalogen brukes av studenter og ansatte til å finne gjenstander.

***Middelalderkort***

*Status pr. 31.12.97:*

Total størrelse	70.000 kort
Konvertert totalt:	70.000 kort

Korrekturlest i delprosjektet:	29.400 kort
Gjenstår å korrekturlese:	40.600 kort
Totalt ferdigstilt materiale:	29.400 kort*

\* Prioriteringen er blitt styrt av magasinrevisjonens behov. Tønsbergundersøkelsene TN, TO, TR, TW og TY er blitt prioritert framfor de andre. Kortene fra lavere prioriterte undersøkelser er ikke ferdig korrekturlest.

#### *Beskrivelse av arbeidet:*

Materialet fra Oslo var i hovedsak allerede innskrevet ved det som var Riksantikvarens utgravningskontor i Oslo, nå NIKU. En del av disse databasene er overført, og materialet fra Mindets tomt er tilrettelagt for arbeidet i museet. En del av materialet fra Oslo ble lest inn på data ved Humanistisk Datasenter i Bergen på 70-tallet, og skal ligge på magnetbånd der. Kontakt med Humanistisk Datasenter har vist at det kan bli like ressurskrevende å lete fram de gamle filene som å lese materialet inn på nytt.

Ved utgangen av 1996 var kortene fra alle større undersøkelser utenfor Oslo innskrevet og korrekturlest i innskrivergruppa. Ved avslutningen av prosjektet er omlag 29.400 kort fra undersøkelser i Tønsberg korrekturlest. Det gjenstår da omlag 37.800 gjenstandskort fra Tønsberg. I tillegg kommer omlag 15.000 kort fra kirkegravninger og mindre undersøkelser. I alt er det rundt 70.000 kort. Arbeidet med å tilrettelegge disse kortene for bruk ved magasinrevisjon er igangsatt. Prioriteringen mellom de ulike middelaldergravningene er gjort ut fra magasinrevisjonens behov.

#### *Hva er gjort med materialet:*

De større middelalderutgravningene skiller seg fra andre undersøkelser ved at hver gjenstand får sitt eget gjenstandskort med katalogiserings- og funnopplysninger. De registrerte opplysningene i hovedkatalogen og trykt tilvekst er gjerne svært summariske. I de tilfeller hvor museet hadde to sett av gjenstandskortene, ble det ene settet brukt ved innskrivningen. I de tilfeller hvor museet bare hadde ett sett, ble dette kopiert. I gruppene ble materialet skrevet inn i en database og korrekturlest. I første omgang ble det skrevet inn i en dBase III-applikasjon. Senere ble dette konvertert til Access, og arbeidet ble fortsatt i en Accessapplikasjon.

#### *Beskrivelse av digitalt arkiv/database:*

Gjenstandskortene ligger på flere Accessbaser og er gjort tilgjengelig for magasinrevisjonen på museet. Enkelte undersøkelser er ferdig korrekturlest og opprettet, for andre utgravningers vedkommende er ikke dette arbeidet fullført. Prioriteringen mellom de ulike undersøkelsene har vært bestemt av behovet på museet, idet ferdig innskrevne baser er stilt til disposisjon for magasinarbeidet.

#### *Teknisk beskrivelse:*

Gjenstandskortene fra bygravningene og andre middelalderundersøkelser ble skrevet inn i Accessdatabaser. Utskriftene som ble gjort for korrekturlesing i gruppene, er også tatt vare på og brukes ved siste korrekturlesing. Filene er samlet og ligger på katalogen "makat". På makat-området ligger underkataloger som grupperer kortene innenfor geografiske områder og emner og også underkataloger som viser hvor langt de er kommet i opprettingen. På denne måten ligger Tønsbergkortene på katalogen "tberg", som har underkataloger for undersøkelsene og en katalog som heter "rettet" hvor de ferdig korrekturleste kortene ligger.

På katalogen "makat" ligger også kataloger for andre databaser som prosjektet har konvertert fra andre systemer.(Svevollen, Halden, Rugtvedt)

*Videre arbeid:*

Accessbasene skal overføres til en Oracle-base hvor alt materialet som er overført fra NIKU allerede er lagt inn, og få et felles brukergrensesnitt. Korrekturlesing og oppretting av de undersøkelsene som ikke har vært prioritert innenfor prosjektet, vil innebære en kvalitetsheving av materialet.

*Bruk av digitalt arkiv/database*

Gjenstandskortene fra middelaldergravningene er tatt i bruk ved magasinarbeidet. Oracle-basen med materialet fra NIKU er allerede i bruk ved museet.

***Fotoarkivets kortarkiv***

*Status pr. 31.12.97:*

Total størrelse:	21.000 kort
Konvertert totalt:	21.000 kort
Korrekturlest i delprosjektet:	Siste korrektur overlates til fotoavdelingen.

*Beskrivelse av arbeidet:*

Kartoteket for fotoarkivet er ferdig innskrevet i en taggeapplikasjon. Det er en tekstdatabase med om lag 21.000 poster. Materialet lå på flere filer som nå er blitt skjøtt sammen, slik at det er bare én base å søke i. Det er skrevet en applikasjon slik at fotoavdelingen nå selv kan komplettere basen med nye bilder. En siste korrekturlesning av det innskrevne materialet er overlatt til fotoavdelingen, ettersom dette krever tilgang til negativarkivet.

*Hva er gjort med materialet:*

Kartoteket ble kopiert opp. Det ble skrevet en databaseapplikasjon som ble brukt ved innlegging og korrekturlesing i gruppa og skrevet inn i en database i innskrivergruppa.

*Beskrivelse av digitalt arkiv/database:*

Materialet foreligger i en tekstdatabase med om lag 21.000 poster.

*Teknisk beskrivelse:*

Kartoteket fra fotoarkivet ble skrevet inn som flere småfiler som er satt sammen til en fil. Denne basen ligger nå på katalogen på maskinen Ariadne. Den blir løpende oppdatert av fotoseksjonen.

*Videre arbeid:*

Søke- og rapporteringsrutinene kan bedres. Det vil være ønskelig å kunne skrive ut kartotek kort.

*Bruk av digitalt arkiv:*

Databasen er i bruk i Fotoarkivet.

***Konserveringsavdelingens arkiv***

*Status pr. 31.12.97:*

Total størrelse: omlag	15.000 sider
Konvertert totalt:	15.000 sider
Korrekturlest i delprosjektet:	Siste korrektur overlates til konserveringsavdelingen



*Beskrivelse av arbeidet:*

Innskrivningen av konserveringsavdelingens arkiv er fullført. Det er også laget rutiner slik at konserveringsavdelingens rapporter kan skrives direkte inn og lagres i tekstbehandling fram til en databaseapplikasjon for denne avdelingen er ferdigutviklet. Det gjenstår en del programmeringsarbeid før konserveringsavdelingen kan registrere sitt løpende arbeid direkte i en database knyttet til det innskrevne materialet. Korrekturlesning av dette materialet overlates konserveringsavdelingen.

Arkivet ble kopiert opp. Det ble utarbeidet en taggeapplikasjon som ble brukt ved innleggingen av materialet. Materialet ble skrevet inn i taggeapplikasjonen og korrekturlest i innskrivergruppa.

*Beskrivelse av digitalt arkiv/database:*

Konserveringsavdelingens arkiv omfatter 15.000 sider. I samarbeid med konserveringsavdelingen er det utarbeidet en database for fortløpende innlegging av nye konserveringsrapporter. Denne databasen skal knyttes sammen med de konverterte rapportene.

*Teknisk beskrivelse:*

Konserveringsavdelingens arkiv er samlet på katalogen "konservering"

*Videre arbeid:*

De taggete tekstene bør samordnes med det nye skjemaet som skal brukes ved konserveringsavdelingen. Filene skal overføres Oracle-basen, som også muliggjør kobling direkte til gjenstandsregisteret.

*Bruk av digitalt arkiv:*

Det konverterte arkivet er i bruk i Konserveringsavdelingen.

***Annet, Oldsaksamlingen***

Registeret over faste fornminner er overført til NIKU, slik at denne koblingen oppnås gjennom samarbeid mellom Dokumentasjonsprosjektet og Norsk Institutt for Kulturminneforskning. Dokumentasjonsprosjektet har også samarbeidet med prosjekter ved Oldsaksamlingen i forbindelse med databaser og kartprogram til bruk under arkeologisk feltarbeid.

Det er også utarbeidet en database for Hamarundersøkelsene. Delprosjektet konverterte data samlet inn på papir i felt, og databasen er brukt ved det videre arbeidet med katalogisering og rapportskrivning fra undersøkelsene ved Domkirkeruinen på Hamar.

Ustinowsamlingen har de siste årene fått midler til en gjennomgang av materialet, og har også fått en stilling ved Universitetets Oldsaksamling. Delprosjektet har utarbeidet en database som brukes ved katalogisering og dokumentasjon av samlingen.

Det er gitt brukerstøtte i forbindelse med digital innmåling i felt og kartproduksjon i forbindelse med etterarbeid til pågående prosjekter ved Oldsaksamlingen.

Arbeidet med bruk av EDB i felt er videreført. Det er arbeidet med opplæring, igangsetting og brukerstøtte til de utgravningsprosjektene som har ønsket å ta i bruk elektronisk innmåling av strukturer i felt. Det har også vært arbeidet med GIS i forbindelse med etterarbeid, både ved å levere ferdige kart og ved å gi brukerstøtte til prosjekter som ønsket å arbeide på egen hånd med dette analyseverktøyet. Dette gjelder

særlig Gardermoprojektet, Dobbeltsporprosjektet, Veienundersøkelsene, Riksantikvarens undersøkelser på Sørenga, Gamlebyen i Oslo, Dobbeltsporundersøkelsene i Vestfold, og utgravningene på Moer i Ås. Delprosjektet har også levert kartmateriale til rapporter som er blitt trykt i Oldsaksamlingens VARIA-serie.

*Videre arbeid:*

For alle deler av materialet gjelder at basene skal forenes i en felles gjenstandsdatabase for hele samlingen, med koblinger mot konservering og eventuelt nye deler av samlingen som blir konvertert, f.eks. Topografisk arkiv.

**Generelt: Fagenhet for arkeologi, Tromsø museum, Universitetet i Tromsø**

Fagenhet for arkeologi har som sine hovedoppgaver å drive forskning og formidling av forhistorie og middelalder i hele Nord-Norge nord for Polarsirkelen. I dette området har Tromsø Museum ansvaret for utgravninger, med Fagenhet for arkeologi som fagenhet. Alle jordfunne gjenstander eldre enn 1537 som kommer fram i området, tilfeldig eller ved utgravning, skal inn til museet for faglig forsvarlig oppbevaring, innbefattet konservering.

Delprosjektansvarlig: Stipendiat Tora Hultgren, Tromsø museum , UiTø

Fagkonsulent: Toini Thommessen

Korrekturleser/tagger: Bjørg H. Andersen

IT-faglig ansvar: Jon Holmen

Antall fagkonsulentårsverk i perioden 1995 - 1997: 3,5

Kontaktperson i fagenheten fra 1.1.98: Fagenhet for arkeologi, Tromsø museum

***Fotolister***

*Om materialet:*

Fotolistene er en oversikt over slides og svart/hvitt-fotos som er dokumentasjon over arkeologiske utgravninger og registreringer. De originale listene foreligger dels som håndskrevne lister, dels som maskinskrevne.

*Status pr. 31.12.97:*

Total størrelse på materialet:	ca 35.000 bilder
Konvertert totalt:	ca 35.000 bilder
Korrekturlest totalt:	ca 35.000 bilder
Ferdigstilt totalt:	ca 35.000 bilder

*Beskrivelse av arbeidet:*

Materialet ble skrevet inn i innskrivingsenheten, korrekturlest der og i delprosjektet.

*Beskrivelse av digitalt arkiv/database:*

Fotolistene ble lagt inn på en Accessbase i 1995. Denne basen er ferdig konvertert, men trenger noe redigering.

## *Tilveksten/Funnkatalogene*

### *Om materialet:*

Tilveksten ved Tromsø Museum foreligger primært som håndskrevne protokoller fra 1874 og frem til i dag. Den er etter ca 1940 ikke ført etter årganger, kun etter løpende museumsnumre (Ts.-nr.). De originale katalogene er ført dels som "fritekst", dels på ulike typer skjema, dels på små arkivkort. Av dette materialet er noe arkivert i form av håndskrevne protokoller, noe i ringpermer og noe som innbundne stensiler og/eller kopier.

Ts.-nr. 1 - ca 4.000 finnes i trykt form som redigerte utgaver av de håndskrevne protokollene i følgende fora: "Aarsberetning for Foreningen til norske Fortidsmindesmerkens Forening" (1879 - 1901), "Tromsø Museums årshæfter" (1902 - 1940) og "Oldtiden, tidsskrift for norsk forhistorie" (1910 - 1940). Tilveksten er etter den tid ikke utgitt i trykt form.

### *Status pr. 31.12.97:*

Total størrelse på materialet: ca 17.000 katalogsider, tilsvarer 8.300 filsider

Konvertert totalt:	ca 16.500 katalogsider, tilsvarer 8.000 filsider
Gjenstår å konvertere:	ca 500 katalogsider, tilsvarer 300 filsider*
Korrekturlest totalt:	ca 16.500 katalogsider, tilsvarer 8.000 filsider
Gjenstår å korrekturlese:	ca 500 katalogsider, tilsvarer 300 filsider
Ferdigstilt totalt:	ca 16.500 katalogsider, tilsvarer 8.000 filsider
Kodet:	3.200 filsider (størrelsen er angitt i ukodede sider)

Gjenstår å kode: 5.100 filsider\*\*

\* restansen gjøres ferdig ved Registreringsentralen i Indre Salten 1998.

\*\* Kodingen av katalogen fortsatte i 2 mnd i 1998 finansiert av Dokumentasjonsprosjektets etterorganisasjon.

### *Beskrivelse av arbeidet:*

Innholdsmessig varierer katalogene sterkt, både i form og terminologi, og ikke to kataloger er like. Dette skyldes fravær av en overordnet standard for katalogiseringspraksis. Den enkelte har stort sett katalogisert det arkeologiske materialet etter sitt eget system. Man vil oppdage variasjoner og inkonsekvenser både innenfor og mellom de enkelte katalogene. Det er også "hull" i rekkefølgen av Ts.-nr., ettersom en del numre står ubrukt (materialet er ennå ikke katalogisert). Disse faktorene har gjort det svært problematisk å holde oversikten under konverteringsprosessen. Det var også nærmest umulig å anslå omfanget av materialet.

Den håndskrevne versjonen av tilveksten ble kopiert i sin helhet, og en del av dette materialet måtte tilrettelegges og redigeres grunnet den store variasjonen i form, kvalitet, språk, lesbarhet og terminologi. Ts.-nr. 1 - 4.000 er for en stor del konvertert etter den trykte utgaven (Tromsø Museums årshæfter) grunnet vanskelig håndskrift i protokollene. Da dette er redigerte utgaver av originalene, vil man kunne finne mer utfyllende opplysninger i de håndskrevne protokollene.

En del av de håndskrevne protokollene ble i sterkt redigert og normalisert form lagt inn på en database på 1980-tallet i Tromsø. Selve basen i elektronisk form er borte, men det foreligger en hel del matriseutskrifter. Utvalget av Ts.nr. og tidsepoker virker tilfeldig og er ikke gjort i rekkefølge. Disse listene ble kopiert og sendt til Registreringsentralen i Fauske. Det viste seg imidlertid at disse listene var ufullstendige og dessuten

svært tidkrevende å skanne og registrere for innskriverne. Arbeidet med disse listene ble derfor stanset. Det innskrevne materialet av denne typen vil bli lagt på et eget område i den ferdige basen, som en slags dokumentasjon på hvordan synet på tilveksten har variert opp gjennom årene. Det finnes altså inntil tre ulike versjoner av visse deler av tilveksten; håndskrevet, trykt og som matriseutskrifter.

Selve konverteringen har blitt utført mest mulig opp mot originalteksten. Det er et valg som ble gjort av både bevaringsmessige, praktiske og tidsmessige årsaker. Å redigere alle tilvekstkatalogene var en umulig oppgave og lå ikke innenfor prosjektets rammer. Materialet er dessuten historiske dokumenter med en stil og et innhold som bør bevares.

Tilrettelegging og redigering ble utført av fagkonsulent i samarbeid med delprosjektansvarlig. En del av teksten måtte merkes, tydes og til dels også oversettes fra engelsk. Svensk og dansk tekst måtte tydes, forklares og i visse tilfeller rettes opp, men i størst mulig grad bevares som originaltekst. Fagkonsulenten formulerte et standardoppsett for innskriverne. Standarden omfattet en overskrift med rekkefølge på opplysninger som fantes øverst på de skjemaene som er mest brukt i katalogiseringen. Dette opplegget fungerte greit.

Problemet under arbeidet var tilvekstmaterialets heterogene karakter, som skapte vansker med tyding, tolkning og tilpasning til standarden for innskriverne. Å anslå omfanget av materialet og å få opp tempo og motivasjon på registreringsarbeidet var av disse årsakene hele tiden problematisk.

Aktuelle arbeidsoperasjoner på registreringsentralen var skanning og revidering av de deler av teksten som var egnet for dette, samt vanlig registrering og korrektur av håndskrevet materiale, og alt ble lagt inn på egne filer som ren tekst. Filene ble deretter returnert til Tromsø Museum både på diskett og som utskrifter for korrektur, som primært ble utført av fagkonsulenten. Tilbakemeldinger til sentralen ble gitt på et eget skjema. Alle filene ble korrekturlest og ført opp på disse skjemaene.

En DTD, eller en manual, for taggingen av tilveksten på Tromsø Museum ble utarbeidet i samarbeid mellom programmerer og fagkonsulenten. Denne manualen var stort sett identisk med Bergensmanualen.

Taggingen av materialet skulle etter den opprinnelige planen utføres på Registreringsentralen i Fauske. Det viste seg imidlertid at det ville ta for lang tid med opplæring og tagging på sentralen. Innskrivingsarbeidet måtte prioriteres for å komme i mål. Høsten 1996 ble derfor en erfaren person, Bjørg H. Andersen i Registreringsgruppa i Sandefjord, engasjert til å utføre så mye tagging som mulig. Dette arbeidet var i full gang fra årsskiftet 1996-97. Korrektur og tilbakemeldinger ble utført av fagkonsulenten. Tekstbehandlingsprogram (WP Windows og Word) samt Author Editor er anvendt under konverteringsarbeidet. Access er anvendt i forbindelse med innleggingen av fotolistene.

#### *Beskrivelse av digitalt arkiv/database:*

Materialet foreligger som delvis SGML-formatert tekst og vil foreløpig bli lagt ut som fritekstsøk.

#### *Teknisk beskrivelse:*

Tekstfilene ble kalt opp etter det museumsnummer som var først på filen, og deretter status. F.eks. Ts.\_5777.Ko3 betyr at dette er tredje korrektur fra sentralen. Skannede tekstfiler ble merket slik: Ts.\_5777.Rev, som betyr revidert. Filnavnet ble justert etter antall korrekture. Hver fil kan inneholde ett eller flere museumsnumre, alt etter antall undernumre. For oversiktens skyld var filene sjelden større enn 5 utskriftsider. Filene er lagret på Dokpros maskiner i Oslo. Backup finnes på disketter ved Dokpro i Oslo.

#### *Videre arbeid:*

De 4500 utskriftsidene som nå mangler manuelt utførte detaljtagger vil kunne autotagges av etterorganisasjonen. Dette innebærer en bedring av grensesnittet, fordi en slik grovtagget versjon av denne delen av teksten i første omgang vil bli lagt ut på nettet som fritekstsøk med en del søkbare kategorier.

Selve katalogiseringen ved arkeologisk avdeling på Tromsø Museum bør tilpasses en form som letter digitalisering av arkivet. Det vil si at man bør utarbeide en standard med oversiktlige opplysninger om funn og funnkontekst, slik at konvertering kan utføres på en smidig og rask måte. Korrekt utfylling av overskriften på de allerede eksisterende skjemaene ved avdelingen vil kunne være tilstrekkelig. Man bør ved arkeologisk avdeling diskutere om flere opplysninger skal med i denne overskriften.

#### *Bruk av digitalt arkiv/database:*

Brukere av tilvekst- og fotoliste-basene vil primært være ansatte og studenter ved Arkeologiseksjonen på Universitetet i Tromsø, samt ansatte ved Tromsø Museum og ved Troms Fylkeskommune. Andre potensielle brukere vil være lekfolk med interesse for funn og gjenstander tilknyttet hjemsted og lokalmiljø, samt lærere for undervisningsformål i skoleverket. Fotolistebasen vil trolig være mindre interessant for de to sistnevnte brukergruppene.

Når det gjelder markedsføring av basene har det i Tromsø-miljøet vist seg at jevnlig og hyppig gjentatte informasjonsopplegg i de aktuelle miljøene (foredrag, demonstrasjon av demo/baser, seminarer) vil fungere bedre enn et par sporadiske seanser i året. F.eks. kan månedlige seminarer der folk kan prøve seg på nettet selv, være en god måte å sette folk i gang på. Utførlig opplæring og oppfølging i basenes konkrete bruk og muligheter er helt avgjørende. Dette gjelder først og fremst innenfor arkeologimiljøene. Interesserte lekfolk vil ha behov for informasjon om hva som finnes tilgjengelig, og hvor de kan finne basene. Bruk av media vil være aktuelt i denne sammenheng.

#### **Generelt: Institutt for arkeologi og kulturhistorie, Vitenskapsmuseet, NTNU**

Institutt for arkeologi og kulturhistorie driver grunnforskning og anvendt forskning i kulturhistoriske fag, samt tverrfaglig forskning på miljøendringer. Instituttet er inne i en utviklingsfase innenfor forskning i konserveringsteknologi. Instituttet utfører forsknings- og utredningsoppgaver hjemlet etter lov om kulturminner. Ivaretagelse av de vitenskapelige samlingene er en viktig del av instituttets virksomhet.

Delprosjektansvarlig: forskningsassistent Geir Grønnesby, Vitenskapsmuseet

Fagkonsulent: Anne Birgitte Høy-Petersen

IT-faglig ansvar: Jon Homen

Tilknyttede enheter/personer: Avdelingsingeniør, edb Jon-Arne Reitan, Vitenskapsmuseet

Antall fagkonsulentårsverk i perioden 1996 - 1997: 1,25

Kontaktperson i fagenheten fra 1.1.98: Anne Birgitte Høy-Petersen, Geir Grønnesby, underdirektør Kurt Alterskjær

### ***Vitenskapsmuseets tilvekster***

Om materialet:

Vitenskapsmuseet har ført protokoll over tilveksten i årene 1871-1981. Størstedelen av disse tekstene (fortrinnsvis det arkeologiske materiale) har blitt utgitt hvert år i trykt form i forskjellige tidsskrifter. En mindre del - især beskrivelser av gjenstander fra nyere tid eller av utenlandsk opprinnelse - er aldri utkommet i trykt form, så her finnes kun de håndskrevne tekster.

*Status pr. 31.12.97:*

Total størrelse på materialet:	3.504 trykte sider fordelt på 110 hefter 1.700 håndskrevne sider, i alt: 5.204 sider
--------------------------------	---

Konvertert totalt:	5.204 filsider
Kodet totalt:	460 filsider
Gjenstår å kode:	3.968 filsider*
Korrekturlest totalt:	1.475 filsider, 60 sider kodet tekst
Gjenstår å korrekturlese:	2.850 filsider*

\* Delprosjektet fortsetter i 1998 finansiert av Vitenskapsmuseet.

### ***Beskrivelse av arbeidet:***

Jobben med tilveksten måtte begynnes med en bred gjennomgang av katalogtekstene, såvel de trykte som håndskrevne, for å gi et inntrykk av hva tekstene inneholdt og hvorledes tekstene var satt opp. Da ikke alle tilvekstnumre var å finne i de trykte tilvekster, forestod et større arbeide med å sammenligne den trykte og håndskrevne teksten for å finne ut hvilke numre som ikke var tatt med i den trykte tilveksten. For å få et overblikk over det faglige innholdet av tekstene med henblikk på senere koding, var det nødvendig å nærlese representative deler av katalogene. Foruten hovedkatalogen har Vitenskapsmuseet også diverse særprotokoller som ikke finnes i trykt form. De viktigste av disse er likeledes blitt gjennomgått for innhold og oppsetning.

Det er lang vei fra Trondheim (delprosjekt) til Drammen (innskrivingsgruppe), så det har vært nødvendig med mange skriftlige instruksjoner, slik at innskriverne bestandig har kunnet holde seg til disse og kunnet klare den daglige driften på egen hånd. I tillegg har det sikret en ensartet behandling av materialet. Innskriverne har først og fremst hatt et slags sjekklister, hvor de har kunnet følge arbeidsprosessene fra start til slutt. I tillegg har de hatt forskjellige rettledninger til bruk for konvertering av diverse tekster.

Den trykte delen av hovedkatalogen er blitt skannet og redigert etterpå, mens den håndskrevne delen er blitt skrevet inn for hånd etter samme oppsetning som den trykte. Til slutt er de to versjonene satt sammen til en og samme, slik at tilveksten for hvert år er blitt komplett. Til bruk i dette arbeidet er det blitt utarbeidet forskjellige veiledninger.

Foruten hovedkatalogen er museets særprotokoller (håndskrevne) skrevet inn. Dette har skjedd ut fra en veiledning til hver protokoll, da hver av dem er satt opp på en egen måte.

#### *Koding:*

Vitenskapsmuseet er kommet med i Dokumentasjonsprosjektet som det siste av museene, hvilket har vært en stor fordel for VM, da alle arbeidsprosesser har blitt utprøvd. Dette gjelder naturligvis også kodingen. Sammen med en grundig gjennomgang av tilvekstteksten til VM har kodeforlegget til Bergen, Oslo og Tromsø dannet utgangspunkt for kodingsforslaget til VM. Det ferdige kodeforslaget inneholder elementer fra Bergen, Oslo og Tromsø samt særlige innslag som er spesielle for Vitenskapsmuseet.

Materialet er kodet slik at opplysningene faller i 6 deler: Administrative data, innledning, nummeravsnitt, flernummeravsnitt, umerket, etnonym. Under hver av de 4 første hovedavsnitt finnes underordnede tagger i 1-4 nivåer. De siste to avsnittene har ingen underordnede tagger, da de kun avmerker tekst som må viderebearbeides senere.

#### *Innskrivingsenheten:*

Innskrivingsgruppene har utført det egentlige konverteringsarbeid med skanning, redigering, innskrivning og korrektur av tilvekstteksten.

Hver gang en ny arbeidsprosess har skullet starte, har fagkonsulentene reist til Tønsberg/Drammen for å lære opp innskriverne og gått igjennom veiledningen og forskjellige eksempler samt vært til stede, mens gruppen har startet opp på den nye oppgaven. Etterpå tok gruppelederen over. Har det oppstått problemer, har gruppelederen tatt kontakt med fagkonsulentene, og problemet blitt løst pr. tlf. eller faks.

De trykte tekstene er blitt skannet (Omnipage Professional Windows Version 5) og dernest lest inn i Word (version 6), hvor skannerens lesefeil er blitt rettet opp. I tillegg er det blitt lest korrektur 3 ganger på tekstene, som har bestått av ca. 10 tekstsider på hver fil, med etterfølgende oppretting. De håndskrevne tekstene er blitt skrevet inn for hånd i Word (10 tekstsider pr. fil), korrekturlest og rettet opp 3 ganger. Etter 3. korrekturen er både de "trykte" og "håndskrevne" filene sendt til Vitenskapsmuseet, hvor de er blitt korrekturlest. Hvis feilprosenten har vært lav, er feilene blitt rettet opp i Trondheim, men hvis det har vært mange feil har filene blitt returnert til Drammen, hvor en ny korrektur er gjennomført. Filene er atter blitt sendt til Trondheim, hvor siste tekstkorrektur er utført. Dernest er de opprettede, korrekturleste enkeltfilene fra de forskjellige tilvekstår blitt sendt til Drammen, hvor de er blitt satt sammen år for år til hele årgangsfiler.

I løpet av hele innskrivingsprosessen har hver enkelt innskriver ført et statusskjema, slik at man kunne følge hvor langt behandlingen av hver enkelt fil er kommet, og hvilken person som har behandlet den. Taggingen av de sammensatte årgangsfilene er påbegynt og fortsetter i 1998. På nåværende tidspunkt er 21 årganger ferdigtagget, men ikke korrekturlest.

De rene tekstfilene inneholdende hele årganger, er blitt tagget helt fra grunnen i Drammen i motsetning til andre delprosjekter, hvor en har automatisk tagget en del av opplysningene og latt innskriverne kontrollere disse taggene og dernest selv påført resten av taggene. Det har vært et ønske fra innskriverne i Drammen, da de syntes det ble for mye feil ved autotaggingen. Etter endt tagging er filene blitt korrekturlest og opprettet to ganger og deretter sendt til Trondheim. Her er det i første omgang lest stikkprøvekorrektur, og filene er blitt sendt til oppretting og korrekturlesing i Dram-

men. Status over filbehandlingen finnes direkte på filene under 'Adminstrasjonsdata'. Gruppelederen har hver måned utarbeidet et skjema over hvilket og hvor mye materiale som er blitt behandlet, og hvor mange personer gruppa har bestått av. Videre har gruppelederen hatt ansvaret for forsendelse av filer til Trondheim.

#### *Delprosjektet:*

På delprosjektet i Trondheim har størstedelen av det forberedende arbeidet foregått: Manualer og veiledninger til innskrivning og tagging er skrevet her. Tilretteleggingen av DTD'en har foregått henholdsvis i Trondheim og Oslo i samarbeide med programmereren. Den trykte tekst er sendt til Drammen i original eller i fotokopiert utgave. Kopieringen er foretatt av en skoleelev.

Videre er det gjort et større arbeid med å sammenligne den trykte og den håndskrevne tekst til hovedkatalogen for å finne beskrivelser av gjenstander som ikke er med i den trykte tilveksten. Det er blitt laget lister over manglende numre, og 1-2 skoleelever har kopiert tekstene til disse. Kopiene er blitt sendt til Drammen, hvor de er blitt skrevet inn. Videre er de viktigste særprotokollene blitt kopiert og sendt til Drammen. Fagkonsulenten har lest korrektur på de håndskrevne tekstene fra 1871-1906, idet håndskriften for disse årene er vanskelig å lese. Innskriverne har foretatt sistekorrektur på resten av den håndskrevne teksten, mens fagkonsulenten har lest stikkprøvekontroll. Langt de fleste filene er blitt sendt i retur til Drammen for oppretting og fornyet korrekturlesing.

På den skannede, redigerte teksten er det veldig lite feil, så her har det kun blitt foretatt stikkprøvekorrektur ved å lese hver 5. side. På de taggede tekstene er det foretatt stikkprøvekorrektur. Sammen med korrekturene, som er sendt til Drammen, har det fulgt med en tilbakemelding til hver enkelt innskriver om hvilke ting som må rettes, og hvorfor.

De ferdig sammensatte årgangsfilene som mottas fra Drammen, kikkes igjennom innen de endelig lagres. Disse ferdige tekstfilene vil legges ut på museets interne nett slik at alle ansatte kan ha mulighet for å lage tekstsøk.

En liten del av de ferdige årgangsfilene er blitt tagget i Drammen, og fagkonsulenten har lest stikkprøvekorrektur på en fil fra hver av innskriverne. De har alle fått en tilbakemelding på denne, slik at de har kunnet tagge videre uten å begå store feil. Foruten direkte korrekturarbeid og filadministrasjon er det laget en oversikt over innhold og omfang av arkivene til Vitenskapsmuseet, som danner grunnlaget for det videre arbeidet med arkivene.

#### *Beskrivelse av digitalt arkiv/database:*

Hovedkatalogen er skrevet inn 100% og finnes som ren tekst, men deler av tekstkorrekturen gjenstår. Særprotokollene er skrevet inn og finnes som ren tekst, men størstedelen av tekstkorrekturen og tagging gjenstår. Det konverterte materialet er lagret som delvis SGML-tagget tekst.

#### *Teknisk beskrivelse:*

Alle filer som inneholder den trykte teksten, har fått navn etter det pågjeldende tilvekstår samt et løpenummer, f. eks.: 1915\_1 (inneholder de første 10 sider av 1915). Alle filer som inneholder den håndskrevne teksten, har fått navn etter det pågjeldende tilvekstår samt en bokstav, f. eks.: 1915\_a (inneholder de første 10 sider av den håndskrevne teksten, som mangler i den trykte teksten). Når de "trykte" og "håndskrevne" filene for et tilvekstår er satt sammen til én årgangsfil, får den navn etter det



pågjeldende år. Det samme gjelder for de taggedede filene - disse lagres i en annen mappe. Status for arbeidsgangene for de rene tekstfilene skjer på manuelle skjemaer, mens status ved taggingen føres direkte på filene. Filene er lagret som rene tekstfiler på serveren til Vitenskapsmuseet. De finnes i to forskjellige mapper, en for årgangsfilene i ren tekst og en for de taggedede filene.

*Videre arbeid:*

- Ferdiggjøring av tekstkorrekturen.
- Ferdiggjøring av taggingen og korrekturen på denne.
- Utarbeidelse av opplegg til tagging av etnografisk materiale og andre ikke jordfunne gjenstander. I originalprotokollene finnes det massevis av tilføyelser som er kommet til etter at originaltekstene ble skrevet inn. F. eks. opplysninger om funnsted som man ikke var kjent med ved anmeldelsen, opplysning om deponering, senere bestemmelse av oldsakstype, opplysning om journalnr. osv. Disse opplysningene er også viktige å få registrert.
- Innskrivning og tagging av katalogene til Myntsamlingen, samt omlegging av museets øvrige arkiver.
- Det er viktig at det tas hensyn til kobling mot museets øvrige baser: Kulturminneregisteret, som er under revisjon i øyeblikket, samt Magasinbasen (tekst og bilder) og Konserveringsbasen, som er under oppbygging.
- Når korrekturen av den rene teksten er ferdig - i løpet av januar 1998 - vil alle de filene som ikke er blitt tagget på nåværende tidspunkt, bli autotagget med 'museum-snr', slik at det blir mulig å søke frem den rene, utaggede tilvekstteksten til hvert enkelt nummer. Når korrekturen på dette er overstått, vil fagkonsulenten muligens autotagge litt flere opplysninger innen den primære tagging gjøres ferdig.

*Bruk av digitalt arkiv/database:*

Fagkonsulenten har holdt forelesninger for studentene på Instituttet i bruk av Dokumentasjonsprosjektets databaser. Allerede så snart den rene tekst legges ut på nettet, vil det bli en liten introduksjon til hvordan det vil være mulig å gjøre fritekstsøk ved hjelp av Words søkefunksjon. Når teksten er tagget, vil det bli kurs for museets medarbeidere og evt. oppsøking av kolleger, som må ha individuell opplæring. Distribuering av manual.

Markedsføring av bruk av databasen utenfor museet vil være helt avhengig av hvilke begrensninger denne gruppe brukere vil få. Det avventes resultater fra den nasjonale diskusjonen om dette emnet.

Sør-Trøndelag Fylkeskommune forbereder et forsøk i Ørlandet kommune, hvor alle opplysninger i Kulturminneregisteret for denne kommune legges ut på Internett og blir tilgjengelig for alle og enhver. Vitenskapsmuseet vil nøye følge utfallet av dette forsøket.

## **Runologi**

### **Generelt: Runearkivet, Oldsaksamlingen, IAKN, UiO**

Runearkivet ved Universitetets Oldsaksamling i Oslo er et landsomfattende arkiv og eneste instans med opplysninger om alle norske runeinnskrifter.

Runearkivets samlinger består av forskjellig slags dokumentasjonsmaterialer for innskriftene. Kjernen i samlingene er dokumentasjonsmapper for hver norsk runeinnskrift eller mindre gruppe av innskrifter. Denne samlingen er delt inn i grupper og ordnet etter publiseringsnummer, fortløpende registreringsnummer eller geografisk beliggenhet/funnsted.

Dokumentasjonsmappenes innhold varierer en del, også fra delarkiv til delarkiv, men består stort sett av:

- 1) originalsider av omtalen i Niyr (for alle innskrifter som allerede er utgitt der),
- 2) kortfattede registreringer,
- 3) forskningsnotater og utkast til artikler,
- 4) kopier av vitenskapelige artikler (enten artikler som presenterer innskriften eller de som kan være av interesse ved tolkning av innskriften) eller litteraturhenvisninger,
- 5) enkelte dokumenter om gjenstandsforvaltning, samt
- 6) eventuelle tegninger og
- 7) bilder i svart/hvitt av gjenstandene eller innskriftene. I tillegg har Runearkivet en stor, uoversiktlig samling av vitenskapelige avtrykk, et lite antall gjenstandskopier, etterlatte papirer etter tidligere runeforskere (som f.eks. Magnus Olsen, Carl J. S. Marstrander og Aslak Liestøl) inkludert en viktig samling "forskjellige" notater, hovedsakelig fra forrige århundre, et håndbibliotek og småtrykksamling, samt en omfattende diassamling som nå blir ordnet.

Delprosjektansvarlig: Professor James Knirk, Institutt for arkeologi, kunsthistorie og numismatikk, UiO

Fagkonsulenter i delprosjektperioden: Anne Asplin, Mette Ekker og Knut Lunde

Korrekturleser: Jonas Nordby

IT-faglig ansvar: Christian-Emil Ore

Tilknyttede enheter/personer: Humanistisk datasenter, Universitetet i Bergen

Antall fagkonsulentårsverk i perioden 1994 - 1996: 2,2

Kontaktperson i fagenheten fra 1.1.98: Professor James Knirk

### ***Runer***

#### *Om materialet:*

Det totale antallet norske runeinnskrifter registrert ved Runearkivet i Oslo er ca. 2.000. Dette inkluderer også et relativt stort antall innskrifter med runelignende tegn (ca. 5%). I tillegg kommer rapporter om "runer" som er blitt avkreftet. Totalantallet fordeler seg på følgende grupper: N-innskrifter: 773, B-innskrifter: 500, A-innskrifter: 330, Y-innskrifter: 300, Æ-innskrifter: 60, Avkreftede: 300

N-, B- og A-innskrifter utgjør vikingtids- og middelalderinnskrifter. N-innskrifter er de som er trykt i korpusutgaven Norges innskrifter med de yngre runer (NiyR) bd. 1-6, totalt ca. 1800 store trykksider. Nesten alle norske runeinnskrifter fra vikingtiden og middelalderen som var kjent i midten av 1950-årene, ble trykt i bd. 1-5 (1941-60). B-innskrifter er innskrifter fra Bryggen i Bergen som ble funnet etter 1955, dvs. for sent til å bli inkludert i de første fem bind av NiyR. Bind 6 av NiyR (1980-90) utgjøres av 171 innskrifter fra Bryggen som således er blitt N-innskrifter.

A-innskrifter er andre innskrifter fra resten av landet, funnet for sent til å bli inkludert i NiyR bd. 1-5. Det er noen store grupper fra utgravninger i middelalderbyer, f.eks. 110 fra Trondheim, 60 fra Gamlebyen i Oslo, 30 fra Tønsberg. Flere av disse er sitert eller presentert i oversiktsartikler; flere av de andre er trykt i mindre artikler.

Y-innskrifter er de yngste innskriftene, dvs. etterreformatoriske (moderne) runeinnskrifter. Registreringene inkluderer en del etterreformatoriske innskrifter med latinske bokstaver også.

Æ-innskrifter er de "ældre" eller urnordiske runeinnskrifter, skrevet med den eldre runerekken i tidlig jernalder fram til vikingtiden.

"Avkreftede" betegner innrapporteringer av "runeinnskrifter" som ved nærmere undersøkelse har vist seg å være naturlaginger, plogspor o.l., altså ikke runer.

*Status pr. 31.12.97:*

Total størrelse på materialet:	6 bind av NiyR, totalt 1.799 sider
Konvertert totalt:	6 bind av NiyR, totalt 1.799 sider
Korrekturlest totalt:	6 bind av NiyR, totalt 1.799 sider
Ferdigstilt totalt:	6 bind av NiyR, totalt 1.799 sider

Total størrelse på materialet:	1.200 skjemaer
Konvertert totalt:	1.200 skjemaer
Korrekturlest totalt:	1.200 skjemaer
Ferdigstilt totalt:	1.200 skjemaer

Total størrelse på materialet:	773 innholdsfortegnelser
Konvertert totalt:	773 innholdsfortegnelser
Korrekturlest totalt:	773 innholdsfortegnelser
Ferdigstilt totalt:	773 innholdsfortegnelser

*Beskrivelse av arbeidet:*

Materialets tilstand har vært bestemmende for opplegget for dataføring innenfor Dokumentasjonsprosjektet. Runearkivet har vært mer eller mindre et én-manns foretagende i alle år. Bare N-innskriftene egnede seg til umiddelbar dataføring, og med fem innskrivere var dataføringen av hele NiyR et tiltalende prosjekt. Da ville man både ha korpusutgaven tilgjengelig elektronisk, samtidig med at alle de viktigste opplysninger om nesten halvparten av de ekte, gamle innskriftene var lagt inn - mer eller mindre ustrukturert - på data. Som supplement ble materialet i arkivmappene gjennomgått og ordnet, og innholdsfortegnelser for N-mappene ble laget. At man satset på dataføringen av den mindre strukturerte teksten i NiyR førte med seg behovet for en strukturering av de viktigste opplysninger fra teksten i en oversiktlig skjemaform eller database. Dataføring av kartotek kortene for A- og B-innskriftene ville ikke ha vært tilstrekkelig siden de var ment å bli brukt hovedsakelig av lederen for Runearkivet.

Det ble bestemt at skjemaer måtte lages til dataføring av disse og av Y-innskriftene; skjemaene inneholder også innholdsfortegnelse for mappene. For Æ-innskriftene er registrering av et forkortet skjema pluss innholdsfortegnelse for mappene tilstrekkelig. Eldre runer (Æ-innskriftene) er, i tillegg til å være et fellesnordisk fenomen, i videre perspektiv også et germansk fenomen, og det arbeides nå med en omfattende database for de eldre runeinnskriftene i Kiel. Det er mer hensiktsmessig å støtte prosjektet i Kiel. Korpusutgaven Norges Indskrifter med de ældre Runer er foreldet og ikke egnet til skanning eller registrering. Som korpusutgave fungerer for tiden Wolfgang Krauses Die Runeninschriften im älteren Futhark (1966). For avkrefte rapporter er registrering av innholdsfortegnelsene i seg selv tilstrekkelig.

Prosjektet har som overordnet mål å lage en strukturert database av opplysninger om Norges runeinnskrifter, spesielt de med yngre runer. Helt sentralt er opplysninger om skrift og språk i innskriftene, spesielt: translitterering av runene, inkludert variasjon i tegnformer, og tolkning av innskriftene, dvs. både normalisering og oversettelse. Disse opplysningene må være knyttet til en kortfattet beskrivelse av innskriften. I tillegg kommer nødvendige opplysninger om gjenstanden, om gjenstandens tilstand, om funnforholdene, om nåværende oppbevaringssted, eventuelt om den kulturhistoriske konteksten osv. At Niyr er gjort tilgjengelig i elektronisk form, er et positivt biprodukt av metoden som er valgt for å forberede opprettelsen av databasen. Som ren tekst er det en mindre strukturert datakilde med overveldende mange opplysninger.

*Beskrivelse av digitalt arkiv/database:*

NiYR er lagret som SGML-tagget tekst, mens skjemaene og innholdsfortegnelsene ble skrevet inn i en spesialtilpasset applikasjon og foreligger i database.

*Videre arbeid:*

Runearkivet vil fortsette kontrollarbeidet for å sikre kvaliteten av opplysningene i databasen.

## **Numismatikk**

### **Generelt: Universitetets myntkabinett, IAKN, UiO**

Myntkabinettet er en avdeling under Institutt for arkeologi, kunsthistorie og numismatikk (IAKN) ved Universitetet i Oslo. Myntkabinettet har riksomfattende ansvar og forpliktelser, innenfor faglig-vitenskapelig, museal og formidlingsmessig virksomhet, og forvaltningsansvar for myntfunn gjort i Norge. Videre er det en undervisningsinstitusjon, og landets eneste numismatiske fagbibliotek blir forvaltet ved avdelingen.

Delprosjektansvarlig: Professor Kolbjørn Skaare

Fagkonsulenter i delprosjektperioden: Svein Gullbekk, Jon Anders Risvaag, Henning Laugerud

Korrekturlesere/andre medarbeidere i delprosjektperioden:  
Marit Opsal

IT-faglig ansvar: Jon Holmen

Antall fagkonsulentårsverk i perioden 1992 - 1995: 6

Kontaktperson i fagenheten fra 1.1.98: Professor Kolbjørn Skaare

## **Myntarkivet**

### *Om arkivet:*

Myntkabinettet har gjennom årenes løp opparbeidet en betydelig samling numismatica hvor de fleste myntpregede kulturer gjennom de siste 2.500 år er representert. Myntkabinettet besitter et særdeles viktig arkeologisk kildemateriale i form av mynter fra norske funn gjennom de siste 170 år. Hovedvekten av disse stammer fra vikingtid og middelalder. Samlingen teller over 200.000 numismatica. Arkivene over samlingene utgjør ca. 90.000 håndskrevne arkivkort.

Arkivene er utarbeidet av en lang rekke personer fra Myntkabinettets opprettelse og til i dag. Det er særlig i tiden etter 1920 at samlingene har blitt arkivført på kort. Opplysningene på arkivkortene er av ymse omfang og kvalitet. Det er alt fra enkle primærdata som land, myntherre, nominal og periode til opplysninger om myntmestre, vekter, metallanalyser, utførlige funnopplysninger, kildehenvisninger og diverse kommentarer. Disse håndskrevne opplysningene er påført arkivkortene på forskjellige måter. I tillegg til arkivkortene finnes ca 30.000 gjenstander oppført i funnlistene, inventeringskataloger og publikasjoner av deler av Myntkabinettets samling. Dette materialet foreligger for det meste i maskinskrevet form og er bedre strukturert enn arkivkortene.

### *Status pr. 31.12.97:*

Total størrelse på materialet:	ca 130.000 gjenstander
Konvertert totalt:	ca 130.000 gjenstander
Korrekturlest totalt:	ca 130.000 gjenstander
Ferdigstilt totalt:	ca 130.000 gjenstander

### *Beskrivelse av arbeidet:*

Konverteringen ble satt i gang uten noe særlig omfattende forutgående kontroll av det papirbaserte arkivet, som var utarbeidet over ganske lang tid, av mer eller mindre kyndige personer, med mer eller mindre tydelig håndskrift. Gale eller foreldete bestemmelser og andre feil i det håndskrevne papirarkivet ble derfor videreført av innskrivingsgruppen.

Alle arkivkort og lister over samlingen ble nummerert, kopiert og klargjort før innskrivningen tok til. Registreringen av samlingene ble utført i spesialkonstruerte registreringsdatabaser utarbeidet for de ulike myntgruppene: 1) Antikke greske mynter 2) Antikke romerske mynter 3) Middelaldermynter 4) Nyere tids mynt 5) Moderne mynt 6) Funnmynter 7) Medaljer og ordener.

For å få registrert de ulikartede håndskrevne arkivkortene så enhetlig som mulig, ble det opprettet referansetabeller hvor all kjent informasjon om myntene ble systematisert og lagt til grunn for all registrering. Hver medarbeider fikk generell innføring i det mynthistoriske kildematerialet og ble opplært i bruken av databasen og hvordan materialet forholdt seg innenfor den enkelte myntgruppe. I neste trinn ble det i delprosjektet foretatt korrekturlesning direkte i databasen mot selve gjenstanden og gammel katalog, slik at mange feil og unøyaktigheter kunne rettes opp.

### *Beskrivelse av digitalt arkiv/database:*

Følgende numismatiske materiale foreligger i Oracle-databasen: greske mynter, deler av samlingen av romerske mynter, norske og andre europeiske mynter eldre enn ca 1850, medaljer, ordener, dekorasjoner, norske numismatiske kirkefunn og en del andre myntfunn.

#### *Videre arbeid:*

Det er gjort enkelte forsøk med innlasting av bildemateriale. En slik innlasting i større skala er svært ønskelig. Første versjon av en relasjonsdatabase for forsknings- og gjenstandsforvaltning vil være klar i 1998.

#### *Bruk av digitalt arkiv/database:*

Myntkabinettet har stor og jevn pågang fra forskere i andre fag, fra institusjoner, forlag, medier og det brede publikum. Ved enkeltsøk i databasen er det nå lettere, raskere og mindre risikofyllt å takle de mange forespørlene. Mange museale rutiner kan forenkles betydelig. Ved søk i databasen vil gjenstandsmagasinet kunne skånes for hyppig trafikk. Ved siden av den daglige ekspedisjon av eksterne henvendelser vil særlig avanserte søk og diverse krysskoblinger være aktuelle i det viktige myntfunnmaterialet og i de norske og de mange andre rikt forekommende myntserier i samlingene. Ved raskt å kunne etablere en oversikt over et så kvantitativt rikt materialet vil grunnlaget kunne legges for videre forskning. Myntdatabasen vil også lenkes opp mot Oldsak-samlingens gjenstands- og funndatabase. Databasen vil være en solid støtte for og et potensiale i undervisning og folkeopplysning.

#### *Utstilling: Norsk mynt i 1000 år*

Høsten 1995 åpnet Universitetets Myntkabinett jubileumsutstillingen Norsk mynt i 1000 år. Som et supplement til jubileumsutstillingen laget Dokumentasjonsprosjektet, i samarbeid med Myntkabinettet, en elektronisk utstilling som kan ses i en publikums-kiosk i utstillingslokalet. Utstillingen er også tilgjengelig via Dokumentasjonsprosjektets hjemmesider.

Utstillingen inneholder følgende temaer: viktige begivenheter og utsnitt fra Norges 1000-årige mynthistorie, utvalg av enkeltfunn og ulike typer myntfunn, myntproduksjon til forskjellige tider, konger og erkebiskoper som har preget mynt i Norge, oversikt over mynter som er blitt preget i Norge, norske steder der det har blitt preget mynt og om Universitetets Myntkabinett.

## **Kulturhistorie**

### **Generelt: Fagenhet for samisk-etnografi og Fagenhet for nyere kulturhistorie, Tromsø museum, UiTø**

Fagenhet for samisk-etnografi ble etablert i 1949. Enheten bygger i stor grad på de samiske samlinger som har vært ved Tromsø Museum siden 1870-årene. Enheten arbeider med forskning, innsamling, dokumentasjon, ivaretagelse av samlinger og arkiv, utadrettet virksomhet, formidling og undervisning primært innenfor samisk kultur. I tillegg kommer prosjekt knyttet til oppdragsforskning, også med hovedvekt på samisk forskning. Faggruppen deltar i forskning, dokumentasjon og formidling, også i fellesskap med 4. verdens folkeslag der særlig urbefolkningene i det subarktiske området vektlegges.

Fagenhet for nyere kulturhistories hovedoppgave er dokumentasjon og forskning innenfor norsk og kvensk folkekultur i Nord-Norge og Svalbard. Dette omfatter både materiell kultur, sosialt liv, folkeminne og folkemusikk. Arbeidsområdet omfatter også populærvitenskapelig virksomhet, utstillingsarbeid og generell service og veiledning i landsdelen.

Delprosjektansvarlig samisk-etnografi: konservator Dikka Storm, Tromsø museum, UiTø

Delprosjektansvarlig nyere kulturhistorie: arkivar Sveinulf Hegstad, Tromsø museum, UiTø

Fagkonsulenter i prosjektperioden: Astri Fremmerlid, Tore Einar Johansen, Kjetil Kvalsvik og Lena Aarkol

Korrekturlesere/andre medarbeidere i prosjektperioden: Ottar Kristoffersen, Morten Johansen, Åse Sjørgård, Egil Akselvoll, Trude Aga Brun

IT-faglig ansvar: Jon Holmen

Antall fagkonsulentårsverk i perioden 1994 - 1997: 3,75

Kontaktperson i fagenheten fra 1.1.98: konservator Dikka Storm og arkivar Sveinulf Hegstad

### ***Fotoarkivene, Samisk-etnografi og Nyere kulturhistorie***

#### *Om materialet:*

Fotosamlingen på Fagenhet for samisk-etnografi består av ca 31.000 enkeltopptak i svarthvitt og 21.000 dias. I tillegg kommer 20.000 uregistrerte fotografier. På Fagenhet for nyere kulturhistorie (NYK) består samlingen av ca. 24.000 svarthvitt bilder og ca. 13.500 fargedias. Delprosjektet har konvertert et utvalg av fotosamlingene på begge fagenhetene.

#### *Status pr. 31.12.97:*

##### Samisk-etnografi:

Total størrelse på materialet:	8.000 papirkopier og 6.400 dias
Skannet og registrert totalt:	8.000 papirkopier og 6.400 dias
Korrekturlest totalt:	8.000 papirkopier og 6.400 dias
Ferdigstilt totalt:	8.000 papirkopier og 6.400 dias

##### Nyere kulturhistorie:

Total størrelse på materialet:	23.900 papirkopier
Skannet og registrert totalt:	23.900 papirkopier
Korrekturlest totalt:	23.900 papirkopier
Ferdigstilt totalt:	23.900 papirkopier

#### *Beskrivelse av arbeidet:*

Arbeidet startet i 1995 med forberedelser til å digitalisere arkivmaterialet ved Tromsø Museum. Det ble foretatt en gjennomgang og kartlegging av arkivmaterialet ved avdeling for Nyere kulturhistorie (NYK) og Samisk-etnografisk avdeling, og forberedelser til digitaliseringen av fotosamlingene. Outlines klassifiseringssystem ble tilrettelagt for dette formålet, og en brukte tid på å utrede de ulike sidene ved arbeidet med å skanne inn foto til digitalt format for å finne den mest hensiktsmessige metoden og arbeidsprosessen til dette formålet. Det ble utarbeidet et notat på hvilke ulike løsninger som er aktuelle i forbindelse med hvordan internett-versjonen av fotodatabasen bør utformes. I oktober 1995 startet arbeidet med opplæring ved registreringsenheten ved Fauske. Det ble utarbeidet et instruksjonsopplegg som ble brukt i opplæringa.

*Innskrivingsenheten:*

- Skanne inn papirkopier, lagre dette i hensiktsmessige filsystemer og brenne til CD-plater.
- Registrering av opplysninger om bildene fra papirkort (både papirkopier og dias) i Access, første gangs korrekturlesing og sende de ferdige filene til fagkonsulent.

*Delprosjektet:*

Fagkonsulentens arbeid har bestått i:

- Opplæring og kvalitetskontroll av innskriverne i registreringsenheten.
- Slå sammen filer som har kommet fra registreringsenheten i Fauske og lese korrektur på det materiale som er registrert, kvalitetskontroll.
- Ansvar for en person som har vært på hospitering for å skanne dias og negativer.
- Lage hensiktsmessige rutiner for skanning av dias og negativer ved Tromsø Museum.
- Kvalitetskontroll av ferdigbrente CD-plater fra registreringsenheten fra Fauske og fra Tromsø Museum.

"Skannerens" arbeidsoppgaver har bestått i:

- Skanne dias og negativer.
- Sette i gang en arbeidsprosess hvor de skannede bildene blir konvertert til jpg-format og det blir laget småbilder.

Det eksisterer "huller" i nummerseriene. Årsaken til dette er for Samisk-etnografi at bildene eller kortene ikke var i arkivet på det tidspunkt de ble tatt ut av arkivet for skanning eller registrering. Bildene var lånt ut til personer på avdelinga, i noen tilfeller mangler rett og slett bildene. Deler av dette ble rettet opp mot slutten av perioden, men fordi alt materialet ikke har kommet til rette i arkivet, har ikke alle "hullene" blitt rettet opp.

I nyere-kulturhistorisk fotosamling ble det foretatt en "kvalitetssiling" i forkant av skanningen av bilder, og det har resultert i "huller" i nummerserien. Det er også laget "huller" i nummerserien i arkivarbeidet på NYK, for å kunne sette inn eventuelle flere bilder i samme serie.

*Beskrivelse av digitalt arkiv / database:*

Materialet er registret i Access 2.0. Det er utarbeidet to applikasjoner for registrering av opplysninger om bildene, en til papirkopier og en til dias. Applikasjonen til dias-registreringen er en utvidet versjon av applikasjonen til papirkopiene. Dias-applikasjonen gjør det mulig å se på diasene på skjermen ved hjelp av fremviserprogrammet ACDSsee, samtidig som opplysningene på papirkortene blir registrert i Access.

Papirkopiene er skannet inn på flat-skanner i 2700 dpi, konvertert til jpg og deretter er det laget småbilder. Negativer er skannet inn i 2700 dpi, konvertert til jpg-format, og deretter er det laget småbilder. Storformater og småformater er brent på CD-plater. Diasene er skannet inn i 230 dpi, konvertert til jpg-format og deretter brent til CD-plater.

*Teknisk beskrivelse:*

Nummerering av bilder og registreringer: Hvert bilde i arkivet ved NYK og Fagenhet for samisk-etnografi har et unikt nummer som også er brukt i registreringen og skanningen av bildene: Det innebærer at man ikke har nummerert arkivmaterialet på nytt. Bildene har nummer etter institusjon (TS= Tromsø Museum), avdeling (N=NYK eller L=Sam.Etn.) og type (F=foto, D=dias og P=postkort). Dette innebærer at skannede bilder og registreringene til hvert enkelt bilde har et unikt nummer, slik at hvert bilde kan kobles til de opplysningene som er registrert om bildet.



Filnavn: registreringene er lagret i mapper på ca 500-1.000 enkeltposter. Filene har fått navn etter hvilken dato de er slått sammen av fagkonsulenten ved Tromsø Museum (f.eks: 2011, 0110). Bildene er lagret i mapper i fortløpende rekkefølge som er brent til nummererte CD-plater (nr 1-50 er skannet ved Tromsø Museum og nr 51-73 er skannet ved Registreringssentralen i Fauske). Småformatene av bildene er lagret sammen med storformatene.

Bildene er brent på CD-rom og lagret i magasinet på Tromsø Museum, ved NYK og Fagenhet for samisk-etnografi og ved Dokumentasjonsprosjektet. Samtlige CDer er deretter overført til Dokumentasjonsprosjektets servere for å sikre varig lagring og rask tilgang. Databasefilene som inneholder registreringene, ligger foreløpig i Access-format på Dokumentasjonsprosjektets servere.

#### *Videre arbeid:*

Det bør foretas en vurdering om hvorvidt de resterende "hullene" i materialet i Samisk-etnografisk fotosamling skal rettes opp eller ikke. Et moment som må avklares, er opphavsrett og eventuelle klausuler på bildene før de distribueres på nettet. Fagenhet for samisk-etnografi har allerede diskutert dette problemet, men har ennå ikke kommet frem til en endelig løsning. Mulig løsning er at bildene som vil bli lagt ut på Internett, vil være en "reduert" utgave av den ferdige databasen. Bildene som vil bli lagt ut, vil sannsynligvis kun være småbildeformatene. Denne "internett-versjonen" vil fungere som en katalog over et utvalg av de bildene som ligger i den fullstendige databasen ved Tromsø Museum. Tilgang til den fullstendige databasen vil være mulig gjennom henvendelser til Tromsø Museum og de respektive fagenhetene.

Det er laget en type ordliste for NYK-bildene, som forsøker å knytte sammen og begrense en del sentrale begreper. Dette for å gjøre basen mer funksjonell, fordi Outline-systemet ikke er helt tilpasset Tromsø Museums fotomateriale. Det er fortsatt behov for at Outline-systemet tilpasses Tromsø Museums materiale ytterligere. Det er særlig viktig at det jobbes sammen med fagmiljøet slik at mer spesifikke samiske begreper knyttes sammen med mer allmenne norske begreper og vice versa. For å lette søk ved hjelp av Outline kan det også være en mulighet å inkludere stikkordene til hvert nummer i Outline til den lista over Outline som ligger i databasen. Årsaken til at dette bør gjøres, er at Outline benytter samlebetegnelser som ikke alltid "vil si" brukeren noe.

#### **Generelt: Institutt for kulturstudier, Universitetet i Oslo**

Delprosjektansvarlig: Professor Birgit Hertzberg, Institutt for kulturstudier, UiO og Velle Espeland, Norsk visearchiv (ballader)

Fagkonsulenter i delprosjektperioden (minneoppgaver): Line Esborg og Siv-Bente Grongstad

Fagkonsulenter i delprosjektperioden (ballader): Åge Skjelborg og Mette Skreslet

IT-faglig ansvar: Christian-Emil Ore

Tilknyttede enheter/personer: Anne Marie Wiersholm og Ørnulf Hodne, begge Institutt for kulturstudier

Antall fagkonsulentårsverk i perioden 1992 - 1997: 9

Kontaktperson i fagenheten fra 1.1.98: Professor Birgit Hertzberg og fagkonsulent Siv Bente Grongstad (minneoppgaver) og Velle Espeland, Norsk visearkiv og Åge Skjelborg (ballader)

### *Norske ballader*

#### *Om materialet:*

Siktemålet med Balladeprojektet er å lage en vitenskapelig, diplomatarisk utgave av norske ballader på data. Materialet finnes i flere arkiv og omfatter dels originaloppskrifter av balladetekster, dels eldre siffer- og noteoppskrifter samt lydopptak av eldre og nyere balladesang. Norge er representert med 240 ulike balladetyper. Mesteparten av balladematerialet er innsamlet i tidsperioden 1840-1950 og skriver seg hovedsakelig fra Telemark og Vest-Agder.

#### *Status pr. 31.12.97:*

##### *Tekstmaterialet:*

Total størrelse på materialet:	3.899 balladetekster
Konvertert totalt:	3.668 balladetekster
Korrekturlest totalt:	3.668 balladetekster
Gjenstår å konvertere:	231 balladetekster*
Ferdigstilt totalt:	3.668 balladetekster

##### *Noteoppskriftene:*

Total størrelse på materialet:	281 melodifiler
Konvertert totalt:	281 melodifiler
Korrekturlest totalt:	281 melodifiler
Ferdigstilt totalt:	281 melodifiler

\* Delprosjektet fortsatte i 5 måneder i 1998 finansiert av Dokumentasjonsprosjektet, Institutt for kulturstudier og Det historisk-filosofiske fakultet, UiO. I tillegg til restansen av balladetekster, vil det arbeides videre med melodimaterialet.

#### *Beskrivelse av arbeidet:*

##### *Tekstmaterialet:*

Balladematerialet ble tilrettelagt, kopiert opp og forsynt med de nødvendige signaturer og sendt en registreringsgruppe i Horten for innskriving. Innskriverne kodet tekstopp-skriftene i SGML og leste selv korrektur mot renskrevne eller håndskrevne originaloppskrifter. Registreringsgruppen returnerte balladematerialet til fagkonsulenten, som stod for korrekturlesing mot originalene. Mens innskriverne korrekturleste mot foreleggene, har fagkonsulenten tilbakeført tekstene til originalens tekstform og korrekturlest mot originaloppskriftene.

##### *Noteoppskriftene:*

Denne delen av balladematerialet består foreløpig av ca. 600 eldre og nyere siffer- og noteoppskrifter med underlagt tekst. Melodioppskriftene blir konvertert av fagkonsulentene via musikkprogrammet Finale som maskinlesegelige notebilder i både diplomatariske og rekonstruerte former.

##### *Lydmaterialet:*

Ca. 200 norske ballader er dokumentert på lydopptak og arkivert i Norsk Folkeminnesamling, Norsk Folkemusikksamling og i NRKs folkemusikkarkiv. Balladefremføringer på lydopptak blir avlest og transkribert til et håndskrevet notebilde og innskrevet

via Finale-programmet i både diplomatariske og modifiserte former. I motsetning til noteoppskriftene dokumenterer det lydfestede materialet individuelle detaljer i sangstil, rytme og tempo hos den enkelte balladesangeren.

*Beskrivelse av digitalt arkiv:*

Det konverterte materialet er lagret som SGML-tagget tekst.

*Teknisk beskrivelse:*

Datafilene er knyttet til balladeoppskriftene systematisk etter løpenummer samt typenummer. Det ferdigstilte dokumentet er merket med innskriver-signaturer og autorisasjonen LMO ÅS (Lest Mot Originalen). Tekstfilene er lagret på MacHedda/Ballader/Tekster/ i ren-tekst-format.

*Videre arbeid:*

Arbeidet fortsatte i 1998 med konvertering av balladetekstene som gjensto og videre arbeid med melodimaterialet.

*Bruk av digitalt arkiv/database:*

To hovedgrupper brukere peker seg ut: for det første forskere og de som vil bruke balladene i undervisning og for det andre de som er interessert i balladene i kunstnerisk sammenheng. Balladene representerer også et rikt språklig materiale med ulike dialektformer fra 1840 og framover. Navnegranskere vil også kunne bruke personnavnene som finnes i teksten.

***Utstilling: Folkeviseautomaten***

Dokumentasjonsprosjektet lagde i samarbeid med Norsk visearkiv en WEB-basert utstilling over norske ballader. Utstillinga presenterer et lite utvalg norske folkeviser, og formålet er å fokusere på folkevisene eller balladene som fellesnordisk kulturarv.

Utstillingen er tilgjengelig på WEB via Dokumentasjonsprosjektets hjemmesider, men er også vist på vandretstillingen i anledning 600 års-jubileet for innstiftelsen av Kalmarunionen, på Forsvarsmuseet fra desember 1997.

***Minneoppgavene***

*Om materialet:*

Minneoppgavene er kort fortalt livshistorier (selvbiografier). Materialet er resultatet av en landsomfattende konkurranse i 1964 og 1981, hvor folk fra alle lag av befolkningen ble invitert til å skrive ned sine minner.

De som deltok i konkurransen, var i gjennomsnitt 75 år gamle. Det er altså eldre mennesker som selv skriver om sine liv og sine erfaringer fra de første barndomsminner til i dag. Minneoppgavene karakteriseres derfor som et erindringsmateriale, enten de beskriver livsløpet eller deler av det. Konkurransen ble arrangert av Nasjonalforeningen for folkehelsen, siste gang i samarbeid med Landslaget for bygde- og byhistorie. En tilsvarende innsamling er nettopp avsluttet (1996). Denne gang var det Landslaget for lokall historie i samarbeid med Norsk Etnologisk Granskning og Institutt for Kulturstudier ved UiO som stod bak. Materialet er arkivert i Norsk Folkeminnesamlings arkiver og utgjør den største tilvekst til arkivene i nyere tid. Minnematerialet har stor betydning som utgangspunkt for forståelsen av vår nære fortid, fordi det dreier seg om vanlige menneskers kulturhistorie skrevet av dem selv. Minneoppgavene

dekker alle lag av befolkningen og alle landets fylker. Den kjønnsmessige fordeling er i balanse (1829 kvinner og 1672 menn), og det dekker en tidsperiode på over 100 år. Det er med andre ord et omfattende og relativt nytt kildemateriale som er egnet for tverrfaglig grunnforskning.

Lengden på besvarelsene varierer fra en håndskrevet side til 1.000 maskinskrevne. Materialet utgjør til sammen rundt 160.000 manuskriptsider, fordelt på ca. 3500 oppgaver.

I Dokumentasjonsprosjektet er det for konvertering valgt ut materiale fra seks fylker (begge årganger er tatt med). Utvalget består av Troms, Nordland, Aust-Agder, Rogaland, Buskerud og Oslo. Utvalget er basert på at alle regioner skal være representert, og at de utvalgte fylkene er differensierte med henhold til geografi, næringsliv og ulike kulturhistoriske hensyn. Det er dessuten valgt ut fylker hvor tekstmassen ikke er for omfattende, men som likevel er godt dekket. Alle steds- og personnavn tagges av hensyn til personvernet og for å lette søk i materialet, når det blir lagt ut som en del av Universitetenes databaser for språk og kultur. Tekstene registreres for øvrig nøyaktig slik de ble skrevet, (det vil si med den målform etc som den enkelte bidragsyter fant naturlig) for å bevare den enkeltes stemme. Konverteringen av minneoppgavene er viktig både når det gjelder synliggjøring av denne type materiale og ikke minst for gjøre det mer tilgjengelig for et større publikum enn i dag.

*Status pr. 31.12.97:*

Total størrelse på materialet:	1.269 oppgaver, 25.524 originalsider
Konvertert totalt:	1.251 oppgaver, 24.863 originalsider
Gjenstår å konvertere:	18 oppgaver, 661 originalsider*
Korrekturlest totalt:	1.088 oppgaver, 20.254 originalsider
Gjenstår å korrekturlese:	181 oppgaver, 5.270 originalsider*
Korrekturlest totalt:	1.088 oppgaver, 20.254 originalsider

\* Delprosjektet fortsetter med midler fra Institutt for kulturstudier fram til 31.3.98. Utvalgt materiale fra innsamlingen i 1996 vil også konverteres i delprosjektets siste fase.

*Beskrivelse av arbeidet:*

Kopier av minneoppgavene er sendt ut til registrering til ulike registreringsentraler. Det har vært foretatt både opplæring og oppfølging av innskrivere. Materialet har blitt skrevet inn nøyaktig slik forfatterne har skrevet det, men dialektord, tegnsetting og skrivefeil. Det har med andre ord blitt kopiert inn. Det eneste tillegget som er blitt gjort, er bruk av tagger for å anonymisere og lette søk i materialet. Taggene merker ut minneoppgavens årgang, fylke og oppgavenummer, samt sidenummer, bilde, tegning eller trykt materiale, bildetekst, stedsnavn, personnavn og mannlig og kvinnelig forfatternavn.

Innskriverne har skrevet inn materialet og lest korrektur på dette, både på det de selv har skrevet inn og for andre. Deretter har de lagret det på disketter som de har sendt til fagkonsulent i Oslo sammen med en sluttrapport over det innskrevne materialet. Fagkonsulenten har lest korrektur og ferdigstilt materialet etter hvert som det har kommet inn. Fagkonsulenten har skrevet kvalitetsrapporter som er sendt tilbake til registreringsenhetene som en tilbakemelding. Informantregisteret er også skrevet inn og korrekturlest.

#### *Beskrivelse av digitalt arkiv/database:*

Totalt ferdigstilt i det digitale arkivet er det 1.088 oppgaver fordelt på 20.254 originalsider. Samtidig er det ferdig informantregister med opplysninger om årstall, oppgavenummer, fylke, kjønn, klausul, navn, fødested, nåværende bosted, yrke og eventuelle andre merknader. Det er også laget et stikkordregister som ligger klart.

Samtidig med at materialet blir lagt ut i en database, må stikkordregister og personregister legges ut. Det må være mulig å søke på for eksempel person, fylke og stikkord samtidig. Kryssreferanser bør absolutt være en mulighet.

Materialet vil bli lagt ut anonymt. Det vil si at forfatternavn og andre navn som forfatteren bruker, vil bli holdt skjult. Stedsnavn vil bli beholdt. For å få tilgang på navn, vil det være nødvendig for den enkelte forsker å henvende seg til instituttet spesielt om dette. Det konverterte materialet er lagret som SGML-tagget tekst.

#### *Teknisk beskrivelse:*

Materialet er lagret på fagkonsulentens hjemmeområde, og etter som hvert enkelt fylke og årgang har blitt ferdig, er alle oppgavene kopiert over på disketter som sikkerhetskopier og lagt i en brannsikker safe i kjelleren.

Hver enkelt oppgave er lagret separat med filnavn som indikerer årstall, fylke og oppgavenummer. Eks: Oppgave 27 fra Nordland hentet fra årgang 1964 lagres med dette filnavnet: 64NOR.027. Det er laget en mappe til hvert fylke og hver årgang. Filene ligger under dette. Det finnes egne mapper for helt ferdigstilt materiale og for materiale som ennå ikke er korrekturlest.

#### *Videre arbeid:*

Det ferdige materialet er ennå ikke lagt inn i en database. Fordi delprosjektet skal fortsette i tre måneder til, så vil dette arbeidet starte opp først etter nyttår.

#### *Bruk av digitalt arkiv/database:*

Materialet er relevant for forskere fra ulike fag som arbeider med sosial- og kulturhistorie fra tiden rundt århundreskiftet og opp mot vår egen tid.

## **Historie**

### **Generelt: Tingbokprosjektet, Institutt for historie, UiO**

Tingbokprosjektet ble til i 1987 ved Historisk institutt, Universitetet i Oslo under ledelse av professor Sølvi Sogner. Formålet er å tilgjengeliggjøre de rike kilder som tingbøkene representerer, og stimulere til forskning på denne kildetyper. Prosjektet er tverrfaglig og favner blant andre historikere, jurister, etnologer, folklorister, kriminologer og kvinneforskere. Aktivitetene er: forskningssamarbeid, bokutgivelser, transkripsjonsarbeid, seminarvirksomhet og undervisning.

Delprosjektansvarlig: Professor Sølvi Sogner, Institutt for historie, UiO

Fagkonsulent: Bodil Chr. Erichsen

Korrekturlesere/andre medarbeidere i delprosjektperioden: Thor Johnsen, Bjørg Kleven, Eva Merethe Møller.

IT-faglig ansvar: Christian Emil Ore

Tilknyttede enheter/personer: Førsteamanuensis Hilde Sandvik, Historisk institutt, / Prosjekt for Nettbaserte Hovedfagsstudier Riksarkivets kjeldeskriftavdeling, ved Knut Sprauten og Kirsten Julien

Antall fagkonsulentårsverk i perioden 1992 - 1997: 6

Kontaktperson i fagenheten fra 1.1.98: professor Sølvi Sogner og fagkonsulent Bodil Erichsen

### ***Tingbøkene***

#### *Om materialet:*

Tingbøkene er rettsprotokoller fra laveste domstol, der byfogd eller sorenskriver var dommer. Rettspraksis, normer og holdninger, kjønnsroller, maktstrukturer og livet i nærmiljøet er blant temaene som utforskes ved hjelp av disse kildene.

Tingbøkene refererer de rettslige forhandlingene som fant sted ved domstolen og gir oss til dels frodige beskrivelser av folks levevis. Her finnes gjeldtvister og nabo-krangler, brutte ekteskapslofter og alvorlige kriminelle tildragelser. Protokollene er nedtegnet under rettsmøtene, og bærer ofte preg av å være skrevet i høyt tempo. Håndskriften er gotisk og skiller seg fra både moderne håndskrift og gotiske trykte bokstaver. Dessuten er språkføringa er uvant; preget av dansk og til dels tysk kansellispråk. Tekstene inneholder mange merkelige tegn og forkortelser.

De eldste norske bevarte tingbøkene er fra Rogaland (Jæren og Dalane 1613 og Ryfylke 1616) og Finmark (1620). Med kongelig forordning 1633 ble det påbudt å føre tingbok ved underrettene i Norge. Likevel finnes ikke særlig mange fra første halvdel av 1600-tallet. Etter innføringen av eneveldet i 1660 ser påbudet ut til å være etterlevd i stor grad, men mange tingbøker fra den tiden er gått tapt. Fra tiden etter 1700 er tingbøkene godt bevart i sammenhengende rekker, men også her finnes lakuner.

#### *Status pr. 31.12.97:*

Total størrelse på det utvalgte materialet:	28 årsprotokoller, ialt 2.764 sider*
Konvertert totalt:	28 årsprotokoller, ialt 2.764 sider

Korrekturlest totalt:	28 årsprotokoller, ialt 2.764 sider
Ferdigstilt totalt:	28 årsprotokoller, ialt 2.764 sider

\* De 28 utvalgte årsprotokollene er fra Bragernes, Lier/Røyken/Hurum, Aker og Kristiania og alle er fra 1680-årene.

Ytterligere 3 årsprotokoller er skrevet av, 445 sider. Det er tingbøkene fra Kristiania 1695-97. 2 av disse, 255 sider, er også korrekturlest. Det betyr at dette materialet også kan ferdigstilles ved en relativt liten innsats på et senere tidspunkt. Nok en tingbok, fra Voss, ble delvis bearbeidet i registreringsgruppa og kan ferdigstilles relativt lett. Dessuten er 6 årsprotokoller tilrettelagt, dvs. xerox-kopi av original foreligger. Det gjelder Lier/Røyken/Hurum 1695-99, og Aker 1688.

#### *Beskrivelse av arbeidet:*

##### Forberedelser:

Fagkonsulenten valgte ut tingbøker på bakgrunn av protokollenes geografiske opphav og fysiske standard. Tingbøker fra innskrivernes lokalmiljø ble foretrukket.

*Tilrettelegging:*

Fagkonsulenten eller timeansatte tok xerox-kopier av hver tingbokside, fra microfilmer.

*Transkripsjon:*

Tingbøkene ble skrevet av i registreringsgruppa. Avskriftene foreligger i wp 5.1 (DOS) format, og ble forsynt med kodelinje mellom hver sak. Første korrektur og oppretting ble foretatt i registreringsgruppa. Andre korrektur: utført av fagkonsulenten eller korrekturleser, mot originalprotokollen. Noe korrektur er utført av Kjeldeskriftavdelingen på Riksarkivet.

*Konsultasjon:*

Kjeldeskriftavdelingen ble forelagt tvilstilfeller og problemer. Et par sider fra hver skriver ble på et tidlig tidspunkt levert avdelingen for kontroll.

Endelig oppretting og klargjøring: I tillegg til retting av feil, ble en del standardiseringsarbeid gjort i denne fasen. Dessuten ble kodelinjene sjekket særskilt. Oppretting er i stor grad foretatt av fagkonsulenten.

*Beskrivelse av digitalt arkiv/database:*

Avskriftene er lagt inn i fri-tekst-søk-database.

*Bruk av digitalt arkiv/database:*

Prosjekt for Nettbaserte Hovedfagsstudier bruker tingbokdatabasene i undervisningen. Tingbokdatabasene er allment tilgjengelige gjennom PNHs WEB-sider. Ved Universitetet i Bergen er rettslig materiale fra en senere periode tatt i bruk i grunnfagsundervisningen. Tingbokbasene kan også brukes i grunnskolen, f.eks. i forbindelse med lokalhistorieundervisning. I videregående skole kan tingbokbasene knyttes til undervisning om norsk-dansk 1600-tall mer generelt.

For historikere er tingbøkene en attraktiv og stadig mer benyttet kildetype. Slektsgranskere og lokalhistorikere er også en stor brukergruppe, men også språk- og kulturforskere og navnegranskere viser interesse for materialet. Museumsfolk kan finne gjenstander i bruk så vel som beskrivelser av bygninger og miljøer i by og bygd. For juristene er også de eldste rettsprotokollene fortsatt aktuelle. Med økt tilgjengelighet vil stadig flere vite å ta dette kildematerialet i bruk.

Digitaliseringen vil også gjøre det mulig å ta i bruk tingbokmaterialet på nye måter. Betydningen av å kunne "skumme" tingbøker ved hjelp av søkeord bør framheves. Slik leting etter spesielle saker kan være av essensiell betydning for eksempel i startfasen av et prosjekt.

I forvaltningssammenheng vil tingbokbasene særlig kunne brukes til å dokumentere gamle eiendomsrettigheter og -grenser. Også spørsmål om navnerettigheter kan belyses ved hjelp av tingbokmaterialet.

**Generelt: UB Tromsø, avd. Tromsø museum**

Delprosjektansvarlig: Tor Sveum, UB Tromsø, avd. Tromsø museum

Fagkonsulent: Øyvind Eide

Korrekturlesere/andre medarbeidere i delprosjektperioden: Nina Einevoll

IT-faglig ansvar: Øyvind Eide og Christian-Emil Ore

Tilknyttede enheter/personer: Dikka Storm, Samisk-etnografisk avdeling, Tromsø museum

Antall fagkonsulentårsverk i perioden 1995 - 1997: 3,25

Kontaktperson i fagenheten fra 1.1.98: Tor Sveum

### ***Reinbeitekommisjonen av 1907, bind II***

#### *Om materialet:*

Etter unionsoppløsningen mellom Norge og Sverige i 1905 ble det satt i gang forhandlinger om svenske samers beiterett i Troms. I den sammenheng ble det pekt ut to sakkyndige, Just Knud Qvigstad og Karl Bernhard Wiklund, som skulle samle informasjon fra eldre kildekrifter, som brev, skatteprotokoller, avtaler, grenseprotokoller o.l. Qvigstad og Wiklund utgav i 1909 et to-bindsverk Reinbeite-kommisjonen bind I og II : *Dokumenter angaaende flytlapperne*. I sitt private eksemplar av bind II gjorde Qvigstad mange notater og understrekninger, slik at hans eksemplar gir informasjon som ikke er tilgjengelig andre steder.

#### *Status pr. 31.12.97:*

Total størrelse på materialet:	532 sider + notater
Konvertert totalt:	532 sider + notater
Korrekturlest totalt:	532 sider + notater
Ferdigstilt totalt:	532 sider + notater

#### *Beskrivelse av arbeidet:*

Boken er skannet, revidert og tagget i henhold til en spesielt utarbeidet DTD. Notatene i margin er også transkribert og integrert i den trykte bokteksten med et spesielt taggesett. Dette gjør det enkelt å lage en versjon uten notatene ved behov. Materialet er korrekturlest i registreringsenheten og i delprosjektet.

#### *Beskrivelse av digitalt arkiv/database:*

Det konverterte materialet er lagret som SGML-tagget tekst og vil gjøres tilgjengelig over Internett

#### *Bruk av digitalt arkiv/database:*

Materialet vil være interessant for bl.a. historikere, geografer og arkeologer, fordi det representerer et utvalg av de mest relevante kildekriftene for reindriftssamers bruk av områdene. Også aktuelt for reindriftsutøvere i området, reindriftsadministrasjonen og annen offentlig forvaltning.

### ***Schnitlers protokoller, bind I***

#### *Om materialet:*

Schnitler, Peter: Major Peter Schnitlers grenseeksaminasjonsprotokoller 1742-1745 / utgitt av Kjeldeskriftfondet. - Oslo : Norsk historisk kjeldeskrift-institutt, 1929-1985. - 3 b. 1: ved Kristian Nissen og Ingolf Kvamen. 1962. - L, 478 s.

Protokollene er Schnitlers dagbøker og rapporter fra grenseoppgangen fra Røros til Øst-Finnmark i 1742-45. Bind 1 dekker et annet geografisk område enn Reinbeite-kommisjonens bind 2 og er et av de viktigste kildene for kunnskap om datidens forhold, gjennom Schnitlers forhør av lokalbefolkningen.



*Status pr. 31.12.97:*

Total størrelse på materialet:	528 sider
Konvertert totalt:	528 sider
Korrekturlest totalt:	528 sider
Ferdigstilt totalt:	528 sider

*Beskrivelse av arbeidet:*

Boken er skannet, revidert og tagget i henhold til en spesielt utarbeidet DTD. Materialet er korrekturlest i registreringsenheten og i delprosjektet.

*Beskrivelse av digitalt arkiv/database:*

Det konverterte materialet er lagret som SGML-tagget tekst og vil gjøres tilgjengelig over Internett

*Bruk av digitalt arkiv/database:*

Brukerne av dette materialet vil være det samme som Reinbeitekommissjonen bind 2.

### ***Qvigstads brev***

*Om materialet:*

260 brev skrevet av rektor Just Qvigstad, som dokumenterer vitenskapelige diskusjoner og generelle samfunnsforhold i perioden 1887 - 1956. Det er valgt ut samlingene av brev til tre forskerkollegaer i Norge, Sverige og Finland: professor K.B.Wiklund: 96 brev fra perioden 1891-1934, professor Emil N. Setälä: 96 brev fra perioden 1887-1935 og professor Magnus Olsen: 64 brev fra perioden 1909-1956.

*Status pr. 31.12.97:*

Total størrelse på materialet:	257 brev, totalt 687 sider
Konvertert totalt:	257 brev, totalt 687 sider
Korrekturlest totalt:	257 brev, totalt 687 sider
Ferdigstilt totalt:	257 brev, totalt 687 sider

*Beskrivelse av arbeidet:*

Brevene er transkribert og SGML-tagget, og faksimiler av hver brevside ble skannet. Materialet er korrekturlest i registreringsenheten og i delprosjektet. I disse brevene er det mange litteraturhenvisninger. Da brevene er skrevet med tanke på kun én innforstått leser, er titlene på de omtalte verkene ofte svært knappe. Her brukes kilde-tagging for å hjelpe leserne til å finne fram til materialet som nevnes i brevene.

*Beskrivelse av digitalt arkiv/database:*

Brevene foreligger i SGML-tagget tekstformat og vil gjøres tilgjengelig over Internett

*Bruk av digitalt arkiv/database:*

Brevene er relevante for studenter og forskere av samisk språk, historie, etnografi og religionshistorie. Den nye tilgjengeligheten vil også føre til at det allmenne publikum kan få vite mer om nordnorsk kulturhistorie, både i faglig og muntlig form.

### ***Bibliografi - nordnorsk topografi***

*Om materialet:*

Det har blitt utgitt flere topografiske og lokalhistoriske bibliografier i Norge, den første kom i 1907. Dokumentasjonsprosjektet har konvertert de delene av disse bibliografiene som dekker Nord-Norge.

*Status pr. 31.12.97:*

Total størrelse på materialet:	6.596 poster
Konvertert totalt:	6.596 poster
Korrekturlest totalt:	6.596 poster
Ferdigstilt totalt:	6.596 poster

*Beskrivelse av arbeidet:*

Bibliografiene er skrevet inn i en bibliografisk database, MikroMARC. I de trykte bibliografiene var postene organisert under geografiske overskrifter. Disse overskriftene ble konvertert til geografiske emneord i databasen. Dette innebærer visse problemer fordi de administrative enhetene stadig forandrer seg. Kommuner deles opp og slås sammen. Dette problemet ble løst ved å indeksere postene i et system med tidsdimensjon, i dette tilfellet Elvestrand (1976). I dette systemet har alle kommuner som noen gang har eksistert en kode.

*Beskrivelse av digitalt arkiv/database:*

Dataene foreligger i en bibliografisk database, MikroMARC.

*Bruk av digitalt arkiv/database:*

Bibliografien vil være et hjelpemiddel for bibliotekarer i universitets- og folkebibliotekene. Samtidig vil historikere og andre interesserte i nordnorsk historie, språk og kultur nå lett få tilgang på en samlet oversikt over litteratur om Nord-Norge.

### ***Friis' etnografiske kart***

*Om materialet:*

I dette prosjektet har det blitt digitalisert tre kartserier som viser etnisitet og språkferdighet i Nord-Norge i siste halvdel av forrige århundre. Disse kartseriene er av interesse fordi det like etter at kartene ble laget, skjedde store endringer i etnisiteten i området.

Følgende kartserier er konvertert:

- Friis, Jens Andreas: Ethnografisk Kart over Finmarken. No. 1-5. 10 Bl. Udg. af Videnskabselskabet i Christiania, med Bidrag af Oplysningsvæsenets Fond 1861-62.
- Friis, Jens Andreas: Ethnografisk Kart over Finmarkens Amt. Udg. paa offentlig Bekostning. 1888. 6 Bl. Norges geograf. Opmåling.
- Friis, Jens Andreas: Ethnografisk Kart over Tromsø Amt. 6 Bl. 1890. Norges geograf. Opmåling.

*Status pr. 31.12.97:*

Total størrelse på materialet:	22 kartblad
Konvertert totalt:	22 kartblad
Ferdigstilt totalt:	22 kartblad

*Beskrivelse av arbeidet:*

Kartbladene er skannet i høy oppløsning.

*Bruk av digitalt arkiv/database:*

Brukerne er først og fremst geografer, historikere og arkeologer, men etter tilgjengeliggjøring over Internett vil det vise seg at kartbladene er av interesse for mange flere grupper.

## Leksikografi

### Generelt: Seksjon for leksikografi og målføregransking, INL, UiO

Universitetet i Oslo har store samlinger av norsk språkmateriale i tradisjonelle seddelarkiv. Det meste av dette finnes ved Det historisk-filosofiske fakultet, Institutt for nordistikk og litteraturvitenskap, Seksjon for leksikografi og målføregransking. Her er tre underavdelinger, som avspeiler ei fortid som enkeltstående ordboksverk: Underavdeling for nynorsk, Underavdeling for bokmål og Underavdeling for gammalnorsk og mellomnorsk.

#### *Nynorsk:*

Delprosjektansvarlig i prosjektperioden: Førsteamanuensis Dagfinn Worren og Professor Lars Vikør, Institutt for nordistikk og litteraturvitenskap, UiO

Fagkonsulenter i delprosjektperioden: Kristin Hagen, Lars Vikør, Jens M. Haukaas, Anne Løvland, Svein Sanden, Oddrun Rangsæter, Unn Røyneland, Oddmund Vestenfor, Gudrun Dahler Vik, Erlend Hammer, Karl Johan Sæth, Liv Stenersen og Ålov Runde

Korrekturlesere/andre medarbeidere i delprosjektperioden: Karianne Bjellås Gilje, Hege Eikenes Randen, Jorunn Fløtra, May Irene Fløtre, Eva Langstrand, Sylvia Foss, Gunnhild Tvinnerheim

IT-faglig ansvar Setelarkivet, Norsk Ordbok: Lars-Jørgen Tvedt

IT-faglig ansvar skanna tekster og Grunnmanuskriptet: Christian-Emil Ore

Antall fagkonsulentårsverk i perioden 1991 - 1997: 46

Kontaktperson i fagenheten fra 1.1.98: professor Lars Vikør og fagkonsulent Karl-Johan Sæth

#### *Bokmål:*

Delprosjektansvarlig: Professor Dag Gundersen, Institutt for nordistikk og litteraturvitenskap, UiO

Fagkonsulenter i delprosjektperioden: Jørn-Otto Akø, Anne Engø, Ellen Skolseg, Bodil Kjenstad, Hanne Lillebo og Hanne Lauvstad.

Korrekturlesere/andre medarbeidere i delprosjektperioden: May-Britt Larsen, Britt Oda Fosse, Undis Vatvedt Fjeld, Tore Christian Myre, Trude Skjelten, Ragnhild Mork, Elisabeth Lysnes, Borghild Goddal

IT-faglig ansvar: Christian-Emil Ore

Antall fagkonsulentårsverk i perioden 1991 - 1997: 21

Kontaktperson i fagenheten fra 1.1.98: fagkonsulent Anne Engø

#### *Gammalnorsk:*

Delprosjektansvarlig: Amanuensis Bjørn Eithun, Institutt for nordistikk og litteraturvitenskap, UiO

Fagkonsulenter i delprosjektperioden: Marit Krogtuft, Kristian Emil Kristoffersen, Tor Ulset, Svein Nestor, Mette Gismerøy Ekker, Bodil Kjenstad og Knut Lunde

Korrekturlesere/andre medarbeidere i delprosjektperioden: Eva Strand, Bjørg Harvey, Håvard Benum

IT-faglig ansvar: Christian-Emil Ore

Antall fagkonsulentårsverk i perioden 1991 - 1997: 15,5

Kontaktperson i fagenheten fra 1.1.98: førsteamanuensis Erik Simensen og amanuensis Bjørn Eithun.

### ***Setelarkivet, Norsk Ordbok***

#### *Om arkivet:*

Arbeidet med Norsk Ordbok tok til i 1930-åra. Undertittelen, Ordbok over det norske folkemålet og det nynorske skriftmålet, presiserer ei av hovudmålsetjingane for ordboka: å føre vidare arbeidet til Ivar Aasen og samtidig inkludere nytt tilfang frå norsk talemål og den nynorske skriftkulturen. Norsk Ordbok er altså ei kombinert dialekt- og skriftmålsordbok som byggjer på to typar materiale: på den eine sida sitat frå nynorsk-litteraturen (både skjønnlitteratur og sakprosa), og på den andre sida tilfang frå dei ulike dialektane rundt om i landet. Den tida ordboka dekkjer, er i prinsippet heile den såkalla "nynorske" perioden i språkhistoria, som vi reknar frå 1525. Men det finst nesten ikkje tilfang frå før 1600, og det er svært tynt fram til 1800-talet. I praksis dekkjer ordboka tida frå ca. 1840 og fram til i dag. Etter planen skal Norsk Ordbok ha 12 band, kvart band er sett saman av 5 hefte. Det første heftet kom ut i 1950. Det første bandet vart ferdig i 1966, band II kom i 1978 og band III i 1994.

I 1930-åra vart det bygd opp eit omfattande nettverk av frivillige medarbeidarar frå heile landet, både fagfolk og interesserte lekfolk. Desse personane har sendt inn setlar til arkivet. Dei har dels skrive ned ord og uttrykk frå sine respektive målføre, dels har dei skrive ned (ekserpert) ord frå den nynorske litteraturen. Setelarkivet danner grunnlaget for mellom anna arbeidet med Norsk Ordbok, og i tillegg blir det brukt til å lage mindre ordbøker og ordlister. Arkivet har også sjølvstendig verdi som informasjonskjelde for norsk språk, historie og kultur. Heile arkivet inneheld om lag 3 millionar setlar.

Setlane i arkivet er ordna alfabetisk etter 1938-normalen, og dei er henta både frå munnlege og skriftlege kjelder. Arkivet inneheld dermed to hovudtypar av setlar: talemålssetlar, det vil seie setlar med dialektopplysningar frå ulike delar av landet, og litteratursetlar, det vil seie setlar med opplysningar frå skriftlege kjelder. I arkivet finst det også ein tredje kategori, kombinerte setlar, det vil seie setlar som inneheld både litteraturopplysningar og talemålsopplysningar.

#### *Status pr. 31.12.97:*

Total storleik på materialet:	2.916.426 ordsetlar
Konvertert totalt:	2.703.000 ordsetlar
Registrering som står att:	213.426 ordsetlar *
Korrektur totalt:	868.609 + 638.334 dobbeltregistrerte og like
Korrektur som står att:	1.409.483 *
Ferdigstilt totalt:	1.506.943 ordsetlar fordelt på: 868.609 korrekturlesne 638.334 dobbeltregistrerte og like

\* Delprosjekt nynorsk held fram til 30.6.98 med fullføring av konvertering, korrekturlesing og slutføring av materialet.

Ordsetlar på bokstavane L, M og N vart SGML-koda, og L vart korrekturlese av fagkonsulentane.

*Skildring av arbeidet:*

Hovudoppgåva til Delprosjekt nynorsk har vore å konvertere setelarkivet frå papir til elektronisk form. Arkivet har vorte digitalisert ved at ein har laga faksimilar (elektroniske bilete) av alle setlane og skrive inn basisopplysningane frå kvar enkelt setel, dvs. oppslag, grammatisk merking og kjelde. På den måten kan brukarane søkje etter ord dei er interesserte i, gruppere og hente ut ord frå dei ulike ordklassene eller bruke kjelda til å avgrense søk til td. ulike forfattarskap, tidsskrift og/eller geografiske område.

*Fase I (1992-1994) (SGML-tagging):*

Førebuing/tilrettelegging: Forprosjekt med utarbeiding av tidsplan. Nummerering, kopiering og seinare også sortering av ordsetlar.

Koding: SGML-koding av heile teksten på ordsetlane etter fastsett mønster.

I innskrivingseininga: Innskriving og tagging.

Delprosjektet: Opplæring og oppfølging av innskrivarane. Korrekturlesing og oppretting.

*Fase II (1994-98) (faksimilar og registrering):*

Førebuing/tilrettelegging: Utarbeiding av nye arbeidsrutiner og ny tidsplan. Fotografiering av alle ordsetlane hos Kodak. Utarbeiding av applikasjonar for innskriving og korrekturlesing i samarbeid med programmerarane. Utarbeiding av ein database over kjeldene i setelarkivet.

Koding: Innskriving av oppslag, grammatikk og kjelder.

Innskrivingseininga: Innskriving av oppslag, grammatikk og kjelder.

Delprosjektet: Opplæring og oppfølging av innskrivarane. Korrekturlesing og oppretting.

*Skildring av digitalt arkiv/database:*

Faksimileregisteringa er lagra i ein Oracle-database. Det er også utarbeidd eit førebels brukargrensesnitt til denne databasen. L, M, N er i tillegg lagra som SGML-koda filer.

*Teknisk skildring:*

Setelarkivet vart først digitalisert ved at kvar setel i arkivet vart skanna og lagra som digitale bilete på 4 CD-plater. I tillegg finst dei same databileta på 8 mm Exabyte-tape, og dei finst og på mikrofilm. Databileta er komprimerte og lagra i TIFF, CCITT Gruppe 4 format, ein fil for kvart bilete.

Faksimilane er lagra på mikrofilm, digitale kassetar, cd-ar og på disk (Dina). Oracle-databasen ligg også på Dina. Til faksimilerregisteringa utvikla programmerarane ein innskrivingsapplikasjon og ein korrekturapplikasjon. L, M, N som er SGML-koda ligg lagra på Hedvig/D8/Nynorsk/Setlar.

#### *Vidare arbeid:*

Det er fleire måtar å vidareutvikle materialet på. Særleg viktig er:

- normering av oppslagsord
- konsekvent grammatikkmerking
- digitalisering av uleselege setlar
- dataføring og innkorporering av nye ordsetlar
- vidareutvikling av grensesnittet
- utnytting av det SGML-tagga materialet

#### *Bruk av digitalt arkiv/database:*

Brukarene er først og fremst redaktørane av Norsk Ordbok, dessutan forskarar, studantar, lærarar, elevar og andre interesserte.

#### ***Nynorske tekstar***

##### *Om materialet:*

Det skanna materialet frå Delprosjekt nynorsk er i hovudsak tekstar som ein finn brokkar av på setlane i arkivet. Desse tekstane vil difor supplere seteldatabasen. Materialet er oppbygd av heile litterære verk av ymse nynorske forfattarar, 1921-utgåva av Bibelen, eit utval bøker frå NFL-serien (Norsk folkeminnelag) og nokre årgangar av fagtidsskriftet Syn og Segn. I tillegg har ein skanna eit utval eldre ordsamlingar. Det skanna materialet vil danne grunnlaget for eit større tekstkorpus der ein vil kunne gjere fritekstsøk.

##### *Status pr. 31.12.97:*

Total storleik på materialet:	ca 30.000 sider
Konvertert totalt:	ca 30.000 sider
Ferdigstilt totalt:	ca 16.200 sider
Står att å ferdigstille:	ca 13.800 sider *

\* Delprosjekt nynorsk held fram til 30.6.98 med fullføring av dette materialet.

##### *Skildring av arbeidet:*

Førebuing/tilrettelegging: I samarbeid med referansegruppa frå fagavdelinga vart det utarbeidd ei liste over skjønnlitteratur, sakprosa og eldre ordsamlingar som skulle skannast inn. Dei nødvendige tekstane vart skaffa og skanna. Det vart utarbeidd instruksjonshefte for merking og korrekturlesing av dei ulike tekstane.

Koding: SGML-koding etter eit utarbeidd mønster.

Innskrivingseininga: Koding og korrekturlesing.

Delprosjektet: Oppfølging av arbeidet med skanna materiale. Korrekturlesing, oppretting og ferdigstilling. Fleire av ordsamlingane blei tagga opp med normerte oppslagsord.

##### *Skildring av digitalt arkiv/database:*

Materialet er lagra som SGML-tagga tekst. 5 fulltagga ordsamlingar er tilgjengelege. Grensesnitt for skjønnlitteratur/sakprosa er under utarbeiding.

##### *Teknisk skildring:*

Dei skanna tekstane er lagra på Hedvig/D8/Nynorsk/Skanna.

*Vidare arbeid:*

Det føreligg planar om å skanne meir materiale, m.a. Aasen: Brev og dagbøker, Ross: Norsk Ordbok og Torp: Nynorsk etymologisk ordbok. Nynorskavdelinga er interessert i å få skanna inn nyare dialektsamlingar. Det er også ønskjeleg å normere fleire av ordsamlingane.

*Bruk av digitalt arkiv/database:*

Brukarene er først og fremst redaktørane av Norsk Ordbok, dessutan forskarar, studentar, lærarar, elevar og andre interesserte.

**Grunnmanuskriptet**

*Om materialet:*

Grunnmanuskriptet er eit gammalt ordboksmanuskript frå 1930-åra (ca 13.500 maskinskrivne A4-sider). Oppslaga er henta frå ordbøkene til Aasen, Ross, Schjøtt, Vidsteen, Torp o.a., men dei har alle fått nynorske definisjonar. Manuskriptet vart aldri utgitt som eiga ordbok, men det danna grunnlaget for arbeidet med Norsk Ordbok.

*Status pr. 31.12.97:*

Total storleik på materialet:	ca 13.500 sider
Konvertert totalt:	ca 13.500 sider
Korrekturlest totalt:	ca 13.500 sider
Ferdigstilt totalt:	ca 13.500 sider

*Skildring av arbeidet:*

Førebuing/tilrettelegging: Forprosjekt med utarbeiding av koding og strukturmerking. Det vart utarbeidd instruksjonshefte for merking og korrekturlesing av dei ulike tekstane.

Koding: SGML-koding etter eit utarbeidd mønster

Innskrivingseininga: Innskriving av tekst og merking av oppslagsord, etter kvart også strukturmerking.

Delprosjektet: Opplæring og oppfølging av innskrivarar. Strukturmerking og korrekturlesing av det innskrivne materialet.

*Skildring av digitalt arkiv/database:*

G-manus er lagra som SGML-tagga tekst.

*Teknisk skildring:*

Dei SGML-koda filene er lagra på Hedvig/D8/Nynorsk/Gmanus.

*Vidare arbeid:*

G-manus er ferdig, eventuelle rettingar kan først inn etter kvart. Brukargrensesnittet kan også forbetrast.

*Bruk av digitalt arkiv/database:*

Brukarene er først og fremst redaktørane av Norsk Ordbok, dessutan forskarar, studentar, lærarar, elevar og andre interesserte.

## *Nyordsarkivet*

### *Om arkivet:*

Nyordsarkivet til Seksjon for Leksikografi og målføregransking, Institutt for litteraturvitenskap og nordistikk, Universitetet i Oslo inneholder sitater fra aviser, tidsskrifter og ukeblad (i alt om lag 300.000 sitater) fra 174 forskjellige kilder. Innsamlingen har foregått over flere tiår. Derfor må begrepet "nyord" her forstås i et historisk perspektiv. Ordet var nytt, hadde en ny betydning eller ble brukt på en ny måte den gang det ble registrert.

### *Status pr. 31.12.97:*

Total størrelse på materialet:	95 arkivskuffer à ca. 2.500 sedler
Totalt ferdigstilt materiale:	67 arkivskuffer à ca. 2.500 sedler (70,5 %).

Dessuten er det blitt ferdigstilt noe av det materialet som er produsert ved underavdeling for bokmål. I alt er ca. 195.000 ekserpter blitt lastet i database og er tilgjengelig over Internett.

### *Beskrivelse av arbeidet:*

Sedlene som skulle skrives inn, ble kodet og kopiert og sendt til registreringsgruppene i Tønsberg, Larvik og Båtsfjord. Ekserptene ble kodet etter det systemet som allerede var i bruk ved underavdeling for bokmål. Bare det ordet som var ført opp som oppslagsord på seddelen, ble kodet. De øvrige ekserperte ordene på samme seddel er blitt kodet av fagkonsulentene og timebetalte korrekturlesere i korrekturfasen. Koding i så stort omfang avdekket inkonsekvenser og mangler i kodesystemet. Gjennom flere møter med de vitenskapelig ansatte ble kodesystemet justert. Opprettingen av korrektoren er gjort ved hjelp av underavdeling for bokmål (Elisabeth Holm), gruppa i Larvik og (fra uke 43 i 1997) av May Brit Larsen i Larvik. Opprettingen er sjekket av fagkonsulentene. Før filene blir lastet i en database, har fagkonsulentene foretatt en ekstra sjekk av alle de ekserperte ordene. Det er blitt kontrollert at de lemmatiserte grunnformene følger gjeldende rettskrivningsregler og at kodene stemmer.

### *Beskrivelse av digitalt arkiv/database:*

Nyordsdatabasen, som er tilgjengelig over Internett, preges av at det hovedsakelig er seddelarkivet som er lastet inn. Produksjonen til seddelarkivet var spesielt stor i tida 1968 til 1972. Derfor blir antall treff på søk i basen ofte uforholdsmessig store fra denne 5-årsperioden. Fra 1972 ble nyordsmaterialet registrert både på data og sedler. Det er bare en liten del av det datalagrede materialet som er produsert ved underavdelingen, som er blitt tilgjengelig i databasen. Materialet som ikke ligger i base, er tilgjengelig på et unix-område.

### *Teknisk beskrivelse:*

Alle nyordsfiler blir nummerert fortløpende og har endelsen .txt. Uferdige filer er lagret på unix-området /hf/hedvig/d9/nyord/Anne/nyord, de ferdige filene ligger på /hf/hedvig/d9/nyord/Anne/nyord/ferdig.

### *Videre arbeid:*

Det nyordsmaterialet som underavdeling for bokmål har digitalisert, bør korrigeres etter det reviderte kodesystemet slik at databasen får en jevnere kronologisk fordeling av ekserptene. Det materialet som ikke ble ferdig på prosjektet (29,5 % av de innskrevne sedlene) bør også ferdigbehandles. (Det er særlig ekserperte ord som begynner på t og u som ikke er blitt ferdig.) Databasen for nyordsmaterialet er ferdigutviklet, men noen justeringer kan bli aktuelt.



### ***Litterære tekster - bokmål***

*Status pr. 31.12.97:*

Total størrelse på materialet: 59.100 boksider  
Totalt ferdigstilt materiale: 48.500 boksider (82 %).

#### *Beskrivelse av arbeidet:*

I stedet for å digitalisere sedlene i disse arkivene er bøkene som seddelarkivene bygger på, blitt dataført. Bøkene ble kjøpt inn fra antikvariat eller lånt fra bibliotek. Før bøkene kunne sendes ut til registreringsgruppene i Fauske og Larvik, måtte de kopieres (utført av timebetalt hjelp). De fleste bøkene er skannet inn. Bøker med gotisk skrift er blitt tastet inn. Også skuespill er tastet inn fordi denne teksttypen blir ekstremt tett tagget. Skanningen er utført i Fauske og Larvik, inntasting bare i Larvik. Gruppene har tagget tekstene og lest minst to korrekturer. Fagkonsulentene laget generelle veiledninger for SGML-tagging av prosa, brevsamlinger, skuespill og dikt. De laget også mange veiledninger for hvordan enkelte bøker skulle tagges. DTDer for bruk i editor-programmet Author/Editor er utviklet av Silvester Pozarnik og Christian-Emil Ore. Etter hvert ble registreringsgruppene opplært i bruk av Author/Editor. (Ved å validere tekstene ved hjelp av Author/Editor får man sjekket at en tekst er korrekt tagget.) Gruppene kunne da utføre mye av sluttkontrollen av tekstene. Stikkprøvekontrollen som fagkonsulentene har foretatt, har hatt som retningslinje at mer enn én tekstfeil per 10 sider, førte til at innsendt produksjon måtte korrekturleses på nytt.

#### *Beskrivelse av digitalt arkiv/database:*

Arbeidet med et framvisningsverktøy for de ferdigbehandlede skjønnlitterære tekstene er påbegynt. Mange av forfatterskapene er ferdigbehandlet og er klare til å legges ut på Internett.

#### *Teknisk beskrivelse:*

Alle tekstfilene er sortert etter forfatter, alle verkene til en forfatter er samlet i én mappe. De ferdige tekstene er skilt ut i egne mapper med navn som begynner med FERDIG. Alle filene er lagret på unix-området /hf/hedvig/d9/Bokmaal. Det eksisterer to versjoner av de ferdige filene. Word-filene er lagret i rent tekstformat og har som regel endelsen .ren, Author/Editor-filene har endelsen .ae.

#### *Videre arbeid:*

Innskanna tekster som ikke er blitt ferdigbehandlet i prosjektperioden (ca.10.600 boksider), bør ferdigstilles.

Arbeidet med å utvikle et framvisningsverktøy for de elektroniske tekstene er så vidt påbegynt. Noen av forfatterskapene kan det være aktuelt å viderebehandle. For eksempel kan det være aktuelt å lage en konkordans med lemmatiserte former bygd på Wergelands skrifter.

#### *Bruk av digitalt arkiv/database:*

Når de digitaliserte tekstene blir tilgjengelig over Internett, er det mange som kan få nytte av dem, både leksikografer, andre fagmiljøer (særlig litteraturforskere), bibliotek og allmennheten.

En ev. anvendelse av basene i forbindelse med fjernundervisningsopplegget ved Institutt for nordistikk og litteraturvitenskap vil være av betydning for videre utvikling og vil tydeliggjøre bruksmulighetene.

## ***Gammalnorsk seddelarkiv***

### *Om arkivet:*

Dette er et tradisjonelt, papirbasert ordarkiv over norske mellomaldertekster skrevet med latinsk skrift, ca. 600.000 sedler. Seddelarkivet utgjøres av litterære tekster i vid forstand, et par mindre lovhåndskrift, samt noe diplommateriale. Sedlene er oppstilt både strengt alfabetisk og etter en finurlig grammatisk, nærmest redigert orden.

### *Status pr. 31.12.97:*

Total størrelse på materialet:	600.000 sedler
Konvertert totalt:	600.000 sedler (seddel eller tekstgrunnlag)
Gjenstår bearbeiding:	65.000 sedler
Ferdigstilt totalt:	535.000 sedler

### *Beskrivelse av arbeidet:*

Teksten sedlene er basert på, er korrigerede utgaver av eldre trykte tekstutgaver av de originale håndskriftene. Det var derfor ikke mulig å laste inn de trykte tekstene. Det kompliserte oppgaven. På den annen side inneholder sedlene både linje og sidehenvisninger til den trykte utgaven. Det var derfor mulig å legge sedlene inn i et database-system på en slik måte at den originale teksten kun blir skrevet en gang. Opplysningene på de enkelte sedlene ble "ført inn" i teksten. Til dette brukte vi et system basert på Claris HyperCard. Vi ville dermed ha tekster på til sammen omkring 535.000 løpende ord, der hvert ord er påført grammatiske opplysninger og normalisert grunnform.

Arbeidet startet våren 1991 med et par tidsstudier med innleid hjelp. Samme høst kom den første registreringsgruppa i Hokksund i gang. Arkivet ble kopiert opp og manuelt nummerstemplet. Materiale som kom fra fagkonsulentene, ble lagt til rette for hver innskriver, og ferdig materiale ble sendt tilbake etter bestemte regler. Utskrifter av materialet ble korrekturlest (flere ganger) og rettingene ført inn på fil (disketter), for til slutt å bli sendt til fagkonsulentene til kontroll. Det ble gjort forsøk med optisk lesing av seddelkopiene. Men det var såpass mye "uren" tekst at skannerprogrammet aldri greide å "læres" tilstrekkelig opp (høy feilprosent) og med mye oppretting som følge. Det gikk altså raskere og bedre å skrive inn direkte på maskin.

I 1997 ble det dataført materiale direkte etter eldre trykte tekstutgaver. Disse tekstfilene (som får lagt inn markering av tilhørende seddelformat) skal eksporteres til databaseprogram (i første omgang HyperCard eller FileMaker Pro). Tekstene vil etter vanlig korrektur og oppretting bli påført ordboksverkets opplysninger om hver enkelt ordforms ordklassemarkering, grammatisk bestemmelse, tekstreferanse osv. Først når dette er gjort, er konverteringen utført.

### *Beskrivelse av digitalt arkiv/database:*

Registreringssystemet gjorde det mulig på en effektiv måte å føre inn informasjonen på sedlene på riktig plass i tekstene. Resultatet er som nevnt over, tekster på omkring 550.000 løpende ord der hvert ord er påført grammatiske opplysninger og normalisert grunnform. Tekstgrunnlagene for de resterende 65.000 sedlene samt alle de andre gammelnorske tekstene (diplomer, lover, se også nedenfor) er klare til bruk i såkalte lemmatiseringsprogrammer. Dette er programmer som på en svært effektiv måte kan brukes til å lage elektroniske ekvivalenter til de tradisjonelle sedlene. Seddelmaterialet er tilgjengelig for bruk i form av FileMaker databaser. Materialet bli lagt inn i felles leksikalske databaser for norsk (bokmål, nynorsk og gammelnorsk).

## **Norske diplomer**

### *Om materialet:*

Dette er ordboksverkets egne håndskrevne avskrifter av originaldiplomer fra forskjellige arkiv, samt noen i privat eie.

### *Status pr. 31.12.97:*

Total størrelse på materialet:	1.829 diplomer, årgangene 1339 - 1405
Konvertert totalt:	1.829 diplomer
Korrekturlest totalt:	661 diplomer
Gjenstår å korrekturlese:	1.168 diplomer
Gjenstår å rette:	1.739 diplomer

Årgangene 1309-1338 er dataført i Gammalnorsk Ordboksverk.

### *Beskrivelse av arbeidet:*

Tilretteleggingen har bestått i utarbeiding av standard for innskriving. I gruppene bestod arbeidet av innskriving av tekstene. Tekstene er blitt SGML-kodet. Fagkonsulentene har kun delvis korrekturlest materialet. Alle diplomavskriftene er dataført, men korrekturlesing og ferdigstilling ble nedprioritert i forhold til arbeidet med *Diplomatarium Norvegicum*. Dette arbeidet blir nå tatt over av Ordboksverkets ansatte.

### *Beskrivelse av digitalt arkiv/database:*

De korrekturleste tekstene vil bli lemmatisert slik dette er beskrevet i slutten av avsnittet om seddelarkivet.

## ***Diplomatarium Norvegicum***

### *Om materialet:*

*Diplomatarium Norvegicum I-XXII. Christiania/Oslo 1847-* er en kildeserie som ordrett og på originalspråket gjengir innholdet i dokumenter som er eldre enn 1570. I 1998 er det 150 år siden det første bindet kom ut, og i alt er det gitt ut 22 bind med ca. 19.000 dokumenter. *Diplomatarium Norvegicum* er selve flaggskipet blant norske kildeutgaver og det viktigste verket for alle som arbeider med middelaldermateriale.

### *Status pr. 31.12.97:*

Total størrelse på materialet:	21 bind, 20.000 trykte sider
Konvertert totalt:	21 bind, 20.000 trykte sider
Korrekturlest totalt:	21 bind, 20.000 trykte sider
Ferdigstilt totalt:	21 bind, 20.000 trykte sider

Bind. XXII er dataført ved Riksarkivet, Kildeskriptavdelingen.

### *Beskrivelse av arbeidet:*

Tilretteleggingen har bestått i utarbeiding av standard for innskriving. I gruppene bestod arbeidet enten i innskanning, OCR-behandling og oppretting av den innskannede teksten, eller av innskriving av tekst som ikke lot seg skanne på grunn av trykkkvalitet. Teksten er blitt SGML-kodet. Fagkonsulentene har korrekturlest materialet.

### *Beskrivelse av digitalt arkiv/database:*

Basen ligger i SGML-tagget tekstformat og er gjort tilgjengelig over Internett.

*Videre arbeid:*

I et mer langsiktig perspektiv kunne en ønske at også tekstene i Diplomatarium Norvegicum-basen kunne gjøres tilgjengelig med normaliserte oppslagsformer som søkekriterium. Slik det nå er, kreves det betydelig kunnskap og tålmodighet om en vil gjøre søk i unormalisert tekst som dette.

Det elektroniske diplomatariet er ikke ferdig slik det nå foreligger. For det første gjenstår det en siste korrekturlesing av 5.000 trykksider. Dette er en viktig, men prosaisk og kostnadskrevende aktivitet som kan fortsette parallelt med andre videreutviklinger. Hva som kanskje kan være viktigere, er å koble resultatene fra de store nysatsningene innenfor arbeidet med diplomer sammen med det elektroniske diplomatariet. Vi tenker her spesielt på nyavskriftene og ordopplysningene ved Gammelnorsk Ordboksverk, Regesta Norvegica med de nye sammendragene og kildeopplysninger, og ikke minst på samlingen av diplomfoto ved Riksarkivet og Gammelnorsk Ordboksverk.

Det er etablert et samarbeid mellom Riksarkivet Gammelnorsk Ordboksverk og Dokumentasjonsprosjekts etterorganisasjon for å lage en elektronisk kombiutgave av Diplomatariet. I første omgang undersøker vi mulighetene av å digitalisere svart-hvitt-fotografiene ved Riksarkivet og ved Gammelnorsk ordboksverk. Det ideelle vil etter manges mening være nye digitale farvereproduksjoner, men de eksisterende svart-hvitt-fotografiene er gode nok for kontroll av teksten i det trykte diplomatariet.

*Bruk av digitalt arkiv/database:*

Via Internett er det adgang til Diplomatarium Norvegicum. Prøvedatabasen er under kontinuerlig oppdatering, etter hvert som materialet ferdigstilles. Dette er et godt tilbud i første rekke til dem er kjent med diplomatariet og som søker opplysninger i dette materialet. I ordboksverket er det gjort bruk av denne tilgangen til et stort korpus. I andre miljø, ikke minst blant historikere, har det vært uttrykt stor tilfredshet med denne nye tilgjengeligheten til norske mellomaldertekster. Det er oppslag i basen fra flere kanter av verden.

**Generelt: Institutt for nordistikk og litteraturvitenskap, NTNU**

Delprosjektansvarlig: Professor Arnold Dalen, Institutt for nordistikk og litteraturvitenskap, NTNU

Fagkonsulenter i delprosjektperioden: Tor Erik Jenstad, Laila Sakshaug, Mette Vaag og Anne Grethe Wold

Korrekturlesere/andre medarbeidere i delprosjektperioden: Britt Karin Holthe, Vebjørn Sletten

IT-faglig ansvar: Lars-Jørgen Tvedt

Antall fagkonsulentårsverk i perioden 1994 - 1997: 6

Kontaktperson i fagenheten fra 1.1.98: Tor Erik Jenstad

## **Trønderordboka**

### *Om arkivet:*

Ordbok over trøndermåla er et prosjekt som startet 1981 ved Nordisk institutt, UNIT. Målet er å lage ei ordbok over den samlede ordmassen i målføra i det trøndske dialektområdet. Det er gjort et omfattende innsamlings- og ekserperingsarbeid, og til sammen har en nå samlet inn ca 187. 000 sedler som viser eksempler på trøndske dialektvarianter hentet fra litteratur og talemål. Sedlene er av samme type som sedlene ved Norsk Ordbok (se Delprosjekt nynorsk). De består av et oppslag, opplysninger om/eksempel på bruk av dette oppslaget og kildeinformasjon. En finner her mye stoff som verken er registrert i ordbøkene til Aasen og Ross eller i arkivene ved Seksjon for leksikografi og målføregransking, UiO.

### *Status pr. 31.12.97:*

Total størrelse på materialet:	187.433 sedler.
Konvertert totalt:	187.344 sedler
Korrekturlest i delprosjektet totalt:	186.899 sedler
Ferdigstilt totalt:	186.899 sedler

445 sedler gjenstår som av ulike årsaker ikke kan eller bør rettes opp. Oppslagene er indeksert og korrekturlest. Norsk Ordbok finansierte dette arbeidet med kr. 10.000

To bøker er skannet:

- Karl Braset: "Hollraøventyra og Svanøventyra" -Gamalt paa Sparbumaal - 1910  
Kristine og Jo Skarland: Dialektord fra Høylandet
- Papirutgaven av Trønderordboka (redigert av Arnold Dalen og Tor Erik Jenstad) kom ut i 1997.

### *Beskrivelse av arbeidet:*

Sedlene som skulle tagges, ble stemplet og kopiert (1994, bokstavene allmenn-d). Seddelarkivet ble i sin helhet sendt til Kodak våren 1995, hvor det ble skannet og lagret på CD-plater.

Materialet ble så registrert inn i en innskrivingsapplikasjon tilrettelagt for vårt behov. Delprosjektet hadde to innskriverlag på 8 personer som gjorde dette arbeidet på Unireg i Mo i Rana. Disse måtte læres opp, og opplæringsmateriale måtte utarbeides.

Det ble etter hvert også tilpasset en korrekturapplikasjon. Korrektur på innskrivernes arbeid ble startet opp, samt tilbakemeldinger på kvalitet. Etter at konverteringen var avsluttet i juni 1996, ble arbeidsoppgavene færre. De siste 18 månedene var korrektur primæroppgave.

### *Beskrivelse av digitalt arkiv/database:*

Overgang fra tagging til faksimile, se Norsk Ordbok, setelarkivet.

### *Teknisk beskrivelse:*

Seddelarkivet ble først digitalisert ved at hver seddel i arkivet ble skannet og lagret som digitale bilder på 4 CD-plater. I tillegg finnes de samme databildene på 8 mm Exabyte-tape, og de finnes også på mikrofilm. Databildene er komprimert og lagret i TIFF, CCITT Gruppe 4 format, en fil for hvert bilde. For hvert bilde har en skrevet av hovedinformasjonen på kortet, og denne informasjonen er lagret i en relasjonsdatabase. Pr. 31.12.1997 vil disse dataene ligge i en Oracle-database på ei maskin ved Universitetet i Oslo.

Informasjonen som er registrert, er oppslagsord, grammatisk informasjon og kildeinformasjon. Kildeinformasjonen er representert ved et fritekstfelt som inneholder forklaring av en kilde. Kildeinformasjonen er også representert ved en peker inn i en referansetabell med oversikt over alle kilder som er brukt i arkivet. I fritekstfeltet ligger også koder for heimpling av seddelen. Denne informasjonen er foreløpig ikke representert på andre måter i relasjonsdatabasen.

All innskriving og korrektur er gjort mot databasen, og alle sedler er fortløpende merket med korrekturstatus. Koblingen mellom digitale bilder og konverterte data ligger også i databasen. Sammen med de digitaliserte bildene utgjør dataene i Oracle-databasen den komplette samlinga med data som er konvertert. Det finnes ingen andre lagringsformer.

Etterorganisasjonen må i samarbeid med fagavdelinga konvertere dataene i innskrivingsbasen til en mer formålsrettet form for database til bruk i vitenskapelig arbeid.

Indeks over oppslagsord det kan søkes på i materialet foreligger på Word-fil (tabell). Redaksjonen i Norsk Ordbok har fått kopi av denne, både i papirutskrift og på diskett. Denne vil kunne oppdateres kontinuerlig etter som nytt materiale kommer inn.

#### *Videre arbeid:*

Det må gis muligheter for å utnytte heimplings-taggen og til å legge inn tilvekst

#### *Bruk av digitalt arkiv/database:*

Materialet blir brukt av redaksjonen i Norsk Ordbok og av andre i forskningsmiljøet, etter hvert også av en del privatpersoner.

## **Navnegransking**

### **Generelt: Seksjon for navnegransking, INL, UiO**

Seksjon for navnegransking er en seksjon under Institutt for nordistikk og litteraturvitenskap ved Universitetet i Oslo. En av seksjonens viktigste oppgaver er å samle inn og arkivere navn, særlig stedsnavn og personnavn.

Stedsnavn samles ved informantintervju i lokalmiljøene, men seksjonen arbeider også med å bygge opp samlinger av gamle skriftformer fra trykte og utrykte kilder (bøker og kart). Personnavn hentes stort sett fra skriftlige kilder. Materialet har tradisjonelt vært arkivert på kartotek kort (Norsk stadnamnarkiv).

Delprosjektansvarlig i prosjektperioden: Førsteamanuensis Bottolv Helleland og førsteamanuensis Tom Schmidt, Seksjon for navnegransking

Fagkonsulenter i perioden 1991 - 1997: Tore Storebø, Bjørg Spørck, Ivar Lund, Solveig Wikstrøm og Morten Holtet

Systemansvar: Lars-Jørgen Tvedt

Antall fagkonsulentårsverk i perioden 1992 - 1997: 15,25

Kontaktperson i fagenheten fra 1.1.98: Førsteamanuensis Tom Schmidt og fagkonsulent Morten Holtet.

## *Etymologiregisteret*

### *Om arkivet:*

Nordens største navnerregister finnes i Ortnamnsarkivet i Uppsala. Siden Ivar Modéer startet arbeidet i 1931, er gammel og ny faglitteratur systematisk blitt gjennomlest for å hente ut opplysninger om navn og navneledd. I alt består det av ca. 240.000 arkivkort, og tallet øker etter hvert som ny faglitteratur blir ekserpert. I registeret finner man nordiske navn og navneledd med litteraturhenvisninger. Foruten det nordiske kjerneområdet dekkes "koloniområder" som Baltikum, Normandie og De britiske øyer samt Nord-Amerika.

Hele spekteret av relevant litteratur er representert: De store standardverkene, tidsskrifter, hovedoppgaver, samt oppslagsbøker, bygdebøker, avisartikler og mer overraskende kilder så som brev, avisnotiser og lokalradioprogrammer.

### *Status pr. 31.12.97:*

Total størrelse på materialet:	240.000 kort
Konvertert totalt:	240.000 kort
Korrekturlest totalt:	240.000 kort
Ferdigstilt totalt:	240.000 kort

### *Beskrivelse av arbeidet:*

#### Forberedelser:

Forslag til skjema for innskriving av materialet utarbeidet av fagkonsulent og delprosjektansvarlig på bakgrunn av korteksempler. Endelig innskrivingsmal utarbeidet av fagkonsulent i samråd med programmerergruppa. Kopiering av hele originalmaterialet ved Ortnamnsarkivet i Uppsala av fagkonsulent. Utarbeidelse av litteraturliste. Nummerering av titler i litteraturoversikt fra Ortnamnsarkivet.

Tilrettelegging: Nummerering og markering av kort. Operasjonen ble først utført av fagkonsulenten og korrekturlesere, seinere av innskriverne i registreringsenheten.

#### Innskrivingsenheten:

Registrering. Korrektur i tre omganger. Nummerering av kort.

Regelmessig kvantitetsrapportering til registreringsleder og fagkonsulent.

#### Delprosjektet:

Korrektur og oppretting av registrert materiale. Vedlikehold og oppdatering av litteraturlista. Utarbeidelse og oppdatering av opplæringsmateriell. Regelmessig kvalitetsrapportering til innskrivingsenheten og registreringsleder.

#### *Teknisk beskrivelse:*

Filene er sortert med nummer innenfor den mappa (disken) de ligger i. Det har ikke vært operert med statusmerking av selve filene. Originalmaterialet har blitt flyttet over i magasinet ved Seksjon for navnegransking etter hvert som det har blitt ferdigstilt. En mappe (disk) med et nummer som ikke refererer til en bunke papirmateriale på fagkonsulentens kontor har blitt ansett som ferdigstilt. Filene er fordelt i mapper (disker) 1-190 på pchfdoc\uppskorr.

#### *Videre arbeid:*

Kortene vil bli overført til en relasjonsdatabase. Databasen bør kunne oppdateres/utvides med nye ekserpter. Dette kan gjøres ved at flere navnegranskingsinstitusjoner får tilgang til en oppdateringsfunksjon i databasen. Det bør være rom for mindre rettelser, rettelser kan foretas i sammenheng med den funksjonen som er nevnt ovenfor.

*Bruk av digitalt arkiv:*

Konverteringen av dette registeret har gjort det mulig å dra inn et langt større komparativt materiale enn det som ellers ville være mulig. Databasen vil også føre til en effektivisering av granskingsprosessen.

***Matrikkelutkastet fra 1950***

*Om arkivet:*

Matrikkelutkastet består av ca. 81.000 lister over fast eiendom i Norge ordnet etter stigende gårdsnummer innenfor hver kommune, utarbeidet av Finansdepartementet. Matrikkelrevisjonen ble aldri fullført. Derfor mangler Finnmark fylke, og derfor snakker man om et utkast.

Foruten navn på gårdsbruk finner vi navn på villaer, fritidseiendommer, parseller, offentlige og private institusjoner osv. Matrikkelutkastet er et sentralt hjelpemiddel for Statens navnekonsulenttjeneste i arbeidet med normering av stedsnavn, og siden det også inneholder grunneierens navn, kan det være til nytte for personnavnforskere. Samtidig representerer det ofte vår eneste kjente og tilgjengelige samlede dokumentasjon av navn på nyere enheter.

*Status pr. 31.12.97:*

Total størrelse på materialet:	81.300 lister
Konvertert totalt:	81.300 lister
Korrekturlest totalt:	30.000 lister (stikkprøvekorrektur)
Ferdigstilt totalt:	81.300 lister

*Beskrivelse av arbeidet:*

Kopiering av originalmaterialet ved Representralen SV.

*Innskrivingsenheten:*

Registrering: materialet er skrevet inn i Word-tabeller. Korrektur i tre omganger. Regelmessig kvantitetsrapportering til registreringsleder og fagkonsulent.

*Delprosjektet:*

Korrektur og oppretting av registrert materiale. Utarbeidelse og oppdatering av opplæringsmateriell. Regelmessig kvalitetsrapportering til innskrivingsenheten og registreringsleder.

*Beskrivelse av digitalt arkiv/database:*

På bakgrunn materialets enkle og regelmessige karakter ble det besluttet at materialet bare skulle gjennomgå stikkprøvekorrektur i delprosjektet, derfor er omlag 30% av materialet korrekturlest i delprosjektet. Derimot har hele materialet vært igjennom et feilsøkingprogram som har gjort det mulig å luke vekk feil i materialets struktur. Programmet ble utarbeidet i samarbeid med programmerer.

*Teknisk beskrivelse:*

Ett filnavn består av fylkesnr., de fire første bokstavene i herredsnavnet og filnummeret. Statusmerking på filene: Se samme punkt under Etymologiregisteret. Filene er lagret i mapper, en for hvert fylke, på pchfdoc\navn\matrikke. Filene er lagret som ren tekst.



#### *Videre arbeid:*

Materialet vil bli overført til en relasjonsdatabase. Databasen bør utformes på en måte som muliggjør kopling med andre matrikler o.l. materiale. Brukere ved Seksjon for navnegransking har tilgang til originalmaterialet og vil gis mulighet til å foreta rettelser.

#### ***Bustadnavnregisteret***

##### *Om arkivet:*

Registeret finnes ved Seksjon for navnegransking, Universitetet i Oslo og omfatter navn på bosteder (gårder, bruk og plasser) fra ti fylker i landet, ordnet etter stigende gårdsnumre innenfor hvert herred.

Registeret består av ca. 109.000 arkivkort med opplysninger om skriftform, uttaleform med preposisjon og evt. dativform, eldre skriftformer, andre skriftlige eller muntlige former av navnet og ofte kommentar, f.eks om topografi, særtrekk i dialekten og navnetolkning. Opplysningene er samlet inn av Kåre Hoel og Alv Muri, fra 1950 til 1975.

##### *Status pr. 31.12.97:*

Total størrelse på materialet:	100.000 kort
Konvertert totalt:	100.000 kort
Korrekturlest totalt:	100.000 kort
Ferdigstilt totalt:	100.000 kort

##### *Beskrivelse av arbeidet:*

Forberedelser: Analyse av materialet, gjort i delprosjektet i 1992 og 1993 og før registreringsstart våren 1995. Analysen gikk ut på å:

1. Finne en egnet metode for markering av hovedledd og utmerkingsledd i oppslagsord.
2. Se på mulighetene for transkribering av lydskrift. Lydskriften viste seg for avansert for innskriving; også hvis man evt. skulle erstatte med tegn eller entiteter som er lett konverterbare. Kortene er derfor skannet.
3. Klassifisere typer opplysninger på kortene for å finne en felles mal og hvilke koder som er hensiktsmessige.
4. Samarbeide med programmerer for utvikling av applikasjon til innskrivingen.

##### *Tilrettelegging:*

Kopiering av originalkort (1993) og skanning (fotografering) av materialet ved Kodak Norge, sommeren 1995. Markering av kortkopier: analyse av oppslagsord i hovedledd og utmerkingsledd. Utarbeiding av koder 1992. Kodesystemet revidert før oppstart i 1995, da fagkonsulentene mente det ville være vanskelig for innskriverne å bruke en del av kodene (koder som forutsatte språklig kompetanse og kunnskap om lydskrift). Kodene er av to typer: De som sorterer ut realopplysninger om navnet, hovedledd og utmerkingsledd i dette, gårds- og bruksnummer, lokalitetstype, nøyaktig stedfesting med evt. kartfesting, parallelle skriftformer og eldre skriftformer, evt. informant og år, oppskriver og år. Videre er det et kommentarfelt der den løpende tekst på kortet skrives inn. En del av realopplysningene har blitt hentet ut fra dette kommentarfeltet.

#### Innskrivingsenheten:

1. Innskriverne får utdelt kortene fra ett herred som de lager dokumenter - 50 kort - av.
2. Innskriving av kort i applikasjon som programmereren har laget spesielt til denne arkivtypen, "Bustreg". I denne applikasjonen, som har visuell likhet med kortene, skrives opplysningene inn i felt. Innskrivingen har fordret grundig opplæring og stiller i tillegg krav til godt skjønn hos innskriverne.
3. Korrekturlesing, først av eget materiale, og to korrekturer for andre deltakere i gruppa. Hvert dokument skal ha gjennomgått tre korrekturer før forsendelse. Kvalitetskravet har vært på maksimum 5 feil pr. dokument, som i Word-utskrift utgjør i snitt 30 sider. Oppretting av feil mellom hver korrektur og ny utskrift.
4. Forsendelse av ferdig registrerte filer på diskett samt kopi av kort til fagkonsulenten hver 14. dag.

#### Delprosjektet:

1. Utarbeidelse av manualer, forberedelse av undervisning og opplæring av innskrivere.
2. Korrektur på filene og tilbakemelding til hver enkelt innskriver hver 14. dag. I tillegg skrives kvalitetsrapport til hele gruppa med gjennomgang av feiltyper og samlet snitt på produksjonen.
3. Oppretting av filene og validering i Bustreg, slik at filene i praksis ikke skal inneholde kodefeil. Deretter blir filene lagt ut på definert fellesområde (se nedenfor).
4. Kopling av ferdig registrerte enheter (hvert registrerte kort) mot skannede kort i egen applikasjon.

#### Teknisk beskrivelse:

Filene har først fylkesnummer (Tilsvarende ikke bindnummer i Norske Gaardenavn), deretter de fire eller fem første bokstavene i herredsnavnet og to siffer som angir filasnummer innenfor herredet. Stigende nummerrekkefølge etter stigende matrikelnummer. Filene er lagret som ren tekst og kalt .bsn. Eks: 1arema01.bsn. Der herredsnavnet har skandinaviske bokstaver, brukes bokstavene ae, oe og aa. Der det står vestre eller østre foran, vil dette bli forkortet til en (to) bokstaver. Eks.: Østre Moland = oemol.

Alle filene som ligger på fellesområdet, er klare for innlasting i database. Enheter der ikke alle opplysninger har latt seg skrive inn, har fått koden <merk>, og hensikten er å ha en funksjon med "se kortet" i grensesnittet til databasen. Lagring: pchfdoc\navn-\bustreg\fylke01, 02 osv. Filene er lagret som ren tekst.

#### Videre arbeid:

Integrering av andre av seksjonens dataregistrerte arkiver er ønskelig. Man bør også gjøre en vurdering av hvordan man kan registrere lydskrift ut fra en felles mal (v.h.a. entiteter e.l.). De ansatte ved avdelingen som arbeider mest med materialet, vil få tilgang til å rette opp eventuelle feil i materialet. Alt materiale er korrekturlest av fagkonsulentene, men senere stikkprøver viser at deler av materialet ikke er feilfritt. Dette gjelder helst fylke 1 og 2 (der er også originalene håndskrevne og ikke maskinskrevne som lenger ut i materialet), og fylke 6 (7) og 8 kan også ha feil. Feilmengden er likevel ikke stor; stikkprøver har vist maks ca. 2 feil pr. fil.

#### Bruk av digitalt arkiv/database:

Databasen gir mulighet til å søke opp ulike opplysninger, samlet eller geografisk avgrenset, f.eks:

- enkeltnavn
- navneledd (forledd eller etterledd, f.eks -rud)
- forekomst av ulike preposisjoner
- eldre skriveformer av navnet, fra midten av 1600-tallet fram til 1900-tallet
- navn innenfor et avgrenset område (herred, fylke)
- lokalitetstype, f.eks navn på husmannsplasser

Bustadnavnmaterialet bør ha tett kopling til Norske Gaardnavne, da det tematisk, faglig og topografisk har svært mange likhetstrekk med dette materialet. Bustadnavnregisteret bygger på Norske Gaardnavne og bør sees på som et naturlig supplement til verket: De registrerte gårdsnavnene med evt. andre navneforklaringer og ofte annen (og mer nøyaktig) lydskrift, men viktigst er den store oversikten over bruksnavn som representerer en annen, yngre navnegenerasjon som i større grad er bygd på folkekulturen (de eldre gårdsnavnene er i overveiende grad rent terrengbeskrivende). Dette materialet vil gi et bredere og tematisk noe annerledes bilde av navnetradisjonene i landet. Dessuten er også dette materialet interessant for slektsgranskere, siden f.eks. også bruksnavn blir brukt som slektsnavn.

Stedsnavnene passer godt som materiale i tverrfaglige problemstillinger, for leksikografien er det interessant å se på ordtilfanget som stedsnavnene representerer, både forekomst, form og betydning. Stedsnavn er også en viktig hjelpedisiplin for historie, agrarhistorie, religionshistorie og arkeologi. Ved å kople navnedata til lokalitetene har en bedre både kvantitativt og kvalitativt grunnlag for å dra slutninger om menneskelig aktivitet i eldre tid. I et religionshistorisk perspektiv er det interessant å se etter arkeologiske registreringer på og topografisk distribusjon av lokaliteter med navn som kan tolkes som uttrykk for førkristne religionsforhold. En lang rekke andre kulturhistoriske og naturhistoriske opplysninger kan trekkes ut av stedsnavntilfanget, som beite- og dyrkingsforhold, husdyrhold, forekomster av visse tre- og plantearter, dyrearter, jordsmonn, landskapsformer.

Navnedatabasene gir et godt grunnlag for nettbasert fjernundervisning på flere nivå. Stedsnavn og personnavn er ofte benyttet som emner for særoppgaver i den videregående skolen. Navnematerialet kan analyseres bl.a. ut fra språklige, sosiologiske og lokalhistoriske synsvinkler. Videre er navnedatabasene til stor nytte for navnekonsulenttjenesten, kartverket og andre vedtaksorgan når det skal fastsettes skrivemåte etter lov om stadnamn.

## ***Norske Gaardnavne***

### *Om verket*

Arbeidet med Norske Gaardnavne (NG), 18-bindsverket som regnes som hovedverket i norsk navnegransking, ble påbegynt ved slutten av forrige århundre. Foranledningen var revisjon av landets matrikkel, og det ble nedsatt en komité med prost Johan Fritzner og professorene Oluf Rygh og Sophus Bugge for å revidere skriftformen av gårdsnavnene. Komitéen fikk i 1896 statsstøtte for å publisere arbeidet, begrunnet med at "Oplysningerne ere af Interesse for nær sagt enhver Gaard i Landet, foruden at de har stor Betydning for det nationale Sprogstudium" (Forord 1898:VIII).

NG består av 18 bind, ett for hvert av landets fylker, og inneholder opplysninger om samtlige av landets matrikelgårder og en del bruk under disse, ca. 55.000 artikler til sammen. De er ordnet herredsvis og etter stigende gårdsnumre. Første bind, Smaa-

lenenes amt, utkom i 1897, og arbeidet avsluttet med Finmarkens amt som det siste i 1923. Oluf Rygh døde i 1899, og hans bror, Karl Rygh, fortsatte arbeidet på bakgrunn av etterlatte notater.

I tillegg til de atten bindene med gårdsnavn fylkesvis, ble det også publisert et Forord og Indledning (1898) til verket, samt et Fællesregister (1936).

- Artiklene til hver matrikelgård gir opplysninger om:
- Matrikelnummer
- Navneform
- Uttaleform med eventuell dativform
- Eldre skriveformer hentet fra middelalderkilder og senere kilder fram til 1723
- Etymologi, ofte med gno. normalform, som tar utgangspunkt i navnesammensetninger (for- og etterledd), navneslektskap i omegnen, særtrekk i målføret og beskrivelse av terrengforhold

Hvert bind inneholder videre registre over de oppgitte gårds- og bygdenavn, elve- og innsjønavn, person- og gudenavn og sammensetninger i forhold til etterledd. Forord og Indledning presenterer bakgrunnen for arbeidet og navneforskning som fag. Videre er det liste med forklaringer av gammelnorske appellativer som ofte forekommer i navn, ordnet alfabetisk. Fællesregister er register over de fleste gårdsnavnene i verket; både oppslagsformer og middelalderformer. Dessuten er det mindre registre over andre stedsnavn og person- og innbyggernavn.

"Elektroniske Norske Gaardnavne" er støttet av Norsk kulturråd og følgende fylkeskommuner: Østfold, Vestfold, Akershus, Rogaland, Hordaland, Møre og Romsdal, Sogn og Fjordane, Sør-Trøndelag og Nord-Trøndelag.

*Status pr. 31.12.97:*

Total størrelse på materialet:	20 bind totalt. I snitt 400 sider pr. bind, samt "Indledning" (100 s.)
Konvertert totalt:	13 bind (inkl. "Indledning")
Korrekturlest totalt:	13 bind
Ferdigstilt totalt:	13 bind
Gjenstår å konvertere:	7 bind*

\* delprosjektet fortsatte i 6 måneder i 1998 på bevilgninger fra ulike fylkeskommuner og Språkrådet

*Beskrivelse av arbeidet:*

Forberedelser: Analyse av materialet, gjort av Delprosjekt bokmål i samråd med Seksjon for navnegransking og Delprosjekt navnegransking, høsten 1995. Delprosjekt bokmål analyserte materialet med tanke på bruk av SGML-baserte koder, og Navnegransking gav råd i forhold til enkel transponering av lydskrift etter Helge Sandøys system for transkribering (Transkripsjonssystem til Norsk stadnamnbase). Valg av SGML-baserte koder og entiteter for diverse tegn.

Innskrivingsenheten:

1. Skanning (optisk lesning) av originalmateriale. Oppretting av arbeidsfiler á 2 bok-sider.
2. Kopiering av originalmateriale.
3. Revidering: innsetting av SGML-baserte tagger og entiteter vha. makro og enkel transponering av lydskrift.

4. Korrekturlesing innad i laget. Kvalitetskravet ligger på under 1 feil pr. 10 sider, noe som fordrer 5-6 korrekturer. Filene korrekturleses først i tagget versjon og deretter i validert versjon, uten tagger.
5. Ferdig tagging og sammenliming av arbeidsfiler til ferdige filer, som hver tilsvarer ett kapittel (ett herred) i Norske Gaardnavne.
6. Validering: Tolking og validering i Author/Editor (fauske.rls).
7. Kopiering av filer til diskett og forsendelse av diskett samt kopi av NG til fagkonsulenten 1 gang i uka.

**Delprosjektet:**

1. Forberedelse av undervisning og opplæring av innskrivere (Fauske, Tønsberg, Bardu).
2. Korrekturlesing av alt mottatt materiale.
3. Oppfølging dels ved å svare på telefonhenvendelser og dels ved ukentlige filstatistikker til hvert lag. Feilsnitt for hver enkelt fil blir satt i tabell, og hvis samlet snitt overstiger 0,1 (1 feil pr. 10 sider), sendes materialet i retur. Videre inneholder filstatistikkene beskrivelse av feilene og gjennomgang av ting som later til å være uklare.
4. Oppretting av ferdig korrekturleste filer. Diverse søk/erstatt-valg i Word.
5. Ferdig tagging av filene. Taggingen fagkonsulenten utfører, dreier seg om oppdeling av større enheter innenfor hver fil slik at man kan skille ut sogn innenfor hvert herred samt tillegg til navneartiklene.
6. Validering: Tolking og validering i Author/Editor (rygh.rls).
7. Uventede tegn og andre eventualiteter blir fortløpende notert i en egen fil som programmereren har tilgang til (se nedenfor).
8. Flytting av ferdig tilrettelagte filer til definert fellesområde (se nedenfor).

*Teknisk beskrivelse:*

Filene har først bokstavene ryg, deretter bindnummer (1, 2 osv), deretter kapittelnummer (01, 02 osv) og lagret som ren tekst. (Eks.: ryg1\_01.txt). De ligger i mapper ordnet bindvis (bind01, 02 osv). Alle filer som ligger på pchfdoc\rygh\ under mappene bind01, 02 osv., er klare for innlasting i database. I tillegg ligger ei fil kalt status.txt under pchfdoc\rygh som programmereren har tilgang til. Denne gir nærmere status for hver enkelt fil og gjør rede for fremmede tegn og annet som er vanskelig å konvertere. status.txt blir fortløpende oppdatert.

*Videre arbeid:*

Grovskisse til grensesnittet, utarbeidet i samråd med delprosjektansvarlig, er gitt til programmereren. Lydskriften er foreløpig for en del av tegnene. Derfor bør man erstatte den med bedre egnede tegn eller og i tillegg ha brukerveiledning i grensesnittet. Transkripsjon til IPA har vært foreslått, men dette vil ikke gi et riktig bilde av den svært enkle lydskriften i NG, da vokalismene, slik de er notert, er såvidt omtrentlige at transkripsjon ville være villedende for både norske og utenlandske brukere.

Det er stor interesse ved seksjonen for kobling til andre databaser, som ord- og arkeologidatabasene. Alle filene er validerte, og dermed burde kodingen være tilfredsstillende. Det kan likevel forekomme taggefeil, da filene er korrekturlest av fagkonsulenten i tolket versjon (uten tagger). Tekstkvaliteten ventes å være svært god, i forhold til stikkprøver på ferdig tilrettelagt materiale.

*Bruk av digitalt arkiv/database:*

Siden alle tekstdeler blir SGML-tagget, er det mulig å søke etter og sortere navn fra hele landet eller innenfor et avgrenset område.

Norske Gaardenavne fungerer som en base for Bustadnavnarkivet, som kan supplere med yngre bruksnavn under de enkelte gårdene og ofte med en noe annen tilnærming i tolkningen av navnene, og Matrikkelutkastet fra 1950, som supplerer med flere moderne navn på bruk og parseller under hver enkelt gård.

Norske Gaardenavne er knyttet opp mot flere av Dokumentasjonsprosjektets databaser (arkeologi og Diplomatarium Norvegicum), siden verket henviser til kildemateriale og ordledd, er godt geografisk avgrenset med gårds- og bruksnumre og dekker hele landet. Skjæringspunktet mellom språk- og kulturhistorisk materiale gjør verket interessant for såvel språkforskere som arkeologer og lokalhistorikere.

Seksjon for navnegransking har begynt å ta databasen i bruk. Fagkonsulenten har demonstrert basen og muligheter for søk for noen av de vitenskapelig ansatte. Fagavdelingen har en del kontakter som bruker basen.

Store grupper brukere vil kunne finne basen nyttig, og i tillegg til markedsføring via fagavdelingen, burde man ta kontakt med enheter som fylkeskulturkontorene, lokalhistorielag, grunn- og videregående skoler, slektsgranskerforeninger, Bondelaget o.l., for å markedsføre basen. Søk på World Wide Web gav for øvrig 35 tilslag på søkerstrengen "Norske Gaardnavne" og 20 på "Norwegian Farm Names", så gjennom Internett (også ved søk etter lokalhistorie- og slektsgranskerforeninger) kan man kartlegge grupper som ville ha nytte av Elektroniske NG.

### **Generelt: Det humanistiske fakultet, UiTø**

Talemåls- og navnearkivet ved Det humanistiske fakultet, Universitetet i Tromsø, inneholder i tillegg til norsk stedsnavnmateriale, også samiske og finske stedsnavnsamlinger. En rekke navneinnsamlingsintervjuer ble gjennomført i landsdelen på 1980- og 1990-tallet. Videre omfatter materialet kopier av eldre samlinger fra andre institusjoner. Det er gjort prioriteringer i samlingen i forhold til konverteringen i Dokumentasjonsprosjektet.

Delprosjektansvarlige i prosjektperioden: førsteamanuensis Gulbrand Alhaug og førsteamanuensis Eira Söderholm, HUM-fak., Universitetet i Tromsø

Fagkonsulent: Siri Sørland

Korrekturlesere/andre medarbeidere i delprosjektperioden: Karin Mydske, Kjell Erik Steinbru, Linda Hamrin Nesby og Trond Søbstad.

IT-faglig ansvar: Lars Jørgen Tvedt

Antall fagkonsulentårsverk i perioden 1994 - 1997: 3,75

Kontaktperson i fagenheten fra 1.1.98: førsteamanuensis Eira Söderholm

### ***Talemåls- og navnearkivet***

#### *Om materialet:*

Ved Talemåls- og navnearkivet finnes det alfabetiske heradsregisteret med norskspråklige, finske og samiske stedsnavn. I tillegg er det gjort lydbandopptak av flere tusen navn som i hovedsak ikke er utskrevet. Arkivet forvalter også to dataførte samlinger med norsk-samiske navnepar og samiske navn hentet fra kartserien N50 og sjøkart. Delprosjektet ved HUM-fak. har dataført heradsregisteret i Tromsø, et utvalg lydbandavskrifter og Skulebarnoppskriftene fra Nordland og Troms.

*Status pr. 31.12.97:*

Total størrelse på materialet:	25.000 stedsnavn (arkivkort)
Konvertert totalt:	25.000 stedsnavn (arkivkort)
Korrekturlest totalt:	25.000 stedsnavn (arkivkort)
Ferdigstilt totalt:	25.000 stedsnavn (arkivkort)

*Status pr. 31.12.97:*

Total størrelse på materialet:	4.000 stedsnavn (lydbandavskrifter)
Konvertert totalt:	4.000 stedsnavn (lydbandavskrifter)
Korrekturlest totalt:	4.000 stedsnavn (lydbandavskrifter)
Ferdigstilt totalt:	4.000 stedsnavn (lydbandavskrifter)

G. Indrebøs skolebarnoppskrifter (nordnorsk del)

*Status pr. 31.12.97:*

Total størrelse på materialet:	100.000 stedsnavn, ca 3.000 lister
Konvertert totalt:	100.000 stedsnavn, ca 3.000 lister
Korrekturlest totalt:	100.000 stedsnavn, ca 3.000 lister
Ferdigstilt totalt:	100.000 stedsnavn, ca 3.000 lister

*Beskrivelse av arbeidet:*

Delprosjektet har brukt innskrivingsopplegg og -applikasjon fra Delprosjekt navnegransking i Oslo. Innskrivingen av de 25.000 stedsnavnene i heradsregisteret tilsvarer arbeidsprosedyrene på Bustadnamnregisteret.

Innskrivingsenheten:

Innskriverne har skrevet etter kopier av arkivkortene, i tekstbehandling eller i spesielt utviklet innskrivingsapplikasjon. Deretter er materialet korrekturlest og rettet opp.

Delprosjektet:

Fagkonsulenten har utarbeidet manualer og lært opp innskrivingsgruppen. I tillegg er det lest korrektur på innsendt materiale og foretatt oppretting. Kortene er også skannet inn i delprosjektet.

*Beskrivelse av digitalt arkiv/database:*

Heradsregisteret foreligger i SGML-kodete tekstfiler og faksimileversjon av arkivkortene. Koblingen mellom faksimilene og tekstfilene foreligger i database. Lydbandavskriftene foreligger i tabellformat med faksimileversjon av kortene. Indrebø-materialet foreligger i SGML-tagget tekst og faksimiler. Alt materialet vil overføres til relasjonsdatabase.

*Videre arbeid:*

En enhetlig navnedatabase med mulighet for søk på tvers av arkivene er under utarbeidelse.

## **Folkemusikk og - dans**

*Generelt: Rådet for folkemusikk og folkedans, NTNU*

Rådet for folkemusikk og folkedans (Rff) er ei offentlig stifting. Institusjonen Rff-sentret er utøvande organ for stiftinga. Stiftinga sitt hovudmål er at Rådet for folkemusikk og folkedans vil fremja, verna om og føra vidare norske folkemusikk- og folke-danstradisjonar som eit uttrykk for kulturell identitet og som bærar av særneigne kvalitetar.

Delprosjektansvarleg: Førsteamanuensis Egil Bakka og førsteamanuensis Bjørn Aksdal, Rådet for folkemusikk og -dans.

Fagkonsulentar: Kari Margrete Okstad og Bente Hjellum

IT-fagleg ansvar: Norsk folkemusikksamling

Antall fagkonsulentårsverk i perioden 1994 - 1997: 3,3

Kontaktperson i fageininga frå 1.1.98: Kari Margrete Okstad

### ***Samlingane ved Rff-sentret***

#### *Om arkivet:*

Arkivmaterialet ved Rff-sentret byggjer på innsamla materiale om folkedans og folkemusikk frå 1960-talet og fram til i dag. Innsamlingane er for det meste gjort av dei fast tilsette ved sentret. I tillegg til film og lydopptak finst det store samlingar med intervju om folkemusikk og folkedans, ei stor samling informantskjema med opplysingar om og frå dei som er filma og tatt opp på lyd eller intervju på anna vis, og ei fotosamling som inneheld både papirfoto og dias.

Rff-sentret har også fått som depositum samlingar frå andre institusjonar. Dette gjeld t.d. ei lydsamling frå Musikkvitskapleg institutt ved NTNU, ei samling av Klara Semb sine papir, notar, notat, bilete m.m. og ein del notesamlingar osb.

#### *Status pr. 31.12.97:*

Total storleik på materialet:

Ei samling på 4.148 informantskjema

Konvertert totalt: 4.148 skjema

Korrekturlese totalt: 4.148 skjema

Ferdigstilt totalt: 4.148 skjema

Total storleik på materialet:

Ei samling 8 mm film på 643 rullar med ca.1.400 kutt

Konvertert totalt: 643 rullar

Korrekturlese totalt: 643 rullar

Ferdigstilt totalt: 643 rullar

Total storleik på materialet:

Ei samling 16 mm film på 1.012 rullar med ca. 4.500 kutt

Konvertert totalt: 1.012 rullar

Korrekturlese totalt: 1.012 rullar

Ferdigstilt totalt: 1.012 rullar

Total storleik på materialet:

Ei samling på 1.789 lydband med ca. 18.247 kutt (Rff-serien)

Konvertert totalt: 1.789 lydband

Korrekturlese totalt: 800 lydband (nr.1-800) tredjekorrektur

Gjenstår å korrekturlesa: 989 lydband (nr. 800-1.789) tredjekorrektur\*

Ferdigstilt totalt: 800 lydband



\* tredjekorrekturen vert teken av fagkonsulenten i folkemusikk ved Rff-sentret etter kvart som materialet vert teke i bruk.

Total storleik på materialet:

Ei samling på 608 lydband med 4.848 kutt (MIT-serien)

Konvertert totalt:	608 lydband
Korrekturlese totalt:	608 lydband
Ferdigstilt totalt:	608 lydband

*Skildring av arbeidet:*

Førebuing:

Dei ulike samlingane ved Rff-sentret vart sjekka i ulik grad i høve originalmaterialet. Informantskjema trong ikkje noka spesiell sjekking. Rff-serien for spoleband, Rff-lyd vart sjekka ved å gå gjennom kvart enkelt band for å sjå at nummereringa var rett. Dette hadde tidlegare vore gjort, og vart vidareført parallelt med innskriving av sivilarbeidarar. Rff 16 mm film hadde delvis vore sjekka og klassifisert av ein tidlegare medarbeidar, og vart no ferdig klassifisert i lag med innskriving av fagkonsulenten. Rff-serien 8mm film vart klassifisert og sjekka av fagkonsulenten parallelt med innskriving.

Tilrettelegging:

Alle skjema og lister over dei samlingane som skulle skrivast inn, vart kopierte og sett i permar. Alt materialet vart delt i porsjonar mellom innskrivarane.

Koding:

Koding for dette materialet har vore å lage klassifikasjon for folkedans og folkemusikk. Det fanst alt utarbeidde system for klassifikasjon av både musikk og dans, men like før innskrivingsperioden vart det utarbeidd nye nordiske klassifikasjonar, og det vart vedteke å bruka desse i klassifikasjonen av både dans og musikk. Desse klassifikasjonane var i utgangspunktet kodifiserte med bokstavar, men i tillegg vart eit gammalt system for tal brukt. Begge delane vart sette inn i datasystemet.

Anna koding som vart gjort, var å laga kodar i datasystemet for kva dei ulike samlingane skulle heite, RffFs, RffFå, RffLs osv.

Innskrivingseininga:

Innskrivingsbasen var i same lokalitetar som fagkonsulentane. Det vart difor ein tett og god kontakt mellom innskrivarane og fagkonsulentane. Ein god del av instruksane vart difor ikkje sett ned i skriftleg form, men vart gitt direkte til innskrivarane etter som behov oppstod. (Dei skriftlege instruksane ligg ved som vedlegg). Dei oppgåvene innskrivarane hadde, var:

- Innskriving av informantskjema
- Innskriving av innsamlingane
- Innskriving av lydserien
- Innskriving av 16mm filmserien
- Innskriving av 8mm filmserien

I tillegg til innskrivinga hadde dei ansvar for å lesa korrektur for seg sjølv og for dei andre.

*Beskrivelse av digitalt arkiv/database:*

Katalogen er bygget rundt noen grunnleggende dataelementer eller filer. Hensikten er at opplysningene skal lagres så kompakt og konsistent som mulig; dvs. at hver enkelt

opplysning (som navn på en utøver) bare finnes på ett enkelt sted i databasen, slik at man unngår at et navn er skrevet på flere måter ulike steder i databasen. De grunnleggende dataelementene er:

- Avsnitt (det enkelte opptaket, noteoppskrift, transkripsjon...)
- Identitet (personer og grupper)
- Rolle (personens/gruppas funksjon og evt. instrument ved fremføringen/ utførelsen av avsnittet)
- Innsamling (tid, sted, etc. for innsamlingsaktiviteten)
- Medium (f. eks. lydbånd, med tekniske opplysninger og katalogsignatur)
- Distrikt; som settes sammen av én eller flere kommuner

Avsnitt er den sentrale enheten. Med et avsnitt menes et enkelt opptak på et lydbånd; et kutt på en plate; en enkelt visetekst i et manuskript, et danseopptak eller en noteoppskrift av en melodi. Dette betyr at katalogen bare er beregnet på medier som inneholder materiale av `avsnittskarakter'. Generelle bøker om folkemusikk og folkedans samt foto faller utenfor det katalogen er innrettet på å brukes til.

Et avsnitt er samlet inn en eller annen gang, og som regel sammen med andre avsnitt. Opplysningene om omstendighetene rundt innsamlingen (hvor, når, hvem) vil derfor som regel være felles for flere avsnitt. I stedet for å legge inn opplysninger om innsamlings-omstendighetene for hvert avsnitt, legges disse opplysningene inn i en egen fil som heter Innsamling. Hver post i denne fila får et eget identifikasjonsnummer.

Et avsnitt er samlet inn en eller annen gang, og som regel sammen med andre avsnitt. Opplysningene om omstendighetene rundt innsamlingen (hvor, når, hvem) vil derfor som regel være felles for flere avsnitt. I stedet for å legge inn opplysninger om innsamlings-omstendighetene for hvert avsnitt, legges disse opplysningene inn i en egen fil som heter Innsamling. Hver post i denne fila får et eget identifikasjonsnummer.

Hvert avsnitt forekommer på minst et medium. Avsnittet kan være kopiert opp, og dermed finnes på flere medier. For hvert avsnitt kan det med andre ord vært én eller flere forekomster. På hvert medium kan det være flere ulike forekomster; mens hver forekomst finnes på bare ett medium, og representerer bare ett avsnitt. Det er egne filer for Forekomst og Medium.

Tekniske og arkivmessige opplysninger om selve mediet (lydbåndet, filmrullen, manuskriptet...) er samlet i Mediefila, mens opplysninger om hvor på mediet forekomsten er plassert, ligger i Forekomstfila. Personopplysninger er samlet i fila Identitet. Her samles opplysninger både om utøvere, innsamlere og eventuelle grupper av utøvere. Forbindelsen mellom Identitet og Innsamling gjøres gjennom fila Samlere. I hver Innsamling kan flere innsamlere ha deltatt, og at hver person kan ha deltatt i flere innsamlinger. Samlerfila gir også plass til å opplyse om hva den enkelte innsamler har gjort (vært ansvarlig; fotograf, lydmann, etc.)

Personer (Identiteter) knyttes til Avsnitt gjennom Roller. Rollebegrepet handler om hvordan en person medvirker på et avsnitt (f. eks. som utøver), hva slags instrument som eventuelt brukes, og hva slags geografisk tilknytning person har i denne sammenhengen. Hver person kan ha flere roller.

Distrikter tar utgangspunkt i kommuneinndelingen. Hver kommune er et distrikt; i tillegg kan man definere distrikter som er sammensatt av flere kommuner, som f. eks. "Setesdal", sammensatt av kommunene Bygland, Rysstad, Valle og Bykle. En rolle kan knyttes til ett distrikt.

Om man ønsker å definere nye distrikter må man ta kontakt med NFS. For å unngå rot i distriktshåndteringen i databasen, må alle distrikter defineres sentralt, slik at alle har de samme distriktene definert i sine databaser.

I databasen for Rff-sentret, har det vore behov for å knytte ein del av samlingane saman til same avsnittet. I utgangspunktet ligg det opp til at kvart enkelt opptak av lyd eller bilete skal ha kvar sitt avsnitt i databasen.

Ved Rff-sentret finst det ein god del tilknytingspunkt mellom delar av samlingane, tilknytningar som er viktige å få samla for at databasen skal vere så funksjonell som råd i bruk. Dette gjeld spesielt for 8- og 16mm filmmaterialet, som begge har parallelle lydopptak på spoleband. For oversikten sin del har det vore viktig å knyte desse saman, så for 16mm film er spolebandopptaket som høyrer til, sett inn som forekomst i same avsnittet. Her vil altså kvart avsnitt ha to forekomster, filmreferanse og lydreferanse. For 8mm-samlinga er det ikkje knytta opp direkte som forekomst, men her er lydreferansennummer sett inn i kommentarfeltet.

#### *Teknisk beskrivelse:*

Databasen er laga i systemet 4 Dimension for Macintosh og har namnet FIOL. Datafila for FIOL ved Rff-sentret er lagra på PROFFEN, ein server ved NTNU på Dragvoll i Trondheim. Ein sikringskopi av datafila ligg på server hos Norsk folkemusikksamling, NFS, i Oslo. Den eine maskina ved Rff-sentret er definert som hovudmaskin. Ein gong kvar veke blir datafilene lagra frå denne hovudmaskina til sikring på PROFFEN. Frå PROFFEN blir datafila lasta ned til dei andre maskinane som er i bruk, slik at ein til ei kvar tid kan sitje med den oppdaterte fila. I tillegg blir ein sikringskopi send til lagring hjå NFS ein gong i månaden, og det er under diskusjon å lage avtale med NFS om at dei kvart halvår brenn datafila på CD. Det blir ført logg over lagring hjå NFS.

#### *Videre arbeid:*

Ved Rff-sentret finst det fleire samlingar enn dei som vart med i denne omgangen av innskrivinga. For at Sentret fullt ut skal ha glede av databasen, bør alle samlingane bli lagt inn i same programmet.

Dei samlingane som står att for innskriving er:

- ei samling lydband (NFMI-serien) på omlag 930 rullar med ca. 25.000 kutt
- ei samling på om lag 1.600 videoband med ca 13.000 kutt (samlinga veks stadig)
- ei samling på om lag 100 DAT-kassettar (samlinga veks stadig)
- ei samling på om lag 6.500 foto og 3.700 dias
- ei samling handskrivne notar på knapt 100 eks. med melodiar
- ei ekserptsamling på om lag 800 nr.

#### *Bruk av digitalt arkiv/database*

Bruk av databasen vil i første omgang vere for dei tilsette ved Rff-sentret. I andre omgang vil det kome brukarane til gode, på den måten at dei som ønskjer oversikt over delar av samlingane, kan få meir oversiktleg informasjon.

## **Etnografi**

### **Hovedkatalogen, Etnografisk museum**

Faglig ansvarlig: førsteamanuensis Tom Svensson

Fagkonsulenter i prosjektperioden: Svein Sanden og Alfhild Birkeland

IT-faglig ansvar: Jon Holmen

Etnografisk museum startet som et tilknyttet prosjekt under Dokumentasjonsprosjektet sommeren 1997, og er fullfinansiert av Institutt og Museum for Antropologi.

Museet besitter i dag ca. 46.000 gjenstander, alt fra eskimoiske harpunspisser til likhåndklær fra India. Inntil midten av 1980-tallet ble alle innkomne gjenstander registrert med et gjenstandsnummer i en håndskrevet katalog; hovedkatalogen. Denne går tilbake til 1850-tallet og er i seg selv et klenodium. Den første gjenstanden som ble registrert i katalogen er en "Sittende mandlig Figur af Bronze, med langt, smalt Hageskjæg, og en bred Hat". Den kom fra "Ægypten", og ble skjenket til museet i 1853 av Kaptein Lund.

Siden har gjenstandene strømmet inn, og hovedkatalogen har vokst til 19 tykke bind. Disse ligger godt nedlåst i museet, og er dermed vanskelig tilgjengelig for de ansatte på museet, forskere og andre interesserte. Konverteringsprosjektet går i korthet ut på å reprodusere dette håndskrevne materialet til elektronisk tekst, som gjør det mulig å ta teksten inn i en database, slik at brukeren kan søke på gjenstandsnumre, stedsnavn og andre kriterier. Dette gir nye brukermuligheter. I dag må forskere som er interesserte i gjenstander, for eksempel fra Lumholtz' Borneo-samling, bla seg gjennom katalogen for å finne informasjon litt her og litt der, spredt utover hele katalogen. Det kan være en tidkrevende prosess. Om kort tid vil det derimot være mulig å søke gjenstander fra Borneo, fetisjer fra Kongo osv. direkte i en database, noe som vil lette arbeidet betydelig både for forskere, publikum og for de som til daglig arbeider i museet. Den første fasen av prosjektet vil altså resultere i et elektronisk verktøy som vil øke tilgjengeligheten til informasjon om museets samlinger betraktelig.

I tillegg til hovedkatalogen finnes det på museet et seddelarkiv over gjenstandene og et fotoarkiv med fotografier av gjenstandene. Seddelarkivet inneholder en del informasjon som ikke er registrert i hovedkatalogen, for eksempel vakre tegninger av detaljer. Det er derfor viktig at også disse arkivsedlene dokumenteres. En senere fase av prosjektet vil være å skanne inn disse sedlene i et elektronisk arkiv, som dermed blir

tilgjengelig i en database sammen med hovedkatalogen og et elektronisk fotoarkiv. Dermed vil brukeren kunne søke i en database som inneholder informasjon om gjenstanden både i form av tekst og bilde.

Dette materialet vil senere legges ut på Internett. Gjennom å gi publikum økt tilgang til informasjon om gjenstandene, vekkes også appetitten på museet. I tillegg vil materialet for eksempel kunne brukes som en integrert del av utstillingene; med datamaskiner plassert i utstillingslokalet, kan publikum selv søke seg frem til den informasjonen de ønsker.

# Prosjektdirektiv

Dette dokumentet setter opp planen for arbeidet med å etablere Universitetenes databaser for språk og kultur gjennom Dokumentasjonsprosjektet. Dokumentet berører følgende forhold:

- 1) bakgrunn
- 2) målstruktur
- 3) oppgavebeskrivelse
- 4) retningslinjer for prosjektarbeidet
- 5) omfang og tidsrammer
- 6) økonomiske rammer og finansiering
- 7) planlegging, rapportering, oppfølging
- 8) kvalitetssikring
- 9) endringshåndtering

Dette dokumentet er å betrakte som prosjektets faglige og administrative kontrakt mellom oppdragsgiver og Den nasjonale styringsgruppen for prosjektet.

## 1 Bakgrunn

### *1.1 Universitetenes databaser for språk og kultur*

Ved universitetene i Norge finnes det i dag en rekke avdelinger med store registre og samlinger. Disse inneholder informasjon om et bredt spekter av norsk språk og kultur. Dersom denne informasjonen skal kunne bli tilgjengelig for forskeren via dennes PC, er det ikke mulig å gå utenom en krevende konvertering av papirdokumentene til en elektronisk form. Dokumentasjonsprosjektet har nettopp dette som sitt første delmål.

Det endelige mål er et informasjonssystem for samlingsavdelingene, kalt "Universitetenes databaser for språk og kultur". Systemet vil ha tre hovedfunksjoner. For det første skal de fagspesifikke delene tjene som verktøy for forskerne innen de enkelte fagene i deres daglige arbeid. For det andre vil systemet bli laget slik at det er mulig å foreta søk blant data fra flere fagfelt samtidig. Den tredje hensikten med systemet er å øke allmennhetens tilgang til arkivmaterialet dog slik at opphavsretter og samlingenes sikkerhet blir ivaretatt.

På denne måten vil "Universitetenes databaser for språk og kultur" spille en viktig rolle i grunnforskningen innen de humanistiske fagmiljøene. Systemet muliggjør også en tverrfaglighet som i sin tur kan initiere ny forskning og innsikt.

### *1.2 Historikk*

Oppbyggingen av registrene ved de ulike samlingsavdelingene har foregått over lang tid, for noen avdelingers vedkommende helt fra begynnelsen av forrige århundre. I årenes løp er det blitt samlet til dels meget omfattende manuelle registre og arkiver. Først rundt 1980 begynte enkelte fagavdelinger å ta i bruk datamaskiner i arkivarbeidet. Den daværende teknologiske virkelighet gjorde at man forestilte seg felles, sentraliserte databaser som kunne betjene flere fagmiljøer på ulike steder i landet. Disse idéene ble av forskjellige årsaker ikke realisert. Dels kunne dette skyldes datidens lite fleksible og svært kostbare tekniske løsninger, dels den totale mangel på tilgjengelig (edb-) personell med tilstrekkelig kompetanse.

Den massive innføringen av små datamaskiner (PCer) i annen halvpart av 1980-årene betydde en vanliggjørelse av datateknologien og at oppmerksomheten omkring edb økte. Det ble derfor adskillig vanligere at avdelinger med registre av noen størrelse prøvde å bygge opp edb-baserte registre. Man fikk dermed en rekke spredte mindre databaser av forskjellige typer og kvalitet. Men disse representerte ofte en voldsom innsats av enkeltmennesker i fagmiljøene.

I første halvår av 1990 tok daværende prodekanus Bjarne Hodne ved Det historisk-filosofiske fakultet ved Universitetet i Oslo i samarbeid med professor Herman Jervell initiativet til et forprosjekt for å kartlegge fakultetets egne arkiver og samlinger med henblikk på en mer samordnet og massiv innføring av datateknologi.

Gjennom et forprosjekt, ledet av førsteamanuensis Christian-Emil Ore, ble det fremkaffet en samlet oversikt over registre og arkiver ved fakultetets tradisjonelle samlingsavdelinger der man la vekt på de ulike registre og arkivers karakter, fysiske utforming og hvilke datamengder de representerte. Det ble også gjort overslag over arbeidsmengden forbundet med en eventuell dataregistrering.

Forprosjektet ble avsluttet sommeren 1990. Prosjektets referansegruppe med representanter for de ulike avdelingene besluttet at arbeidet skulle fortsette etter de linjer som ble trukket opp i forprosjektrapporten. Det forberedende arbeidet med hovedprosjektet ble tatt opp igjen høsten 1990. I februar 1991 vedtok så fakultetsrådet ved Hf-fakultet ved Universitetet i Oslo at man skulle gå i gang med et seksårig Dokumentasjonsprosjekt.

I samråd med Kirke-, utdannings- og forskningsdepartementet ble det sommeren 1991 arbeidet for å gjøre prosjektet nasjonalt. Under denne forutsetningen ble det avsatt øremerkede midler til prosjektet på statsbudsjettene for 1992 og 1993. Den formelle igangsettelsen av Det nasjonale dokumentasjonsprosjektet ble vedtatt på det nasjonale dekanmøtet i november 1991. Våren 1992 ble det så oppnevnt en styringsgruppe for prosjektet med representanter fra universitetene i Bergen, Oslo, Trondheim og Tromsø. Leder for styringsgruppen var inntil 31/12-92 førsteamanuensis Tellef Kvifte og fra 1/1-93 førsteamanuensis Knut Aukrust.

### ***1.3 Konverteringsarbeid***

Dokumentasjonsprosjektet har så langt fulgt to strategier for å få utført konverteringsarbeidet.

#### ***1.3.1 Spredte tiltak***

Prosjektet har siden 1991 benyttet "Arbeid for trygd"-programmet til å finansiere deler av konverteringsarbeidet. Innskriverene er organisert i grupper med 5 til 10 medlemmer, i alt 70 personer pr. januar 1993. Gruppene er plassert ulike steder og har en innleid lokal administrasjon (for det meste det lokale Folkeuniversitetet). Tiltakene fungerer godt, men dekker ikke kostnadene til oppfølgingsarbeidet (en fagkyndig for hver fjerde innskriver).

#### ***1.3.2 "Omstilling gjennom målrettet utdanning"-prosjektene***

Til tross for at AFT-opplegget fungerer godt, ble det likevel vurdert som lite realistisk å basere hele registreringen på slike tiltak idet universitetene ikke har midler til det tilstrekkelige antall assistenter. Dessuten har arbeidsmyndighetene ytret ønske om å utforme tiltak som tilfører deltakerne mer relevant kompetanse enn det rene registreringsarbeidet kan gi.

I forståelse med den sentrale ledelse ved Universitetet i Oslo har derfor Dokumentasjonsprosjektet i samarbeid med universitetets kurssekretariat og Statens datasentral A/S (SDS nord) utarbeidet prosjektet "Omstilling gjennom målrettet utdanning i Mo i Rana og Kirkenes". Prosjektet går ut på å etablere kombinerte opplærings- og arbeidspraksissentre der deltakerne får undervisning rettet mot de lokale behov kombinert med praktisk arbeid lagt opp etter modellen utprøvd i det AFT-baserte prosjektet beskrevet over. Hvert slikt tiltak er beregnet å vare i fire år.

I september 1992 ble det første av sentrene etablert i Mo i Rana. Sentret er organisert som et arbeidsmarkedstiltak med 75 deltakere. Etableringskostnadene ble dekket av Næringsdepartementet, mens drifts- og undervisningskostnadene dekkes av arbeidskontoret i Rana. Sentret i Kirkenes er planlagt etablert våren 1993.

Gjennomføringen av "Omstilling gjennom målrettet utdanning"-prosjektene hører ikke inn under Dokumentasjonsprosjektets oppgaver. Men disse prosjektene synes å være svært viktige for å få finansiert og gjennomført konverteringsarbeidet.

## **2 Målstruktur**

Hovedmålet for Dokumentasjonsprosjektet er "å gi en samlet og rasjonell tilgang til informasjon om språk og kultur".

Informasjon om språk og kultur finnes i dag spredt i en rekke arkiver og samlinger over hele Norge. Med utgangspunkt i kulturhistoriske og språklige samlinger ved norske universitet skal prosjektet gjøre slik informasjon tilgjengelig for forskere, forvaltning og allmennhet i en form som gjør at informasjonen fra ulike registre kan utnyttes samtidig.

På bakgrunn av hovedmålet er det formulert følgende fire delmål:

### *1 Dataregistrering av arkiv og samlinger*

Dette er den største delen av prosjektet, målt i arbeidsmengde, og omfatter all konvertering av data til maskinleselig form. Dette omfatter dels registre og dels fullstendige materialesamlinger.

### *2 Rasjonelle rutiner for samlinger og arkiver*

Omlegging av arkiver og kataloger til edb medfører endringer i interne rutiner ved avdelingene. I forbindelse med dataføring ved avdelingene gjennomgås interne arkivrutiner for å få størst mulig gevinst ut av prosjektet.

### *3 Optimal tilgjengelighet for forskere, forvaltning og allmennhet*

Brukere av systemet skal ha tilgang til den informasjon de ønsker, på enklest mulig måte og presentert i klarest mulig form. Presentasjonsform, veiledning og avregning av kostnader skal tilpasses ulike brukergrupper. Tilgangen til dataene må likevel kunne reguleres slik at sikkerheten i muséene, opphavsrettigheter og personvern ivaretas.

### *4 Samordnet datarepresentasjon og grensesnitt*

Rasjonell tilgang til informasjon forutsetter at datarepresentasjon og grensesnitt er samordnet. Prosjektet skal definere standarder på bakgrunn av faglige hensyn, eksisterende praksis og etablerte registre, og sørge for kobling mellom ulike registre, eksterne og interne, med henblikk på faglig interesse, samordningskrav og gjennomførbarhet.

### 3 Oppgavebeskrivelse

I løpet av prosjektperioden skal alle registre som inngår i prosjektet, dataføres. Det skal også utvikles systemer for bruk og presentasjon av dataene. Aktiviteten kan deles i følgende punkter:

#### *1 Identifisering av registre som skal inngå i prosjektet*

På grunnlag av utredninger i fagmiljøene avgjør styringsgruppen hvilke registre som skal inngå i prosjektet samt tidspunkt for deltakelse. Kriteriene for å inkludere registre er:

- Registret har materiale av nasjonal interesse
- Det er dokumentert behov for materialet av flere brukere
- Det er materiale som støtter opp om og utfyller øvrig materiale i prosjektet
- Materialet er unikt
- Det er innsatsvilje i miljøene, uttrykt ved prioriteringer innen eksisterende rammer og forarbeid utført ved egne midler
- Det er gjennomførbart å inkludere registret i prosjektet

#### *2 Utredning for interne og eksterne brukerbehov for registrene*

Utredning skal skje i samarbeid med fagmiljøene og etter felles retningslinjer vedtatt av styringsgruppen. Utredningen bør bl.a. omfatte identifisering av interne behov, konsekvenser i forhold til interne rutiner, organisasjonsforhold og ressurser. Dette gjelder både fortløpende bruk ved avdelingene, tilrettelegging for universitetsbrukere utenfor avdelingene og identifikasjon av mer eksplisitte produksjons- og forskningsinteresser. Med hensyn til eksterne brukerinteresser bør brukerbehovene sees i sammenheng med krav til informasjon, tilrettelegging, ressursbehov og ansvarsforhold.

#### *3 Datateknisk analyse av registrene som skal inngå i prosjektet, dokumentert ved en standard metode.*

Det finnes ikke nødvendig ekspertise i de enkelte fagmiljøene til dette. Dataanalysen er en av forutsetningene for en fornuftig utvikling av registret. Analysen samordnes sentralt blant annet for å holde oversikten over felles dataelementer.

#### *4 Fastsetting av nødvendige standarder for dataformater med henblikk på nødvendige koblinger mellom ulike registre*

Universitetenes databaser for språk og kultur vil inneholde data fra en rekke forskjellige fagfelt. For å muliggjøre en kobling av dataene på tvers av faggrensene må det utarbeides standarder for representasjon av felles datatyper. Disse standardene skal så langt råd er, være i samsvar med eksisterende nasjonale og internasjonale standarder.

#### *5 Utvikling av innskrivings- og brukerprogrammer for registrene*

Der vil være mye felles i metodikken. Det er heller ikke her tilstrekkelige ressurser i fagavdelingene til dette arbeidet. Videre må det fra starten av tas hensyn til det endelige felles systemet.



## *6 Innskriving av registre som inngår i systemet*

Prosjektet har ansvar for at innskrivingsmiljøene blir etablert og for koordineringen av dem. Prosjektet står ikke selv for finansieringen av tiltakene, men vil samarbeide med sentrale myndigheter for å få i stand eksterne finansieringsordninger. Eventuelle utdanningstilbud som inngår i innskrivingspakkene, samt teknisk og administrativ drift av tiltakene, kan settes bort på oppdrag til samarbeidspartnere lokalt og sentralt. Ansvars- og arbeidsdeling skal avtales spesielt med de forskjellige aktører i hvert enkelt tilfelle.

Dokumentasjonsprosjektets oppgaver i forbindelse med etablering av innskrivingsmiljøene skal være å føre etableringsprosessen(e) frem til avtaleinngåelse mellom involverte parter. Koordineringsansvaret knyttes til 1) sammenkobling mellom delprosjekter og innskrivingsmiljøer, 2) informasjonsvirksomhet og 3) organisering og oppfølging av innskrivingsmiljøene og 4) fortløpende evaluering av den samlede innskrivingsaktiviteten ut fra kvalitet, effektivitet og fremdrift.

## *7 Prosjektering av videre drift av registrene*

Selve driften er ikke en del av prosjektet og må forankres innenfor de enkelte organisasjonene på permanent basis.

Den videre driften vil innebære tunge, permanente oppgaver etter at selve prosjektet er avsluttet. Driftsfasen bør settes i gang så snart registrene inneholder store nok datamengder til at de er nyttige. Driften av maskiner og nettverk bør innarbeides i planer og rutiner i den normale edb-driften.

## *8 Formidling*

Det enkelte delprosjekt og avdeling bør ha ansvaret for den aktive tilrettelegging for bruk av registrene internt og eksternt. Formidlingen bør skje i tråd med retningslinjer og behov som utredningene under punkt 3 har avdekket.

# **4 Retningslinjer for prosjektarbeidet**

## **4.1 Organisering**

### *4.1.1 Oppdragsgiver*

Det historisk-filosofiske fakultet ved Universitetet i Oslo er oppdragsgiver for prosjektet. De øvrige universitetene delegerer myndighet hit. Det nasjonale fakultetsmøtet kan bare gripe inn i den grad prosjektet ikke styres etter forutsetningene og da bare gjennom dekanus ved Det historisk-filosofiske fakultet ved Universitetet i Oslo.

### *4.1.2 Styringsgruppen*

Den nasjonale styringsgruppen er ansvarlig for prosjektet som helhet og de ulike delprosjektene. Styringsgruppen består av en representant fra hvert av de deltakende universitetene, samt en representant fra den sentrale ledelsen ved Universitetet i Oslo. Faglig leder møter i styringsgruppen. I tillegg har gruppen en sekretær. Styringsgruppen er ansvarlig overfor oppdragsgiver for prosjektet som helhet og har ansvar for den overordnede organiseringen av prosjektet, for fremdrift og for hovedbeslutninger som gjelder prosjektets mål og avgrensning, milepælplaner og prosjektets økonomi. Styringsgruppen har således myndighet til å godkjenne eller forkaste forslag til delprosjekter, se under pkt. 4.1.6 og avsnitt 5 om prosjektets rammer.

#### *4.1.3 Lokale arbeidsgrupper*

Hver av fakultetsrepresentantene i styringsgruppen kan knytte til seg en lokal arbeidsgruppe. Disse arbeidsgruppene skal bistå fakultetsrepresentanten med dennes oppgaver som blant annet omfatter:

- nødvendig koordinering av ressurser lokalt på det enkelte universitet
- drive frem utredninger for prosjekter som ennå ikke er igangsatt
- om nødvendig formidle kontakt til egen institusjon for utdanningsdelen av "Omstilling gjennom målrettet utdanning"-prosjektene (se punkt 1.3 under "Bakgrunn")
- holde kontakten med den lokale fakultetsledelsen.

Det understrekes at de lokale arbeidsgruppene ikke har prosjektstyringsfunksjoner knyttet til de nasjonale prosjektene. All slik styring går gjennom styringsgruppen og prosjektarbeidsgruppen (se under, pkt 4.15). Delprosjekter som ikke har tilknytning utover det enkelte universitet, kan imidlertid underlegges en slik arbeidsgruppe når spesielle forhold tilsier at dette er den beste løsningen.

#### *4.1.4 Referansegruppen*

Det vil bli oppnevnt en referansegruppe for prosjektet, eventuelt også egne referansegrupper for delprosjektene. Referansegruppen trekkes inn ved behov og ved milepæler i prosjektet. Det kan vurderes om referansegruppen skal inneholde personer fra berørte eksterne institusjoner som Statens kartverk og Riksantikvaren.

#### *4.1.5 Prosjektarbeidsgruppen*

Prosjektarbeidsgruppen er prosjektets daglige ledelse. Den består av en faglig leder, en registreringsleder og en driftskoordinator for registreringsentralene. Faglig leder koordinerer det løpende arbeidet i prosjektet og delprosjektene og har rapporteringsansvar til styringsgruppen.

Faglig leder er ansvarlig for den edb- og systemmessige utviklingen av løsningene for de enkelte delprosjektene, og design av det integrerte totalsystemet, se også punkt 4.3 (Strategi for edb-løsninger). Vedkommende leder kjernegruppen som består av prosjektets programmerere/systemerere.

Registreringslederen er ansvarlig for innskrivingsvirksomheten i prosjektet. Spesielt gjelder dette informasjon, organisering, oppfølging og evaluering av denne virksomheten samt fordelingen av materialet på de ulike innskrivingstiltakene. Registreringslederen har også personalansvar for det administrative personalet som er tilknyttet prosjektarbeidsgruppen, for prosjektassistentene og har ansvar for registreringsøkonomien.

Driftskoordinatoren har ansvar for administrasjonen av registreringsentralene, ansvar for kontakt og rapportering til de bevilgende myndigheter i forbindelse med sentralene, sammenstilling av økonomien for disse samt kontrakt- og avtaleforvaltning. For prosjektarbeidsgruppens arbeid gjelder at alle strategiske valg skal forelegges styringsgruppen til godkjenning.

#### *4.1.6 Delprosjekt og delprosjektansvarlige*

Dokumentasjonsprosjektet deles i et antall delprosjekt etter faggrenser (som et stedsnavnsprosjekt, et gammelnorskprosjekt, et folkemusikk etc.) der et delprosjekt omfatter alle universiteter med relevante samlinger. Styringsgruppen utnevner en delprosjektansvarlig for hvert delprosjekt etter forslag fra fagmiljøene.

Delprosjektansvarlig har det emnespesifikke fagansvaret for delprosjektet. Til det enkelte delprosjekt skal det utarbeides et prosjektdirektiv som i avtales form setter opp retningslinjene for gjennomføringen. Dette skal gjøres i samråd med faglig leder for dokumentasjonsprosjektet. Et slikt prosjektdirektiv bør inneholde følgende punkter: Mål, oppgavebeskrivelse, omfang og tidsrammer, økonomiske forhold og betingelser, ansvarsforhold, rapportering, kvalitetssikring og endringshåndtering. Prosjektdirektivet skal godkjennes av styringsgruppen.

#### *4.1.7 Ansvarlig ved de enkelte avdelingene*

I delprosjekt som involverer flere universiteter, skal det ved hver deltakende avdeling ved det enkelte universitet være en person som er ansvarlig for delprosjektet ved avdelingen. Denne ansvarlige for avdelingsprosjektet skal ha ansvar for å drive frem planene for avdelingens del av prosjektet. For prosjekt som bare omfatter én avdeling vil avdelingsansvarlig være delprosjektansvarlig.

## **4.2 Strategi for innskriving**

Prosjektets store omfang nødvendiggjør bruk av ufaglært arbeidskraft til selve registreringsarbeidet. Tilrettelegging av materialet, instruksjon av innskriverene og annen oppfølging samt korrektur og kontroll av materialet forutsettes gjort av assistenter med fagspesifikk kompetanse. Ut fra tidligere erfaringer brukes måltallet 1:4, én assistent for fire innskrivere, ved tildelingen av midler til det enkelte delprosjekt. Dette er ikke å regne som et absolutt forhold og kan avvikes i begge retninger. Eksempelvis kan det der det er mer rasjonelt å bruke kvalifisert fagpersonale til selve innskrivningen (f.eks. når dette er en forsvinnende liten del i forhold til den faglige tilretteleggingen), tilføres fagavdelingene ressurser til å leie inn slikt personale.

## **4.3 Prosjektassistentene**

Dokumentasjonsprosjektet skal sikres en fast stamme av vitenskapelige assistenter som står i forhold til antall innskrivere. På grunnlag av de erfaringene som til nå er høstet, er forholdet satt til 1:4 (se punkt 4.2, Strategi for innskrivingen), det vil si ett assistentårsverk for fire innskrivningsårsverk. Det bør tilstrebes kontinuitet i bemanningen for å unngå sløsing med ressurser i forbindelse med opplæringen av prosjektassistentene. Kortids- og deltidsengasjementer av hovedfagsstudenter og andre med tilstrekkelig kompetanse er aktuelt som supplement til de mer langtidsengasjerte prosjektassistentene.

Prosjektassistentene knyttes til fagavdelingene og delprosjektene gjennom det fagspesifikke ansvaret disse har i prosjektet. Prosjektassistentene skal i størst mulig grad integreres i fagmiljøene både gjennom faglig veiledning og gjennom romplassering. Det forutsettes at den enkelte avdeling/institusjon stiller nødvendig arbeidsplass til disposisjon.

Prosjektassistentene er ansatt ved Dokumentasjonsprosjektet. Assistentene skal om mulig ha den enkelte fagavdeling som tjenestested.

Oppdragsgiver har ansvaret for at det nødvendige antall stillingshjemler kan disponeres i henhold til fastsatte planer.

#### **4.4 Strategi for edb-løsninger**

Dokumentasjonsprosjektets behov for systemutvikling er tredelt. For det første må det fremskaffes programmer for registrering av dataene, for det andre må de enkelte avdelingene i prosjektet få utviklet verktøy slik at de kan benytte de registrerte dataene i sitt arbeid, og for det tredje skal tilgjengeligheten økes ved at det lages et felles, samlet informasjonssystem for alle de deltakende avdelingene.

Prosjektet har som erklært målsetting å holde programutviklingen på et minimum og i størst mulig grad utnytte tilgjengelig programvare. Det må imidlertid understrekes at det i dag ikke finnes programvare som dekker alle de særegne behov (spesielle tegnsett, fritekstsøking, hypertekst, multimedia m.m.) man finner innen humaniora. Det må derfor avsettes ressurser til en del programmeringsarbeid og til arbeidet med å tilpasse eksisterende produkter.

Totalsystemet skal være et nettverksbasert databasesystem der de enkelte delprosjektsystemene vil inngå som selvstendige moduler. I tillegg skal det utvikles separate datasystemer som forskerne kan benytte i forbindelse med egen virksomhet og "i felten". Det skal legges vekt på at totalsystemet skal kunne brukes av institusjoner og personer utenfor de enkelte fagmiljøene. Det er en forutsetning at klausulbelagt materiale og annet sensitivt materiale beskyttes mot uautorisert bruk.

For å muliggjøre oppbygning av totalsystemet skal design og utvikling av de enkelte avdelings- og delprosjektsystemene koordineres av Dokumentasjonsprosjektets faglige ledelse.

Daglig drift av ferdige delsystemer og senere vedlikehold av systemene er ikke en del av systemutviklingen i Dokumentasjonsprosjektet. Denne virksomheten bør overlates en driftsorganisasjon.

### **5 Omfang og tidsrammer**

Prosjektets fase 1 har et omfang på 750 årsverk. I utgangspunktet er disse tenkt fordelt med omlag 600 årsverk på Universitetet i Oslo og omlag 150 årsverk fordelt på de andre universitetene. Den endelige fordelingen avgjøres av styringsgruppen når omfanget ved de enkelte universitetene er endelig beskrevet.

Dersom det samlede dokumenterte registreringsbehovet ved de fire universitetene overstiger den nevnte rammen, skal styringsgruppen utarbeide planer for fase 2 av prosjektet.

### **6 Økonomiske rammer og finansiering**

750 årsverk innskrivning vil kreve kostnader i en størrelsesorden av 300 millioner kroner. Prosjektets størrelse umuliggjør derfor en totalfinansiering over universitetenes ordinære budsjetter. Det forutsettes derfor at størstedelen av prosjektet finansieres eksternt. Det overlates til styringsgruppen og prosjektarbeidsgruppen å utarbeide og fremme eksterne finansieringsmåter og å videreutvikle de allerede igangsatte registreringstiltakene (se punkt 1.3 under "Bakgrunn"). Oppdragsgiver vil sørge for å skaffe støtte for disse hos de sentrale universitetsmyndighetene så langt råd er. Oppdragsgiver vil i tillegg arbeide for at universitetenes fagdepartement opprettholder og øker bevilgningen over statsbudsjettet til prosjektet. Oppdragsgiver vil også sørge for tilstrekkelig kontorplass for prosjektarbeidsgruppen.

## **7 Planlegging/rapportering/oppfølging**

I hvert delprosjekt skal det utarbeides en halvårig rapport om fremdrift, status og vurdering av gjenstående arbeid. På basis av disse utarbeider prosjektarbeidsgruppen statusrapporter og annet informasjonsmaterieil for hele Dokumentasjonsprosjektet.

## **8 Kvalitetssikring**

Verdien av Dokumentasjonsprosjektet står og faller med om kvaliteten på de data som legges inn er i overensstemmelse med originalene og at de data som kommer ut til brukerne faktisk dekker behovene innenfor de ulike fag- og interesseområdene. Nøyaktighet og anvendbarhet er to sentrale nøkkelord. Disse vil man forsøke å etterleve gjennom kontinuerlig intern kvalitetssikring. I tillegg vil det være ønskelig med ekstern kvalitetskontroll. Den kan dels sikres gjennom referansegruppen, men det kan også være behov for ytterligere tiltak ved bruk av uavhengige kvalitetsrevisjoner.

## **9 Endringshåndtering**

Dokumentasjonsprosjektet er et svært omfattende prosjekt. Det er i dag ikke mulig å overskue alle problemer og hindringer som vil dukke opp i prosjektperioden. Prosjektarbeidsgruppen har myndighet til å foreta mindre endringer som er nødvendige for den daglige drift. Endringer som har konsekvenser for prosjektets omfang og lengde, skal forelegges styringsgruppen. Spesielt må definering og innlemmelse av nye delprosjekter på forhånd godkjennes av styringsgruppen.