Dokumentasjonsprosjektet ved Universitetsbiblioteket i Tromsø Rapport

Øyvind Eide Tor Sveum

1998

Innhold

1 Innledning **7**

2 Delprosjektets gang **9**

2.1 Forhistorien 9

2.1.1 Oversikt over Qvigstads etterlatte papirer 10

2.1.2 Innleggelse av 60 poster fra *Kvensk bibliografi til 1994* i databasen MicroMARC 10

2.2 Utvalg av materiale 11

2.2.1 Bibliografier 11

2.2.2 Brev 14

2.2.3 Kart 16

2.2.4 Bøker 17

2.3 Opphavsrett 21

2.4 Digitalisering 21

3 Fulltekstregistrert materiale **23**

3.1 Digitalisering av kart 23

3.1.1 Kvalitet 23

3.1.2 Publisering 24

3.2 Tekstlig materiale. Hovedprinsipper 25

3.3 Bakgrunn: SGML 25

3.3.1 Hvorfor SGML 26

3.3.2 Eksempel på bruk 27

3.4 Tekstregistrering 28

3.4.1 Tolkning av håndskrift — normalisering 28

3.4.2 Eksterne hjelpemidler 29

3.4.3 Plassering av håndskrevne notater 30

3.4.4 Forkortelser 31

3.5 Tegnsett og entiteter 31
3.6 Tagging 33
3.6.1 Strukturelle tagger 33
3.6.2 Grafiske tagger 36
3.6.3 Innholdstagger 39
3.6.4 Kryssende tagger 45
3.6.5 Automatisk tagging 47
3.7 Faksimiler 47
3.7.1 Brevsamlingene 47
3.7.2 Schnitlers protokoller 49
3.7.3 *Dokumenter angaaende flytlapperne* 49

4 Meta-materiale 51
4.1 Digitaliseringen 51
4.2 De enkelte trykte bibliografiene 51
4.2.1 *Bibliotheca Norvegica* 51
4.2.2 *Tillegg 1928–1945* 52
4.2.3 “Tillegg til Finnmark fylke” 52
4.2.4 *Bibliografi til Norges historie* 53
4.3 De enkelte feltene 53
4.3.1 Forfatter (*100/*245/*700) 53
4.3.2 Tittel (*245) 53
4.3.3 Utgave og utgiver (*250/*260) 54
4.3.4 Fysisk beskrivelse (*300) 54
4.3.5 Overskrifter som emneord (*653) 54
4.3.6 Kilde (*773) 54
4.3.7 Bibliografi (*850) 54
4.4 Tegnsett 55
4.5 Krav til publisering av MARC-data 56
4.5.1 Sortering og søking — stoppord og alfabetiseringsfelt . 56
4.5.2 Søkefelt 56

5 Oppsummering 57
5.1 Publisering og veiledning for leserne 57
5.2 Kildekritiske betraktninger 57
5.2.1 Bibliografiene 57
5.2.2 Trykt fulltekst 58
5.2.3 Håndskrifter 58

5.2.4 Kartene 59
5.2.5 Kommentarer 59
5.3 *Dokumenter angaaende flytlappernes* registre 59
5.3.1 "Identificering af stedsnavne" 59
5.3.2 "Sagregister" 61
5.4 Innholdsmessige betrakninger 62
5.5 Videre registrering 62
5.6 Modeller for registrering av annet materiale 63
5.6.1 Dokpro-modellen 63
5.6.2 Konvertering som en del av vanlig drift 63
5.6.3 Automatisk massekonvertering 63

Tillegg **65**

A Litteratur **65**
A.1 Om Dokumentasjonsprosjektet 65
A.2 Om meta-materialet 66
A.3 Om fulltekstregistrert materiale 67
A.4 Om SGML og digitalisering generelt 69

B Arkivalia etter Just Knud Qvigstad **71**
B.1 UB Oslo. Håndskriftsamlingen 71
B.1.1 Manuskripter 71
B.1.2 Mottatte brev (fra:) 73
B.1.3 Avsendte brev (til:) 74
B.2 Brev fra Qvigstad i finske arkiver og bibliotek 75
B.2.1 Riksarkivet 75
B.2.2 Finsk-Ugriska Sällskapet 75
B.2.3 Helsingfors universitetsbibliotek 75
B.2.4 Finska litteratursällskapet 76
B.2.5 Uleåborgs landsarkiv 76

C Databeskrivelser **77**
C.1 DTD for tekstlig materiale 77
C.2 DTD for bibliografiske referanser 86

D MARC-delmengden **87**

E Konvertering fra MARC til SGML **89**
E.1 Hvilken SGML? 89

F RTF-konvertering **91**
F.1 rtfhex 91
F.2 rtftag 93
F.3 rtfhenv 94
F.4 rtfvask1 96
F.5 rtfvask2 96
F.6 rtflik 97

G Utredning om Qvigstads felthåndbøker **99**
G.1 Aktuelle for selvstendig registrering 100
G.2 Bør sees i sammenheng med et evt. samisk språkprosjekt . . . 100
G.3 Bør sees i sammenheng med prosjekt navn 101
G.4 Bør sees i sammenheng med nordnorsk topografi 102
G.5 Krever forarbeid 102
G.6 Neppe aktuelle 103
G.7 Ikke på plass 103

H Samisk leksikografi **105**
H.1 Bakgrunn - dokumentasjonsprosjektet 105
H.2 Samisk ordlistemateriale 106
H.3 Samisk tekstkorpus 107
H.3.1 *Bibal* (Kristiania, 1895) 107
H.3.2 *Lappiske eventyr og sagn* / ved J. Qvigstad. (Oslo 1927–1929) 107
H.4 Hvor stort prosjekt? 108

Kapittel 1

Innledning

Dokumentasjonsprosjektet var et samarbeidsprosjekt mellom de fire universitetene i Norge. Prosjektets formål var å konvertere papirbasert materiale til elektronisk form. Prosjektet starter opp i 1991 og ble avsluttet ved utgangen av 1997. Denne rapporten omhandler et delprosjekt ved Universitetsbiblioteket i Tromsø, som gikk fra 1995 til 1997. For en oversikt over litteratur om prosjektet, se tillegg A.

I dette delprosjektet er det blitt konvertert ulike typer materiale. Dette har medført mye prøving og feiling. Alle avgjørelser som ble tatt var nok ikke like fornuftige, men vi håper denne rapporten vil bidra til at delprosjektet gir gode ettervirkninger. Dels håper vi at beskrivelsene av hvordan materialet er registrert vil gi bedre utnyttelse av databasene ettersom de blir publisert, dels håper vi at rapporten kan fungere som inspirasjon og praktisk hjelp ved konvertering av nytt materiale.

I delprosjektet er det blitt digitalisert to ulike former for materiale som behandles i hvert sitt kapittel i denne rapporten:

- historisk kildemateriale — se kapittel 3
- meta-materiale (bibliografier) — se kapittel 4

I forkant av, og parallelt med, delprosjektet har UB Tromsø drevet omfattende registreringer av biblioteks- og arkivmateriale, særlig i forbindelse med samlingene etter Just Qvigstad. Resultatene av dette arbeidet publiseres av UB Tromsø (se avsnitt 2.1).

Fagkonsulent Øyvind Eide stod for den daglige ledelsen av delprosjektet. Tor Sveum var delprosjektansvarlig, noe som innebar at han var prosjektets kontaktperson ved Universitetsbiblioteket. Han fungerte som faglig veileder i arbeidet med bibliografiene, og hadde også nyttige innspill til arbeidet med det andre materialet. Dikka Storm fungerte i praksis som en ekstra delprosjektansvarlig, dels ved å integrere konverteringsarbeidet i Tromsø Museums arbeid, dels ved å ta hovedansvaret for den faglige veiledningen i det historiske kildematerialet. Hovedfagsstudent i litteraturvitenskap, Nina Einevold, var leid inn som korrekturleser i Tromsø.

Avsnitt 2.1, 2.2.1–2.2.3 og tillegg B i denne rapporten er skrevet av Tor Sveum, mens Øyvind Eide har skrevet resten.

Databeskriveler, programmer og lignende som gjengis i denne rapporten finnes i digital form på **http://www.dokpro.uio.no/dokub/**. De er kun ment som inspirasjon for andre som vil utvikle lignende hjelpemidler, og vi tar intet ansvar for eventuelle feil og mangler.

Vi vil gjerne benytte anledningen til å takke alle som har hjulpet til med dette delprosjektet: En rekke kollegaer ved Universitetsbiblioteket i Tromsø, ved Tromsø Museum og ved Dokumentasjonsprosjektet har bidratt i større og mindre grad. Takk også til Trykksentralen i Harstad for hjelp med scanning av kart.

Tromsø og Oslo, 9. oktober 1998

Tor Sveum Øyvind Eide

Kapittel 2

Delprosjektets gang

2.1 Forhistorien

Dette delprosjektet i Dokumentasjonsprosjektet er en oppfølging av konserveringsprogrammet ved Universitetsbiblioteket i Tromsø, avd. Tromsø Museum.

Konserveringsprogrammet gikk fra 1991 til 1993, og bestod av flere forskjellige delprosjekter. Hovedelementene i programmet var konservering av eldre samlinger og konvertering av spesialsamlinger til BIBSYS.

De samlingene som ble registrert i BIBSYS var Qvigstad-samlingen og samlinger av polarlitteratur og litteratur som omhandler Nord-Norge. Under Dokumentasjonsprosjektets gang var det til stor hjelp at Qvigstad-samlingen var registrert i BIBSYS. Oversikt over denne og de andre samlingene fantes før bare på katalogkort.

Hensikten med programmet var tosidig, for det første å restaurere samlingene, for det andre å gjøre litteraturen mer tilgjengelig. I perioden 1991–1993 ble det årlig brukt ca. kr. 500.000,- på programmet.

I denne perioden ble også utlånsystemet automatisert. På den måten fikk vi bedre kontroll med utlånte bøker. Eldre litteratur (før 1980) som lånes ut blir nå automatisk registrert i BIBSYS. På sikt vil hele boksamlingen bli registrert i databasen.

I 1997 begynte HSJ-avdelingen (avd for humaniora, samfunnsfag og jus) ved Universitetsbiblioteket i Tromsø å konvertere alle sine bøker til BIBSYS. Tre bibliotekarer ble tilsatt på ekstraordinære midler for 1997 og 1998.

Gjennom hele 1990-tallet har store deler av Universitetsbiblioteket i Tromsøs boksamlinger blitt konvertert til BIBSYS.

Universitetsbiblioteket i Tromsø kom for alvor med i Dokumentasjonsprosjektet 1. juli 1995. Da ble Tor Sveum midlertidig tilsatt i 2 måneder fram til permanent fagkonsulent tiltrådte 1. september 1995. I de to foregående årene hadde vi bare bidratt med vår finansieringsdel, kr. 50.000,- pr. år.

I disse to månedene var arbeidet konsentrert om to ting, for det første å skaffe en oversikt over papirer etter Qvigstad i arkiver og bibliotek i Norge, Danmark, Sverige og Finland, og for det andre å teste ut MikroMARC-databasen for registrering av *Nordnorsk topografisk og lokalhistorisk bibliografi*. Et utvalg poster fra *Kvensk bibliografi til 1994* ble brukt som testmateriale.

2.1.1 Oversikt over Qvigstads etterlatte papirer

Qvigstad førte en omfattende brevveksling med vitenskapskollegaer i Norge og i utlandet. Ved Håndskriftsamlingen ved Universitetsbiblioteket i Oslo har det blitt satt opp en oversikt over materialet der ved gjennomgang av kortkatalogen og i samarbeid med Håndskriftsamlingens medarbeider Tone Modalsli. I Sverige og Finland ble oversikten utarbeidet gjennom korrespondanse med relevante arkiv og bibliotek. Totalt dreier det seg om ca. 400 brev, foruten diverse manuskripter ved Håndskriftsamlingen. Det finnes antakelig flere brev, f.eks. i Samisk-etnografisk avdeling ved Tromsø Museum, men det er i følge Dikka Storm ved avdelingen ikke mange. Vi gjorde også en skriftlig henvendelse til riksarkivene i Stockholm og København, Landsarkivet i Härnösand og Det Kgl. Bibliotek i København.

Vi vet at Qvigstad bygde opp sin boksamling gjennom bytte av bøker, innkjøp fra antikvariater osv. Brevvekslingen kaster et interessant lys over denne virksomheten. Gjennomlesning av et utvalg av brevene viser etter vår mening at de har stor kulturhistorisk og vitenskapshistorisk verdi. De egnet seg derfor godt til elektronisk konvertering innenfor Dokumentasjonsprosjektet, spesielt med tanke på en samlet oversikt, konservering og senere utgivelse av en trykt versjon. Kopi av oversikten finnes i tillegg B.

2.1.2 Innleggelse av 60 poster fra *Kvensk bibliografi til 1994* i databasen MicroMARC

Arbeidet ble gjort for å bli kjent med datasystemet, spesielt de forskjellige MARC-feltene. Innenfor Dokumentasjonsprosjektet ble en bibliografi over nordnorsk topografisk og lokalhistorisk litteratur registrert i MikroMARC. Et eget registreringsopplegg ble utarbeidet av fagkonsulent Øyvind Eide (se kapittel 4).

Foreløpig er Kvensk bibliografi bare en ren tekstfil. Med støtte fra Kommunal- og arbeidsdepartementet (kr. 10.000,-) ble bibliografien oppdatert og lagt ut på Internett. Den er tilgjengelig fra Universitetsbiblioteket i Tromsøs hjemmeside, med lenke fra Dokumentasjonsprosjektet.

2.2 Utvalg av materiale

Vi skal nedenfor gi en oversikt over materialet som ble registrert i Dokumentasjonsprosjektet, og forsøke å sette det inn i en kulturhistorisk sammenheng.

2.2.1 Bibliografier

Seks bibliografier over nordnorsk topografisk og lokalhistorisk litteratur ble valgt ut til registrering. Bibliografiene er landsdekkende. Vi ekserperte de delene som vedrørte Nord-Norge: Nordland, Troms, Finnmark, Svalbard og Jan Mayen. Grunnen til at vi valgte ut lokale bibliografier for Nord-Norge er at det ikke finnes en samlet bibliografi for hele dette området tilsvarende f.eks. Marklunds bibliografi for Nord-Sverige[1] og Nord-Finlands bibliografi.[2]

Et annet siktemål var at det registreringsopplegget vi utarbeidet, kunne brukes av andre (f.eks. Nasjonalbiblioteket) til å fullføre registreringen for hele landet. Vi tok tidlig kontakt med Nasjonalbiblioteket og Universitetsbiblioteket i Oslo for å informere om arbeidet. Ved Universitetsbiblioteket i Oslos Bibliografiske avdeling fikk vi opplyst at i de nasjonalbibliografiske basene registreres fortløpende norske geografiske emneord i MARC-felt 651 (både for periodika, bøker, tidsskriftanalytter, kart m.m.). I tillegg har Bibliografisk avdeling påbegynt konverteringen av Norsk bokfortegnelses femårs-kataloger til MARC-poster på basis av scannede filer som bearbeides. På poster hvor det framgår at innholdet har lokal interesse tilføyes MARC-felt 651 og geografisk betegnelse (landsdelsbetegnelser, områdebetegnelser, fylkesnavn og kommunenavn).

Innenfor Dokumentasjonsprosjektet ble følgende bibliografier registrert:

[1] E. Marklund: *Övre Norrland i Litteraturen.* 1963.

[2] *Pohjois-Suomen bibliografia vuoteen 1960.* 1972–1976.

- Deichmanske bibliotek. *Register til en del norske tidsskrifter. I. Topografi.* Kristiania, 1908.

  Som grunnlag for utvalget av de nordnorske postene ble det brukt en trykt versjon: *Bibliografi over eldre nordnorsk topografisk litteratur : tidsskriftartikler.* Red. av Tor Sveum. Universitetsbiblioteket i Tromsø, 1991.

- Johan Schweigaard. *Norges topografi : bibliografiske fortegnelser over topografisk og lokalhistorisk literatur.* Kristiania, 1918. [Fram til 1917].

- *Tillegg 1917–1927* ved W.P. Sommerfeldt. Oslo, 1930

- *Tillegg 1928–1945* ved W.P. Sommerfeldt. Oslo, 1955

- *Norsk lokalhistorisk litteratur 1946–1970* [Nordland, Troms, Finnmark]. Oslo 1979, 1980.

- *Bibliografi til Norges historie* 1946–75.

Allerede under Konserveringsprogrammet hadde vi scannet inn disse bibliografiene og lastet dem inn i en fritekst database[3] til internt bruk. Dette var lite tilfredsstillende på lengre sikt, da postene ikke kunne konverteres (importeres/eksporteres) i standard MARC-format eller gjøres søkbare i bestemte bibliografiske felter.

Den første av disse bibliografiene, *Deichmanske bibliotek*, ble utgitt i 1908, og er en indeks over tidsskriftartikler. Utvalget av tidsskrifter utgjorde bare 75 stk. I motsetning til fortegnelsen over norsk biografi som Deichmanske bibliotek utgav 3 år senere og som tok for seg 200 tidsskrifter var den ment som en prøve. Av de 75 er bare et fåtall vitenskapelige fagtidsskrifter (historie, arkeologi, geografi). De fleste av tidsskriftene er av blandet innhold utgitt til opplysning og underholdning.

Utarbeidelsen av denne bibliografien må ses på bakgrunn av den voldsomme veksten i floraen av tidsskrifter etter midten av forrige århundre og utgjør et første forsøk på å skaffe oversikt over mengden av tidsskriftartikler innen humaniora knyttet til bestemte steder i Norge. Ordningen er alfabetisk etter sted, med henvisning fra mindre steder til større. Ingen naturvitenskapelige

[3] Programmet som ble brukt var AskSam.

tidsskrifter er inkludert. I vårt eksemplar fra Qvigstad-samlingen har Qvigstad gjort tilføyelser på innskutte ark for steder i Nord-Norge. Disse notatene er foreløpig ikke lagt inn i basen.

Schweigaards bibliografi ble utgitt ca. 10 år senere. Det regnes som hovedverket i norsk topografisk bibliografi. Det analyserer over det dobbelte antall tidsskrifter og inkluderer også bøker. Spesielt merkbar er veksten av antall fagtidsskrifter innen alle fagområder, også naturvitenskap og anvendt vitenskap. Ordningen er geografisk etter fylke med en egen emneinndeling. Schweigaard har et eget henvisningsystem til *Deichmanske bibliotek*, *Bibliotheca Norvegica* og *Halvorsens forfatterlexikon*. Med unntak av *Halvorsens forfatterlexikon* er disse også registrert i vår database.

Schweigaard ble anmodet av Landskomiteen for lokalhistorie om å utarbeide bibliografien. Bakgrunnen for tiltaket var en voksende interesse for by- og bygdehistorie og lokalhistorisk forskning. Fra århundreskiftet og fram til 1920 hadde det blitt opprettet en lang rekke historielag rundt om i landet. De hadde som en av sine viktigste oppgaver å utgi by- og bygdebøker. Med opprettelsen av en sentral organisasjon (fra 1920 Landslaget for by- og bygdehistorie) ble det lokalhistoriske arbeidet fastere organisert. Det ble utført mer systematisk og planmessig. Bibliografien inngikk som en viktig del av dette arbeidet, og den ble et uunnværlig hjelpemiddel for å skaffe oversikt over litteratur om de enkelte steder.

W.P. Sommerfeldt fortsatte Schweigaards bibliografi i to tillegg som dekker periodene 1917–1927 og 1928–1945, etter de samme linjer som forgjengeren, men med grovere emneinndeling; litteraturen om hvert herred er her samlet i én alfabetisk gruppe. *Tillegg 1928–1945* inneholder bare referanser til *Norsk bokfortegnelse* og *Norsk tidsskriftindeks* og litteratur som er inkludert i disse. Som basis for vår registrering i MikroMARC er derfor *Norsk tidsskriftindeks* og *Norsk bokfortegnelse* benyttet.

Fra *Norsk lokalhistorisk litteratur 1946–1970* har vi registrert Nordland, Troms og Finnmark. Denne bibliografien inneholder bare bøker (ikke avis- og tidsskriftartikler) og omfatter litteratur som er registrert i *Norsk bokfortegnelse* i perioden 1946–1970. Innenfor hvert sted (fylke, region, kommune) er det en emneinndeling etter Deweys klassifikasjon.

Som et supplement til *Norsk lokalhistorisk litteratur* valgte vi ut *Norsk historisk bibliografi* for perioden 1946–1975 fordi en del av den historiske litteraturen har lokalhistorisk interesse eller omhandler bestemte steder. Den omfatter også tidsskriftartikler. Her var imidlertid den geografiske inndelingen mye grovere. Historisk bibliografi har en egen emneinndeling hvorav topografi og lokalhistorie bare utgjør en del.

Med registreringen av disse bibliografiene fra de eldste tider og fram til ca. 1975 skulle den lokalhistoriske og topografiske litteraturen for Nord-Norge være godt dekket. Vi valgte å gå fram til ca. 1975 fordi BIBSYS og UBO:BOK begynner sin registrering på begynnelsen av 1970-tallet. Norsk tidsskriftindeks begynner registreringen i 1980.

Det vesentligste av nordnorsk topografisk og lokalhistorisk litteratur skulle dermed være elektronisk tilgjengelig. Med MikroMARCs webversjon er vår database også tilgjengelig via Internett.

2.2.2 Brev

Just Qvigstads brev til vitenskapskollegaer i nordiske land ble valgt ut til registrering i Dokumentasjonsprosjektet fordi de er sentrale kilder til Qvigstads liv og arbeid, til rekonstruksjon av tid, sted, tanker og følelser i Qvigstads liv. De gir et blikk bak kulissene i hans vitenskapelige produksjon. Gjennom Qvigstads kommentarer får vi også mange interessante glimt inn i samtidas kulturelle og politiske forhold. Registrering og innsamling av brevene kan danne grunnlag for en trykt utgave og en kritisk biografi. Den bør etter vår mening skrives av en som er fortrolig med finsk-ugrisk (samisk) språkforskning og samisk etnografi og kulturhistorie.

Brevene synes ikke være skrevet med tanke på å bli lest i ettertid — noen av dem er skrevet i all hast og med en slurvete skrift som vi hadde problemer med å tolke. Vanligvis har Qvigstad en god, leselig håndskrift.

Når vi foretok dette innsamlingsarbeidet fra biblioteker og arkiver i Norden var det fordi vi ønsket en så fullstendig oversikt over brev- og arkivmateriale etter Qvigstad som mulig. Sammenstillingen av dette materialet vil lette arbeidet for dem som kommer etter oss og som ønsker å bruke dette forskningsmaterialet.

Det er likevel ikke noe fullstendig innsamlingsarbeid vi har gjort.

Den største enkeltsamlingen finnes ved Håndskriftsamlingen, Universitetsbiblioteket i Oslo. Foruten brev finnes der også ordsamlinger og språkprøver. Materiale etter Qvigstad i Sverige fant vi bare ved Uppsala universitetsbibliotek. En henvendelse til Landsarkivet i Hernösand ga negativt resultat, men det er ikke klart hvor grundig søket var. Vi vet at Qvigstad førte korrespondanse med prester i Nord-Sverige, men ikke om brevene er bevart. Det kan finnes mer materiale, f.eks. i Nordiska muséet i Stockholm. I Finland er materialet spredt rundt i forskjellige arkiv og bibliotek (se tillegg B).

Ut av brevene kan vi lese hvordan Qvigstad bygde opp sin boksamling, som vi har registrert i BIBSYS. Det skjedde først og fremst gjennom utveksling av publikasjoner med vitenskapskollegaer og gjennom regulære innkjøp i antikvariater og bokhandlere.

Det er særlig tre brevsamlinger som skiller seg ut, både i omfang og innhold, og som vi har funnet det natulig å digitalisere. Det er brevene til Professorene Magnus Olsen, Karl Bernhard Wiklund og Emil Nestor Setälä.[4] Alle tre er fremtredende språkforskere og filologer som Qvigstad i lange perioder hadde et nært samarbeid med, og etterhvert utviklet et personlig vennskapsforhold til.

Samlingen til Setälä utgjør 96 brev, og spenner over en periode fra 1887 til Setäläs død i 1935. Antall brev til Wiklund er 96 i perioden 1891 til 1934, og til Olsen 64 i perioden 1909 til 1956. Totalt dekker altså brevvekslingen en periode på nesten sytti år.

På det tidspunkt da brevvekslingen med Setälä tok til hadde Setälä allerede utgitt viktige språkvitenskapelige arbeider, bl.a. en dialektavhandling fra Satakunta i Sørvest-Finland, og en doktoravhandling som introduserte de såkalte neogrammatiske språkteorier i Finland. Setälä kom til å bli den ledende representant for denne skolen, som både Qvigstad og Wiklund sluttet seg til. Det filosofiske grunnlaget for denne skolen er positivismen, basert på empiri, det "positivt" gitte og avvisning av metafysisk spekulasjon og spekulativ teoridannelse.

Her skisseres kort hva disse språkteoretiske idéene gikk ut på, siden de kom til å danne de vitenskapsteoretiske forutsetninger og betingelser for Qvigstads og Wiklunds arbeid og forskning.

Positivistiske idéer overført til språkvitenskapen gjorde at utforskningen av det nåtidige, levende folkespråk ble satt i sentrum, og spesielt utforskningen av de enkelte sammenlignbare språkgrupper og dialekter. Lydmaterialet basert på innsamlede språk- og dialektprøver var det som lettest kunne gjengis med eksakt vitenskapelig presisjon. Stor nøyaktighet i prosedyrene for fonetisk transkripsjon er et poeng som gjentatte ganger understrekes i Qvigstads brev. Siden det enkelte språk er foranderlig og resultat av en historisk utvikling i kontakt med andre språk var det nødvendig å dokumentere det enkelte språk i form av historiske språkprøver, tekster og ordbøker. Dette materialet utgjorde grunnlaget for sammenlignende språkstudier basert på

[4]Det er påfallende at vi ikke har funnet brev til Qvigstad fra noen av disse tre personene i noe arkiv eller bibliotek.

den historisk-komparative metode.

Det vitenskapelige forskningsarbeidet til Qvigstad, Setälä og Wiklund består primært i fortløpende innsamling og publisering, og kan kort skisseres i fem punkter:

- innsamling av språkvitenskapelig materiale (dialekt- og språkprøver) basert på feltarbeid og forskningsreiser.
- innsamling av forskjellige typer etnografisk materiale, basert på feltarbeid og forskningsreiser.
- innsamling og avskrift av historiske dokumenter basert på arkivstudier.
- utgivelse av språkvitenskapelige, historiske og etnografiske avhandlinger basert på dette materialet.
- utgivelse av språkvitenskapelig materiale, leksika, fagordbøker og tekstsamlinger.

Den faglige delen av korrespondansen dreier seg derfor i stor grad om spørsmål i forbindelse med språk- og dialektprøver, publisering av historiske dokumenter og utveksling av skrifter.

Korrespondansen gjenspeiler også det samarbeidet disse tre forskerne hadde i lange perioder. Med støtte fra Setälä publiserte Qvigstad sine tidlige arbeider i skriftseriene til Finsk-Ugrisk Selskap, som ble opprettet i 1883. Qvigstad og Wiklund samarbeidet nært om dokumentasjon av samiskspråklige skrifter og samisk bibliografi. Qvigstad og Olsen gav sammen ut Finnmarksbindet af Oluf Ryghs *Norske Gaardnavne*.

2.2.3 Kart

Til registrering i Dokumentasjonsprosjektet valgte vi ut to kartserier av Jens Andreas Friis. *Ethnografisk kart over Finmarken* kom ut i 1861–62. Disse kartene ble omarbeidet i 1888 og utvidet med Tromsø amt, samt Lofotens prestegjeld i Nordlands amt i 1890.

Jens Andreas Friis overtok som professor etter N. Stockfleth ved Universitetet i Christiania, og fikk ansvar for undervisning i samisk og finsk og utgivelse av språklige og etnografiske skrifter. Friis hadde tidligere studert finsk og samisk under Stockfleth og fortsatt studiene i Finland under Elias

Lönnrot. Han er kanskje mest kjent som forfatter av fortellinger med emne fra samenes liv.

De etnografiske kartene er sjeldne. På Tromsø Museum finnes kun en komplett serie (på Samisk-etnografisk avdeling). I biblioteket finnes ukomplette enkeltkart. Fordi de er så vanskelig tilgjengelige ble det på 1970-tallet diskutert å gi dem ut i faksimileutgave på Universitetsforlaget. Dette ble det imidlertid ikke noe av. Grunnet stor pågang på vårt bibliotek fikk vi fotokopiert et sett til utlån.

På grunn av sin detaljrikdom og nøyaktighet er kartene en enestående samtidig kilde til sammensetningen av befolkningen i Finnmark og Troms, dvs. forholdet mellom nordmenn, samer og kvener. I tillegg gir kartene detaljerte opplysninger om språksituasjonen i de enkelte familier og boligforhold etter hustype (jordgamme eller trehus).

Kartserien fra 1861 ble utarbeidet etter en reise til Finnmark og en del av Nordlands amt som Friis foretok med offentlig støtte i 1859. Kartene er utgitt av Videnskabsselskabet i Christiania med økonomisk støtte fra Oplysningsvæsenets Fond. På dette tidspunkt var Friis dosent i samisk ved Universitetet i Christiania. Som grunnlag for arbeidet brukte Friis kystkart i målestokk 1:200.000. Utarbeidelsen av kartene er basert på nøyaktige observasjoner og studier av Friis selv. Som informanter har han også brukt skolelærere og andre lokalkjente folk. De utarbeidede blad ble tilslutt sendt prestene i de enkelte prestegjeld for endelig korrektur og revidering.

Bosetningen i de enkelte lokaliteter er avmerket med 22 ulike tegn. Tegnene er inndelt i 3 hovedkategorier etter norske, samiske og kvenske familier. Disse er igjen oppdelt etter språk og hustype. I 1861-serien er de etniske gruppene skilt fra hverandre gjennom fargelegging.

Selv om de første tallmessige registreringer av den samiske befolkning går helt tilbake til 1500-tallet, er det først fra midten av 1800-tallet at en begynner å registrere de ulike etniske gruppene. Fra 1865 ble det også innhentet opplysninger om språkbruk. Disse statistiske kildene ble brukt av myndighetene for tiltak i skole og undervisning, og senere også som underlagsmateriale for den strenge fornorskningspolitikken som satte inn fra 1880-tallet.

2.2.4 Bøker

Av historisk kildemateriale valgte vi ut dokumenter som allerede er utgitt som bøker. Vi baserte digitaliseringen på bokversjonene, og har ikke foretatt noen sammenligning med originaldokumentene som lå til grunn for bokutgivelsene.

Dokumenter angaande flytlapperne

En av de viktigste dokumentsamlingene som omhandler grensespørsmål i Nord-Norge er *Dokumenter angaande flytlapperne* som ble utgitt av J.K. Qvigstad og K.B. Wiklund i 1909. Samlingen inneholder transkripsjoner av dokumenter og utdrag fra dokumenter fra perioden 1558–1900.

Dokumentsamlingen ble utarbeidet på oppdrag av *Renbeitekommissionen af 1907*. Behovet for en oversikt over de historiske kildene skyltes uenighet mellom Norge og Sverige om hvilken rett til beite svenske samer hadde, særlig i Troms fylke. Fra norsk side hadde det i lang tid vært et ønske om å forby svenske samer beite i Norge. Norske myndigheter mente at bruk av områdene som reinbeiter var til hinder for jordbrukskolonisasjonen. Fra svensk side ble det derimot hevdet at samene hadde historiske rettigheter til slikt beite som hverken svenske eller norske myndigheter kunne se bort fra.

Uoverensstemmelsen ble aktualisert ved unionsoppløsningen i 1905, og reinbeitespørsmålet ble et av fire forhandlingstema i Kalstadforhandlingene. Saken var gjenstand for lange forhanlinger i årene som fulgte.[5] Et ledd i prosessen var nedsettelsen av en norsk-svensk reinbeitekommisjon i 1907. På grunn av et svært omfattende mandat, besluttet den å nedsette to underkomitéer i desember 1907. Her følger et utdrag av instruksen til den ene av disse:

> Paa grundlag af 1ste kodicil til grænsetraktat af 1751 og konventionen af 26de oktober 1905 skal undersøges:
>
> 2. i hvilken udstrækning de svenske lappers renbeitning i Tromsø amt har fundet sted fra gammel tid saavelsom i nyere tid; herunder søges bragt paa det rene tiden for lappernes flytninger vaar og høst, for overgangen over rigsgrænsen, samt for deres ankomst til og flytning fra opholdsstederne; endvidere i hvilke trakter nævnte lappers rener har pleiet at beite. [...]
>
> Hvad kommissionen ønsker, er en fremstilling af materialet uden nogen paavisning af de slutninger, hvortil dette maatte kunne give anledning.
>
> Det er forudsætningen, at alle de oplysninger, som enten den ene eller den anden af de to sagkundige anser for at være af betydning

[5] Dette er beskrevet i detalj i Erling Lae: *Fra Karlstadkonvensjon til reinbeitekonvensjon.*

> for undersøgelsen, medtages i fremstillingen, saaledes at kun det udelukkes, som begge anser for betydningsløst.[6]

Qvigstad og Wiklund ble utpekt som ansvarlige for dette arbeidet. Resultatet ble en dokumentsamling som ble trykket i to bind med tittelen *Dokumenter angaaende flytlapperne*. Det første bindet består hovedsakelig av de deler av Schnitlers grenseeksaminasjonsprotokoller som dekket området konflikten omhandlet.[7] Dette utgjør slutten av volum 2, samt mesteparten av volum 3 og 4. Dette bindet har foreløpig ikke blitt digitalisert.

Bind 2 av dokumentsamlingen består av en omfattende samling større og mindre dokumenter og deler av dokumenter. Dessuten finnes der to omfattende registre, "Identificering af stedsnavne" og "Sagregister". Bindet er en av hovedkildene til forståelse av reinbeitekonfliktens historiske bakrunn. Det eldste dokumentet i samlingen er "Jöns Nilssons räkenskaper för Pite lappmark 1558". Den trykte dokumentsamlingen er delvis kronologisk ordnet, men innholder kun en kortfattet innholdsfortegnelse. For de informasjonsbiter som er indeksert i registre (som stedsnavn) har man hatt en relativt enkel vei inn i materialet, mens det har vært vanskeligere å finne igjen f.eks. bestemte personnavn i den trykte versjonen.

Dette var en av grunnene til at dokumentsamlingen ble sett på som spesielt viktig å digitalisere, og vi har under arbeidet med å digitalisere bind 2 forsøkt å merke ut spesielle innholdskategorier for å gjøre det mulig å strukturere den store informasjonsmengden i dokumentsamlingen (se kapittel 3).

Ved Tromsø Museum finnes et trykt eksemplar av dokumentsamlingens 432 første sider, som er alt unntatt registrene. I dette eksemplaret har Qvigstad ført inn en stor menge notater i margen og mellom linjene. Dette er antakelig et prøvetrykk som Qvigstad blant annet brukte til å gjøre notater som siden dannet grunnlag for registrene. Disse ble så trykket, og utgjør side 433–523. Eksemplaret inneholder dessuten korreksjoner, av hvilke noen gjenfinnes i "Rettelser" på side 524, og store mengder understrykninger utført med blå, rød og grå blyant, samt med blekk. Alt dette er digitalisert sammen med den trykte teksten av boka, og vil således publiseres for første gang som et resultat av dette prosjektet.

---

[6] Fra "post 2 af det af norske og svenske delegerede den 19de januar 1907 afgivne forslag", her sitert etter *Dokumenter angaaende flytlapperne*, bd. I, første side i forordet (upaginert).

[7] Mer om disse protokollene i avsnitt om Schnitlers protokoller nedenfor.

Etter at Qvigstads og Wiklunds arbeid var avsluttet, ble det nedsatt flere reinbeitekommisjoner som bl.a. gjennomførte omfattende befaringer i marken. Rapportene fra disse kommisjonene finnes i trykt form, men er ikke digitalisert (se også avsnitt 5.6.3). Qvigstad skrev i 1911 til professor Setälä i Finland:

> De undersökelser, som foretages fra norsk side og antagelig ogsaa fra svensk side i anledning av renbeitesaken, vil, naar de engang foreligger for offentligheten, skaffe et ualmindelig rikt materiale til oplysning av rendriften og lappernes livsforhold. Saa nöiagtige undersökelser vilde ellers aldrig være blit anstillet.[8]

Schnitlers grenseeksaminasjonsprotokoller

I forbindelse med forhandlingene fram mot grenseavtalen mellom Danmark-Norge og Sverige-Finland i 1751 foretok Major Peter Schnitler en grenseoppgang fra Røros til Øst-Finnmark der han organiserte forhør av befolkningen. Dette arbeidet er dokumentert i hans grenseeksaminasjonsprotokoller, hvis originaler finnes i Riksarkivet i Oslo sammen med kart, korrenspondanse og annet materiale vedrørende hans arbeid med grenseproblematikken.

Protokollene ble ført i felten, og er en svært viktig kilde til forståelse av forholdene i grenseområdene i midten av det attende århundre, og inneholder også opplysninger om eldre historie. Selv om dokumentene må leses i lys av at de er skrevet av en Dansk-Norsk embedsmann med et bestemt oppdrag, inneholder de viktig samisk kildemateriale, ikke minst som bakgrunn for den 1. kodicill som kom med i grenseavtalen av 1751, den såkalte "Lappekodicillen".

Protokollene består av 7 volumer. Av disse ble slutten av volum 2 og mesteparten av volum 3 og 4 sammen med noen andre dokumenter trykt som bind 1 av *Dokumenter angaaende flytlapperne*. Fra et restopplag av dette trykket fjernet man i 1926 de delene som ikke er Schnitlers protokoller, laget nytt tittelblad og utgav det som bind 2 i en samlet utgave av Schnitlers protokoller, *Major Peter Schnitlers grenseeksaminasjonsprotokoller 1742–1745*. Bind 1 av samlingen, bestående av volum 1, de ikke trykte delene av volum 2–4 og volum 5 og 6, ble utgitt i 1962. Endelig ble volum 7 utgitt som bind 3 i 1982.

[8] Brev fra J.K. Qvigstad til E.N. Setälä, datert "Kristiania 15/10 1911".

Det er ingen tvil om at alle tre bind fortjener å konverteres til digital form. Imidlertid hadde vi ikke kapasitet til dette i denne omgang. Vi valgte derfor å konsentrere oss om det bindet som dekker de delene av grensegangen som i liten grad behandles i *Dokumenter angaaende flytlappernes* bind 2, nemlig *Major Peter Schnitlers grenseeksaminasjonsprotokoller 1742–1745*, bind 1. De volumer som her er gjengitt dekker Trøndelag, deler av Nordland, og Finnmark.

Dessverre førte forsinkelser i registreringen til at registrene i bindet (s. 438–) ikke ble digitalisert. Dette er uheldig for mulighetene til gjenfinning av materialet i bindet.

2.3 Opphavsrett

Siden deler av materialet i delprosjektet er opphavsrettslig beskyttet, ble det nødvendig å innhente tillatelse fra rettighetshaverne før konverteringen startet — selve konverteringen til elektroniske form må sannsynligvis regnes som eksemplarframstilling i juridisk forstand, og vil i seg selv kreve tillatelse fra rettighetshaver uavhengig av om materialet publiseres eller ikke.

Tillatelse ble derfor hentet inn fra:

- Just Knud Qvigstads arvinger, representert ved dr. Just Knut Qvigstad (yngste sønn).
- Karl Bernhard Wiklunds arvinger, representert ved Anne Wiklund (datter).
- Kildeskriftfondet, som står ansvarlig for utgivelsen av Schnitlers protokoller.

Tillatelsene til bruk ble gitt under forutsetning av at Tromsø Museum gjør de nødvendige vurderinger av forskningsetiske og personvernmessige aspekter ved å publisere forskningsnotater og brev, og innebærer ikke nødvendigvis at alt materialet gjøres fritt tilgjengelig.

2.4 Digitalisering

Selve konverteringen skjedde ved *Universitetenes registreringssentral i Indre Salten*, med unntak av Friis' etnografiske kart. Dette ga god kvalitet på det digitaliserte materialet, samtidig som registreringshastigheten var akseptabel. At hastigheten ikke var høyere hadde sammenheng med at materialet var vanskelig, men også med stadige endringer i arbeidsstokken, som naturlig er i et sysselsettingstiltak.

Opplæring av nye arbeidslag, eller overgang til ny type materiale, skjedde alltid ved at fagkonsulenten besøkte Fauske i 1–3 dager. I starten var han noe tilbakeholdende med besøkene, noe som førte til at kontakten med lagene ikke var god nok. Kostnadene ved å registrere såpass komplisert materiale ved en sentral langt unna det faglige miljøet ved universitetet må bli store. Selv om man ikke skal overdrive reisningen, viser det seg at det ikke er noe å tjene på å reise for lite, da dette fører til lavere produksjonstempo og økt risiko for systematiske feil.

Kvalitetssikringen skjedde i to trinn. Først ble det lest flere korrekturer i Fauske, deretter ble materialet sendt til Tromsø for sluttkorrektur. Hvis materialet ved ankomst Tromsø inneholdt for mange feil, ble det returnet Fauske for ny korrektur. Dette systemet fungerte ikke så godt i starten, da materiale ble sendt for sjelden fra Fauske. Dette førte til unødig ekstraarbeide for fagkonsulenten. Etterhvert kom sendingene inn i fastere former, og fagkonsulenten kunne raskt sende tilbake materiale som ikke var av tilfredsstillende kvalitet.

Det fantes ikke utstyr som kunne konvertere Friis' etnografiske kart i Tromsø. Når det gjaldt de to senere seriene, 1888 og 1890, fantes en scanner i Bergen som kunne ta jobben. Denne scanneren kunne ta store formater, men kun bitonalt (svart/hvitt). Den kunne derfor ikke brukes på 1861-serien.

I Bergen fantes dessuten eksemplarer av alle kartene i 1888 og 1890-seriene. Dette gjorde det mulig for Dokumentasjonsprosjektets delprosjekt ved Bergens Museum å scanne kartene.

1861-serien var vanskeligere. Vi var ikke i stand til å finne mer enn ett komplett sett av kartene. Dette gjorde at vi ikke kunne ta sjansen på å sende kartene i posten. Etter noe leting fant vi ut at en bedrift i Harstad, Trykksentralen, hadde utstyr til å gjøre jobben. Fagkonsulenten reiste da med kartene til Harstad og fikk dem scannet i farger i høy oppløsning.

Kapittel 3

Fulltekstregistrert materiale

3.1 Digitalisering av kart

3.1.1 Kvalitet

Først etter mye fram og tilbake kom Friis' etnografiske kart med i delprosjektet, som beskrevet i siste del av avsnitt 2.4.[1] Kartseriene fra 1888 og 1890 ble scannet i svart-hvitt med en oppløsning på 300 punkter pr. tomme (DPI). Selv om kartene er trykt i svart-hvitt og har klare kontraster, viser det seg likevel av kvaliteten ikke ble så god. Kartserien fra 1861 er scannet i 24-bits farger med en oppløsning på 400 punkter pr. tomme, noe som gir svært god kvalitet. Det viser seg at selv rene svart-hvitt kart krever enten gråtoner eller høy oppløsning for at nyansene ikke skal forsvinne. Dette fenomenet er velkjent i litteraturen:

> There is by now a good deal of consensus as to what does constitute a research quality image. An important study conducted at Colombia University, the Oversized Color Image Project, concluded that, in the case of most maps, all the essential details could be captured at 200 dpi in 24-bit color. The Library of congress has adopted a 300 dpi 24-bit standard for its map digitization projects, and this is likely to become the de facto standard for similar projects. [...] When bitonal scanning is used, materials

[1] En av kvadrantene på kartblad 10 i 1861–62-serien ble ikke scannet på grunn av tidspress i Harstad. Denne kvadranten inneholder kun en nordstjerne, og kan scannes inn i lavere kvalitet, f.eks. på Universitetsbibliotekets svart-hvitt A3-scanner.

are scanned at a higher resolution (between 600–800 dpi) to gain image quality comparable to 300 dpi 24-bit color scanning, but the size of the bitonal files is still much smaller than the color.[2]

3.1.2 Publisering

Vi la arbeid i å finne en fornuftig måte å publisere kartene på. Vi kunne tenke oss å koble kartmaterialet mot moderne GIS-kart, slik at informasjonen på Friis' kart kunne legges på elektronisk kart. Dette kunne så sammenholdes med annen informasjon på slike kart, slik som kommunegrenser, veier, stedsnavn, og også andre data fra Dokumentasjonsprosjektet og lignende prosjekter, f.eks. arkeologiske funn.[3]

Det viste seg dessverre at dette ikke var mulig. Kartene har ikke tilstrekkelig nøyaktighet til at en slik kobling vil være forsvarlig. Det er f.eks. umulig å fastslå eksakt hvor gårder ved elver i innlandet ligger. Det betyr ikke at man ikke kan anta omtrent hvor stedene ligger i dag, og kartene fungerer godt som veiviser i marken.[4] Men man kan ikke plassere stedene eksakt over på et moderne kart. Å legge dem over på et moderne digitalt kart ville gitt inntrykk av en presisjon vi ikke kan stå inne for.[5]

I stedet bør kartbladene legges ut hver for seg med angivelse av hvilket område de dekker. Senere kan det tenkes en mellomløsning der man deler skjermen i to og legger det moderne kartet på den ene siden og Friis-kartet på den andre. En grov lenking kan gjøre at når man beveger seg rundt på det moderne kartet følger Friis-kartet bevegelsene som en skygge. Det kan også være fornuftig å lage et stedsnavnregister, jfr. avsnitt 5.5.

[2] David Yehling Allen: "Creating and Distributing High Resolution Cartographic Images".

[3] Se Hans Peter Blankholm: "Geografiske informasjonssystemer" for flere opplysninger om hvilke muligheter som finnes.

[4] Opplyst av J.A. Kalstad.

[5] Mye av dette arbeidet er basert på 1861–62-serien. Foreløpig undersøkelser av 1888–90-serien var noe mer lovende, og det kan være mulig å arbeide fram løsninger her. Disse sistnevnte kartene skal være basert på en Lambertprojeksjon (en kjegleprojeksjon) i følge muntlige opplysninger fra Roald Aanerud, en pensjonist som tidligere arbeidet i Statens Kartverk. Troms-kartet fra 1890 er muligens basert på amtskartet fra 1874?

3.2 Tekstlig materiale. Hovedprinsipper

Arbeidet med tekstene fra bøkene begynner med at sidene scannes inn og kjøres gjennom et tekstgjenkjenningsprogram.[6] Resultatet av dette var en digital tekst som ble tatt opp i en tekstbehandler.[7] Denne teksten ble korrekturlest flere ganger.

Håndskreven tekst ble skrevet inn i tekstbehandlingsprogrammet. For notatene i *Dokumenter angaaende flytlapperne* ble det skrevet inn på grunnlag av fotokopier, mens sistekorrekturer ble lest på grunnlag av originalen. Brevene ble skrevet inn og korrekturlest på grunnlag av utskrift fra mikrofilm.

I tillegg til den rene teksten ble det lagt inn i tilleggsinformasjon (metainformasjon) som står mellom '<' og '>' og kalles tagger (se avsnitt 3.6). Spesialtegn som ikke kan gjengis i "vanlig" tegnsett er kodet som uttrykk som står mellom '&' og ';' og kalles entiteter (se avsnitt 3.5).

Spesialtegnene og taggene er lagt inn i henhold til den internasjonale standarden SGML (se avsnitt 3.3). Materialet registreres i SGML i henhold til en spesialskrevet databeskrivelse (DTD) som er gjengitt i tillegg C.1. Alt tekstlig kildemateriale, både brev, bøker og notater, bruker samme databeskrivelse. Denne registreringen kan sammenlignes med at det legges et databaseskjema over teksten, slik at tekstfragmenter havner i ulike felt samtidig som de er en del av den løpende teksten.

Her følger en gjennomgang av hvordan selve teksten registreres, av entitetene og av taggene. Det som primært beskrives er selve kodingen av tekstene, mens det legges mindre vekt på hvordan de elektroniske utgavene bør utformes. Eksempler fra de ulike materialtypene trekkes inn der det er relevant. Ideen bak en slik strukturering av denne rapporten er å vektlegge de generelle idéene som ligger til grunn for konverteringsarbeidet. Men aller først en kort innledning om SGML.

3.3 Bakgrunn: SGML

På samme måte som en trykt bok er noe mer enn et maskinskrevet manuskript, er en digital utgivelse noe mer enn en datafil. For å utvide tekstens bruksmuligheter, må man legge inn tilleggsinformasjon i tillegg til den rene teksten.

[6] Optical Character Recognition, OCR. Programmet som ble brukt var OmniPage.

[7] Programmet som ble brukt var Microsoft Word 6.0.

3.3.1 Hvorfor SGML

Innen Dokumentasjonsprosjektet har dette vært gjort ved hjelp av den internasjonale standarden SGML (Standard Generalised Markup Language, ISO 8879:1986). For de som ønsker mer informasjon om SGML og Dokumentasjonsprosjektet enn gjennomgangen nedenfor, henvises det til litteraturlista i tillegg A.

Fordelene ved å bruke SGML-baserte systemer er sammenfattet av Peter Robinson på følgende måte:

- SGML is an international standard. It is therefore exactly and consistently defined.
- SGML is widely supported, and implementations of great sophistication exist.
- The fundamental premise of SGML, that it describes what text is rather than declares how it should be processed, matches the scholarly model.
- SGML is a powerful tool for describing the structure of a text and ensuring that a particular text fits that description.
- SGML is not dependent on any particular software or hardware implementations: the one file will display and print idetically on the different systems.
- Because SGML is a ratified standard and because it is system-independent, SGML files may be considered a permanent, or even an archival, medium. As long as there are computers, SGML files can be read and used.[8]

Det finnes retningslinjer for hvordan humanistisk kildemateriale i digital form bør struktureres (TEI guidelines). Disse baserer seg på SGML. Selv om man i Dokumentasjonsprosjektet ikke følger disse retningslinjene direkte,[9] vil det, fordi man står på "felles grunn", være mulig å konvertere tekstene maskinelt til TEI-format.

---

[8] Peter Robinson: *The Transcription of Primary Textual Sources Using SGML*, s. 33.

[9] Dette skyldes at Dokumentasjonsprosjektet startet så tidlig som 1991. På den tiden var ikke TEI ferdig med å utarbeide sine retningslinjer. Imidlertid var man i kontakt med TEI ved utarbeidelsen av Dokumentasjonsprosjektets databeskrivelser.

Å bruke standardiserte metoder i digitaliseringsarbeid gjør det lettere å utveksle tekster mellom ulike institusjoner. Det gjør det også lettere for ansatte å forstå strukturen i andre institusjoners digitale tekster, og det gir tilgang til en vidt spekter av dataprogrammer for manipulering av tekstene med priser fra gratis til noen få tusen kroner.[10] Slike programmer kan bistå i korrekturlesing av materiale. Bruk av SGML gjør det også enkelt å produsere elektroniske og trykte utgaver av tekstene, for eksempel kan man koble tekstene opp mot såkalte stilark (style-sheets).

Når man har skrevet inn disse taggene i tillegg til selve teksten, blir det mulig å svare på spørsmål som: "Hvilke stedsnavn i *Dokumenter angaaende flytlapperne* er understreket med blå blyant, hvilke er understreket med rød?" "Finnes det stedsnavn som noen steder er understreket med rødt, andre steder med blått?" Man kan si at et slags databaseskjema er lagt over teksten. Det er opp til den elektroniske utgaven[11] hvilke deler av denne tilleggsinformasjonen som blir synlig i teksten, hvilke deler som brukes til å produsere lister over f.eks. stedsnavn, hva som brukes til å generere hypertekstlenker, osv.

3.3.2 Eksempel på bruk

Dette eksempelet er hentet fra stedsnavnsregisteret i *Dokumenter angaaende flytlapperne*, og viser hvordan SGML brukes i dette delprosjektet. Innførsler fra registeret ser omtrent slik ut i den trykte boka:

> **A.**
> *Aakie-vare* I 354, 382, 387, 396 ved rigsrøs 312 mellem Finland og Koutokæino.
> *Aasfield* I 315 ovenfor Vaagsnæs V for Ulfsfjordens munding i V paa AkNO.

Her kan ulike typer informasjon merkes ut. Først legger vi inn tagger som beskriver hirarkiet og layouten i orignaldokumentet, så som <ukap> (underkapittel), <avsn> (avsnitt) og <kur> (kursiv):

[10] I dette delprosjektet brukte vi først og fremst SoftQuads RulesBuilder og Author Editor, og open source-editoren GNU-emacs.

[11] Eller de elektroniske utgavene. Man kan tenke seg flere ulike utgaver som legger vekt på ulike sider ved materialet.

```
<UKAP><UKAPTIT>A.</UKAPTIT>
<AVSN><KUR>Aakie-vare</KUR> I 354, 382, 387, 396 ved rigsrøs
312 mellem Finland og Koutokæino.</AVSN>
<AVSN><KUR>Aasfield</KUR> I 315 ovenfor Vaagsnæs V for Ulfs-
fjordens munding i V paa AkNO.</AVSN>
```

Videre legger vi inn informasjon om spesielle innholdskategorier: <snavn> (stedsnavn), <henv> (kryssreferanser til sider i samme bok), og <kilde> (referanser til andre dokumenter):

```
<UKAP><UKAPTIT>A.</UKAPTIT>
<AVSN><KUR><SNAVN>Aakie-vare</SNAVN></KUR> <HENV>I 354,
382, 387, 396</HENV> ved <SNAVN>rigsrøs
312</SNAVN> mellem <SNAVN>Finland</SNAVN> og
<SNAVN>Koutokæino</SNAVN>.</AVSN>
<AVSN><KUR><SNAVN>Aasfield</SNAVN></KUR> <HENV>I
315</HENV> ovenfor <SNAVN>Vaagsnæs</SNAVN> V for <SNAVN>Ulfs-
fjordens</SNAVN> munding i V paa
<KILDE LOPENR="5800002">AkNO</KILDE>.</AVSN>
```

Mer om de enkelte elementene følger i avsnitt 3.6.

3.4 Tekstregistrering

De trykte tekstene legges inn slik de står, uten noen slags normalisering eller feilretting. Heller ikke åpenbare trykkfeil rettes. Vi har ved registrering og korrekturlesing kun forholdt oss til de trykte versjoner av Schnitlers protokoller og *Dokumenter angaaende flytlapperne*, og ikke konferert håndskriftene som ligger til grunn for bøkene.

3.4.1 Tolkning av håndskrift — normalisering

Det foretas heller ikke noen egentlig normalisering av de håndskrevne tekstene som er digitalisert. Imidlertid er detaljer ved manuskripter ofte usikre. Er det mellomrom etter komma? Kanskje er det egentlig et åpent rom rett før kommaet? I slike tilfelle transkriberes teksten i henhold til grammatikalske regler dersom det ikke framgår tydelig at Qvigstad bryter reglene.

En bør også være klar over den normalisering som ubevisst skjer ved at man tolker håndskrifter i et språk man kjenner i retning av fornuftige utsagn.

Særlige problemer oppstår i to situasjoner: Det ene er diskusjoner av samiske dialekter i brevene, der Qvigstad bruker svært mange ulike spesialtegn (se avsnitt 3.5). Det andre er margnotatene i *Dokumenter angaaende flytlapperne*, særlig forkortelsene. I disse notatene er det f.eks. svært vanskelig å skille mellom 'o' og 'a'. 'I' og 'J' kan også være vanskelige, f.eks. som initial til Fellmann. De vanskeligste tilfellene er merket med <usikker>.

Noen steder i brevene til Magnus Olsen, for eksempel i vedleggene til brev til Magnus Olsen datert 4/2 1924, er det skrevet kommentarer med noe som ser ut til å være en annen hånd. Dette er kommentert i <usikker>.

Tallene i notatene i *Dokumenter angaaende flytlapperne* kan ha forskjellig betydning. De er innholdstagget med stor forsiktighet, siden det er lett å ta feil av hva som er hva. En videre tagging vil kreve grundigere undersøkelser av hvert enkelt notat for å sikre at det blir gjort riktig.

Numrene i notatene kan f.eks. være:

- Gårds/bruksnummer. Disse er ikke tagget opp spesielt, men man kan i ettertid går gjennom notatene og tagge dem, slik at man for eksempel kan lage lenking inn til den elektroniske Rygh.[12]

- Kartnummer. Dette er samme slags henvisninger som i registeret (se avsnitt 5.3.1) og kunne være tagget <kilde>. Disse er heller ikke tagget foreløpig.

- Sidetall. Dersom det står 's.', 'I' eller 'II' etterfulgt av et tall, er dette ansett som et sidetall og er tagget som en intern henvisning (<henv>).

3.4.2 Eksterne hjelpemidler

Vi har for brevene og margnotatene i *Dokumenter angaaende flytlapperne* benyttet visse eksterne kilder som støtte i tokningen av håndskriften:

1. Bibliografiske baser (særlig Bibsys og Libris) benyttes som støtte når personnavn og titler skal tolkes.

2. Atlas og leksika benyttes tilsvarende for stedsnavn.

[12]Rygh, Oluf: *Norge gaardnavne* er under digitalisering og vil være ferdig i løpet av 1998.

3. Forekomster ellers i dokumentet sjekkes dersom man er i tvil om en stavemåte. F. eks.: Man er i tvil om en utydelig bokstav er "e" eller "æ". Bokstaven står i ordet "prestegjeld/gjæld". Alle andre steder i brevsamlingen staves ordet med e, dermed går vi ut fra at det er "e" det står.

   Dersom det derimot hadde stått en tydelig "æ" ville ikke denne blitt skriftet ut med en "e" uansett hva som brukes i dette ordet ellers i brevsamlingen.

3.4.3 Plassering av håndskrevne notater

De håndskrevne notatene til *Dokumenter angaaende flytlapperne* er plassert på ulike steder rundt den trykte teksten, i mange tilfeller er det intet entydig sted å putte dem inn i teksten. I en transkribering av teksten er det ikke praktisk mulig å gjenspeile notatenes reelle plassering i skriftbildet. Dette innebærer at vi har valgt et punkt for hvert notat.[13] Dette gjøres etter følgende kriterier:

1. Dersom det ved hjelp av understrekninger eller logisk sammenheng kan knyttes til et bestemt sted i teksten plasseres det der.

2. Hvis ikke, plasseres notatet først eller sist i det nærmeste avsnitt.

Ved plassering av håndskriftene legges stor vekt på understrekninger i teksten, da notatene ofte refererer til understrøkne stedsnavn. I tillegg vurderes navnelikhet dersom notatet er et stedsnavn, og også den trykte "Identificering av stedsnavne" er brukt i noen grad.

Vertikale streker er også tagget. I visse tilfelle ville det være fornuftig å lage en kobling mellom vertikale streker og notater, som kunne brukes f.eks. på 'NB'ene på side 72 og 73.

Der det finnes et notat til et stedsnavn er det ikke konsekvent hvorvidt <gblyant> står slik:

```
stedsnavn<gblyant>...</>,
```

[13]Det er teknisk mulig å koble et notat til et området i stedet for et eksakt punkt, men det ville heller ikke si noe om plasseringen i tekstbildet. En løsning kunne være å oppgi en eller annen form for koordinater i forhold til teksten, men dette ville bli altfor komplisert. Den beste løsningen vil være å digitalisere faksimiler av boksidene, jfr. avsnitt 3.7.

eller slik:

```
stedsnavn,<gblyant>...</>
```

Dette er heller ikke konsekvent ved anførselstegn, punktum eller andre skilletegn.

I ett tilfelle (side 202) henvises det til samme notat flere steder i teksten. Det lages da én kopi av notatet for hvert sted det skal inn.

3.4.4 Forkortelser

Notatene i *Dokumenter angaaende flytlapperne* preges av forkortelser. Dette er naturlig, da de var ment til eget bruk. Noen av forkortelsene finner man igjen i "Identificering af stedsnavne" (særlig kartreferansene). Her følger en oversikt over noen forkortelser, med antydning av mulig betydning:

Bls (Balsfjorden?), Fl (Finland?), fn. (Finland?), gl. Mno (Gammelt matrikkelnummer?), Gno (Gårdsnummer?), Handl. (Handlingar...?), Ib (Ibestad?), Kfj. (Kåfjord?), Kout. (Kautokeino?), Krl. (Karlsø?), Kt (Kautokeino?), Köng., Ly (Lyngen?), Mno (Matrikkelnummer?), Qv (Qvigstad?), Sfj.

Det kunne være nyttig med en tagging av alle forkortelser med antatt betydning i et attributt, men dette er foreløpig ikke gjort.

3.5 Tegnsett og entiteter

Tegnsettet som er brukt til innskriving av alt det historiske materialet er ISO 8859-1, også kalt ISO latin 1. Dette er i praksis identisk med de tegnsett som benyttes i UNIX og i Microsoft Windows/Windows 95. Plasseringen av de tegn som benyttes i engelsk (a-z og A-Z) er sammenfallende med DOS og MAC-tegnsettene, mens tegn fra andre skriftspråk (f.eks. æ, ø, å, á, ô) er plassert i andre posisjoner i DOS og MAC-tegnsettene. Bokstaver som ikke benyttes i vesteuropeiske språk finnes ikke i tegnsettet. For vårt prosjekt betyr det at vi mangler følgende grupper av tegn:

1. Greske bokstaver (*Dokumenter angaaende flytlapperne*)
2. Grafiske tegn for måleenheter som pund, mark, skilling, etc. (*Dokumenter angaaende flytlapperne* og Schnitlers protokoller)

3. Samiske bokstaver og lydskrifttegn (registrene i *Dokumenter angaaende flytlapperne*, Qvigstads brev og notatene til *Dokumenter angaaende flytlapperne*)

Alle slike tegn registreres som uttrykk på formen '&navn;'. For eksempel 'č' for c med caron, '&agr;' for gresk alfa. En oversikt over disse tegnene finnes bakerst i databeskrivelsen, se tillegg C.1. Der er det også gjort en kobling mot UNICODE, som er den kommende standarden for tegnsett som dekker de aller fleste språk som er i bruk i dag.

For kategori 3 er det et tilleggsproblem: Vi har ikke vært i stand til å oppdrive en samlet oversikt over Qvigstads lydskriftsystem.[14] Det har derfor vært vanskelig å gruppere tegnene. Her følger eksempler på tegn som har voldt problemer:

caron eller breve Vi har ikke klart å skille mellom disse to variantene. Over vokaler setter vi breve, over konsonanter caron, med andre ord: č, ikke c̆; ĕ, ikke ě.

c eller e Et problem er noen steder å skille c fra e, da Qvigstad av og til skriver c'en med en lite krøll slik at den kan ligne en e. De vanskeligste tilfellene er merket med <usikker>.

n′ kunne også vært fortolket som ń.

\` ′ ´ skilles ikke mellom som frittstående tegn. Alle kodes som ′.

— Det skilles ikke mellom tankestrek og "lang bindestrek", dvs ikke mellom '—' og '–', begge skrives '—'. '-', derimot, gjengis som seg selv.

Brøker skrives som 'tall/tall', dvs. at '$\frac{3}{8}$' skrives inn som '3/8'.

@ Tegnet '@' brukes der håndskrevet tekst ikke er leselig.

Det er litt ulogisk at noen av tegnene fra ISO 8859-1 er gjengitt som 8-bits-koder, f.eks. 'ö', 'è', mens andre gjengis som entiteter, f.eks. 'â' ('â').

[14]Han bygger et stykke på vei på G. Storm, men ikke fullt ut, og vi har ikke funnet igjen alle tegnene Qvigstad bruker i noen av oversiktene over Storms lydskriftsystem, se f.eks. Joh. Storm: *Kortere Ordliste med Forklaring af Lydskriften* og Oddvar Nes: *Storms norsk lydskriftsystem (med tillegg)*. Qvigstad hadde også behov for et system forskjellig fra Storms, da han beskrev andre språk enn Storm.

3.6 Tagging

De historiske dokumentene og brevene er tagget i henhold til en spesialskrevet databeskrivelse, DOKUB.DTD (se tillegg C.1). Taggene som er benyttet kan deles inn i tre kategorier, selv om de glir noe over i hverandre:

1. strukturelle
2. grafiske
3. innholdsmessige

Her følger en gjennomgang av de viktigste elementene fordelt på disse tre gruppene.

3.6.1 Strukturelle tagger

Disse taggene merker opp kapittelstrukturen i bøker og strukturerer brevsamlingene. I tillegg markerer de avsnitt, tabeller, osv.

Figur 3.1 viser hovedstrukturen i dokumentene. For bøker inneholder dokumentet ei bok. Denne starter med ei tittelside og en innholdsfortegnelse. Deretter kommer kapitler på ett til tre nivåer (kap, mkap (ikke alltid), ukap (ikke alltid)), mens selve teksten finnes i avsnitt og tabeller.

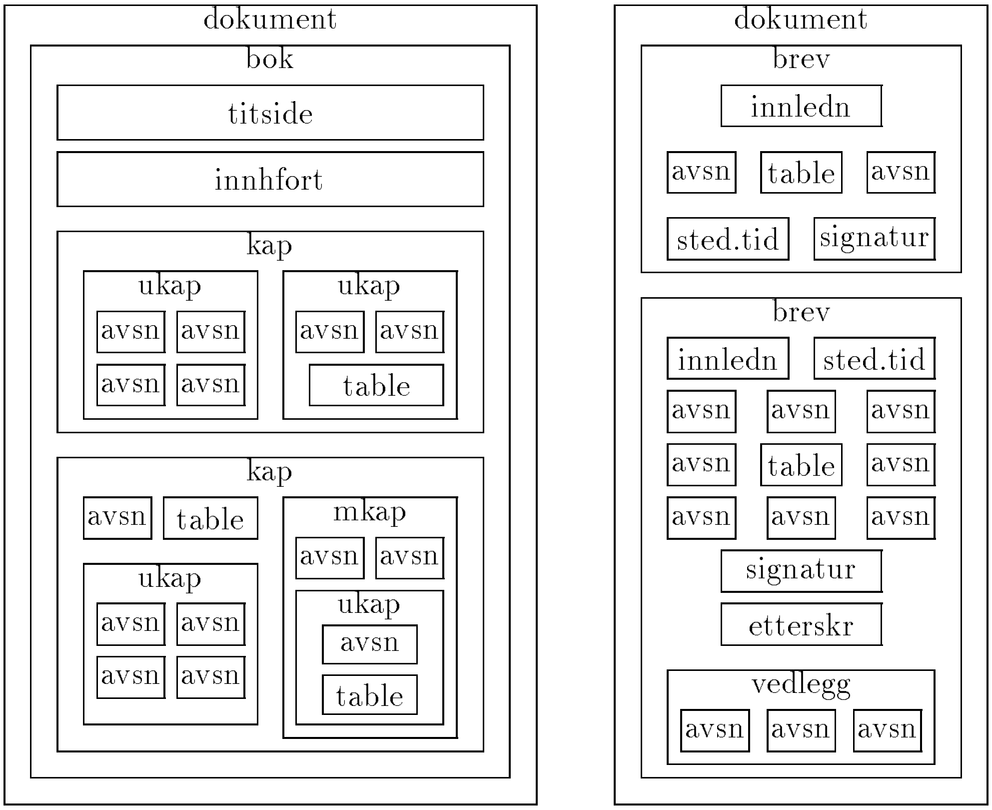
For brev inneholder dokumentet et antall brev. Disse starter normalt med en innledning, deretter kommer oftest dateringen (sted.tid). Etter dette følger teksten i en avsnitt og evt. tabeller, deretter kommer i noen tilfelle dateringen og vanligvis signatur, og i visse tilfelle etterskrift og ett eller flere vedlegg.

brev

Hver <brev>-tag inneholder ett brev. Det finnes to attributter, **avs** og **mott**, hvor henholdsvis avsender og mottaker av brevet legges.

innhfort: innholdsfortegnelsen

I *Dokumenter angaaende flytlapperne* er innholdsfortegnelsen en enkel tabell. I Schnitlers protokoller er dette ikke så enkelt. I bind 1 av Schnitlers protokoller la man ved utgivelsen i trykt form vekt på å ikke legge noe til teksten. Dette innebar at det ikke ble satt inn overskrifter som ikke sto i det



Figur 3.1: Strukturen i dokumenter. Bøker til venste, brevsamlinger til høyre.

håndskrevne manuskriptet. Imidlertid lagde man en fyldig innholdsfortegnelse som strukturerte innholdet.

I den digitale utgaven ble disse overskriftene flyttet ut til sin rette plass i teksten, men på en måte som viser at dette er informasjon som er lagt til teksten: som attributter. Starten på et underkapittel i *Dokumenter angaaende flytlapperne* kan se slik ut:

```
<ukap><ukaptit>13. juni. I Tromsen</ukaptit>
<avsn>Denne dagen...
```

En tilsvarende start på et underkapittel i Schnitlers protokoller vil se slik ut:

```
<ukap overskr="13. juni. I Tromsen">
<avsn>Denne dagen...
```

Legg merke til at i det siste eksempelet står overskriften mellom '<' og '>' og dermed er en del av meta-informasjonen, ikke av selve teksten, mens i det første eksempelet er overskriften en del av selve teksten. Dette gjenspeiler altså en forskjell i den trykte utgaven: I det første eksempelet er det snakk om en overskrift som sto i den trykte teksten, og som derfor er lagt inn som vanlig tekst mellom <ukaptit> og </ukaptit>. I det andre tilfellet en overskrift som ikke sto i selve teksten, men som var hentet fra innholdsfortegnelsen først i boka, og som er lagt inn som attributt.

Da dette opplegget ble laget, fant vi det mest naturlig å ikke legge inn innholdsfortegnelsen samlet når den var spredt rundt i teksten på denne måten. Imidlertid er det intet problem å gjøre dette i ettertid; det er bare å samle alle **overskr**-attributtene. Innholdsfortegnelsen som er trykt i første bind av Schnitlers protokoller dekker imidlertid både første og andre bind, mens kun første bind er digitalisert.

avsn: avsnitt

Utgaven skiller ikke mellom ulike typer avsnitt, eller om avsnittene skilles fra hverandre med mellomrom eller innrykk. Dette innebærer at en eksakt gjengivelse av de trykte bøkenes lay-out ikke kan gjenskapes ut fra den digitale versjonen alene.

Merkingen av avsnitt er problematisk i brevene; Qvigstad lager ikke tydelige skiller. Avsnittsskillene er derfor heftet med en viss usikkerhet.

Når han skifter tema i brevene, fortsetter han gjerne på samme linje etter en tankestrek. Dette markeres ikke spesielt i den digitale versjonen.

table: tabeller

Dette elementet inneholder et antall underelementer med sine attributter. Dette er en forenklet versjon av table-systemet i HTML 3.2. Denne standarden finnes på World Wide Web Consortiums hjemmesider. Tabellene, som er lagt inn på denne måten, vil vises korrekt i vevlesere som Netscape, Opera og Internet Explorer.

I tillegg til typiske tabeller, er også <table> brukt for å vise oppsett i den trykte boka som det er vanskelig å fremstille på annen måte, for eksempel flere linjer som slås sammen med klammer.

side: sideskift og snr: sidetall

Sideskift legges inn øverst på alle sider, både i bøker og brev. Det finnes også et attributt som brukes ved faksimiler, se avsnitt 3.7.

Sidetall settes inn på alle sider i de trykte bøkene, også de sider der det ikke er eksplisitt angitt i originalen. I denne taggen legges også evt. topptekst inn i et eget attributt. I brevene legges sidetall kun inn når de originale brevsidene er påført sidetall.

Om årsakene til at det tagges side*skift* og ikke sider, og om forskjellen mellom disse to mulighetene, se avsnitt 3.6.4.

overskr: overskrifter

Denne taggen benyttes til overskrifter som ikke tagges i overskriftshirarkiets tre nivåer. For *Dokumenter angaaende flytlapperne* er dette avsnittsoverskrifter på lavere nivå enn underkapitteloverskrifter, i den trykte boka er disse ikke satt med halvfeit, men gjerne i kursiv og/eller sperret skrift.

I Schnitlers protokoller vil alle overskrifter stå i denne taggen, fordi de strukturelle overskrifttaggene (ktittel, mkaptit, ukaptit) ikke brukes — overskriftene legges jo i attributt, jfr. avsnittet om innholdsfortegnelsen ovenfor.

Denne taggen brukes også på eventuelle overskrifter i brevene.

fotnoter

Fotnotene settes inn i teksten der henvisningsmerket står. Når samme fotnote henvises til flere steder, lages det flere kopier av fotnota.

3.6.2 Grafiske tagger

Disse taggene merker ut tekst med spesielle kjennetegn.

Stiltyper på trykt tekst

kur Kursivert tekst

hfeit Halvfeit tekst

mstilt Midtstilt tekst

hstilt Høyrejustert tekst

sperring Sperret tekst, dvs. tekst med øket mellomrom mellom bokstavene

hevet Hevet tekst (tekst som et satt høyere enn normallinja)

senket Senket tekst (tekst som et satt lavere enn normallinja)

Understrekinger

gstrek Tekst som er understreket med grå blyant

rstrek Tekst som er understreket med rød blyant

bstrek Tekst som er understreket med blå blyant

blekkstr Tekst som er understreket med blått eller svart blekk

I brevene er all understreket tekst tagget <blekkstr>. Vi har ikke kontrollert om all understreking faktisk er foretatt med blekk.

I *Dokumenter angaaende flytlapperne* er alle fire taggene brukt, slik at man skiller mellom de ulike typene. Det skilles imidlertid ikke mellom svart og blått blekk.

Taggingen av understrekningene brytes ikke av tegn, kun av bokstaver. Dette betyr at

```
``Nordland, Troms og Finnmark''
```

tagges

```
``<gstrek>Nordland, Troms</gstrek> og
<gstrek>Finnmark</gstrek>''
```

Understrekinger brytes når håndskrevne notater kommer inn (dersom disse ikke er understreket), og ved '—'. De brytes ikke ved sideskille, ved ',', '.', ':', ';', '?', anførselstegn o.l. De brytes heller ikke når en enkelt bokstav korrigeres med overstr/erstatt (jfr. *Dokumenter angaaende flytlapperne* s. 153, linje 7).

Når noe er understreket med både grå blyant og en annen type strek tagges ikke <gstrek>. Ved kombinasjon av to andre typer understreking tagges begge.

Vertikale streker

Disse strekene er i margen på boka.

gvert Tekst som det er en vertikal grå blyantstrek ved siden av

rvert Tekst som det er en vertikal rød blyantstrek ved siden av

bvert Tekst som det er en vertikal blå blyantstrek ved siden av

blekkvrt Tekst som det er en vertikal strek av blått eller svart blekk ved siden av

Håndskrevne notater

Disse taggene markerer håndskrevne notater i *Dokumenter angaaende flytlapperne*:

gblyant Notater skrevet med grå blyant

rblyant Notater skrevet med rød blyant

bblyant Notater skrevet med blå blyant

blekk Notater skrevet med blått eller svart blekk

Hva plasseringen av håndskrevne notater angår, se avsnitt 3.4.3.

Den mest naturlige måten å vise de håndskrevne notatene i *Dokumenter angaaende flytlapperne* på er å presentere dem som en slags “fotnoter”. Dette gjør at mange interne henvisninger presenteres på en måte som ikke er særlig opplysende. Eksempler på problemene kan sees på side 334–6: Her er det en hovedhenvisninger til s. 37–39, mens enkeltlinjer henviser til de numre som er påført med blyant. Øverst på side 336 står henvisningen “S. 37 ff.” Den første posten i tabellen er “Anders Andersen Matt”. Foran dette navnet står tallet 16. På side 38 finner vi tallet igjen, foran en tabellinnførsel på “Anders Anderson Matti, äldre”. Det kunne være ønskelig å kode slike håndskrevne notater som kryssreferanser, men dette krever en type tagging som ennå ikke er gjort.

Korreksjoner

Den siste gruppen grafiske elementer brukes både i *Dokumenter angaaende flytlapperne* og i brevene. Det er et system for avmerking av korreksjoner.

Tekst som er strøket over tagges <overstr>. Dersom noe er satt inn i stedet for denne teksten, settes dette inn før sluttaggen, plassert i en egen tag <erstatt>. Dette kan f.eks. se slik ut:

```
Ved <overstr>viken<erstatt>odden</erstatt></overstr>
så jeg at...
```

I *Dokumenter angaaende flytlapperne* er det også markert i attributtet **type** hvilken type overstrykning det er snakk om. Typene som er brukt er visket (2 ganger), gblyant (61 ganger) og blekk (14 ganger) og betyr at den overstrøkne teksten er henholdsvis visket ut, strøket over med grå blyant og strøket over med blekk.

3.6.3 Innholdstagger

Disse taggene merker ut utvalgte innholdskategorier i teksten.

dato

Taggen brukes til å merke ut datoer. En sorteringsform av datoen legges i attributtet **standard**:

```
Det var ikke før <dato standard=1865-02-24>24. februar
1865</dato> at vi...
```

Dersom dato og/eller måned ikke er spesifisert, brukes “00”.

For at noe skal tagges som dato, kreves det en viss grad av presisering. “Oktober” vil kunne tagges i en kontekst der årstallet er åpenbart, mens begrep som “til vinteren” og “til jul” ikke tagges. Tidspunkt for årlige hendelser, f.eks. angivelse av når kalving normalt skjer, tagges ikke. Vage henvisninger til større tidsspenn, f.eks. “det attende aarhundrede” tagges heller ikke, mens “1700–1800” tagges. Ved slike spenn tagges hele spennet, mens første dato legges i attributtet:

```
Dette foregikk i <dato standard=1865-00-00>1865-74</dato>.
```

Mange tidspunkt som i den trykte teksten i *Dokumenter angaaende flytlapperne* er oppgitt etter kirkeåret, har et notat der Qvigstad har skrevet inn datoen for aktuell høytid. Selv om disse høytidene av og til tidfester bestemte hendelser, settes aldri <dato> rundt datoene han har skrevet inn — de regnes som generelle tidfestinger av en kirkelig begivenhet som gjentas hvert år.

henv og kilde: henvisninger

En vitenskapelig tekst inneholder stadige henvisninger til andre tekster. Det har vært et mål for delprosjektet å prøve ut ulike måter å hjelpe leserne å aktivisere slike henvisninger, typisk i form av hypertekstlenker.[15] Grunntrekkene i arbeidet slik det framstår i de versjonene som nå foreligger er følgende:

- Alle konkrete henvisninger intert i samme dokument er tagget <henv>. Konkret betyr i denne sammenheng at sidetall er oppgitt. Det mest markante eksempel er registrene i *Dokumenter angaaende flytlapperne*, der hvert sidetall eller hver serie av sidetall er tagget.
- Et utvalg av henvisninger fra *Dokumenter angaaende flytlapperne*, Schnitlers protokoller og brevene til andre skrifter er tagget <kilde>. Dette dreier seg om skrifter der vi har kunnet lokalisere en forekomst i elektroniske form eller en referanse til papirversjoner i Bibsys. Dessuten har vi tatt med enkelte andre viktige skrifter, for eksempel matrikler. Enhver henvisning til matrikler peker til en felles referanse, der man i publiseringen bør lage en peker videre til generell informasjon om de ulike matriklene.

<kilde> har vært en konseptuelt vanskelig tag. Lenge forsøkte vi å være konsekvente i bruken slik at alle dokumenter av visse typer ble merket ut. Dette viste seg å være et svært vanskelig arbeid. Når Qvigstad nevner tittelen på verk i brev beregnet på én inneforstått leser er gjerne henvisningen så knapp at det er umulig å finne ut hvilket skrift den referer til.

Til slutt kom vi fram til at vi må se på praktisk nytte av oppmerkingen. Dette innebærer at kun utvalgte referanser som er mulig å identifisere entydig merkes som kilder.

[15] Dette er diskutert grundigere i Øyvind Eide: "Re-integrasjon av kildemateriale : nye dokumentformer på tvers av tradisjonelle skiller" og Christian-Emil Ore: "Making Multidiciplinary Resources".

<kilde> brukes slik i de ulike tekstene:

1. *Dokumenter angaaende flytlapperne*: <kilde> er tagget for 12 ulike verk med til sammen 2398 referanser. Dette er først og fremst henvisninger til kart i registrene.

2. Schnitlers protokoller: <kilde> er tagget for 5 ulike kilder med til sammen seks referanser. Dette er matrikler og kartene fra forordet til registrene (som er den eneste delen av registrene som er digitalisert).

3. Brevene: <kilde> er tagget for 299 ulike kilder med 930 henvisninger. Dette er en del av de trykte skrifter Qvigstad omtaler. Ingen utrykte arkivmanuskripter er tagget <kilde>.

I kildetaggen finnes et attributt **lopenr** som inneholder et sjusifret nummer. Dette nummeret finnes igjen i filen `kildene.sgm`, som er en SGML-fil som har sin egen databeskrivelse (se tillegg C.2). I tillegg til **lopenr**-nummeret inneholder hver post kortfattede bibliografiske opplysninger og en peker til ytre kilder — de aller fleste inneholder ett eller flere objektid som kan finnes igjen i Bibsys. Dette kan presenteres som hypertekstpekere til Bibsys-postene.

I den elektroniske utgaven bør ikke kildene representeres ved direkte hyperteksthopp, man bør bruke et mellomtrinn der en slags fotnote kommer opp, slik at leseren ikke mister oversikten over hvilket dokument hun befinner seg i.

Oversikten over kildene ble utarbeidet i MikroMARC og siden konvertert til SGML. Et avsnitt om slik konvertering finnes i tillegg E.

I Schnitlers protokoller ser det normalt ut til at henvisningene er til sider i den trykte boka, men vi har ikke funnet noen presisering av at henvisninger ikke kan være til sider i de håndskrevne protokollene. Dette bør kontrolleres.

Bruken av begge disse henvisningstaggene kan utvides i ettertid dersom dette er ønskelig. <kilde> kan f.eks. settes inn på flere skrifter som kan lokaliseres i andre kataloger enn Bibsys, mens <henv> f.eks. kan brukes til henvisninger som ikke er presisert ved sidetall.

Kildetaggen avmerker hvilket verk som er omtalt, mens evt. sidetall normalt holdes utenfor. Et unntak fra dette er taggingen av henvisninger fra brev til Magnus Olsen til *Norske gaardnavne*. Her omtales arbeidet med Finnmark-bindet av *Norske gaardnavne* på avsnittsnivå, med kommentarer som ikke finnes i den trykte boka. Dette åpner for at det kan lages toveis hypertekstlenker på avsnittsnivå.

pnavn: personnavn

Her legges alle navn på personer som finnes i tekstene. Flere personer med felles etternavn tagges sammen dersom de ulike navnene kun skilles med komma og ord som "og", "och" og "eller". Dersom navnene skilles med flere eller andre ord, skilles navnene i separate tagger:

```
Dette dreier seg om brødrene <pnavn>Per</pnavn> eller,
som han også kalles, <pnavn>Peder, Hans och Gunnar Anders-
son</pnavn>
```
[16]

I en liste over personnavn bør man liste opp navnene sortert på alle ord i navnet, slik at for eksempel "Peder Hansen Matt" kommer i lista både under 'P', 'H' og 'M'.

Dersom en enke er oppgitt med eget navn tagges:

```
<pnavn>Peder Hansens</pnavn> enke <pnavn>Josefine Joens-
datter</pnavn>
```

men dersom navnet hennes ikke er er oppgitt, tagges

```
<pnavn>Peder Hansens enke</pnavn>.
```

I brevsamlingene merkes ikke navnet på avsender og mottager ut i selve brevteksten, de legges kun inn som attributter i <brev>-taggen.

For at en person skal merkes ut med <pnavn> kreves at enten fornavn eller etternavn står i teksten — "Deres svigerfar", "din hustru" o.l. tagges ikke.

Tidligere ble en sorteringsform av navnet lagt i attributtet **sorter** i formen *etternavn, fornavn*. Dette viste seg å være lite fornuftig, fordi begrepet "etternavn" ikke kan defineres på en enkel og entydig måte. Disse **sorter**-attributtene er derfor fjernet. Skulle man gjort denne type inndeling, måtte man ihvertfall skilt mellom fornavn, farsnavn og gårdsnavn, noe som ville blitt for komplisert for dette delprosjektet.

[16] Dersom man i tilfelle som dette ønsker å koble de to <pnavn>-elementene sammen, kan dette gjøres ved hjelp av attributter. En nærmere beskrivelse av slik sammenkobling finnes i *TEI guidelines*, kapittel 31 ("Multiple Hirarchies").

snavn: stedsnavn

Denne taggen brukes til å merke ut alle stedsnavn. Å definere hva som er stedsnavn og hva som ikke er det er imidlertid ikke så lett for det materialet som er konvertert i dette delprosjektet. Særlig skaper *Dokumenter angaaende flytlapperne*, med dokumenter som er skrevet over en periode på 350 år, problemer. Vi har ikke klart å finne enkle regler, men har måttet basere oss på skjønn. Nedenfor følger noen av de vurderingene som er gjort.

Først lister vi opp noen av de navnene vi har vært i tvil om skal tagges og viser hvilke av dem som har blitt tagget <snavn>:

tagget stedsnavn Iishavet, Indiager, Ishafvet, Lappland, Nordhafwet, Nordsjön, Nordsjö-kusten, Norra Oceanen, Norrwästra Byen, Södra og Norra Luledahlen, Wästerhafwet.

ikke tagget stedsnavn fiellryggen, Kiöhlen, Lands-kiöhlen, Landz-råen, Lappmarken, Norrskakusten, Norrska sidan, Rans-Lappar, Riksgränsen, Rågången, Sjökusten, Sjöstranden, Swenska Prætensions Linien, Wapstby-lappar, Wäster hafsfiordarne.

Ord som står etter stedsnavn tagges dersom det en snakk om en presisering: kirke, sogn, amt, gård osv. tagges. Naturformer tagges der det er en naturlig del av stedsnavnet. Etterledd som beskrive landskapsform tagges:

```
<snavn>Altens Fiord</snavn>
```

Ledd som betegner plassering i dette landskapet tagges ikke:

```
<snavn>Altens Fiords</snavn> Botten
```

Rigsrøs tagges der nummeret står rett etter ("Tårestj. 221"), men ikke der det står "ved" ("Lille Umevtn. ved 221"). Riksrøsenumre som står uten annet stedsnavn tagges <snavn> når det står "rigsrøs" foran.

</snavn> settes ikke inn slik at ord deles, heller ikke ved bindestrek.

Genitivformer av stedsnavn tagges, mens adjektiv laget av stedsnavn ikke tagges. Appellativer som "fjällryggen" o.l. tagges ikke.

Visse sammensetninger med stedsnavn utelates, for eksempel tagges ikke personer; "Lyckselelapperne" tagges ikke <snavn>. Heller ikke "Norske" i "Norske bygde- eller fjeldlapper" tagges <snavn>. Men når stedet er løsrevet fra personenene, tagges <snavn>:

```
<snavn>Beierns</snavn> og <snavn>Saltdalens</snavn>
lapper
```

Avgrensende ord foran tas kun med i den grad de er en del av stedsnavnet.

```
Søndre <snavn>Salten</snavn>
<snavn>Vestre Rostojaure</snavn>
```

Flere stedsnavn som har felles for- etter etterledd legges i samme <snavn> dersom delene kun skilles med ordene og tegnene: 'och', 'og', ',', 'eller', '(', ')', 'med', 'samt':

```
<snavn>Opp-, Ned- og Bortdalen</snavn>
```

Matrikkelens gårdsnummer o.l. er ikke tagget <snavn>.

Gårdsnavn vil være tagget enten som en del av personnavnet i <pnavn> eller som <snavn>. Ved søking/browsing av stedsnavn er det verdt å gjøre oppmerksom på at navn på gårder også kan forekomme som "etternavn" på personer.

samisk: setninger som inneholder samisk språk

De setninger eller avsnitt i brevene som inneholder samisk språk tagges <samisk>. Dette er gjort ut fra praktiske hensyn, men er ikke helt uproblematisk — de delene av teksten som inneholder norsk språk tagges jo ikke <norsk>. Andre språk, som latin og finsk, tagges heller ikke. Dersom man videre har noe nytte av slikt tagging, enten til å skille ut delene som skaper størst korrekturproblemer, eller til å forenkle tekstsøking for de som er interessert i Qvigstads språkdiskusjoner, kan man jo bruke den. Taggen kan eventuelt fjernes fra materialet.

usikker: problematiske deler av teksten

Spesielt problematiske deler av teksten er merket med <usikker>. En kort beskrivelse av problemet ligger i fritekst i attributtet **komm**.

Begreper som ikke er tagget

Det er flere informasjonskategorier i tekstene som kunne vært merket ut, f.eks. flere typer stedsnavn, personers alder, pengebeløp og andre mål. Men det måtte gjøres et skjønnsomt utvalg for at ikke arbeidet skulle forsvinne i tagger, et kompromiss mellom muligheten til strukturering og den praktiske registreringen av ei bok.

Vi håper de erfaringen man gjør når disse dokumentene kommer i bruk kan bidra til at man ved senere elektroniske utgaver kan velge de kategorier for tagging som det er mest bruk for.

3.6.4 Kryssende tagger

Dette avsnittet er delvis basert på pågående forskning som utføres av C.M. Sperley-McQueen, en av redaktørene for *TEI guidelines*, og Claus Huitfeldt ved Wittgensteinprosjektet ved Universitetet i Bergen.

En av de fundamentale egenskapene ved SGML er at kryssende tagger ikke er tillatt. Å forklare årsakene til dette blir for omfattende for denne rapporten, men det har sammenheng med at en struktur er langt enklere å behandle maskinelt dersom man kan gå ut fra at den er uten kryssende tagger.[17]

Behov for en kryssende tag kan oppstå i følgende tilfelle:

`Vitnet var Per Hansen` *Tomasjord i Altens prestegjeld*

I dette uttrykket ønsker vi å tagge personens navn inkludert gårdsnavnet, men også tagge det kursiverte området. Det vi ønsker oss er altså

```
Vitnet var <snavn>Per Hansen <kur>Tomasjord</snavn> i Altens
prestegjeld</kur>
```
(andre tagger er utelatt i dette eksempelet)

Dette representerer imidlertid kryssende tagger, fordi <snavn>-elementet starter utenfor, men avsluttes inni <kur>-elementet.

Vi har løst dette problemet ved hjelp av såkalt fragmentering, hvor grafiske elementer splittes opp, mens innholdselementer (snavn, pnavn, kilde, osv.) ikke splittes opp. Løsningen blir slik:

[17] Både fordelene og ulempene med SGML har sammenheng med at SGML-applikasjoner kan uttrykkes som kontekstfrie grammatikker, jfr. Noam Chmoskys språkhirarki. En god gjennomgang av ulike grammatikker i Chomskys hirarki finnes i Grune & Jacobs: *Parsing techniques*.

```
Vitnet var <snavn>Per Hansen <kur>Tomasjord</kur></snavn>
<kur>i Altens prestegjeld</kur>
```

Lignende problemer med understrekinger og vertikale streker ved siden av teksten løses på samme måte. At innholdselementer aldri fragmenteres, fører til at spørsmål som "Hvor mange personnavn finnes i dokumentet" kan besvares.[18] Imidlertid vil spørsmål av typen "Hvor mange blå undertrekinger finnes i dokumentet" ikke kunne besvares meningsfullt på grunnlag av den elektroniske teksten. Derimot kan spørsmål som "Hvor mange ord i teksten er understreket med blått" besvares, dersom begrepet "ord" defineres entydig.[19]

Et annet overlappingsproblem er forholdet mellom sider og elementer som avsnitt, stedsnavn o.l. Problemet oppstår når et avsnitt startes på en side og avsluttes på neste. Vi kunne ønsket oss en oppmerking som denne:

```
<side>
<avsn>...
...
</side>
<side>
...
</avsn>
<avsn>...
...
</side>...
```

Som man ser, oppstår samme problem som ovenfor. I dette tilfellet løses problemet ved hjelp av milepeler (eng. milestones). I stedet for å merke hver side som noe som starter øverst og slutter nederst på sida, er det sideskiftene som merkes ut, på denne måten:

```
<side>
<avsn>...
...
<side>
```

[18]Selv om dette krever litt arbeid, fordi flere personer kan være tagget sammen i ett element.

[19]Det er mulig å gjøre en sammenkobling av slike fragmenterte elementer ved hjelp av attributter, men dette er ikke gjort i vårt delprosjekt. En nærmere beskrivelse av slik sammenkobling finnes i *TEI guidelines*, kapittel 31 ("Multiple Hirarchies").

```
...
</avsn>
<avsn>...
...
```

Man går da implisitt ut fra at en side varer til neste side starter.

3.6.5 Automatisk tagging

Taggingen av *Dokumenter angaaende flytlapperne* er gjort manuelt av innskriverne for side 1–432. Imidlertid er resten av boka, som består av registre, av en slik art at det ikke var nødvendig å gjøre taggingen manuelt. Innskriverne registrerte i stedet materialet i Word-format med linjeskift, avsnittsskiller og kursiv lagt inn som Word-koder. Tekstkorrektur ble lest i Word-format. Deretter eksporterte vi materialet til RTF-format og skrev et sett programmer som konverterte kodene til våre SGML-tagger, og som la inn taggene for stedsnavn, henvisninger og kilder (karthenvisninger). For nærmere beskrivelse av prosessen, se tillegg F.

Etter at programmene var kjørt, ble det funnet ca. 15 logiske feil og en rekke ting som måtte rettes etter manuell korrekturlesning. Mye av dette var ting det ikke er mulig å autotagge, andre ting kunne vært gjort automatisk, men ble heller overlatt til den manuelle korrekturen — det var tross alt under hundre sider som ble autotagget.

3.7 Faksimiler

3.7.1 Brevsamlingene

Under arbeidet med brevene ble det klart at det var utilfredsstillende å offentliggjøre kun en transkripsjon av teksten. Dette skyldes flere faktorer:

- Håndskrift er i seg selv et usikkerhetsmoment.

- Brevene ble skrevet inn på grunnlag av utskrifter fra mikrofilm. Vi har heller ikke hatt tilgang til originalene under korrekturlesning, da disse befinner seg i Uppsala, Helsingfors og Oslo.

- I brevene forekommer stadige diskusjoner av samiske dialektformer. I slikt materiale er det særlig viktig at man ikke legger inn feil i materialet. Hverken fagkonsulenten eller delprosjektansvarlige hadde tilstrekkelig kunnskap i samisk til å vurdere hvilke former som er korrekte der håndskriften er utydelig.

Da vi opprinnelig ikke hadde planlagt å ta med faksimiler, ble det nødvendig å finne en løsning som var så billig og enkel som mulig. Vi konkluderte derfor med at det beste var å gå via mikrofilm. Vi sendte samtlige brev på mikrofilm til Nasjonalbibliotekets avdeling i Rana for scanning. Vi fikk en CD-plate med innscannede bilder i TIFF-format; hvert bilde inneholder 1–4 brevsider. Deretter klipte vi ut en og en brevside, lagret disse på hver sin fil, og la filnavnene inn i den transkriberte teksten. Hver ny side i teksten blir således markert på denne måten:

```
<side faks=wik023.tif>
```

der "wik023.tif" er navnet på fila som inneholder bildet av denne sida.[20]

Kvaliteten på faksimilene er lav. De er scannet fra mikrofilm på en 1-bitsscanner.[21] Det ville vært en fordel, særlig der det er overstrykinger eller der det finnes blekkflekker, å ha scannet materialet i farger.[22] Imidlertid ville dette ført til en dyrere scanning, og billedstørrelsen ville økt betraktelig.

Vi håper tilbakemeldinger fra brukerne vil gi oss indikasjoner på om det er fornuftig å satse på faksimiler av så lav kvalitet eller om man heller bør gå inn for kun å lage faksimiler når man har ressurser til å gjøre det i god kvalitet.

Hva er med?

All teksten som er transkribert skal finnes på faksimilene. I tillegg er baksiden av noen av postkortene og brevkortene med — vi har tatt med faksimiler av

[20]Denne måten å gjøre det på er ikke helt god. Selve filnavene bør helst ikke ligge ute i selve dokumentet, det er bedre å bruke attributter, og ha en samlet oversikt over hvilke attributter som referer til hvilke filer. Særlig hadde det vært en fordel dersom det blir aktuelt å skifte ut faksimilene med bedre scanninger.

[21]En slik scanner operer kun med to nyanser, svart og hvitt. Den gir derved dårligere gjengivelse av gråtoner enn svart/hvitt-fotografi.

[22]Dette anbefales også i litteraturen, se f.eks. Peter Robinson: *The Digitization of Primary Textual Sources*, side 28–29.

alt som ble mikrofilmet, men de ulike institusjonene hadde ulik praksis. Det mangler enkelte sider og deler av sider, dette er merket med <usikker>-taggen.

3.7.2 Schnitlers protokoller

Det kunne ha vært argumentert for å scanne faksimile av den trykte versjonen av Schnitlers protokoller for bedre å vise oppsettet der taggingen ikke er spesifikk nok (justeringer ut over mstilt/hjust, asynkrone spalter, etc.) Imidlertid vil det være for mye jobb i forhold til gevinsten.

Derimot kunne det være aktuelt å digitalisere håndskriftene i Riksarkivet. Dette er er nokså stor jobb, og i første omgang bør man kanskje konsentrere seg om de deler av teksten som foreløpig ikke er publisert: kartene. I Riksarkivet finnes rundt 20 kart som er sentrale for forståelsen av protokollene.

3.7.3 *Dokumenter angaaende flytlapperne*

Her burde man lage en digital faksimile av alle sidene for å vise hvor i skriftbildet de håndskrevne notatene er plassert. Dette ville også gi leseren mulighet til å kontrollere tolkningen av vanskelig håndskrift.

Faksimilene bør produseres i 2–300 DPI med 24-bits fargedybde enten ved hjelp av digitalt kamera eller ved avfotografering på fargefilm som deretter scannes. Ved publisering kan faksimilene legges ut i lavere oppløsning (f.eks. 72, 100 eller 150 DPI), men med mulighet til å fremskaffe fullversjoner, enten on-line eller ved bestilling på CD-plater.

Kapittel 4

Meta-materiale

4.1 Digitaliseringen

Arbeidet med å digitalisere bibliografiene bestod i å samle de bibliografiske opplysninger fra flere kilder (se avsnitt 2.2.1) og registrere disse i NORMARC-formatet, som er norsk standard for utveksling av bibliografiske data. Postene legges inn i henhold til en ekte delmengde av NORMARC som er gjengitt i tillegg D.

Selve innskrivingen ble gjort i bibliotekssystemet MikroMARC. Basen ble deretter eksportert i MARC-format, og er søkbar som en onlinebase under WWW.[1] Den kan også danne grunnlag for trykte utgivelser.

I utgangspunktet brukes samme postinndeling som i den trykte bibliografien. Imidlertid benyttes det der forkortede skriveformer basert på rekkefølgen. F.eks. kan en '—' i en trykt bibliografi bety at posten har samme forfatter som forrige. Slike forkortelser må skrives fullt ut i en database.

4.2 De enkelte trykte bibliografiene

4.2.1 *Bibliotheca Norvegica*

Noen av postene i *Bibliotheca Norvegica* er svært omfattende. Disse forkortes ved at innholdsnotene kuttes ut. Dette gjelder ikke analyserte tidsskrifter der det henvises rett til en enkelt artikkel.

[1]Adressen til basen finnes i tillegg A.

4.2.2 *Tillegg 1928–1945*

Dette tillegget består av henvisninger til *Norsk bokfortegnelse* (NB) og *Norsk tidsskriftindex* (NTI), dessuten er et "Tillegg til Finnmark fylke" behandlet separat (se avsnitt 4.2.3).

Her er kjente forlag satt opp uten sted, men stedet er oppgitt i forlagslista. For disse forlagene kan sted enkelt legges inn dersom dette er ønskelig, men dette er foreløpig ikke gjort.

Redaktør legges inn på tidsskriftserier dersom det har vært samme redaktør hele tiden. Ved skifte legges ingen av redaktørene inn.

Ved henvisninger til NTI går vi ut fra at årstallet for tidsskriftene er NTI'ens der ikke annet er angitt. Ved NTI'er som går over flere år settes årstall inn der det er mulig å utlede det av oversikten over analyserte tidsskrifter.

Overskriftene er her på flere nivåer, fra det generelle til det spesielle. Kun nederste nivå (det mest spesielle) er lagt inn. I visse tilfelle er ikke nederste nivå pregnant, da legges også overskrifter fra høyeste nivå inn i *653 ind. 2 (se avsnitt 4.3.5). I ettertid kan det være aktuelt å sette inn alle overskriftsnivåer for å bedre søkemulighetene.

I enkelte tilfelle er andre kilder eller originaler benyttet. For serien *Nordnorske samlinger* støttet vi oss på katalogiseringen i Bibsys for å få seriene riktig. For artikler fra *Festskrift til rektor J. Qvigstad* (1928) er primærkilden brukt som referanse.

4.2.3 "Tillegg til Finnmark fylke"

Dette er en bibliografi med svært kortfattede innførsler fra fem Finnmarksaviser. Det er ikke helt lett å forstå kriteriene for hvilke artikler som er plukket ut, og bibliografien gir i det hele tatt et tilfeldig inntrykk. Man kan vurdere hvorvidt det er best å fjerne den fra basen.

Poster står stadig som kommentarer til foregående poster; dette er det ikke helt enkelt å overføre til en database. Vi bruker noter og i en viss utstrekning konstruerte titler for å få dette til. Det skaper en viss usikkerhet, men vi har ikke kapasitet til noen omfattende verifisering.

Ved flere poster der det kun er oppgitt tittel på den første benyttes den samme tittelen på de følgende, for eksempel:

```
34/97: Er F.'s befolkning mindreværdig? - 93: av Bjarne
Hofseth. - 102: av Bastard.
```

På slike poster burde tittelen vært verifisert skikkelig, da det kan være at kommentarer til tidligere artikler har hatt avvikende tittel. Dette har vi ikke kapasitet til. Vi går ut fra at de er mulig å finne igjen artiklene i avisene likevel.

4.2.4 *Bibliografi til Norges historie*

En stor mangel ved denne, i forhold til de andre bibliografiene, er at den ikke er fininndelt på sted; For årene 1946–65 er den en felles kategori "Nord-Norge og Svalbard", for årene 1966–75 er inndelingen gjort etter fylke. For friteksts øk skaper ikke dette særlige problemer, men for strukturerte søk og en eventuell integrering med andre baser er det problematisk. Det ville vært svært nyttig dersom noen kunne gjøre en bedre stedsklassifikasjon her. Dette er ikke en uoverkommelig jobb, et hjelpemiddel er at stedsnavn i titler er satt med sperret skrift i innførslene i den trykte bibliografien.

4.3 De enkelte feltene

4.3.1 Forfatter (*100/*245/*700)

Når kun forfatterens initialer er oppgitt, f.eks. "C.O." legges disse inn uten å være invertert, både i *100$a og i *245$c. De inverteres ikke selv om de var invertert i originalen.

Det lages ikke biinnførsel på oversetter.

Flere énbokstavs initialer legges inn uten mellomrom. Dette gjelder også initaler som står med små bokstaver, f.eks. "L.v. Buch".

Dersom etternavnet er satt opp sortert bruker vi samme regler som i den trykte bibliografien, f.eks "von der Lippe, D.", ikke "Lippe, D. von der".

I de tilfelle der henvisninger legges inn i note, f.eks. "jfr."-henvisninger og anmeldelser, legges forfatter og tittel inn i *700$a og $t der de er oppgitt.

4.3.2 Tittel (*245)

I noen av bibliografiene er tittelen på innførsler uten forfatter invertert etter datidens sorteringsregler:

```
Beskrivelse, Norges topografiske
```

Disse titlene er lagt inn i invertert form.

4.3.3 Utgave og utgiver (*250/*260)

Dersom det i en post bemerkes at det dreier seg om et nytt opplag, legges denne opplysningen i note. Dersom det nye opplaget hadde egen post i den trykte bibliografien, lages det ny post.

Institusjoner som står som "Utg. av" regnes som forlag/distributør og legges i *260$b.

4.3.4 Fysisk beskrivelse (*300)

Plater (pl.) og tavler legges i *300$b, mens normale blad (bl.) legges i $a.

I *Bibliotheca Norvegica* innledes ofte sidetallene med 4. eller 8. Dette er formatet (4to, 8to, ...) og legges i *300$c slik det står.

$e benyttes kun der det er oppgitt at blad er løse. Ellers benyttes $b.

4.3.5 Overskrifter som emneord (*653)

I dette feltet legges både geografiske og generelle overskrifer, kun skilt ved indikator 2. Dersom det finnes flere geografiske emneord uten at det finnes noe generelt, vil en feil i MikroMARC gjøre ett av de geografiske om til generelt emneord ved å endre indikatoren. Dette har vi omgått ved at slike poster inneholder et generelt emneord <tom>. Dette bør om mulig fjernes ved publisering.

4.3.6 Kilde (*773)

Det er umulig å se på artikkelposter i *Norges topografi*, *Tillegg 1917–1927* og *Tillegg 1928–1945* om det henvises til et sidetall eller et heftenummer. For tidsskrifter antar vi at det er snakk om sidetall, for aviser antar vi det er snakk om heftenummer. Disse settes da etter årstallet i *773$g.

Ved to årstall legges det første sammen med andre volumopplysninger i $g og det siste (trykkeår) i $i.

Ved kildetitler på over 200 tegn avbrytes tittelen og [...] settes inn.

4.3.7 Bibliografi (*850)

Her legges navnet på bibliografien inn etter følgende koder:

db Deichmanske bibl.: *Register til en del norske tidsskrifter*

sc Schweigaard: *Norges topografi*

s1 *Tillegg 1917–1927* ved W.P. Sommerfeldt

s2 *Tillegg 1928–1945* ved samme

s3 "Tillegg til Finnmark fylke", trykt i **s2**

nh *Norsk lokalhistorisk litteratur 1946–1970*

nl *Bibliografi til Norges historie* 1946–75

Sidetall eller postnummer legges også inn.

Dersom samme post er nevnt i flere bibliografier lages det nye 850-poster for hver ny bibliografi. Hvis samme post nevnes flere steder i samme bibliografi ramses sidetallene eller postnumrene opp skilt med komma.

4.4 Tegnsett

Følgende tegn finnes ikke i tegnsettet og er gjengitt som koder:[2]

): <dvs>

^{o} <grad>

† <dagger>

ϕ <phi>

λ <lambda>

Č <Ccaron>

ń <nacute>

Tankestrek erstattes med "mellomrom-mellomrom".

[2]Disse kodene må ikke forveksles med SGML-tagger.

4.5 Krav til publisering av MARC-data

4.5.1 Sortering og søking — stoppord og alfabetiseringsfelt

Følgende “stoppord” må sees bort fra ved søking og alfabetisering:

```
( ) [ ] " ' - /
```

Artikler skal ikke innføres som stoppord, dette er ordnet ved hjelp av sorteringsfelt der det er aktuelt. Innedende “[” og “"” er også fjernet ved hjelp av sorteringsfelt fordi MikroMARC sorterer poster med disse tegnene feil.[3]

Data leveres i 8-bits format, codepage 850 eller ISO 8859-1 etter valg. Bokstaver med aksenter og andre spesialtegn må tas hensyn til ved sortering og søking.

4.5.2 Søkefelt

Følgende søkefelt er aktuelle i tillegg til MikroMARCs predefinerte:

bi hvilken bibliografi posten er hentet fra

el klassifikasjon etter Elvestrand.

[3] Dette skjedde ihvertfall i vårt oppsett.

Kapittel 5

Oppsummering

5.1 Publisering og veiledning for leserne

Det må lages brukerveiledninger til alt materiale som skal publiseres. I tillegg bør man søke å publisere kortere innledningstekster som setter materialet inn i en sammenheng. Dette kan gjøres ved å konvertere artikler, eller ved å nyskrive innføringer. For Schnitlers protokoller kan dette være oppfylt ved at Kristian Nissens innledning er konvertert.

Dersom noe av materialet skal beskyttes bak passord, bør man benytte anledningen til å sende ut brukerveiledning og bakgrunnsinformasjon sammen med passordet brukeren tildeles.

5.2 Kildekritiske betraktninger

Skulle det etter publiseringen bli oppdaget feil av brukerne, bør det finnes tilbakemeldingsrutiner som gjør det enkelt å rapportere mulige feil, og man må avklare hvem som har ansvaret for å behandle slike meldinger.

5.2.1 Bibliografiene

Bibliografiene er ikke gjengitt som de trykte versjonene for enhver pris. Imidlertid har det ikke vært mulig å foreta utstrakte rettelser og forbedringer, kun omforming til digital versjon.

Korrekturlesningen garanterer at det finnes svært få feil i overføringen av tekst fra de trykte bibliografiene til databasen. Imidlertid er det ikke kontrollert systematisk om de trykte bibliografiene er korrekte. Det er likevel grunn til å tro at det store flertall av postene vil fungere både i søk og til gjenfinning av dokumentene de referer til.

5.2.2 Trykt fulltekst

Ved siste korrektur, som ble utført av fagkonsulenten og av den innleide korrekturleseren i Tromsø, ble det i gjennomsnitt funnet ca. 0,2 feil pr. utskriftsside, noe som tilsvarer ca. én feil på annenhver originalside. Disse feilene ble rettet, og det antas å gjenstå svært få feil.

Av de feil som ble funnet, var det kun én som ville ført til betydelig endring av meningsinnholdet dersom den ikke hadde blitt korrigert. De fleste feil førte til at de feilaktige formene ikke var meningsfylte,[1] for eksempel tekstfeil av typen forvekslet 'c'og 'e', forvekslet 'a' og 'å', og lignende.

Taggingen er korrekturlest på tilsvarende måte, men det er ønskelig at man for alle tekstene tar ut lister over alt som er tagget med innholdstagger som stedsnavn og personnavn. Disse listene gåes så gjennom av kvalifisert personale for å fjerne eventuelle mistolkninger.

5.2.3 Håndskrifter

I transkribering av håndskriftene er det særlige problemer, og selv om de har vært gjennom omfattende korrekturlesning, vil de inneholde feil og mistolkninger. Særlig gjelder dette brevene, som kun er korrekturlest mot utskrifter fra mikrofilm. Jeg anbefaler derfor at man gjør følgende:

1. Faksimiler av *Dokumenter angaaende flytlapperne* produseres (jfr. avsnitt 3.7.3)

2. Brevene korrekturleses mot originalene av en person som behersker norsk, samisk og gjerne finsk.

Taggingen kontrolleres på tilsvarende måte som for trykt fulltekst.

[1] De fleste feilene som ble funnet var slike som gjerne kalles "scannefeil", selv om de oppstår i tegngjenkjenningen (OCR-lesningen). De gir ofte en annen type misforståelser enn de som oppstår ved feil skapt av mennesker.

5.2.4 Kartene

Digitaliseringen av kartene innebærer ingen kildemessige problemer.

5.2.5 Kommentarer

Innskriverne i Fauske la i noen tilfelle inn kommentarer til materialet de skrev inn. Til dette brukte de standard SGML-kommentarer. Disse ser slik ut:

```
<!--Dette er en kommentar-->
```

Dels dreide dette seg om systeminformasjon (filnavn, navn på innskriver, osv.), dels dreier det seg om galgenhumoristiske kommentarer ved særlig vanskelig materiale. Et godt ekstempel på sistnevnte finnes i Qvigstads brev til Sätäle 2/11 1890 der han over flere sider diskuterer samiske dialekter fra Arjeplog, Arvidsjaur og Sorsele før han avslutter med:

```
<AVSN>Det er min Mening, at Dialekterne i <SNAVN>Arje-
plog</SNAVN>, <SNAVN>Arvidsjaure</SNAVN> og
<SNAVN>Sorsele</SNAVN> fortjener en nærmere
<!-- NÆRMERE!? --> Undersögelse.
```

Disse kommentarene er ikke slettet for brevenes vedkommende fordi vi mener de er en del av tekstenes digitaliseringshistorie. Men de må selvsagt ikke være synlige ved publisering!

5.3 *Dokumenter angaaende flytlappernes* **registre**

5.3.1 "Identificering af stedsnavne"

Autotaggingen og den etterfølgende validering og korrektur har tagget opp kapitlet på samme måte som resten av boka.[2] Imidlertid har et register spesielle muligheter til strukturering som vi her kort vil komme inn på.

Ved utgivelsen av den digitale versjonen vil det være to stedsnavnregistre: En rå liste over alle stedsnavn basert på taggingen, og en vurderende laget av Qvigstad. Hver innførsel i sistnevnte er tagget som ett avsnitt.

[2]For nærmere beskivelse av autotaggingen, se avsnitt 3.6.5

Himmelretninger

Karthenvisninger er markert med en tag <kilde> rundt forkortelsen av navnet på kartbladet, f.eks.

```
<kilde lopenr="5800001">L11</kilde>,
<kilde lopenr="5800002">AkSV</kilde>
```

Imidlertid finnes også en nærmere bestemmelse av hvor på kartet stedet ligger angitt med himmelretning, evt. "midt paa". Oppsettet av dette er så fritt at vi ikke fant mulighet til å registrere informasjonen automatisk; f.eks.

```
<kilde lopenr="5800001">L11</kilde> S
```

er et strukturert oppsett, men det kan like gjerne stå

```
i S paa <kilde lopenr="5800002">AkSV</kilde>
```

Man kan koble disse opplysningene til moderne kart, f.eks. ved å legge inn lengde- og breddegrader på hvert stedsnavn, men dette er et omfattende og krevende arbeid. Dersom man senere digitaliserer faksimiler av selve kartene, kan man også lage en kobling dit. En annen mulighet ville være en gjennomgang av alle stedsnavn med kobling til moderne kommune der det er mulig å fastslå, slik at navnene lettere kan kobles til navn i annet kildemateriale.

En mulig vei å gå er å arbeide mot et digitalt historisk stedsnavnregister koblet mot et digitalt historisk atlas.

Relasjoner mellom stedsnavn

Stedsnavnene i registeret kan deles i to kategorier:

1. Stedsnavn som står i kursiv først i hvert avsnitt er selve innførselen; disse er sortert alfabetisk. De står normalt i den formen de har i selve boka, men minimalt avvikende former kan være slått sammen uten at dette er spesielt avmerket dersom disse likevel ville havnet rett under hverandre. Etter dette stedsnavnet kommer det vanligvis en henvisning til én eller flere sider der navnet er nevnt. Bindnummer angis som "I" eller "II".

2. Etter dette kommer det en beskrivelse av stedsnavnet. For steder forfatterne har hatt problemer med å finne ut av består beskrivelsen av "= ?" og av og til en kommentar. For lokaliserte steder kan det forekomme ett eller flere stedsnavn og en eller flere kartreferanser i beskrivelsen. Dersom det kommer et "=" og deretter et stedsnavn er dette navnet i den form man finner på de kartene Qvigstad hadde tilgang til i 1907–9.

Det forekommer også se-henvisninger, disse kjennetegnes ved at det står 'se' (eventuelt 'jfr') rett før et stedsnavn, slik:

```
se <snavn>Altze tresk</snavn>
```

Se-henvisninger er ikke tagget som henvisninger. Det forekommer også se-henvisninger av typen "se foreg.", se innførselen på "Aivos-Vuöbme".

Vanlige fraser i beskrivelsene er "stedsnavn i stedsnavn" og "stedsnavn ved stedsnavn". I noen tilfelle skrives flere former av oppslagsordet i kursiv, se innførselen på "Achefieldet". Disse skilles gjerne med ';' eller ','.

Avsnitt kan avsluttes med "og</avsn>" (se innførselen på "Aduer"), med ",</avsn>" (se innførselen på "Akelsvatnet"), med "gaar gjennem</avsn>" (se innførselen på "Gub(b)elaaen") eller med "ant S for</avsn> (se innførselen på "Grux-dahlen"). I alle disse oppsettene pekes det mot innførselen som står i neste avsnitt.

Himmelretninger brukes også i relasjon til stedsnavn, se innførslene på "Affuijoware" og "Grebstad".

5.3.2 "Sagregister"

Sakregisteret er tematisk ordnet i mellom- og underkapitler.

Innførsler i sakregisteret består av en beskrivende tekst, et ":", og deretter en henvisning. I mange av postene er det listet opp ett eller flere årstall med en henvisning for hvert årstall. Mellom henvisningene kan det komme kommenterende paranteser, slik at henvisningene blir brutt opp og det vil bli nødvendig å se bakover for å finne bindnummeret. Eksempel fra side 508:

```
<HENV>I 204, 205</HENV> (personalier), <HENV>212</HENV>
(pers.), <HENV>217</HENV> (pers.), <HENV>245 ff.</HENV>;
```

Det forekommer også henvisninger til andre steder i registrene. Disse er uten bindnummer.

```
Se nedenfor <HENV>sid. 512</HENV>.
```

5.4 Innholdsmessige betrakninger

I brevene til Wiklund er det et opphold fra 1906 til 1911. Dette kan neppe være reelt; Qvigstads og Wiklund samarbeidet nettopp i denne perioden om arbeidet for *Renbeitekommissionen af 1907*. Vi har ikke funnet noen forklaring på dette. Kan det kanskje finnes brev fra denne perioden i departementsarkiver? Et tilsvarende opphold finnes i brevene til Setälä, her fra 1906 til 1910.

Den elektroniske versjonen er et godt verktøy til å finne ut mer om hva understrekningene i ulike farger betyr, selv om den selvsagt ikke kan gi slike svar alene; det krever kyndige spørsmål, og svarene må fortolkes. Noen eksempler på spørsmål man kan stille og som den elektroniske versjonen kan være en hjelp til å finne svar på:

- Man kan be om å få en liste over alle stedsnavn som er understreket med rødt og en liste over alle som er understrøket med blått. Hvorfor denne forskjellen? Er de ene i Norge og de andre i Sverige? Har fargene sammenheng med hvordan navnene er ført inn i registeret?

- Man kan studere de notatene som er skrevet med grå blyant og de som er skrevet med blekk. Hvorfor denne forskjellen? Er den tilfeldig, eller samsvarer den med en forskjell i innhold? Er den ene skrifttypen eldre enn den andre? Viser de at notatene er mer eller mindre sikre?

5.5 Videre registrering

Under arbeidet har materiale som ikke ble konvertert pekt seg ut som aktuelt dersom man kan skaffe ressurser til videre arbeid. Aktuelt materiale kan være:

- Utfyllende historisk kildemateriale: *Dokumenter angaaende flytlapperne* bd. 1 og Schnitler bd. 2 og 3, dessuten registret i Schnitler bd. 1. Ellers er det nok kildemateriale å ta av!

- For kartene som allerede er digitalisert bør man skrive inn stedsnavnene på kartet med en form for identifikasjon, slik at det kan lages registre for kartene.

- Flere Qvigstad-håndskrifter: Flere brev, annet kildemateriale, særlig felthåndbøker (se tillegg G).

5.6 Modeller for registrering av annet materiale

I dette delprosjektet er ulike former for materiale konvertert. Imidlertid finnes det andre måter å organisere konvertering på enn den arbeidsintensive måten som er benyttet. Vi vil her risse opp noen muligheter for framtida.

5.6.1 Dokpro-modellen

"Dokpro-modellen" betyr ikke nødvendigvis at man bruker samme slags organisering av arbeidskraften gjennom sysselsettingstiltak. Vi bruker det her som et begrep som dekker større konverteringsarbeider, der flere personer utfører konverteringen på hel- eller deltid under faglig ledelse. Slikt arbeid vil normalt være avhengige av av ekstraordinære bevilgninger.

5.6.2 Konvertering som en del av vanlig drift

Dette er en modell der en eller flere av de fast ansatte gjennomfører konverteringsprosjekter, eventuelt med medarbeidere innleid på timebasis. Dersom dette skal foregå ved siden av andre oppgaver sier det seg selv at man må satse på mindre prosjekter. Det er rimelig å se på slik konvertering som en videreføring av den tradisjonelle utgivelsen av arkivmateriale. Man kan tenke seg kortere avhandlinger og oppsatser som ikke er trykt, Qvigstads felthåndbøker, og lignende. Arbeidet vil ligne på tidligere utgivelser på papir, og må selvsagt foregå med like strenge kvalitetskrav. Samtidig bør man bruke et registreringsopplegg som gjør det enkelt å publisere en elektronisk versjon, enten i tillegg til eller i stedet for papirversjon. Vi vil selvsagt peke på erfaringene fra Dokumentasjonsprosjektet som viktige i et slikt opplegg.

5.6.3 Automatisk massekonvertering

Dette har ikke vært mulig tidligere, men med den raskt økende kapasitet vi nå ser på datamaskiner og platelagre, kan det bli aktuelt i løpet av kort tid.

Et eksempel på hvordan dette kan gjøres: Man tar et eksemplar av rapporten fra Renbeitekommssionen av 1913 og skjærer av ryggen (bøkene kan om ønskelig bindes inn på nytt etter at jobben er ferdig). Deretter fôrer man boksidene i en rask scanner med arkmater i to omganger (for- og bakside). Disse kan et dataprogram siden sette sammen i riktig rekkefølge.

Hver bokside vil nå eksistere som en elektronisk faksimile, et bilde som kan tas opp på en dataskjerm. Neste trinn i prosessen er å kjøre disse bildene gjennom et tekstgjenkjenningsprogram slik at en digital tekst genereres. Denne teksten vil inneholde en viss menge feil, men med godt forarbeid og moderne programvare bør man kunne komme ned i 10–20 feil pr. side.

Denne teksten korrekturleses *ikke*, da den kun skal brukes som et søkeverktøy. Hvis man f.eks. skulle vært interessert i alle steder i teksten ordet "myr" nevnes, kan man gjøre søk i den digitale teksten, men når man får treff, dukker ikke denne teksten opp, men bildet (faksimilen) av siden. Dette gjør at feilene som ikke ble rettet ikke vil skade lesingen — det eneste som skjer, er at dersom ordet "myr" etpar steder har blitt til "mur" vil man ikke finnes disse stedene i søket.

Dette er selvsagt ikke et like godt hjelpemiddel som de elektroniske tekstene som er laget i Dokumentasjonsprosjektet, men for materiale der en grundig konvertering ikke er økonomisk forsvarlig, kan dette være en mulig mellomløsning.

Tillegg A

Litteratur

A.1 Om Dokumentasjonsprosjektet

Man kan lese om Dokumentasjonsprosjektet generelt følgende steder:

- *Fra skuff til skjerm* / Knut Aukrust og Bjarne Hodne (red.). Universitetsforlaget, 1998
- Dokumentasjonsprosjektet: *Sluttrapport 1992–1997* / Christian-Emil Ore og Nina Kristiansen (red.). Universitetet i Oslo, 1998.
- Prosjektets hjemmesider: http://www.dokpro.uio.no/

Øyvind Eide har skrevet følgende artikler om delprosjektet:

- "Re-integrasjon av kildemateriale : nye dokumentformer på tvers av tradisjonelle skiller". I: *Human IT* nr. 1 (1998), s. 7–20. Elektronisk versjon på http://www.hb.se/bhs/ith/humanit.htm.
- "The Quest for Re-Intergration : Creating New Documents across Traditional Borders", i: *Shaping the Knowledge Society : Proceedings of* 6th *International BOBCATSSS Symposium january 26–28 1998, National Széchénui Library, Budapest* / red.: Camilla Moring, Anders Christiansen, Nanna Kann-Rasmussen. København, 1998, s. 280–91. Finnes i elektronisk versjon under http://www.uio.no/~oeide/artikler/.
- "Hvor ble det av sammenhengen" i: *Klassekampen* 28/2 1998.

Tor Sveum har skrevet:

- "Fra skuff til skjerm og bok til base" i: *Synopsis* 2/1998, s. 165–167.

Dessuten er det publisert artikler om flere andre delprosjekter. Oversikt over disse finnes etter artiklene i *Fra skuff til skjerm*. Følgende artikkel utyper viktige punkter i denne rapporten:

- Ore, Christian-Emil: "Making Multidiciplinary Resources" i: *The Digital Demotic: A Selection of Papers from DRH97 (Digital Resources for the Humanities)* / Edited by Lou Burnard, Marilyn Deegan and Harold Short. London, 1998, s. 65–74.

Databeskrivelser, programmer og lignende som gjengis i denne rapporten finnes i digital form på: http://www.dokpro.uio.no/dokub/

A.2 Om meta-materialet

Bibliografien er tilgjengelig fra:
http://webhotel.mikromarc.no/uito/nor/mikromarc/main.idc?dbAlias=uito

Den er basert på følgende utgaver:

- Deichmanske bibliotek: *Register til en del norske tidsskrifter. I. Topografi.* Kristiania 1908.
- *Bibliografi over eldre nordnorsk topografisk litteratur : tidsskriftartikler.* Red. av Tor Sveum. Universitetsbiblioteket i Tromsø, 1991
- Johan Schweigaard: *Norges topografi : bibliografiske fortegnelser over topografisk og lokalhistorisk literatur.* Kristiania, 1918. [Fram til 1917].
- *Tillegg 1917–1927* ved W.P. Sommerfeldt. Oslo 1930
- *Tillegg 1928–1945* ved W.P. Sommerfeldt. Oslo 1955
- *Norsk lokalhistorisk litteratur 1946–1970* [Nordland, Troms, Finnmark]. Oslo 1979, 1980.
- *Bibliografi til Norges historie* 1946–75.

- Pettersen, Hjalmar: *Bibliotheca Norvegica.* Christiania, 1899–1924.

I tillegg er følgene litteratur relevant for konverteringen:

- Halvorsen, J.B: *Norsk forfatter-lexikon 1814–1880 paa grundlag af J. E. Krafts og Chr. Langes "Norsk forfatter-lexikon 1814–1856".* Kristiania, 1885–1908.
- Marklund, Erik: *Övre Norrland i litteraturen : en bibliografi över Norrbottens och Västerbottens län.* Umeå, 1963–1966.
- Uuttu, Leena-Kaarina: *Pohjois-Suomen bibliografia vuoteen 1960.* Oulu, 1972–.
- NORMARC-formatet (1993), som er tilgjengelig fra http://info.rbt.no/normarc/Welcome.html
- *Norsk bokfortegnelse, Årskatalog*
- *Norsk tidsskriftindex ... : systematisk fortegnelse over indholdet av ... norske tidsskrifter*

A.3 Om fulltekstregistrert materiale

Det digitaliserte materiale som finnes i trykte versjoner er:

- *Dokumenter angaaende flytlapperne m.m. : samlede efter renbeitekommissionens opdrag* / af J. Qvigstad og K.B. Wiklund. — Kristiania, 1909. — 2 b.
  Over tittelen: Renbeitekommissionen af 1907.
  Bind 2: 1909. — 523 s.

  I tillegg til den trykte teksten er også de notater Qvigstad har gjort i margen på sitt private eksemplar lagt inn.

- Schnitler, Peter: *Major Peter Schnitlers grenseeksaminasjonsprotokoller 1742–1745* / utgitt av Kjeldeskriftfondet. — Oslo : Norsk historisk kjeldeskrift-institutt, 1929–1985. — 3 b.
  Bind 1: ved Kristian Nissen og Ingolf Kvamen. 1962. — L, 478 s. (s. I–L, 1–437 er ferdigstilt, mens registrene, s. 438–478, dessverre ikke er ferdig digitalisert)

- Friis, Jens Andreas: *Ethnografisk Kart over Finmarken*. No. 1–5. 10 Bl. Udg. af Videnskabsselskabet i Christiania, med Bidrag af Oplysningsvæsenets Fond 1861–62.

- Friis, Jens Andreas: *Ethnografisk Kart over Finmarkens Amt*. 6 Bl. / Udg. paa offentlig Bekostning. 1888. Norges geograf. Opmåling.

- Friis, Jens Andreas: *Ethnografisk Kart over Tromsø Amt*. 6 Bl. 1890. Norges geograf. Opmåling.

Originalene til de digitaliserte brevene finnes i følgende arkiver:

- Universitetsbiblioteket i Oslo, håndskriftsamlingen, brevsamling 375 Olsen Magnus (Ms4^{o}2355).

- Riksarkivet i Helsinki, prof. Setälä privatarkiv

- Universitetsbiblioteket i Uppsala. Håndskriftsamlingen.

I tillegg er følgende titler aktuelle som bakgrunnsmateriale:

- Blankholm, Hans Peter: "Geografiske informasjonssystemer" i: *Ottar* nr. 3, 1998, s. 34–42.

- *Dokumenter angaaende flytlapperne m.m. : samlede efter renbeitekommissionens opdrag* / af J. Qvigstad og K.B. Wiklund. — Kristiania, 1909. — 2 b.
  Over tittelen: Renbeitekommissionen af 1907.
  Bind 1: 1909. — 538 s.

- Hansen, Lars Ivar: *Kulturforskeren Just Qvigstad*. [Tromsø], 1991. (Prøveforelesning 24.05.1991 for den filosofiske doktorgrad, Universitetet i Tromsø)

- Hansen, Lars Ivar: "J.A. Friis' etnografiske kart over Troms og Finnmark" i: *Ottar* 3/1998, s. 43–48.

- Johnsen, Oscar Albert: *Finmarkens politiske historie : aktmæssig fremstillet*. Kristiania, 1923.

- Lae, Erling: *Fra Karlstadkonvensjon til reinbeitekonvensjon : forhandlingene i reibeitesaken mellom Norge og Sverige fram til reinbeitekonvensjonen av 1919.* Hovedoppgave i historie, Universitetet i Oslo, 1977.

- Nes, Oddvar: *Storms norske lydskriftsystem (med tillegg) : definert ved hjelp av IPS's lydskriftsystem.* Universitetet i Bergen. Fonetisk institutt, 1982 (Skriftserie fra insituttet, serie B, nr. 8).

- Registreringssentralen for historiske datas hjemmesider: http://www.sv.uit.no/seksjon/rhd/

- Rygh, Oluf: *Norske gaardnavne : Oplysninger samlede til brug ved Matrikelens Revision.* Kristiania, 1897–1936.

- Storm, Joh.: "Kortere Ordliste : med Forklaring af Lydskriften". Bilag til *Norwegia*, 1. hefte. Kristiania, 1884.

A.4 Om SGML og digitalisering generelt

Christian-Emil Ores artikkel i *Fra skuff til skjerm* gir en kortfattet innføring på sidene 35–39.

Om bruk av SGML innen humaniora generelt henvises til Text Incoding Initiatives hjemmeside (http://www-tei.uic.edu/orgs/tei/) og til deres retningslinjer. Kapittel 2 er en svært god innføring i SGML.:

- *Guidelines for electronic text encoding and interchange : TEI P3 /* edited by C.M. Sperberg-McQueen and Lou Burnard ; [published by] Text Encoding Initiative, Chicago, Oxford, 1994. 2 b. (XXVI, 1290 s.). Finnes også på CD-rom og online på TEIs hjemmeside.

For studier av SGML anbefales ikke selve ISO-standarden dersom man ikke er vant til å lese slike standardiseringspublikasjoner. Det finnes imidlertid flere gode bøker, for eksempel:

- Charles F. Goldfarb: *The SGML handbook.* Oxford : Clarendon Press, 1990.

En grei innføring i formelle grammatikker finnes i:

- Grune, Dick: *Parsing techniques : a practical guide* / Dick Grune, Ceriel J.H. Jacobs. New York, 1990.

To gode innføringer i digitalisering av kildemateriale er skrevet av Peter Robinson:

- Robinson, Peter: *The Digitization of Primary Textual Sources.* (Office for Humanities Communication Publications ; 4) Oxford, 1993.
- Robinson, Peter: *The Transcription of Primary Textual Sources Using SGML.* (Office for Humanities Communication Publications ; 6) Oxford, 1994.

En god beskrivelse av digitaliseringsprosesser med hovedvekt på den tekniske siden av scanneprosessen er:

- Kenney, Anne R.: *Digital Imaging for Libraries and Archives* / by Anne R. Kenney and Stephen Chapman. Itacha, NY, 1996.

Karsten Kynde har skrevet en interessant artikkel om bakgrunnen for tekstkoding:

- Kynde, Karsten: "Appearance, Rendeling, and the Abstract Intention of the Text" i: *Kierkegaard Studies. Yearbook 1996* / ed. by N.J. Cappelørn and H. Denser. Berlin, New York, 1996.

I denne samlingen finnes det flere interessante artikler basert på innlegg fra en konferanse i Oxford i 1997:

- *The Digital Demotic: A Selection of Papers from DRH97 (Digital Resources for the Humanities)* / Edited by Lou Burnard, Marilyn Deegan and Harold Short. London, 1998.

En artikkel om digitalisering og publisering av kart:

- Allen, David Yehling: "Creating and Distributing High Resolution Cartographic Images", i: RLG DigiNews vol. 2 nr. 4 (15. august 1998). Finnes på http://www.rlg.org/preserv/diginews/

World Wide Web Consortiums hjemmesider finnes på http://www.w3c.org/

Tillegg B

Arkivalia etter Just Knud Qvigstad

B.1 UB Oslo. Håndskriftsamlingen

B.1.1 Manuskripter

Ms8^o 2891 "Etymologisk ordliste". Samisk. 2 bind
1. A-N. 240 bl. Bl. 232-237 ubeskr.
2. O-Æ. 99 beskr. bl. Ms film 147

Ms8^o 2892 Oppslagsbok samisk ordlære. Oppslagsbok I.
Sammensatte ord. Samisk. 1+93 bl. + 1 innlagt løst bl. "sammensatte ord" fyller bl. 1-38. Resten av bindet inneholder andre språklige opptegnelser. Ms film 147

Ms4^o 1669 Lappiske sprogprøver.
a) "Breve, samlede av provst Balke." Avskrifter
b) "Breve fra Lars Johnsen Gaino, Kautokeino [og] S. Samuelsen, Smalfjord i Tana" til J.K. Qvigstad. Originaler. (1880-1923). 9 brev.
c) "Brev fra Johan Aikio, Nesseby" til J.K. Qvigstad. Originaler. 72 brev, 1890-1926, + 2 fra John Aikio. Ms film 148

Ms8^o 3300 Notater om noen samiske manuskripter sendt til ham fra Finnmarksbiblioteket til gjennomsyn mars 1934. Egenh. 2 dobbeltark

Ms fol. 1765 Grammatiske optegnelser. 2 h.
 a) Kvænangen. Vestfinnmark
 b) Varanger.
 Ms film 151

Ms fol. 1822 “Sproget i Graans Manuale Lapponicum” 22 s. Tr. i “Studia Septentrionalia” III

Ms8^o 1465 Lappiske ordlister a-k :
 a) Arjeplog-dialekt (Beiarn, Saltdalen, Rana). 1.2.
 b) Enare-lappisk
 c) Kautokeino. Karasjok. Kistrand. Polmak. Lebesby. Tana. Karesuando.
 d) Kvænangen. 1-3.
 e) Lenvik. Ibestad. Ofoten.
 f) Lyngen. 1.2.
 g) Skoltelappisk
 h) Tysfjordlappisk. 1.
 Tysfjordlappisk. 2. Kaldfjord. Vesterålen.
 i) Varanger. 1-3.
 k) Vestfinnmarken
 I alt 17 b.

Ms.fol. 953 Lappiske navne paa stjerner. S. 1-10. Ms film 150

Ms8^o 1466 Lappiske personnavn og slektsnavn

1467 Utvalgte stykker av “Lærebog for Folkeskolen og Folkehjemmet”. Overs. til lappisk av lærer S. Nilsen.

Ms4^o 2355 Fortegnelse over ku- og geitnavn fra Troms og Nordland. Egenh. 60 bl. Vedlegg: Brev fra Qvigstad til Magnus Olsen 7. mai 1923 og fra Aug. Nielsen (Nilssen ?) til Qvigstad 19. mars 1923, 1 blad med håndskr. fortegnelse over kunavn i Enger kontrollforening, S. Land, samt en trykt katalog fra en husdyrutstilling i Vefsn 1914. Ms film 149

Ms4° 3051 Se Nissen, Kristian
Etterlatte papirer
G. Samer
29. Avskrift av rektor Qvigstads optegnelser fra gamle justisprotokoller m.v. vedr. Finmarkens historie m.v.
H. Biografiske artikler
2. J.K. Qvigstad 1 mappe

Hovedoppg. 5129 k Eddy Wighus. J.K. Qvigstad og hans samling av samisk folkeminne. Mag.avh. 1976. Høst. Unversitetet i Oslo

B.1.2 Mottatte brev (fra:)

Brevs. 707 Aas, Arne 1940 apr. 6. (Vedl. brev fra J.Q. til Harry Fett 10.4.1940) [Opplysninger fra Arne Aas, sorenskriver i Varanger, om John Savio]

Ms4° 1669c Aikio, Johan J. 1890-1926: 72 brev, alle på samisk

Ms4° 1669c Aikio, John 1900 juni 9 (Vedl. brevene fra Johan J. Aikio til J.Q.)

Brevs. 1 Døving, Carl 1920-1932: 10 brev

Ms4° 1669b Gaino, Lars J. 187*-1896 9 brev

Ms fol. 1450 Jacobsen, J. Adrian 1938-1940: 4 brev.
[Selvbiogr. oppl. om...J.A.J.]

Brevs. 378b Kolsrud, Oluf 1944 juni 6.

Ms4° 3924E Nielsen, Konrad 1951 april 1

Ms4° 2355 Nilssen, August 1923 mars 19.

Ms4° 1669b Samuelsen, S. (lærer) 1923 aug. 12-15. 1923 sept. 14.

Brevs. 1 Urbye, Andreas Tostrup (1869-) 1908 apr. 6

B.1.3 Avsendte brev (til:)

Brevs. 469B Bjerknes, Vilhelm 1948 april 13

Brevs. 232A Boeck, Thorvald 1898 mars 4 [Spm. ang. Professor Friis boksamling]

Brevs. 2 Bugge, Sophus 1888 febr. 3 , 1897 jan. 31

Brevs. 569 Alb. Cammermeyers forl. 1911 des. 18 [Takkebrev for tilsendt eks. av Johan Meyer "Fortids kunst i Norges bygder"]

Brevs. 707 Fett, Harry 1940 apr. 10 [Ang. opplysninger om John Savio]

Brevs. 140 Garborg, Arne 1911 mars 30, 1911 april 3 [Ang. ordningen med utdeling av forfatterens stipend]

Brevs. 437 Holm, Olaf Andreas 1883 apr. 20 [I forb. med reise for å studere dialekten til samene i Nordland amt ber Q. om oppl. om hvor samene bor, hvor studere språket, informanter osv.]

Ms.fol.3710:1 Koht, Halvdan 1910 mai 24, (Ms.fol.3713:1) 1919 sept. 25, (Brevs. 386) 1920 jan. 21, 1921 jan. 15, 1921 febr. 28, 1924 nov. 29, 1950 sept. 8

Ms.fol. 2240 Munthe, Wilhelm 1946 juni 16 (avskrift) [Ang. fortegnelse over uteksaminerte elever ved Tromsø Lærerskole]

Ms.fol. 1924 6b,3 Nansen, Fridtjof 1910 juni 1, des. 7

Brevs. 639 Nielsen, Yngvar (udat.), 1910 april 14, 1911 juli 8

Brevs. 438 Nissen, Kristian (1879-1968) 1915 des. 9, 1922 aug. 8, 1951 jan. 18, april 14, nov. 24, 1952 juni 14, 1953 april 11, des. 18

Brevs. 375 Olsen, Magnus 1909-1956 (64 brev) + 1923 mai 7 (Ms4° 2355)

Brevs. 375 Pettersen, Hjalmar 1899 sept. 16, 1924 mai 10, 1922 sept. 26

Brevs. 86 Storm, Gustav 1893 okt. 20, 1901 mai 19

Ms4^o 1150 Universitetsbiblioteket 1915 mai 13

Ms.film 98 Wiklund, K.B. 1891-1906 66 brev (mikrofilm)
(Mange brev til prof. K.B. Wiklund ble i 1967 gitt til Uppsala Universitetsbibliotek. I alt finnes 96 brev fra Qvigstad til Wiklund ved UUBs håndskriftsamling)

Brevs. 1 Wiig, Margarethe 1938 april 28 (xeroxkopi)

Brevs. 97 Wille, Nordal 1901 april 23, aug. 26, 1902 febr. 28, 1904 febr. 13

B.2 Brev fra Qvigstad i finske arkiver og bibliotek

B.2.1 Riksarkivet

Avsendte brev (til):

Emil N. Setälä 1887-1935 96 brev (i prof. Setäläs privatarkiv)

Nils Isak Fellman 1899-1916 49 brev (i släkten Fellmans samling)

B.2.2 Finsk-Ugriska Sällskapet

1900-1953: 15 brev til Finsk-Ugriska Sällskapet (hvorav to brev til K.F. Krjalainen, ett brev til Selskapets bibliotek, resten til Selskapets ledelse og sekretær).

B.2.3 Helsingfors universitetsbibliotek

Avsendte brev (til):

Theodor Petter Schvindt 1898-1906 3 brev

B.2.4 Finska litteratursällskapet

Avsendte brev (til):

August Ahlqvist 1880-1889 8 brev

Oscar Adolf Hainari (Forsstöm) 1897 1 brev

Helmi Krohn 1948, 1955 2 brev

Kaarle Krohn 1894-1932 23 brev

Martti Rapola 1935, 1940 2 brev (& utkast til Martti Rapolas brev til J.K. Qvigstad 1935)

Finska Litteratursällskapet 1899-1931 6 brev

E.A. Tunkelo 1924 1 brev

B.2.5 Uleåborgs landsarkiv

Avsendte brev (til):

Jenny Paulaharju 1947 1 brev (i Paulaharjus samling)

Tillegg C

Databeskrivelser

C.1 DTD for tekstlig materiale

```
<!– DOKPRO delprosjekt UB Tromsø, avd. Tromsø museum –>

<!– DTD for innskriving av bøker med Qvigstads notater –>

<!– Alle entiteter har UNICODE-verdien for et tegn eller en tegnkombinasjon som kan benyttes ved presentasjon av en-
titeten i kommentarfeltet. De entitetene som ikke ble funnet i Unicode-standarden, har verdien xxxx.–>

<!– Enkelte entiteter er definert for både store og små bokstaver, selv om kun de små bokstavene er brukt. –>

<!– ver. 13.1, Øyvind Eide 22.12.97. –>

<!– ===================================== –>
<!– HOVEDPOST FOR DOKUMENTET –>
<!– ===================================== –>

  <!ELEMENT    dokument   - -      (bok | brev*)
                                   +(vskille | side | snr | spalte | spalter | gblyant | rblyant | bblyant | blekk |
                                   overstr | usikker | samisk) >

  <!ELEMENT    usikker    - -      ((#PCDATA | avsn)*) >
  <!ATTLIST    usikker
               komm       CDATA    #IMPLIED >
<!– Etterhvert bør vel alle usikker som blir stående ha en slik kommentar –>

<!– ===================================== –>
<!– HOVEDPOST FOR BØKER –>
<!– ===================================== –>

  <!ELEMENT    bok   - -   (titside?, (innhfort | kap)+) >


<!– ===================================== –>
<!– SIDER OG SPALTER –>
<!– ===================================== –>

  <!ELEMENT    side   - o      EMPTY >
  <!ATTLIST    side
               fil    CDATA    #IMPLIED >
<!– I fil legges navnet på bildefila som er faksimileutgaven av siden som starter med side-taggen –>

  <!ELEMENT    snr       - o      EMPTY >
  <!ATTLIST    snr
               nr        CDATA    #REQUIRED
               topptxt   CDATA    #IMPLIED >
<!– Brukes i Schnitler fordi topptekstene i visse tilfelle inneholder informasjon i en form som ikke finnes i teksten –>
```

```
 <!ELEMENT    spalter    - o         EMPTY >
 <!ATTLIST    spalter
              sp         CDATA       #REQUIRED >
<!-- Settes inn der antall spalter skifter. Spalteskille markeres med spalte. Tekst uten denne tagen har alltid én spalte -->

 <!ELEMENT    spalte     - o     EMPTY >


<!-- ==================================== -->
<!-- HÅNDSKR. NOTATER UNDER "dokmument" -->
<!-- ==================================== -->

 <!ELEMENT    gblyant    - -         (#PCDATA)
                                     +(table | kilde | henv | snavn | pnavn | dato | hevet | gstrek | rstrek | bstrek |
                                     blekkstr) >

 <!ELEMENT    rblyant    - -         (#PCDATA)
                                     +(kilde | henv | snavn | pnavn | dato | hevet | gstrek | rstrek | bstrek | blekkstr) >

 <!ELEMENT    bblyant    - -         (#PCDATA)
                                     +(kilde | henv | snavn | pnavn | dato | hevet | gstrek | rstrek | bstrek | blekkstr) >

 <!ELEMENT    blekk      - -         (#PCDATA)
                                     +(kilde | henv | snavn | pnavn | dato | hevet | gstrek | rstrek | bstrek | blekkstr) >

 <!ELEMENT    overstr    - -         ((#PCDATA | erstatt)+) >
 <!ATTLIST    overstr
              type       CDATA       #IMPLIED >
<!-- type kan være en av blyanttypene, eller visket dersom noe er visket ut -->

 <!ELEMENT    erstatt    - -     (#PCDATA) >


<!-- ==================================== -->
<!-- TITTELSIDE -->
<!-- ==================================== -->

 <!ELEMENT    titside    - -         (#PCDATA)
                                     +(tittel | del | forf | konf | utg | sted | tid) >

 <!ELEMENT    tittel     - -         (#PCDATA)
                                     +(snavn | pnavn | dato) >

 <!ELEMENT    del        - -         (#PCDATA) >

 <!ELEMENT    forf       - -         (#PCDATA) >
 <!ATTLIST    forf
              sort       CDATA       #REQUIRED >

 <!ELEMENT    konf       - -         (#PCDATA) >
<!-- D.e. konferanse, koroprasjon o.l. -->

 <!ELEMENT    utg    - -    (#PCDATA) >
<!-- D.e. utgiver (forlag) -->

 <!ELEMENT    sted    - -    (#PCDATA) >

 <!ELEMENT    tid     - -    (#PCDATA) >

<!-- ==================================== -->
<!-- INNHOLD -->
<!-- ==================================== -->

 <!ELEMENT    innhfort    - -    ((avsn | overskr | brev | table)+) >


<!-- ==================================== -->
<!-- KAPITLER -->
<!-- ==================================== -->

 <!ELEMENT    kap        - -         (ktittel?, innhfort?, (mkap | ukap | avsn | overskr | brev | table)+) >

 <!ELEMENT    ktittel    - -         (#PCDATA)
                                     +(snavn | pnavn | dato | hfeit | hevet | kur | mstilt | hstilt | gstrek | rstrek |
                                     bstrek | blekkstr) >
 <!ATTLIST    ktittel
              par        CDATA       #IMPLIED >
```

```
<!– Brukes til å lenke parallelltekstede sekvenser, som på s. 52 ff. –>

  <!ELEMENT    mkap       - -      (mkaptit?, innhfort?, (ukap | avsn | overskr | brev | table)+) >

  <!ELEMENT    mkaptit    - -      (#PCDATA)
                                   +(snavn | pnavn | dato | kilde | henv | hfeit | hevet | kur | mstilt | hstilt |
                                   gstrek | rstrek | bstrek | blekkstr) >

  <!ELEMENT    ukap       - -      (ukaptit?, innhfort?, (avsn | overskr | brev | table)+) >

  <!ELEMENT    ukaptit    - -      (#PCDATA)
                                   +(snavn | pnavn | dato | kilde | henv | hfeit | hevet | kur | mstilt | hstilt |
                                   gstrek | rstrek | bstrek | blekkstr) >
  <!ATTLIST    ukaptit
               par        CDATA    #IMPLIED >
<!– Brukes til å lenke parallelltekstede sekvenser, som på s. 132 ff. –>

  <!ATTLIST    (kap | mkap | ukap)
               overskrift          CDATA    #IMPLIED >
<!– Overskriftene hos Schnitler som er hentet fra innholdsfortegnelsen. –>


<!– ==================================== –>
<!– AVSNITT –>
<!– ==================================== –>

  <!ELEMENT    avsn     - -      (#PCDATA)
                                 +(kilde | henv | snavn | pnavn | dato | kur | hfeit | mstilt | hstilt | sperring |
                                 hevet | senket | gstrek | rstrek | bstrek | blekkstr | gvert | rvert | bvert |
                                 blekkvrt | fotnote | innskudd | vertikal) >
  <!ATTLIST    avsn
               par      CDATA    #IMPLIED >
<!– Par brukes til å lenke parallelltekstede sekvenser, som på s. 52 ff. –>

  <!ELEMENT    vskille   - o      EMPTY >
  <!ATTLIST    vskille
               merke     CDATA    #IMPLIED >

  <!ELEMENT    overskr   - -      (#PCDATA)
                                  +(snavn | pnavn | dato | kilde | henv | fotnote | hfeit | kur | mstilt | hstilt |
                                  sperring | hevet | senket | gstrek | rstrek | bstrek | blekkstr) >
  <!ATTLIST    overskr
               par       CDATA    #IMPLIED >
<!– Brukes til å lenke parallelltekstede sekvenser, som på s. 52 ff. –>


<!– ==================================== –>
<!– KILDE OG HENVISNINGER –>
<!– ==================================== –>

  <!ELEMENT    kilde     - -      (#PCDATA) >
  <!ATTLIST    kilde
               lopenr    NUMBER   #IMPLIED
– Brukes til MikroMARC-løpenummer ettersom poster med opplysninger til erstatning for de øvrige attributter legges inn.
Skal evt. byttes ut med ID-lenking senere. –>

  <!ELEMENT    henv     - -    (#PCDATA) >
<!– Brukes til interne henvisninger, foreløpig kun i registrene. –>


<!– ==================================== –>
<!– STEDSNAVN –>
<!– ==================================== –>

  <!ELEMENT    snavn    - -    (#PCDATA | beskr) >

  <!ELEMENT    beskr    - -    (#PCDATA)
                               +(snavn | kur | hfeit | mstilt | hstilt | sperring | hevet | senket | gstrek | gvert | rvert |
                               bvert | blekkvrt | rstrek | bstrek | blekkstr) >
<!– beskr og karthenv kan brukes til videre tagging av registeret "Identificering av stedsnavne" –>
```

```
<!- ==================================== ->
<!- SAMISK ->
<!- ==================================== ->

  <!ELEMENT    samisk    - -    ((#PCDATA | avsn | table)+) >


<!- ==================================== ->
<!- PERSONNAVN ->
<!- ==================================== ->

  <!ELEMENT    pnavn    - -    (#PCDATA) >
<!- attributtet sorter fjernet. ->



<!- ==================================== ->
<!- DATO ->
<!- ==================================== ->

  <!ELEMENT    dato          - -          (#PCDATA) >
  <!ATTLIST    dato
               standard    CDATA    #REQUIRED >

<!- ==================================== ->
<!- DIVERSE GRAFISKE OG HÅNDSKRVNE ->
<!- ==================================== ->

  <!ELEMENT    kur         - -    (#PCDATA) >

  <!ELEMENT    hfeit       - -    (#PCDATA) >

  <!ELEMENT    mstilt      - -    (#PCDATA) >

  <!ELEMENT    hstilt      - -    (#PCDATA) >

  <!ELEMENT    sperring    - -    (#PCDATA) >

  <!ELEMENT    hevet       - -    (#PCDATA) >

  <!ELEMENT    senket      - -    (#PCDATA) >

  <!ELEMENT    gstrek      - -    (#PCDATA) >

  <!ELEMENT    rstrek      - -    (#PCDATA) >

  <!ELEMENT    bstrek      - -    (#PCDATA) >

  <!ELEMENT    blekkstr    - -    (#PCDATA) >

  <!ELEMENT    gvert       - -    (#PCDATA) >

  <!ELEMENT    rvert       - -    (#PCDATA) >

  <!ELEMENT    bvert       - -    (#PCDATA) >

  <!ELEMENT    blekkvrt    - -    (#PCDATA) >

<!- ==================================== ->
<!- BREV ->
<!- ==================================== ->

<!- Her kunne vi kanskje strukturert rekkefølgen noe mer? ->

  <!ELEMENT    brev    - -        (((avsn | overskr | table | samisk)+), etterskr?, vedlegg*)
                                  +(sted.tid | innledn | signatur | vertikal) >
  <!ATTLIST    brev
               avs     CDATA    #IMPLIED
               mott    CDATA    #IMPLIED >
<!- Disse to settes midlertidig til IMPLIED fordi innholdet for Qvigstad-brevene legges inn automatisk i ettertid. ->
```

```
  <!ELEMENT    innledn    - -   (#PCDATA)
                                +(snavn | pnavn | dato | kur | hfeit | mstilt | hstilt | sperring | hevet | senket |
                                gstrek | rstrek | bstrek | blekkstr) >

  <!ELEMENT    sted.tid   - -   (#PCDATA)
                                +(snavn | dato | kur | hfeit | mstilt | hstilt | sperring | hevet | senket | gstrek |
                                rstrek | bstrek | blekkstr) >

  <!ELEMENT    signatur   - -   ((#PCDATA | avsn)+)
                                +(snavn | pnavn | dato | kur | hfeit | mstilt | hstilt | sperring | hevet | senket |
                                gstrek | rstrek | bstrek | blekkstr) >
<!- avsn er puttet inn for å gi mulighet til flerspalter ->

  <!ELEMENT    etterskr   - -   (#PCDATA)
                                +(snavn | pnavn | dato | kilde | henv | kur | hfeit | mstilt | hstilt | sperring | hevet |
                                senket | innskudd | gstrek | rstrek | bstrek | blekkstr) >
<!- Skulle etterskr kanskje inneholde kun (avsn)? ->

  <!ELEMENT    vedlegg    - -   ((avsn | overskr | table)+) >


<!- ==================================== ->
<!- TABELL ->
<!- ==================================== ->

<!- Lagt inn mulighet til å gå rett på PCDATA/AVSN for å gjøre det mulig å registrere foreløpig uten å tagge tabellene. ->

  <!ELEMENT    table      - -        ((#PCDATA | avsn) | tr)+ >
  <!ATTLIST    table
               border     CDATA      #IMPLIED >
<!- Skriv inn "+" som verdi for border dersom den skal settes. Det takler Netscape ->

  <!ELEMENT    tr    - -    (th | td)+ >
<!- At kilde ligger på cellenivå fører til litt "klossete" tagging fordi det blir kun tittel eller forfatter som havner inni
taggen. Det godtas fordi kilde er en typist attributt-tag ->

  <!ELEMENT    (th | td)    - -                   (#PCDATA)
                                                  +(kilde | henv | snavn | pnavn | dato | kur | hfeit | mstilt | hstilt |
                                                  sperring | hevet | senket | gstrek | rstrek | bstrek | blekkstr |
                                                  innskudd | fotnote| vertikal) >
  <!ATTLIST    (th | td)
               align      (left|center|right)     left
               valign     (top|middle|bottom)     middle - Brukes som i HTML 3.2. -
               colspan    CDATA                   #IMPLIED
               rowspan    CDATA                   #IMPLIED >
<!- Disse to attributtene viser om innholdet dekker flere celler. Hvis de ikke brukes, er verdien 1 ->



<!- ==================================== ->
<!- FOTNOTE OG INNSKUDD ->
<!- ==================================== ->

  <!ELEMENT    fotnote     - -   ((#PCDATA | avsn)+)
                                 +(kilde | henv | snavn | pnavn | dato | hevet | senket | gstrek | rstrek | bstrek |
                                 blekkstr) >

  <!ELEMENT    innskudd    - -   ((#PCDATA | avsn)+)
                                 +(kilde | henv | snavn | pnavn | dato | hevet) >
<!- Brukes i Qvigstads brev til å merke tekst puttet inn med hake ->

  <!ELEMENT    vertikal    - -   ((#PCDATA | avsn | etterskr)+)
                                 +(kilde | henv | snavn | pnavn | dato | hevet) >
<!- Brukes i Qvigstads brev til å merke tekst som er skrevet vertikalt i margen og der det ikke finnes noen hake som
indikerer hvor den skal puttes inn. Plasseringen er derved ikke definert presist. ->
```

```
<!- ==================================== ->
<!- ENTITETER ->
<!- ==================================== ->

  <!ENTITY   mdash       SDATA   "[—]"          -=tankestrek (UNICODE 2012-2015?) ->
  <!ENTITY   bull        SDATA   "[·]"          -=prikk midt på linja (UNICODE 00B7) ->
  <!ENTITY   bmdvs       SDATA   "[):]"         -=det vil si (UNICODE xxxx) ->
  <!ENTITY   bmminus     SDATA   "[÷]"          -=minustegn med prikker (UNICODE 00F7) ->
  <!ENTITY   bmpund      SDATA   "[pund]"       -=pund (vektenhet) (UNICODE xxxx) ->
  <!ENTITY   bmmark      SDATA   "[mark]"       -=mark (pengeenhet) (UNICODE xxxx) ->
  <!ENTITY   bmskilling  SDATA   "[skilling]"   -=skilling (pengeenhet) (UNICODE xxxx) ->
  <!ENTITY   bmdagger    SDATA   "[dagger]"     -=Dagger (UNICODE 2020) ->
  <!ENTITY   anflav      SDATA   "[anflav]"     -=Lavt anførselstegn (UNICODE 201E) ->
  <!ENTITY   anfhoy      SDATA   "[anfhoy]"     -=Høyt anførselstegn (UNICODE 201D) ->
  <!ENTITY   anfh        SDATA   "[≪]"          -=Anførselstegn: << (UNICODE 00AB) ->
  <!ENTITY   anfv        SDATA   "[≫]"          -=Anførselstegn: >> (UNICODE 00BB) ->
  <!ENTITY   grad        SDATA   "[°]"          -=Grader(UNICODE 00B0) ->
  <!ENTITY   gt          SDATA   "[>]"          -=Større enn (UNICODE 003E) ->
  <!ENTITY   lt          SDATA   "[<]"          -=Mindre enn (UNICODE 003C) ->
  <!ENTITY   amp         SDATA   "[&]"          -= (UNICODE 0026) ->
  <!ENTITY   larr        SDATA   "[<—]"         -=Pil som peker mot venstre (UNICODE 2190) ->
  <!ENTITY   rarr        SDATA   "[—>]"         -=Pil som peker mot høyre (UNICODE 2192) ->
  <!ENTITY   ominus      SDATA   "[ominus]"     -=Ring med en vannrett strek gjennom midten (UNICODE 2296) ->
  <!ENTITY   oplus       SDATA   "[oplus]"      -=Ring med et kryss i midten (UNICODE 2295) ->
  <!ENTITY   agr         SDATA   "[agr]"        -=liten gresk alfa (UNICODE 03B1) ->
  <!ENTITY   bgr         SDATA   "[bgr]"        -=liten gresk beta (UNICODE 03B2) ->
  <!ENTITY   ggr         SDATA   "[ggr]"        -=liten gresk gamma/ (UNICODE 03B3) ->
  <!ENTITY   dgr         SDATA   "[dgr]"        -=liten gresk delta (UNICODE 03B4) ->
  <!ENTITY   egr         SDATA   "[egr]"        -=liten gresk epsilon (UNICODE 03B5) ->
  <!ENTITY   phgr        SDATA   "[phgr]"       -=liten gresk phi (UNICODE 03D5) ->

<!-Entiteter for bruk til Qvigstads brev, først og fremst samisk lydskrift ->
  <!ENTITY   bue         SDATA   "[bue]"           -=bue (breve) uten bokstav under (UNICODE 02D8)->
  <!ENTITY   caron       SDATA   "[caron]"         -=caron uten bokstav under (UNICODE 02C7) ->
  <!ENTITY   Abue        SDATA   "[Abue]"          -= Buelignende form over bokstaven (UNICODE 0102) ->
  <!ENTITY   abue        SDATA   "[abue]"          -= Buelignende form over bokstaven (UNICODE 0103) ->
  <!ENTITY   Abuestrek   SDATA   "[Abuestrek]"     -= bue øverst, deretter strek over bokstaven
                                                   (UNICODE 0100+0306) ->
  <!ENTITY   abuestrek   SDATA   "[abuestrek]"     -= bue øverst, deretter strek over bokstaven
                                                   (UNICODE 0101+0306) ->
<!- acaron endret til abue ->

<!- acaronvar endret til abuestrek ->
  <!ENTITY   Acirk       SDATA   "[Â]"             -= (UNICODE 00C2) ->
  <!ENTITY   acirk       SDATA   "[â]"             -= (UNICODE 00E2) ->
  <!ENTITY   Acirkbue    SDATA   "[Acirkbue]"      -= cirkumflex øverst, deretter bue (UNICODE 0102+0302) ->
  <!ENTITY   acirkbue    SDATA   "[acirkbue]"      -= cirkumflex øverst, deretter bue (UNICODE 0103+0302) ->
  <!ENTITY   Acirkvar    SDATA   "[Acirkvar]"      -= cirkumflex øverst, deretter strek over bokstaven(UNICODE
                                                   0100+0302) ->
  <!ENTITY   acirkvar    SDATA   "[acirkvar]"      -= cirkumflex øverst, deretter strek over bokstaven(UNICODE
                                                   0101+0302) ->
  <!ENTITY   Adot        SDATA   "[Adot]"          -= Prikk under (UNICODE 1EA0) ->
  <!ENTITY   adot        SDATA   "[adot]"          -= Prikk under (UNICODE 1EA1) ->
  <!ENTITY   Alavgrav    SDATA   "[Alavgrav]"      -= Speilvendt komma under (UNICODE 0104) ->
  <!ENTITY   alavgrav    SDATA   "[alavgrav]"      -= Speilvendt komma under (UNICODE 0105) ->
  <!ENTITY   Aprikk      SDATA   "[Aprikk]"        -= Prikk over(UNICODE 0041+0307) ->
  <!ENTITY   aprikk      SDATA   "[aprikk]"        -= Prikk over(UNICODE 0061+0307) ->
  <!ENTITY   Astrek      SDATA   "[Astrek]"        -= Vannrett strek over (UNICODE 0100) ->
  <!ENTITY   astrek      SDATA   "[astrek]"        -= Vannrett strek over (UNICODE 0101) ->
  <!ENTITY   Astreklg    SDATA   "[Astreklg]"      -= Vannrett strek over, speilvendt komma under (UNICODE
                                                   0100+0328) ->
  <!ENTITY   astreklg    SDATA   "[astreklg]"      -= Vannrett strek over, speilvendt komma under (UNICODE
                                                   0101+0328) ->
  <!ENTITY   Astrekring  SDATA   "[Astrekring]"    -= Å med vannrett strek over (UNICODE 00C5+0304) ->
  <!ENTITY   astrekring  SDATA   "[astrekring]"    -= å med vannrett strek over (UNICODE 00E5+0304) ->
  <!ENTITY   Atilde      SDATA   "[Ã]"             -= (UNICODE 00C3) ->
  <!ENTITY   atilde      SDATA   "[ã]"             -= (UNICODE 00E3) ->
  <!ENTITY   avar        SDATA   "[avar]"          -= Brukes der Qvigstad har skrevet en trykkskriftlignende
                                                   'a' i stedet for sin vanlige håndskrift-a (UNICODE 0061) ->
  <!ENTITY   Avarcirk    SDATA   "[Avarcirk]"      -= strek øverst, deretter cirkumflex over bokstaven
                                                   (UNICODE 00C2+) ->
  <!ENTITY   avarcirk    SDATA   "[avarcirk]"      -= strek øverst, deretter cirkumflex over bokstaven
                                                   (UNICODE 00E2+) ->
<!- avarstrek endret til astreklg ->
```

```
  <!ENTITY   bgrave      SDATA   "[bgrave]"      –= (UNICODE 0062+0300) –>
  <!ENTITY   Bstrek      SDATA   "[Bstrek]"      –= Vannrett strek over(UNICODE 0042+0304) –>
  <!ENTITY   bstrek      SDATA   "[bstrek]"      –= Vannrett strek over(UNICODE 0062+0304) –>
  <!ENTITY   Bstung      SDATA   "[Bstung]"      –= Loddrett strek gjennom hele bokstaven(UNICODE 0042+0338) –>
  <!ENTITY   bstung      SDATA   "[bstung]"      –= Loddrett strek gjennom hele bokstaven(UNICODE 0062+0338) –>
<!– cbue endret til ccaron –>
  <!ENTITY   Ccaron         SDATA   "[Ccaron]"         –= (UNICODE 010C) –>
  <!ENTITY   ccaron         SDATA   "[ccaron]"         –= (UNICODE 010D) –>
  <!ENTITY   cgrave         SDATA   "[cgrave]"         –= (UNICODE 0063+0300) –>
  <!ENTITY   comv           SDATA   "[comv]"           –= lydskrifttegnet som ser ut som en speilvendt 'c'
                                                       (UNICODE 0254) –>
  <!ENTITY   Cstrek         SDATA   "[Cstrek]"         –= Vannrett strek over (UNICODE 0043+0304) –>
  <!ENTITY   cstrek         SDATA   "[cstrek]"         –= Vannrett strek over (UNICODE 0063+0304) –>
  <!ENTITY   dgrave         SDATA   "[dgrave]"         –= (UNICODE 0064+0300) –>
  <!ENTITY   Dstung         SDATA   "[Dstung]"         –= Loddrett strek gjennom hele bokstaven(UNICODE
                                                       0044+0338) –>
  <!ENTITY   dstung         SDATA   "[dstung]"         –= Loddrett strek gjennom hele bokstaven(UNICODE
                                                       0064+0338) –>
  <!ENTITY   dstungacute    SDATA   "[dstungacute]"    –= Loddrett strek gjennom hele bokstaven, acute over
                                                       (UNICODE 0064+0338+0301) –>
  <!ENTITY   ETH            SDATA   "[ETH]"            –= ETH/tradisjonell stungen D(UNICODE 00D0) –>
  <!ENTITY   eth            SDATA   "[eth]"            –= eth/tradisjonell stungen d(UNICODE 00F0) –>
  <!ENTITY   Ebue           SDATA   "[Ebue]"           –= Bue over(UNICODE 0114) –>
  <!ENTITY   ebue           SDATA   "[ebue]"           –= Bue over(UNICODE 0115) –>
  <!ENTITY   Ebuedot        SDATA   "[Ebuedot]"        –= Bue over, prikk under(UNICODE 0114+0323) –>
  <!ENTITY   ebuedot        SDATA   "[ebuedot]"        –= Bue over, prikk under(UNICODE 0115+0323) –>
<!– ecaron endret til ebue –>

<!– ecarondot endret til ebuedot –>
  <!ENTITY   Ecirk        SDATA   "[Ê]"            –= (UNICODE 00CA) –>
  <!ENTITY   ecirk        SDATA   "[ê]"            –= (UNICODE 00EA) –>
  <!ENTITY   Edot         SDATA   "[Edot]"         –= Prikk under(UNICODE 1EB8) –>
  <!ENTITY   edot         SDATA   "[edot]"         –= Prikk under(UNICODE 1EB9) –>
  <!ENTITY   Elavgrav     SDATA   "[Elavgrav]"     –= Speilvendt komma under(UNICODE 0118) –>
  <!ENTITY   elavgrav     SDATA   "[elavgrav]"     –= Speilvendt komma under(UNICODE 0119) –>
  <!ENTITY   Eprikk       SDATA   "[Eprikk]"       –= Prikk over(UNICODE 0116) –>
  <!ENTITY   eprikk       SDATA   "[eprikk]"       –= Prikk over(UNICODE 0117) –>
  <!ENTITY   Estrek       SDATA   "[Estrek]"       –= Vannrett strek over(UNICODE 0112) –>
  <!ENTITY   estrek       SDATA   "[estrek]"       –= Vannrett strek over(UNICODE 0113) –>
  <!ENTITY   Estrekdot    SDATA   "[Estrekdot]"    –= Vannrett strek over, prikk under(UNICODE 0112+0323) –>
  <!ENTITY   estrekdot    SDATA   "[estrekdot]"    –= Vannrett strek over, prikk under(UNICODE 0113+0323) –>
  <!ENTITY   Eumlstrek    SDATA   "[Eumlstrek]"    –= Strek øverst, deretter prikker over bokstaven(UNICODE
                                                   00CB+0304) –>
  <!ENTITY   eumlstrek    SDATA   "[eumlstrek]"    –= Strek øverst, deretter prikker over bokstaven(UNICODE
                                                   00EB+0304) –>
  <!ENTITY   Gstung       SDATA   "[Gstung]"       –= Loddrett strek gjennom hele bokstaven(UNICODE
                                                   0047+0338) –>
  <!ENTITY   gstung       SDATA   "[gstung]"       –= Loddrett strek gjennom hele bokstaven(UNICODE
                                                   0067+0338) –>
  <!ENTITY   Gspir        SDATA   "[Gspir]"        –= Vannrett strek nederst på bokstaven(UNICODE 01E4) –>
  <!ENTITY   gspir        SDATA   "[gspir]"        –= Vannrett strek nederst på bokstaven(UNICODE 01E5) –>
  <!ENTITY   Ibue         SDATA   "[Ibue]"         –= Buelignende form over bokstaven(UNICODE 012C) –>
  <!ENTITY   ibue         SDATA   "[ibue]"         –= Buelignende form over bokstaven(UNICODE 012D) –>
  <!ENTITY   Ibuedot      SDATA   "[Ibuedot]"      –= Buelignende form over bokstaven, prikk under (UNICODE
                                                   012C+0323) –>
  <!ENTITY   ibuedot      SDATA   "[ibuedot]"      –= Buelignende form over bokstaven, prikk under (UNICODE
                                                   012D+0323) –>
<!– icaron endret til ibue –>
```

```
  <!ENTITY   Ilavbue    SDATA   "[Ilavbue]"    -= Circumflexvendt bue under(UNICODE 0049+032F) ->
  <!ENTITY   ilavbue    SDATA   "[ilavbue]"    -= Circumflexvendt bue under(UNICODE 0069+032F) ->
  <!ENTITY   Istrek     SDATA   "[Istrek]"     -= Vannrett strek over(UNICODE 012A) ->
  <!ENTITY   istrek     SDATA   "[istrek]"     -= Vannrett strek over(UNICODE 012B) ->
  <!ENTITY   kacute     SDATA   "[kacute]"     -= (UNICODE 1E31) ->
  <!ENTITY   kgrave     SDATA   "[kgrave]"     -= (UNICODE 006B+0300) ->
  <!ENTITY   lgrave     SDATA   "[lgrave]"     -= (UNICODE 006C+0300) ->
  <!ENTITY   Lstrek     SDATA   "[Lstrek]"     -= Vannrett strek over(UNICODE 004C+0304) ->
  <!ENTITY   lstrek     SDATA   "[lstrek]"     -= Vannrett strek over(UNICODE 006C+0304) ->
  <!ENTITY   Eng        SDATA   "[Eng]"        -= Som i moderne nordsamisk (UNICODE 014A) ->
  <!ENTITY   eng        SDATA   "[eng]"        -= Som i moderne nordsamisk (UNICODE 014B) ->
  <!ENTITY   nacute     SDATA   "[nacute]"     -= (UNICODE 0144) ->
  <!ENTITY   Ncaron     SDATA   "[Ncaron]"     -= (UNICODE 0147) ->
  <!ENTITY   ncaron     SDATA   "[ncaron]"     -= (UNICODE 0148) ->
  <!ENTITY   Nstrek     SDATA   "[Nstrek]"     -= Vannrett strek over(UNICODE 004E+0304) ->
  <!ENTITY   nstrek     SDATA   "[nstrek]"     -= Vannrett strek over(UNICODE 006E+0304) ->
  <!ENTITY   Obue       SDATA   "[Obue]"       -= Buelignende form over(UNICODE 014E) ->
  <!ENTITY   obue       SDATA   "[obue]"       -= Buelignende form over(UNICODE 014F) ->
  <!ENTITY   Obuelg     SDATA   "[Obuelg]"     -= Speilvendt komma under, buelignende form over(UNICODE
                                               014E+0328) ->
  <!ENTITY   obuelg     SDATA   "[obuelg]"     -= Speilvendt komma under, buelignende form over(UNICODE
                                               014F+0328) ->
<!- ocaron endret til obue ->
  <!ENTITY   Odot         SDATA   "[Odot]"         -= Prikk under(UNICODE 1ECC) ->
  <!ENTITY   odot         SDATA   "[odot]"         -= Prikk under(UNICODE 1ECD) ->
  <!ENTITY   oe           SDATA   "[oe]"           -= oe (UNICODE 0153) ->
  <!ENTITY   Okvist       SDATA   "[Okvist]"       -= Strek opp til høyre fra bokstaven(UNICODE 01A0) ->
  <!ENTITY   okvist       SDATA   "[okvist]"       -= Strek opp til høyre fra bokstaven(UNICODE 01A1) ->
  <!ENTITY   Olavbue      SDATA   "[Olavbue]"      -= Circumflexvendt bue under(UNICODE 004F+032F) ->
  <!ENTITY   olavbue      SDATA   "[olavbue]"      -= Circumflexvendt bue under(UNICODE 006F+032F) ->
  <!ENTITY   Olavgrav     SDATA   "[Olavgrav]"     -= Speilvendt komma under(UNICODE 01EA) ->
  <!ENTITY   olavgrav     SDATA   "[olavgrav]"     -= Speilvendt komma under(UNICODE 01EB) ->
  <!ENTITY   Oprikk       SDATA   "[Oprikk]"       -= Prikk over(UNICODE 004F+0307) ->
  <!ENTITY   oprikk       SDATA   "[oprikk]"       -= Prikk over(UNICODE 006F+0307) ->
  <!ENTITY   Ostrek       SDATA   "[Ostrek]"       -= Vannrett strek over(UNICODE 014C) ->
  <!ENTITY   ostrek       SDATA   "[ostrek]"       -= Vannrett strek over(UNICODE 014D) ->
  <!ENTITY   Ostreklg     SDATA   "[Ostreklg]"     -= Speilvendt komma under, strek over(UNICODE 01EC) ->
  <!ENTITY   ostreklg     SDATA   "[ostreklg]"     -= Speilvendt komma under, strek over(UNICODE 01ED) ->
  <!ENTITY   Otilde       SDATA   "[Õ]"            -= (UNICODE 00D5) ->
  <!ENTITY   otilde       SDATA   "[õ]"            -= (UNICODE 00F5) ->
  <!ENTITY   Oumlstrek    SDATA   "[Oumlstrek]"    -= Strek øverst, deretter prikker over bokstaven(UNICODE
                                                   00D6+0304) ->
  <!ENTITY   oumlstrek    SDATA   "[oumlstrek]"    -= Strek øverst, deretter prikker over bokstaven(UNICODE
                                                   00F6+0304) ->
  <!ENTITY   Oumlacute    SDATA   "[Oumlacute]"    -= ´ øverst, deretter prikker over bokstaven(UNICODE
                                                   00D6+0301) ->
  <!ENTITY   oumlacute    SDATA   "[oumlacute]"    -= ´ øverst, deretter prikker over bokstaven(UNICODE
                                                   00F6+0301) ->
  <!ENTITY   pgrave       SDATA   "[pgrave]"       -= (UNICODE 0070+0300) ->
  <!ENTITY   rgrave       SDATA   "[rgrave]"       -= (UNICODE 0072+0300) ->
  <!ENTITY   Sacute       SDATA   "[Sacute]"       -= (UNICODE 015A) ->
  <!ENTITY   sacute       SDATA   "[sacute]"       -= (UNICODE 015B) ->
<!- sbue endret til scaron ->
  <!ENTITY   Scaron     SDATA   "[Scaron]"     -= (UNICODE 0160) ->
  <!ENTITY   scaron     SDATA   "[scaron]"     -= (UNICODE 0161) ->
  <!ENTITY   sgrave     SDATA   "[sgrave]"     -= (UNICODE 0073+0300) ->
  <!ENTITY   Sigma      SDATA   "[Sigma]"      -= (UNICODE 03A3) ->
  <!ENTITY   sigma      SDATA   "[sigma]"      -= (UNICODE 03C3) ->
  <!ENTITY   Sstrek     SDATA   "[Sstrek]"     -= Vannrett strek over (UNICODE 0053+0304) ->
  <!ENTITY   sstrek     SDATA   "[sstrek]"     -= Vannrett strek over (UNICODE 0073+0304) ->
  <!ENTITY   szlig      SDATA   "[ß]"          -= (UNICODE 00DF) ->
  <!ENTITY   tgrave     SDATA   "[tgrave]"     -= (UNICODE 0074+0300) ->
  <!ENTITY   Tstung     SDATA   "[Tstung]"     -= Stungen t som i moderne nordsamisk(UNICODE 0166) ->
  <!ENTITY   tstung     SDATA   "[tstung]"     -= Stungen T som i moderne nordsamisk(UNICODE 0167) ->
  <!ENTITY   Ubue       SDATA   "[Ubue]"       -= Buelignende form over bokstaven (UNICODE 016C) ->
  <!ENTITY   ubue       SDATA   "[ubue]"       -= Buelignende form over bokstaven (UNICODE 016D) ->
<!- ucaron endret til ubue ->
```

```
  <!ENTITY    Ulavbue    SDATA    "[Ulavbue]"    -= Circumflexvendt bue under(UNICODE 0055+032F) ->
  <!ENTITY    ulavbue    SDATA    "[ulavbue]"    -= Circumflexvendt bue under(UNICODE 0075+032F) ->
  <!ENTITY    Uprikk     SDATA    "[Uprikk]"     -=Prikk over(UNICODE 0055+0307) ->
  <!ENTITY    uprikk     SDATA    "[uprikk]"     -= Prikk over(UNICODE 0075+0307) ->
  <!ENTITY    Ustrek     SDATA    "[Ustrek]"     -= Vannrett strek over(UNICODE 016A) ->
  <!ENTITY    ustrek     SDATA    "[ustrek]"     -= Vannrett strek over(UNICODE 016B) ->
  <!ENTITY    Vcaron     SDATA    "[Vcaron]"     -= (UNICODE 0056+030C) ->
  <!ENTITY    vcaron     SDATA    "[vcaron]"     -= (UNICODE 0076+030C) ->
  <!ENTITY    vgrave     SDATA    "[vgrave]"     -= (UNICODE 0076+0300) ->
  <!ENTITY    Zacute     SDATA    "[Zacute]"     -= (UNICODE 0179) ->
  <!ENTITY    zacute     SDATA    "[zacute]"     -= (UNICODE 017A) ->
  <!ENTITY    Zcaron     SDATA    "[Zcaron]"     -= (UNICODE 017D) ->
  <!ENTITY    zcaron     SDATA    "[zcaron]"     -= (UNICODE 017E) ->
  <!ENTITY    zgrave     SDATA    "[zgrave]"     -= (UNICODE 007A+0300) ->
  <!ENTITY    Aestrek    SDATA    "[Aestrek]"    -= Vannrett strek over(UNICODE 01E2) ->
  <!ENTITY    aestrek    SDATA    "[aestrek]"    -= Vannrett strek over(UNICODE 01E3) ->
<!- aecaron endret til aebue ->
  <!ENTITY    Aebue    SDATA    "[Aebue]"    -= (UNICODE 00C6+0306) ->
  <!ENTITY    aebue    SDATA    "[aebue]"    -= (UNICODE 00E6+0306) ->
```

C.2 DTD for bibliografiske referanser

```
<!– DOKPRO delprosjekt UB Tromsø, avd. Tromsø museum –>
<!– DTD for bibliografien til DOKUB.DTD-materiale –>
<!– ver. 1.0, Øyvind Eide 12.12.97 –>

<!ELEMENT  bib                                       - -  (post+)>
<!ELEMENT  post                                      - o  (lopenr, forf?, tittel+, utg?, (bibsys | komm)*)>
<!ELEMENT  lopenr                                    - o  (#PCDATA)>
<!ELEMENT  forf                                      - o  (#PCDATA)>
<!ELEMENT  tittel                                    - o  (htittel, utittel?, ansvar?, del?, deltit?)>
<!ELEMENT  (htittel, utittel, ansvar, del, deltit)   - o  (#PCDATA)>
<!ELEMENT  utg                                       - o  (sted?, tid?)>
<!ELEMENT  (tid, sted)                               - o  (#PCDATA)>
<!ELEMENT  bibsys                                    - o  (objektid, merknad?)>
<!ELEMENT  (objektid, merknad)                       - o  (#PCDATA)>
<!ELEMENT  komm                                      - o  (#PCDATA)>
```

Tillegg D

MARC-delmengden

*082	$a	Dewey (brukes i Norsk lokalhistorisk bibl.)
*100	$a	Forfatter
	$w	Sorteringsfelt
*245	$a	Hovedtittel
	$b	Undertittel
	$c	Ansvarhavende
	$n	Nr. i flerbindsverk
	$p	Tittel på bind
	$w	Sorteringsfelt for hovedtittel
*250	$a	Utgave
	$b	Ansvarlig
*260	$a	Utg. sted. (260 brukes ikke på analyttposter)
	$b	Forlag
	$c	År
*300	$a	Omfang
	$b	Illustrert
	$c	Format
	$e	Vedlegg
*440	$a	Serie
	$n	Nummer i underserie
	$p	Navn på underserie
	$v	Nummer i serien
	$w	Sorteringsfelt
*500	$a	Generell note

*651	$a	Norsk stedsnavn. Foreløpig ikke brukt.
	$1	Klassifikasjon i henhold til Elvestrands standard
*653	$a	Her legges overskrifter. Indikator 2: For geografiske overskrifter 1, for andre overskrifter 2.
*700	$a	Medforfatter el. andre biinnførsler på person
	$e	Funksjon
	$t	Biinnførsel - tittel på verk
	$w	Sorteringsfelt
*740	$a	Biinnførsel - tittel på verk
*773	$a	Forfatter på verk analytten er hentet fra
	$d	Sted, forlag, år på verk analytten er hentet fra
	$g	Vol, dato, o.l. for tidsskrifter/aviser
	$k	Serie
	$i	Årstall i ren form
	$t	Tittel på verk/tidsskrift/avis analytten er hentet fra
	$u	Sorteringsfelt for tittel
*850	$a	TOPOGRAFI
	$b	Katalogiseringskilde. Her brukes et fast sett koder: DB, SC, S1, S2, S3, NL, NH
	$c	Sted i kilden - sidetall eller postnummer

Tillegg E

Konvertering fra MARC til SGML

Det var opprinnelig planlagt å konvertere bibliografien til SGML. Dette fant vi ut at ikke var nødvendig. Derimot ble kildene til det fulltekstregistrerte materialet registrert i NORMARC-formatet ved hjelp av programmet MikroMARC og senere konvertert til SGML. Her følger en gjennomgang av noen grunnprinsipper i en slik konvertering.

E.1 Hvilken SGML?

Det er ikke vanskelig å lage en generell DTD som dekker alle typer MARC-poster. Den kunne for eksempel se omtrent slik ut:

```
<!ELEMENT  biblio  - -      (post+) >
<!ELEMENT  post    - o      (f000 | f001 | ... | f999) >
<!ELEMENT  (f000 | f001 | ... | f999)
                   - o      ((dfa | dfb | ... | dfz | df0 | df1 | ... | df9)+) >
<!ATTLIST  (f000 | f001 | ... | f999)
           ind1    CDATA    #IMPLIED
           ind2    CDATA    #IMPLIED >
<!ELEMENT  (dfa | dfb | ... | dfz | df0 | df1 | ... | df9)
                   - o      (#PCDATA) >
```

Konverteringen vi da bli svært enkel, f.eks. kan

```
*10010$aHansen, Hans
*245 1$aMin bok
$cHans Hansen
```

konverteres til

```
<biblio><post><f100 ind1=''1'' ind2=''0''><dfa>Hansen, Hans
<f245 ind2=''1''><dfa>Min bok
<dfc>Hans Hansen
</biblio>
```

For noen typer bruk er sikkert dette en grei løsning, men denne "SGML i form, MARC i innhold" egner seg dårlig for vårt bruk. Vi har derfor laget en strukturert DTD der vi:

1. Kun definerer felt vi bruker (se tillegg D)
2. Lager en struktur basert på NORMARC-delmengden, men hvor vi kobler sammen felt som naturlig hører sammen

Selve konverteringsprosessen består av enkel erstatning av tekststrenger og blir ikke beskrevet her.

Tillegg F

RTF-konvertering

Disse programlistingene viser et eksempel på hvordan man kan konvertere fra Word-format til SGML. Programmene er ikke spesielt sofistikerte, de tar utgangspunkt i "Identifisering av stedsnavne" slik den ser ut etter at innskriverne har lagt inn Word-koder for formatet, slik at avsnitt, kursiv, halvfet og andre synlige aspekter ved teksten er merket ut. Deretter er teksten eksportert i Words tekstformat RTF (Rich text format), hvis spesifikasjon er tilgjengelig fra Microsofts WWW-sider.[1]

Selve programmene er ment å kjøres fortløpende, slik at utfilen fra ett er innfilen til neste. De kan neppe benyttes til andre konverteringer slik de står, men kan brukes som inspirasjon for andre som vil skrive skreddersydde konverteringsprogram.

Programmene finnes også i modifiserte versjoner for kjøring på "Sagregister".

F.1 rtfhex

```
/*---------------------------------------------------------------------*/
/* rtfhex.c  Konverterer fra en RTF-fil alle spesialtegn,              */
/*           dvs. alle verdier nn i \'nn, til 8-bits tegn.             */
/*           Konverterer også spesialtegnene fra Fauske, [, ], \{, \}. */
/*           For linjeskift settes <n> inn. Dette betyr at følgende    */
/*           substitusjoner må gjøres: LF-->SPACE og <n>-->LF          */
/*           Setter inn punkttagger for \par, \line, \page, \emash,    */
/*           mens \i beholdes. Alle andre \'er fjernes.                */
/*                                                                     */
/*                                                                     */
/*                                                                     */
/* Ver. 1.0 1997-04-28 Øyvind Eide/Dokumentasjonsprosjektet            */
/*                                                                     */
/*---------------------------------------------------------------------*/

#include <stdio.h>
```

[1] Disse sidene finnes på http://www.microsoft.com

```
#include <string.h>

void main(argc, argv)
int argc;
char **argv;
{
    FILE *innfil, *utfil;
    char tegn, tegn2, streng[1000];
    int verdi, i, teller;
    printf("Innfil: %s  Utfil: %s  OK?", argv[1], argv[2]);
    tegn= getchar();
    if (tegn == 'J' || tegn == 'j')
    {
        innfil= fopen(argv[1], "r");
        utfil= fopen(argv[2], "w");
        teller= 0;
        printf("\n\nKode nummer: %5d", teller);
        tegn= getc(innfil);
        while (tegn != EOF)
        {
            if (tegn == '\\')
            {
                tegn2= getc(innfil);
                if (tegn2 == '\'')
                {
                    fscanf(innfil, "%2x", &verdi);
                    putc(verdi, utfil);
                    tegn= getc(innfil);
                }
                else if (tegn2 == '{')
                {
                    fprintf(utfil, "&Cstrek;");
                    tegn= getc(innfil);
                }
                else if (tegn2 == '}')
                {
                    fprintf(utfil, "&cstrek;");
                    tegn= getc(innfil);
                }
                else
                {
                    streng[0]= tegn;
                    streng[1]= tegn2;
                    tegn= getc(innfil);
                    i=2;
                    while (tegn != ' ' && tegn != '\n' && tegn != EOF)
                    {
                        streng[i++]= tegn;
                        tegn= getc(innfil);
                    }
                    tegn= getc(innfil);
                    streng[i]= 0;
                    if (i > 7) ;
                    else if (i == 2 && streng[1] == 'i') fprintf(utfil, "%s ", &streng[0]);
                    else if (i == 4 && streng[1] == 'p' && streng[2] == 'a' &&
                             streng[3] == 'r') fprintf(utfil, "</avsn><n><avsn>");
                    else if (i == 5 && streng[1] == 'l' && streng[2] == 'i' &&
                             streng[3] == 'n' && streng[4] == 'e') fprintf(utfil, "<n>");
                    else if (i == 5 && streng[1] == 'p' && streng[2] == 'a' &&
                             streng[3] == 'g' && streng[4] == 'e') fprintf(utfil, "<n><side><n>");
                    else if (i == 7 && streng[1] == 'e' && streng[2] == 'm' &&
                             streng[3] == 'd' && streng[4] == 'a' && streng[5] == 's' &&
                             streng[6] == 'h') fprintf(utfil, "&mdash;");
                    printf("\b\b\b\b\b%5d", ++teller);
                }
            }
            else if (tegn == '[')
            {
                fprintf(utfil, "&Sstrek;");
                tegn= getc(innfil);
            }
            else if (tegn == ']')
            {
                fprintf(utfil, "&sstrek;");
                tegn= getc(innfil);
            }
            else
```

```
        {
            putc(tegn, utfil);
            tegn= getc(innfil);
        }
      }
    }
    printf("\n\n");
    fclose(innfil);
    fclose(utfil);
}
```

F.2 rtftag

```
/*--------------------------------------------------------------------*/
/* rtftag.c  Skifter ut RTF-formattering med SGML-tager i henhold til   */
/*           DOKUB.DTD. Gjelder kur og snavn + innledende/utledende.    */
/*           Feilmeldinger skrives til rtftag.err                       */
/*                                                                      */
/* Ver. 1.0 1997-04-29 Øyvind Eide/Dokumentasjonsprosjektet             */
/*                                                                      */
/*--------------------------------------------------------------------*/

#include <stdio.h>
#include <string.h>

void main(argc, argv)
int argc;
char **argv;
{
    FILE *innfil, *utfil, *feilfil;
    char tegn, tegn1, tegn2, tegn3, streng[1000];
    int i, teller;
    printf("Innfil: %s  utfil: %s  OK?", argv[1], argv[2]);
    tegn= getchar();
    if (tegn == 'J' || tegn == 'j')
    {
        innfil= fopen(argv[1], "r");
        utfil= fopen(argv[2], "w");
        feilfil= fopen("rtftag.err", "w");
        teller= 0;
        printf("\n\nAvsnitt: %5d", teller);
        fprintf(utfil, "<!DOCTYPE DOKUMENT SYSTEM \"DOKUB.DTD\"><n><dokument><bok><kap><ukap><n><avsn>");
        tegn= getc(innfil);
        while (tegn != EOF)
        {
            while (tegn != '{' && tegn != EOF)
            {
                putc(tegn, utfil);
                tegn= getc(innfil);
            }
            tegn1= getc(innfil);
            tegn2= getc(innfil);
            tegn3= getc(innfil);
            if (tegn1 == '\\' && tegn2 == 'i' && (tegn3 == ' ' || tegn3 == '\n'))
            {
                i= 0;
                tegn= getc(innfil);
                while (tegn != '}' && tegn != EOF)
                {
                  if (tegn == ',')
                    {
                      tegn1= getc(innfil);
                      streng[i]= 0;
                      fprintf(utfil, "<kur><snavn>%s</snavn></kur>%c%c", &streng[0], tegn, tegn1);
                      i= 0;
                    }
                  else if (tegn == '<')
                    {
                      while (tegn != '>' && tegn != EOF)
                    {
                      putc(tegn, utfil);
                      tegn= getc(innfil);
                    }
```

```
                    putc(tegn, utfil);
                    tegn= getc(innfil);
                  }
                else streng[i++]= tegn;
                tegn= getc(innfil);
              }
              streng[i++]= getc(innfil);
              tegn= getc(innfil);
              while (tegn == '{')
              {
                  tegn1= getc(innfil);
                  tegn2= getc(innfil);
                  tegn3= getc(innfil);
                  if (tegn1 == '\\' && tegn2 == 'i' && (tegn3 == ' ' || tegn3 == '\n'))
                  {
                      tegn= getc(innfil);
                      while (tegn != '}' && tegn != EOF)
                      {
                        if (tegn == ',')
                          {
                            tegn1= getc(innfil);
                            streng[i]= 0;
                            fprintf(utfil, "<kur><snavn>%s</snavn></kur>%c%c", &streng[0], tegn, tegn1);
                            i= 0;
                          }
                        else streng[i++]= tegn;
                        tegn= getc(innfil);
                      }
                      streng[i++]= getc(innfil);
                      tegn= getc(innfil);
                  }
                  else
                  {
                      streng[--i]= 0;
                      fprintf(feilfil, "Annet enn kursiv! Sted: %s\n", &streng[0]);
                  }
              }
              tegn1= streng[--i];
              streng[i]= 0;
              fprintf(utfil, "<kur><snavn>%s</snavn></kur>", &streng[0]);
              printf("\b\b\b\b\b%5d", ++teller);
              putc(tegn1, utfil);
              putc(tegn, utfil);
          }
          else fprintf(feilfil, "Annet enn kursiv! Forrige sted var %s, koden er %c.\n", &streng[0], tegn2);
          tegn= getc(innfil);
      }
      printf("\n\n");
      fprintf(utfil, "</avsn><n></ukap></kap></bok></dokument>");
      fclose(innfil);
      fclose(utfil);
      fclose(feilfil);
  }
}
```

F.3 rtfhenv

```
/*----------------------------------------------------------------------*/
/* rtfhenv.c Tagger henvisninger fra registrene til selve teksten i     */
/*           renbeitekommissionen i taggen <henv>.                      */
/*                                                                      */
/*                                                                      */
/* Ver. 1.0 1997-05-12 Øyvind Eide/Dokumentasjonsprosjektet             */
/*      1.1 1997-05-21 Kart legges i <kilde>.                           */
/*                                                                      */
/*----------------------------------------------------------------------*/

#include <stdio.h>
#include <string.h>

void main(argc, argv)
int argc;
char **argv;
```

```
{
    FILE *innfil, *utfil;
    char tegn, streng[1000], tilbake[1000];
    int i, j, teller;
    printf("Innfil: %s  utfil: %s  OK?", argv[1], argv[2]);
    tegn= getchar();
    if (tegn == 'J' || tegn == 'j')
    {
        innfil= fopen(argv[1], "r");
        utfil= fopen(argv[2], "w");
        teller= 0;
        printf("\n\nHenvisning: %5d", teller);
        tegn= getc(innfil);
        while (tegn != EOF)
        {
            if (tegn == 'I')
            {
                i=0;
                streng[i++]= tegn;
                tegn= getc(innfil);
                if (tegn == 'I' || tegn == ' ') /* Tar sjansen på at dette stemmer */
                {
                    printf("\b\b\b\b\b%5d", ++teller);
                    streng[i++]= tegn;
                    tegn= getc(innfil);
                    while ((tegn >= 48 && tegn <= 57) || tegn == ' ' || tegn == 'f' ||
                            tegn == '.' || tegn == ',' || tegn == '\n')
                    {
                        streng[i++]= tegn;
                        tegn= getc(innfil);
                    }
                    j= 0;
                    while (streng[i-1] == ' ' || streng[i-1] == ',' ||
                           streng[i-1] == '.' || streng[i-1] == '\n')
                        tilbake[j++]= streng[--i];
                    streng[i]= 0;
                    fprintf(utfil, "<henv>%s</henv>", &streng[0]);
                    for(i=j-1; i>=0; i--)
                        putc(tilbake[i], utfil);
                }
                else putc(streng[0], utfil);
            }
            /* Topografisk kart over Norge: */
            else if (tegn == 'N' || tegn == 'K' || tegn == 'J' || tegn == 'L' ||
                     tegn == 'H')
            {
                streng[0]= tegn;
                tegn= getc(innfil);
                if (tegn >= 48 && tegn <= 57)
                {
                    i=1;
                    while (tegn >= 48 && tegn <= 57)
                    {
                        streng[i++]= tegn;
                        tegn= getc(innfil);
                    }
                    streng[i]= 0;
                    fprintf(utfil, "<kilde>%s</kilde>", &streng[0]);
                }
                else putc(streng[0], utfil);
            }
            /* Kart over Tromsø Amt i 4 blade: */
            else if (tegn == 'A')
            {
                streng[0]= tegn;
                for (i=1; i<4; i++) streng[i]= getc(innfil);
                streng[4]= 0;
                tegn= getc(innfil);
                if (streng[1] == 'k' && (streng[2] == 'N' || streng[2] == 'S') &&
                    (streng[3] == 'V' || streng[3] == 'O'))
                    fprintf(utfil, "<kilde>%s</kilde>", &streng[0]);
                else fprintf(utfil, "%s", &streng[0]);
            }
            /* Norrlandskarter: */
            else if (tegn >= 48 && tegn <= 57)
            {
                streng[0]= tegn;
```

```
                tegn= getc(innfil);
                i=1;
                while (tegn >= 48 && tegn <= 57)
                {
                    streng[i++]= tegn;
                    tegn= getc(innfil);
                }
                streng[i]= 0;
                fprintf(utfil, "<kilde>%s</kilde>", &streng[0]);
            }
            /* Finland: */
            else if (tegn == 'f')
            {
                streng[0]= tegn;
                for (i=1; i<5; i++) streng[i]= getc(innfil);
                streng[5]= 0;
                if (streng[1] == 'i' && streng[2] == 'n' && streng[3] == 's' &&
                    streng[4] == 'k')
                {
                    for (i=5; i<8; i++) streng[i]= getc(innfil);
                    streng[8]= 0;
                    fprintf(utfil, "<kilde>%s</kilde>", &streng[0]);
                }
                else fprintf(utfil, "%s", &streng[0]);
                tegn= getc(innfil);
            }
            putc(tegn, utfil);
            tegn= getc(innfil);
        }
    }
}
```

F.4 rtfvask1

```
#!/local/bin/perl

$side= 435; #Sjekk startverdi
while (<>) {
s/<avsn><n><side>/<side><n><avsn>/;
s/<side>/<side><snr nr="$side" topptxt="Identificering af stedsnavne">/;
if (/<side>/) {$side++;};
s/\s+/ /g;
print "$_";
}
```

F.5 rtfvask2

```
#!/local/bin/perl
while (<>) {
s/ +<\/avsn>/<\/avsn>/g;
s/ +/ /g;
s/<n>/\n/g;
print "$_";
}
```

F.6 rtflik

```
/*----------------------------------------------------------------------*/
/* rtflik.c Tagger stedsnavn som står etter = se og jfr                 */
/*          Enkelt og banalt; må ryddes opp i korrektur.                */
/*          Det skaper visse problemer at se og jfr av og til benyttes  */
/*          til andre henvisninger (særlig jfr sagregisteret og         */
/*          se <henv>) men dette skaper kryssende tagger som avsløres   */
/*          i Å/E.                                                      */
/*                                                                      */
/* Ver. 1.0 1997-05-21 Øyvind Eide/Dokumentasjonsprosjektet             */
/*                                                                      */
/*----------------------------------------------------------------------*/

#include <stdio.h>
#include <string.h>

void main(argc, argv)
int argc;
char **argv;
{
    FILE *innfil, *utfil;
    char tegn, streng[1000], lik;
    int i, teller;
    printf("Innfil: %s  utfil: %s  OK?", argv[1], argv[2]);
    tegn= getchar();
    if (tegn == 'J' || tegn == 'j')
    {
        innfil= fopen(argv[1], "r");
        utfil= fopen(argv[2], "w");
        teller= 0;
        printf("\n\n=: %5d", teller);
        tegn= getc(innfil);
        while (tegn != EOF)
        {
            if (tegn == ' ')
            {
                putc(tegn, utfil);
                tegn= getc(innfil);
                if (tegn == '=')
                {
                    streng[0]= tegn;
                    tegn= getc(innfil);
                    if (tegn == ' ')
                    {
                        printf("\b\b\b\b\b%5d", ++teller);
                        lik= 0;
                        putc(streng[0], utfil);
                        putc(tegn, utfil);
                        tegn= getc(innfil);
                        i=0;
                        if (tegn == '?') lik= 1;
                        while (lik == 0 && tegn != ' ' && tegn != '.' && tegn != ',' && tegn != EOF)
                        {
                            streng[i++]= tegn;
                            tegn= getc(innfil);
                        }
                        streng[i]= 0;
                        if (lik == 0) fprintf(utfil, "<snavn>%s</snavn>", &streng[0]);
                    }
                    else putc(streng[0], utfil);
                }
                else if (tegn == 's' || tegn == 'j')
                {
                    streng[0]= tegn;
                    for(i=1; i<3; i++) streng[i]= getc(innfil);
                    streng[3]= 0;
                    tegn= getc(innfil);
                    if (streng[0] == 's' && streng[1] == 'e' && streng[2] == ' ')
                    {
                        fprintf(utfil, "%s", &streng[0]);
                        i=0;
                        while (tegn != ' ' && tegn != '.' && tegn != ',' && tegn != EOF)
                        {
                            streng[i++]= tegn;
                            tegn= getc(innfil);
```

```
                    }
                    streng[i]= 0;
                    fprintf(utfil, "<snavn>%s</snavn>", &streng[0]);
                }
                else if (streng[0] == 'j' && streng[1] == 'f' && streng[2] == 'r')
                {
                    streng[3]= tegn;
                    streng[4]= 0;
                    tegn= getc(innfil);
                    fprintf(utfil, "%s", &streng[0]);
                    i=0;
                    while (tegn != ' ' && tegn != '.' && tegn != ',' && tegn != EOF)
                    {
                        streng[i++]= tegn;
                        tegn= getc(innfil);
                    }
                    streng[i]= 0;
                    fprintf(utfil, "<snavn>%s</snavn>", &streng[0]);
                }
                else fprintf(utfil, "%s", &streng[0]);
            }
        }
        putc(tegn, utfil);
        tegn= getc(innfil);
    }
  }
}
```

Tillegg G

Utredning om Qvigstads felthåndbøker

I forbindelse at vi vurderte hvilke skrifter som skulle digitalisertes i delprosjektet, foretok fagkonsulenten en gjennomgang av Qvigstads håndskrevne felthåndbøker, som finnes i original ved Tromsø Museum. Vi så ikke mulighet til å digitalisere noe av dette materiale, men det ville være svært verdifullt om det kunne gjøres ved en senere anledning. Notatet som ble skrevet på grunnlag av gjennomgangen følger her:

Gjennomkikking v/Øyvind Eide 2/5-96. Vurdering av relevans for Dokumentasjonsprosjektet.

Alle opplysninger nedenfor baserer seg på rask blaing og må ikke sees på som dekkende for materialets innhold.

Materialet er delt inn i følgende sju klasser:

1. Aktuelle for selvstendig registrering
2. Bør sees i sammenheng med et evt. samisk språkprosjekt
3. Bør sees i sammenheng med prosjekt navn
4. Bør sees i sammenheng med nordnorsk topografi
5. Krever forarbeid
6. Neppe aktuelle
7. Ikke på plass

G.1 Aktuelle for selvstendig registrering

I. Register over arkivdokumenter 211 s. + fortegnelse over hvilke arkiver som er gjennomgått. Bla. a. en gjennomgang av Missionskollegiets journal 1728-1741 og Finnmarkens amtsjournal 1793-1855. Vedlagt 1 brev fra Johan Frogner 14/4 1908; 4 blad med notater. *Er denne aktuell?*

XIII. Landmaalerprotokollen 1786-1891. Maasø. 60 s. + <1> løsblad. Omhandler i tillegg til Maasø flere andre sogn. Også om stedsnavn i Nord- og Sydvaranger; diverse andre kildehenvisninger. *Er denne aktuell sett i sammenheng med I, eller er den for rotete?*

XIV. Ministerialbog (1721-1850) + noen andre kilder. *Er denne aktuell sett i sammenheng med I, eller er den for rotete?*

1. Reiser. 1871-1944 Løsblad, <20> bl. Oversikt over hvilke reiser Q. gjorde i perioden. *Er denne aktuell, nærmest som en kuriositet?*

10. Lister over lappe-familier og enkeltpersoner i Norge, Finland, Sverige. Upaginert, <ca. 60> s. Biografisk informasjon, andre kildenotater. *Er denne aktuell sett i sammenheng med I, eller er den for rotete?*

12. Skoltelappisk. Borisgleb 1893. <Ca. 40> s. Løsblad, men ser ut til å ha en viss sammenheng. Beskrivelser, stort sett i norsk prosa, av migrasjon/avstamming, historie. *Kan være aktuelt for selvstendig registrering.*

G.2 Bør sees i sammenheng med et evt. samisk språkprosjekt

III. Lappiske ord med to konsonanter i omlyd Liste over ord, 125 s. Vedlegg: 1 brev <7 s.> i A4-format fra Knut Bergsland. *Språk*

IV. "Samisk - norsk" Ordliste, <ca. 250> s. Vedlegg: 2 små utklipp med oversikt over "Landkarter" (se IX) påskrevet "vaaren 1938". *Språk*

V. Opslagsbok II Få av sidene er brukt. Samisk språk: "Navn paa ren" - 10 s.; "Renens kropp og legemsdeler, fys. forhold" - 12 s. (vedlagt trykt ark "Innsamling av tradisjoner um slakt" - 1 s.); "Rendriften" - 13 s.; Noen notater om bokstavveksling (dialektforskjeller?) *Språk*

6. Kvænsk. Div. notater Løsblad: Kildegranskning, finsk, ungarsk, andre språk i familien; Generell og sammenlignende lingvistikk. *Grenser mot språk*

7. Svensk-lappisk ordbok. Tittelen er feil. Inneholder "Beidaren 1923, 1927". Alfabetisk samisk-norsk ordliste på 317 s. *Språk*

11. Skattemandtal for 1745 (Fogders regnskab) Tittelen er neppe riktig. Upaginert, <ca. 100> s. Samisk språk, ord, grammatikk. *Språk*

13. Lappisk ordregister. Upaginert, <ca. 50> s. *Språk*

15. "Historie og geografi". Løsblad - notater: Kilder, språk - sammenlignende lingvistikk, folklore. *Grenser mot språk*

G.3 Bør sees i sammenheng med prosjekt navn

VI. Lappiske stedsnavn: Finnmark fylke 342 s. Vedlagt utklipp fra "Landkarter" (se IX). *Stedsnavn*

VII. Lappiske stedsnavn: Finland og Russland Finland s. 1-102; Russland s. 103-167; Finske apellativistiske stedsnavn s. 169-225. *Stedsnavn*

VIII. Lappiske og finske stedsnavn: Sverige 380 s. Vedlegg: "Plan för indeling af Generalstabens karta öfver Sverige". <1> bl., udatert. *Stedsnavn*

IX. Lappiske stedsnavn: Nordland og Trøndelagen <ca. 150> s. Vedlegg: 1 lapp m/notater; Oversiktskart over gradavd. kart og rekt. kart, jan 1929 (NGO). Deler er klippet vekk (se IV, VI). *Stedsnavn*

X. Lappiske stedsnavn: Troms fylke 364 s. *Stedsnavn*

G.4 Bør sees i sammenheng med nordnorsk topografi

[Unumrert, ikke funnet på arkivkort] Opsatser om Finnmark i tidsskrifter, etc. 81 løpenummer på <15> s. *Kan tas med i bibliografien over nordnorsk topografisk og lokalhistorisk litteratur?*

G.5 Krever forarbeid

II. Etnografi Løsblad, det virker som om innholdsfortegnelsen i omslaget ikke gjelder det nåværende innhold. Løsbladene inneholder bibliografiske og historiske notater, kildegranskning. Brev med lignende innhold. Små etnografiske notater (oppskrifter, fortellinger, ordspråk, ...) Noe "slektsforskning", oversikt over karter. *Rotete - krever kyndig forarbeid*

XI. Storfjorden 1923. Mandal 1922. Løsblad i omslag. Tittelen over er neppe riktig; boka den henviser til er neppe her. Løsbladene inneholder notater vedr. ernæring, historisk, bibliografisk og biografisk arkivmateriale; Samisk språk (lister med ord), stadsnavn; Landmåler- og andre protokoller. *Rotete - krever kyndig forarbeid*

2. Ordbok - notater Ca. <50> s., lite format. Diverse bøker, intervjuer; Folklore, samisk språk, reiseskildringer, ernæring. *Rotete - krever kyndig forarbeid*

3. Notater - (kun på permen) Løsblad. Notatene omhandler historisk og biografisk materiale; Undervisning i Haparanda (<7> s.); Folklore, joik; Avisutklipp fra 1950; Lister over gjenstander, ordspråk. *Rotete - krever kyndig forarbeid*

4. Notater - (kun på permen) Løsblad. Kladd til brev, notater om foto, brevveksling vedr. bestilling av te; Diverse trykk (utklipp), utdrag fra almanakk, avisutklipp. *Rotete - krever kyndig forarbeid*

5. Notater - ordbok m.m. Løsblad. Notater om folklore, oppskrifter, gåter; Samisk språk, bl.a. kildestudier; Om norske dialekter; Stedsnavn; <2> bl. et noteark med melodier til salmer. *Rotete - krever kyndig forarbeid*

9. Private notater - almanakk-blader 1937. Stemplet: Nissens pikeskole, Kristiania. Upaginert, <17> s. Inneholder notater om folklore, oppskrifter, historie, geografi. *Rotete - krever kyndig forarbeid*

14. Finnmarkens amtmanns journal. 1825. Løsblad. Diverse kildenotater; historie, geografi, samisk språk, biografisk; Notat om sjøormen. *Rotete - krever kyndig forarbeid*

G.6 Neppe aktuelle

8. Private brev (finsk/engelsk/tysk). Notater Delvis private notater, noen adresser. Upaginert, <ca. 30> s., lite format. *Neppe aktuelt?*

16. Lyd-notater (kun på permen) *Neppe*

17. Forelæsninger over Heimskringbu. Af professor Sofus Bugge. Forelesningsnotater i løsblad. *Neppe*

18. Russisk bok (liten) Transkribert tittel ca.: "Makjtvjeest ; Pasj-Jevangeli ; Samas ; Jevangelije at Malfeja <...>". Helsingfors, 1878. *Bør kanskje overføres boksamling?*

G.7 Ikke på plass

XII. Lappiske stedsnavn: Vefsn - Hatfjelldal 1927. Tysfjorden 1927. Mangler.

Tillegg H

Samisk leksikografi

I forbindelse med forarbeidet fram mot det som var ment å bli en fase II av Dokumentasjonsprosjektet skrev fagkonsulenten en skisse til et samisk leksikografiprosjekt i april 1996. Denne skissen er gjengitt her:

Dette er et forslag til et delprosjekt under fase II i Dokumentasjonsprosjektet der man bygger opp et register over samiske ord, sammen med et apparat for å analysere de sammenhenger ordene brukes i. Et slikt delprosjekt vil også kunne være nyttig for litteraturforskere. Målet med dette notatet er ikke å legge fram et ferdig opplegg, men å argumentere for at et slikt delprosjekt er mulig og ønskelig.

H.1 Bakgrunn - dokumentasjonsprosjektet

Bokmålsprosjektet har gått som et delprosjekt ved Avdeling for leksikografi (Institutt for nordistikk og litt. vit., Univ. i Oslo) siden 1991. Prosjektet har pr. 31/12-95 registrert noe over 60.000 boksider norsk skjønnlitteratur.

> Databasen vil inneholde tekster fra ca 1550 til ca 1900, de fleste fra 1800-tallet. Det er tatt med både skjønnlitteratur og sakprosa. Henrik Wergelands og Bjørnstjerne Bjørnsons samlede verker utgjør den største delen. I tillegg er det skannet inn (eller er det planer om å dataføre) verker av Absalon Pederssøn Beyer, Petter Dass, Claus Fasting, C.B. Tullin, Camilla Collett, Asbjørnsen og Moe, Jonas Lie, Alexander L. Kielland, Arne Garborg og Ragnhild Jølsen. Barnelitteratur og memoarer kan det

> også bli aktuelt å dataføre. Tekstene er skannet fra standardutgaver og førsteutgaver. De er merket med SGML-tagger slik at de kan gjenskapes lik de originale boksidene. Taggene gjør det mulig å gjøre forskjellige typer søk i tekstene. Tekstene er merket med sjanger (prosa, lyrikk, drama). Samtlige prosatekster er typologisert (artikkel, novelle, roman osv), noen lyrikkformer er merket ut (sonnett, ode osv.). Dessuten er hver del av en tekst (tittel, avsnitt, strofe, kapittel osv.) merket ut. For at tekstene skal kunne brukes til å lage konkordanser, er deler av teksten som ikke er forfatternes ord, tagget (f.eks. sitater og merknader fra utgiver). Det er tatt hensyn til at både språk- og litteraturforskere skal kunne ha nytte av tekstene.

Man bør gjøre noe lignende for samisk språk, men man må tenke annerledes. Her følger noen vurderinger av metode og noen løselige anslag på behov for arbeidskraft.

H.2 Samisk ordlistemateriale

I prosjektet kan man registrere et utvalg av de samiske ordlistene som finnes. Dette vil gjøre det mulig å effektivisere produksjon av nyere ordlister, samtidig som språkforskere vil få bedre verktøyer til å analysere og sammenligne materialet.

Som et eksempel på arbeidskraftbehov brukes Friis' *Ordbog over det lappiske sprog* (1887). Dette er en ganske omfattende bok (ca 900 sider i to spalter) samtidig som man ved Tromsø Museum også har et eksemplar der Qvigstad har gjort notater i margen og på interfolierte blad. Uten å ha talt nøyaktig anslås det til mellom 50.000 og 100.000 enkeltnotater.

Erfaringene fra bokmålsmaterialet er at man kommer opp i ca 8 sider ferdigstilt materiale pr. dagsverk ved scanning og korrekturlesing av vanlig skjønnlitteratur. Dette er selvsagt ikke sammenlignbart med et ordlistemateriale der informasjonen må legges inn i definerte felter. Jeg går ut fra en hastighet på ca 1 ferdigstilt side pr. dagsverk ved manuell innskriving, en noe høyere hastighet vil kunne oppnås dersom man kan lage effektive rutiner for scanning og automatisk oppmerking av felter. Det betyr at boken vil kreve maksimalt 4 årsverk registrering.

I tillegg kommer Qvigstads notater. Jeg vil anslå registratorenes innskriving til å kreve ca 2 årsverk. Senere kan det bli aktuelt å gjennomføre etter-arbeid for å tolke uklare notater, men det må selvsagt gjøres av forskere eller studenter.

H.3 Samisk tekstkorpus

Hvis man skal registrere eldre samiske tekster må man tenke litt annerledes enn man gjør på norsk. Eksempler på aktuelt materiale er bibeloversettelser, nedtegnelser av muntlige tradisjoner, aviser og "vanlig" skjønnlitteratur.

Bokmålsprosjektet er ikke helt fornøyd med at de på grunn av opphavsretten, som nå gjelder 70 år etter en forfatters død, måtte stoppe registreringen før Hamsun og Undset. Det er et arbeid på gang for å lage en standardkontrakt der forfatterne gir tillatelse til registrering av opphavsrettslig beskyttet materiale. Med andre ord kan det bli aktuelt å registrere nyere samisk skjønnlitteratur dersom tilfredsstillende juridiske løsninger finnes.

Utgangspunktet er likevel at opphavrettsspørsmålet er uløst, med ett unntak: Samisk-Etnografisk avdeling ved Tromsø Museum har allerede fått godkjenning fra Qvigstads arvinger til å registrere alt han er opphavsmann til.

Jeg bruker her to eksempler på materiale det kan være aktuelt å registrere:

H.3.1 *Bibal* (Kristiania, 1895)

Ut fra erfaringene fra Dokumentasjonsprosjektet vil jeg anslå registreringshastigheten til å bli ca 3 ferdigstilte sider pr. dagsverk. Samisk tekst vil for norsktalende registratorer føre til lavere registreringshastighet enn norsk tekst. Imidlertid viser erfaringene med materiale på latin og gammalnorsk at feilprosenten blir noe lavere dersom registratorene ikke kan språket de registrerer. Man antar at dette skyldes at man da leser korrektur tegn-for-tegn, ikke ord-for-ord.

Et årsverk burde da ferdigstille ca 750 sider, slik at hele Bibelen på 1740 sider skulle utgjøre ca 2 1/2 årsverk.

H.3.2 *Lappiske eventyr og sagn* / ved J. Qvigstad. (Oslo 1927–1929)

Boka er satt med noe mindre tekst på hver side enn *Bibal*, så registreringshastigheten anslås til 4 ferdigstilte sider pr. dagsverk. 1.150 sider skulle da utgjøre ca 1 1/4 årsverk.

Boken er parallelltekstet på samisk og norsk. Jeg tar utgangspunkt i at kun den samiske teksten registreres. Jeg ser likevel gode grunner til å ta med også den norske teksten. Det vil øke tidsforbruket med ca 50%.

H.4 Hvor stort prosjekt?

Materialet ovenfor vil i sum utgjøre en registreringstid på ca 10 årsverk. Ut fra en normalfordeling (4 registratorer pr. fagkonsulent pr. år) vil det gjøre at et slikt prosjekt vil kunne gjennomføres på 2 1/2 år med én fagkonsulent.

Dokumentasjonsprosjektets fase II er planlagt sluttført i 2001. Hvis en regner en prosjektperiode på fire år ser man at med et delprosjekt med én fagkonsulent kan man få registrert noe mer materiale enn det som er nevnt ovenfor.

Jeg synes likevel det er for knapt. Hvis man nå benytter sjansen og starter opp et større prosjekt vil man kunne få registrert store samlinger som vil utgjøre et glimrende hjelpemiddel for forskningen i lang tid framover.

Selv om størrelsen og formen på et eventuelt delprosjekt må avklares etterhvert, vil jeg allerede nå anbefale at man sikter mot et delprosjekt med minst 3 fagkonsulenter. Det vil gjøre at man kan få registrert et omfattende materiale, uten at kostnadene blir uoverkommelige. Dessuten vil flere fagkonsulenter kunne utfylle hverandre faglig.

Jeg er selvsagt tilgjengelig dersom det dukker opp spørsmål i forbindelse med Dokumentasjonsprosjektet. Når det gjelder erfaringene fra Bokmålsprosjektet, vil jeg i tillegg til Dokumentasjonsprosjektets ledelse også henvise til Anne Engø (tlf. 22 85 43 93), som har vært med som prosjektassistent siden 1991.