

# Tekst- og verktøybase for Medieval Nordic Text Archive: Kodingsstandard, ordbankbygging og Heimskringla-edisjon

## Menota TVB (2003–2006)

Sluttrapport, 31.01.2007, også publisert på <http://www.menota.org/TVB>

Odd Einar Haugen (prosjektleder)

### Bakgrunn

Menota TVB er eit fireårig prosjekt som har vore leia av prof. Odd Einar Haugen, Nordisk institutt, Universitetet i Bergen. Det har hatt nær tilknytning til det elektroniske arkivet *Medieval Nordic Text Archive* (Menota), oppretta i 2001, og til *Medieval Unicode Font Initiative* (MUFI), også oppretta i 2001. Administrativt har prosjektet vore lagt til Avdeling for kultur, språk og informasjonsteknologi (Aksis) i Bergen. Fleire av dei faglege medarbeidarane ved senteret har teke del i prosjektet; i første rekkje Tone Merete Bruvik, Paul Meurer og Vemund Olstad. I periodar har eksterne medarbeidarar vore engasjerte, m.a. Jo Rune Ugulen (Bergen), Michael Irlenbusch-Reynard (Leer, Tyskland), Florian Grammel (København), Michael Everson (Westport, Irland) og Andreas Stötzner (Leipzig). Det har knytt til seg tre aktive samarbeidspartnarar ved Universitetet i Oslo; Christian-Emil Ore ved Enhet for digital dokumentasjon (EDD), Jon Gunnar Jørgensen i *Heimskringla*-prosjektet, og Karl G. Johansson i Vadstena-prosjektet, *Vadstenaklostrets handskrifter elektronisk tillgängliga*.

I søknaden var det sett av ei heil forskarstilling gjennom tre år, med tanke på dr.art. Rune Kyrkjebø, Bergen. Då han fekk fast stilling ved Universitetsbiblioteket i Bergen rett før oppstarten av prosjektet, vart prosjektledaren, Odd Einar Haugen, frikjøpt for dei tre første åra, og Nordisk institutt kompensert med full vikardekning. Rune Kyrkjebø har teke del i prosjektet innanfor rammene av si forskingstid ved Universitetsbiblioteket. Medarbeidarane ved Aksis har vore engasjerte med om lag 6 månadsverk kvar over prosjektperioden, medan dei eksterne deltakarane har vore honorerte på timebasis. For delar av arbeidet med *Heimskringla* har det også vore innarbeidd eit frikjøp av Jon Gunnar Jørgensen. Utover dette har dei tre aktive samarbeidspartnarane bidratt til prosjektet gjennom eigeninnsats.

Hovudmålet for prosjektet har vore å byggje opp ein brei tekst- og verktøybase for Menota. Dei tekstane som går inn i arkivet, dannar no eit koordinert korpus, som gjer at brukarane kan søkje i arkivet etter eit breitt utval av kriterium, samtidig som dei også kan få tilgang til tekstane som litterære verk. Målet har såleis vore at arkivet kan dekkje typiske behov blant både litteraturforskarar, språkforskarar og historikarar.

For å oppnå dette hovudmålet formulerte søknaden tre delprosjekt, som i denne slutt-rapporten skal refererast, drøftast og vurderast. Det tredje delmålet vart nedjustert i 2005, og eit fjerde delprosjekt lagt til. I praksis førte dette til at Menota TVB totalt vart utvida og gjort meir krevjande.

## Delprosjekt 1: utvikling av kodingsstandard og tekstvisning

### (a) Handboksutvikling og visningsproblematikk

#### Målformulering

Delprosjektet skal utvikle robuste visningsformat for mellomalderleg tekst både for HTML og PDF, oppdatere kodingshandboka i samsvar med utviklinga av TEI-standarden, og gje tilbakemelding til TEI på felt som har å gjere med transkripsjon av mellomalderlege primærkjelder.

#### Resultat i prosjektperioden

Delprosjektet byggjer på den kodingshandboka som vart utvikla ved prosjektet *Editing medieval manuscripts* (Centre for Advanced Study, Oslo, 2000–2001) under leiding av Odd Einar Haugen. Handboka, som var forfatta på norsk, vart overført til engelsk språkform, utvida på ei rekkje punkt og gjeve tittelen *The Menota handbook*. Redaktør for handboka var Odd Einar Haugen, med Rune Kyrkjebø (Bergen), Karl G. Johansson (Oslo) og Matthew Driscoll (København) som medarbeidarar. Versjon 1.0 vart publisert 20. mai 2003 og ein omhyggeleg korrigert versjon 1.1 kom 5. mai 2004. Michael Irlenbusch-Reynard gjorde ein stor innsats for å rette opp inkonsistensar i handboka, forutan at han førte den over i ein striktare versjon av HTML. Versjon 1.1 er publisert i Bergen Open Research Archive (BORA).

Etter publiseringa av v. 1.1 har Tarrin Wills (Sydney) kome med i redaksjonen, og handboka er ført over frå HTML til XML. Med tanke på at handboka gjev retningslinjer for bruk av XML i koding av mellomalderhandskrifter er dette eit naturleg steg, men det har også vist seg å vere tenleg på andre måtar. Eitt av problema med v. 1.0 av handboka var rett og slett at ein del av eksempla ikkje var valide i forhold til den *Document Type Definition* (DTD) som var publisert som ein medfølgjande ressurs til boka. Dette feilskjæret vart retta opp med v. 1.1, og gjennom overføringa til XML har vi oppnådd å få til ei vesentleg lettare og fortløpande validering av eksempla; handboka blir slik sett sjølvrefererande konsistent. I tillegg vart det utarbeidd eit grensesnitt for XML-validering over nettet, som gjorde at medarbeidarar andre stader (Oslo, København, Sydney) kunne oppdatere sine kapittel direkte mot serveren i Bergen. Tone Merete Bruvik utvikla desse rutineane, som har vist seg svært robuste og praktiske. Dei same rutineane har vore innførte også for nettsidene til Menota, <http://www.menota.org>.

Etter at handboka vart overført til XML, vart HTML-visinga generert ved bruk av transformasjonsskjema, XSLT. Slike stilskjema kan generere HTML for visning på nettet, og supplert med ein Formatting Object Processor (FOP) kan ein også generere PDF-versjonar av boka. Dette gjev m.a. større typografisk kontroll, og det inneber at ein kan bake inn spesialfontar. Arbeidet med å utvikla slike stilskjema har vore leia av Vemund Olstad, som har dokumentert prosessen gjennom kommenterte stilark publiserte på Menota-sidene. Menota-prosjektet har her kunna dra nytte av tilsvarande utviklingsarbeid for prosjektet *Henrik Ibsens skrifter*.

Versjon 1.1 av handboka var kompatibel med Text Encoding Initiative (TEI) P4, medan den neste versjonen, v. 2.0, skal vere kompatibel med TEI P5. Versjon 2.0 av handboka

er no på det nærmaste ferdig og må reknast som stabil, men inntil TEI P5 blir avslutta seinare i 2007, ønskjer vi å vente med den endelege publiseringa. Handboka er tilgjengeleg som v. 2.0 β ved Aksis, og vil bli publisert på BORA så snart P5 er klar og vi har kunna ta stilling til eventuelle seine endringar i P5. Det er vår vurdering at v. 2.0 β av handboka er stabil, og at det same gjeld dei tilhøyrande ressursane til denne, særleg Menotas *Document Type Definition*. Dette har ført til at ei rekkje Menota-konforme tekststar er publiserte på nettet dei siste par månadene (jf. pkt. 4 nedanfor).

Eit særleg problem for all XML-basert koding gjeld handteringa av overlappande trekk i tekstane; her har vi henta inn råd frå fleire ekspertar, m.a. frå Claus Huitfeldt (Bergen) og Andreas Witt (Tübingen), og vi meiner at vi no har ei så god løysing på dette problemet som det er praktisk mogleg innanfor XML-rammeverket. Dette er drøfta nærmare i kap. 4 og særleg kap. 7 av handboka.

Det har vore til uvurderleg hjelp at Tone Merete Bruvik har så gode kontaktar i TEI-miljøet generelt og blant dei fremste utviklarane av standarden. Dermed har vi kunna halde oss løpande orienterte om utviklinga i TEI, og det har også innbore at utviklinga av Menota-kodinga har kunna gje innspel til TEI-miljøet. Ikkje minst gjeld dette vår modulære koding på tre representasjonsnivå, <facs> (faksimilær attgjeving av teksten), <dipl> (diplomatarisk) og <norm> (normalisert). Denne splittinga har gjort det mogleg å handtere det problematiske forholdet mellom element som <abbr> (forkorta form) og <expan> (utskreven form) på den eine sida, og <sic> (original lesemåte) og <corr> (korrigert lesemåte) på den andre sida. Dette er i praksis ei løysing som TEI no handterer med bruk av det såkalla <choice>-elementet.

### **Vurdering av resultat**

Måla for delprosjektet er oppfylte på alle punkt. Gjennom oppbygginga av ein felles XML-struktur for handboka har vi oppnådd praktiske (og overførbare) resultat som vi ikkje såg for oss i 2002. Vi merkar oss med glede at handboka har vore brukt ved fleire utanlandske prosjekt, m.a. ved universiteta i Hamburg, Cardiff og Sydney.

### **(b) Unicode-tilpassing**

#### **Målformulering**

Delprosjektet skal utvikle ein akseptert standard for bruk av Private Use Area i Unicode som dekkjer behova for transkripsjon av nordiske mellomaldertekstar i det latinske alfabetet på alle hovudnivå (faksimilert, diplomatarisk og normalisert), og så langt mogleg også for nærståande mellomalderlitteraturar, særleg den gammalengelske.

#### **Resultat i prosjektperioden**

Dette arbeidet har vore koordinert med *Medieval Unicode Font Initiative* (MUFI) som sidan etableringa i 2001 har vore leia av Odd Einar Haugen, med Andreas Stötzner (Leipzig) som nestleiar. MUFI koordinerer i praksis alle større akademiske Unicode-baserte prosjekt for mellomalderkskrift i det latinske alfabetet, herunder TITUS-prosjektet ved prof. Jost Gippert (Frankfurt am Main) og Junicode-prosjektet ved prof. Peter S. Baker (Virginia). Det første tilrådinga frå MUFI vart framlagd 8. des. 2003 under tittelen *MUFI character recommendation* v. 1.0, redigert av Odd Einar Haugen.

Denne tilrådinga omfattar godt over 800 mellomalderteikn, der om lag halvparten er identifiserte i ulike tabellar av den offisielle Unicode-standard, og den andre halvparten ligg i det såkalla Private Use Area. Fleire MUFI-kompatible fontar har seinare vore publiserte, m.a. *Alphabetum* (Juan-José Marcos, Plasencia, Spania), *Leeds Uni* (Alec McAllister, Leeds) og *Cardo* (David J. Perry, Rye, N.Y.).

Versjon 2.0 av *MUFI character recommendation* vart publisert 22. des. 2006, også denne redigert av Odd Einar Haugen, og no delt i to band – eitt som presenterer teikna i alfabetisk rekkjefølgje og eitt som presenterer dei i same rekkjefølgje som i Unicode-standard. Denne versjonen kjem opp i drygt 1300 mellomalderteikn, og er koordinert med publiseringa av ein Unicode-font, *Andron Scriptor Web*, som inneheld alle dei over 1300 teikna i tilrådinga. Takk vere finansiell støtte frå Menota TVB (og frå Universitetet i Bergen) er denne fonten publisert på MUFI-sidene til gratis nedlasting. *Andron*-fonten har elles vore vald til bruk i *Gutenberg-Jahrbuch*, og vil også bli brukt i den tyske omsetjinga av *Handbok i norrøn filologi*, som kjem i 2007 på Walter de Gruyter i Berlin. Det er såleis tale om ein font av høg typografisk kvalitet, rett nok berre publisert i rett skrift (“Regular”) på MUFI-sidene.

I tillegg til utarbeidinga av *MUFI character recommendation* har Menota-prosjektet vore aktiv deltakar i utarbeidinga av to søknader (*proposals*) til Unicode-konsortiet, begge redigerte av Michael Everson (Westport, Irland), sjølv medforfattar av Unicode-standard. Dette arbeidet har for ein stor del vore finansiert av Menota TVB og Universitetet i Bergen, formidla gjennom Script Encoding Initiative (leia av Deborah W. Anderson, Berkeley, og Rick McGowan, visepresident i Unicode-konsortiet). Den første av dei to søknadene omfattar drygt 100 mellomalderlege teikn og vart godt motteken av Unicode Technical Committee i februar 2006 og av ISO i april 2006. Den er no på vandring i Unicode fram mot den sannsynlege endelege godkjenninga i neste revisjon av Unicode-standard. Dette er ein prosess som normalt tek eit par år.

Den neste søknaden omfattar eit sett av mellomalderlege interpunksjonsteikn og er no komen til kontroll- og korrekturfase. I skrivande stund er den til vurdering hos Malcolm B. Parkes (Oxford), den fremste autoriteten på feltet gjennom standardverket *Pause and Effect*, og Teresa Webber (Cambridge), redaktør for *Cambridge Studies in Paleography and Codicology*. Etter planen vil søknaden bli fremja for ISO i april 2007.

Både v. 1.0 og v. 2.0 av *MUFI character recommendation* er publiserte på BORA (sjå s. 11 nedanfor). Søknadene til Unicode er publiserte på nettsida til Michael Everson (med lenke frå MUFI-sidene, <http://www.mufi.info>).

### **Vurdering av resultat**

Måla for delprosjektet er overoppfylte. *MUFI character recommendation* representerer i realiteten gjeldande industristandard for mellomalderlege Unicode-teikn i det latinske alfabetet. I tillegg til nordiske og engelske teikn har den også eit større utval av allmenne europeiske teikn, særleg frå det latinspråklege abbreviatur- og interpunksjonsrepertoaret. I tillegg til dette koordineringsarbeidet kjem dei to søknadene til Unicode-konsortiet, som ikkje var planlagde i 2002, men som er med på å gjere teiknnormeringa til ein del av sjølve Unicode-standard. Dette har vidtrekkjande og svært verdifulle

konsekvensar; som eitt av mange døme kan nemnast at det vil løfte heile teikninventaret i *Ordbog over det norrøne prosasprog* (som kan hende må fullførast i elektronisk form) over i Unicode-standard. Det er også svært gledeleg at ein gratis Unicode-font av høg typografisk kvalitet har vore utvikla og publisert på nettet.

## Delprosjekt 2: oppbygging av ordbank og utvikling av lemmatiseringsstrategiar

### Målformulering

Oppbygging av ein norrøn ordbank, og utvikling av ein rutine for halvautomatisert lemmatisering med tilhøyrande grensesnitt, slik at dette kan bli lett tilgjengeleg for tekstleverandørane til Menota. Ein rapport som evaluerer arbeidet på grunnlag av testing på hittil ulemmatiserte norrøne prosatekstar.

### Resultat i prosjektperioden

Dette delprosjektet var det mest opne og usikre då søknaden vart skriven i 2002, og arbeidet med å utvikle det gjekk noko seint og famlande dei to første åra av Menota TVB. I nært samarbeid med prosjektleiaren utvikla Johan Utne Poppe ein modell for leksikografisk søk i *Barlaams ok Josaphats saga* i handskriftet Holm perg 6 fol (ca. 76.000 ord). Her kunne prosjektet dra nytte av den føreliggjande lemmatiserte teksten, konvertert til XML av Christian Emil Ore ved EDD i Oslo. Innsatsen låg først og fremst i å byggje opp eit godt grensesnitt for ulike typar av søking og vising av teksta, både i form av konkordansar og side for side. Dette kravde ei vidareutvikling av kodinga for grammatiske kategoriar på ordnivå i den underliggjande XML-teksten, og vart reflektert i utviklinga av kap. 8 i *The Menota handbook*. Eit lite, men viktig steg framover tok vi då vi gjekk over frå den EAGLES-baserte kodinga av morfologiske kategoriar til full bruk av såkalla *name tokens* på attributtnivå. Dette skjedde etter framlegg frå Tarrin Wills (Sydney), og førte til at det vart vesentleg lettare å byggje opp kombinerte søk på grammatiske kategoriar (genus, numerus, kasus, tempus, modus m.m.). Søkjegrensesnittet for *Barlaams ok Josaphats saga* er i seg sjølv godt, og ligg framleis tilgjengeleg på nettet, <http://www.menota.org/search/search-test.htm>.

Det store og attståande problemet etter dei to første åra på prosjektet låg i utviklinga av rutinar for halvautomatisk lemmatisering. Det viste seg å vere avgjerande for framdrifta på prosjektet at Paul Meurer gjekk aktivt inn i dette arbeidet i 2005. Etter eit symposium i februar 2005, der vi inviterte dei som har arbeidd mest på dette feltet, Hans Fix (Greifswald) og Andrea de Leeuw van Weenen (Leiden), såg vi at vi hadde dei nødvendige verktøya for utvikling av ei slik lemmatiseringsteneste i Bergen. Paul Meurer kunne ta utgangspunkt i det Corpus Workbench-baserte korpussystemet han har bygd opp ved Aksis, og gjennom eit nært samarbeid med Edda-utgjevingsprosjektet på Island, har han utvikla ein avansert modul for halvautomatisk lemmatisering av nordiske mellomaldertekstar, *Menota Lemmatisation Assistant* (MLA). Denne lemmatiseringstenesta byggjer på tekstar som i utgangspunktet er lemmatiserte og morfologisk analyserte. Vi hadde berre *ein* slik tekst tilgjengeleg, den før nemnde *Barlaams ok Josaphats saga* (norsk), og vi rekna med at denne ville vere til relativt liten nytte for lemmatiseringa av eddadikta i GKS 2365 4to (islandsk). Til vår overrasking

viste det seg at *Barlaams ok Josaphats saga* var til langt større hjelp enn tenkt, og per i dag har vi bygd opp eit lemmatisert korpus som omfattar begge desse tekstane pluss tre nyttilkomne tekstar frå Oslo, *Strengleikar*, *Thómass saga erkibiskups* og den legendariske *Óláfs saga ins helga*. Om kort tid vil korpuset bli utvida med resten av dei lemmatiserte tekstane frå Gammelnorsk ordboksverk i Oslo. MLA tilbyr eit avansert grensesnitt for halvautomatisk lemmatisering av nordiske mellomaldertekstar, og er i dag utan konkurranse internasjonalt – det nærmaste måtte vere prosjektet Cultural Heritage Language Technologies (<http://www.chlt.org/>), men dette har etter vår vurdering kome mykje kortare.

MLA vart først utvikla for lemmatisering av norrøne tekstar. Men modulen vart snart utvida til også å omfatte lemmatisering av austnordiske, nærmare bestemt svenske tekstar. For dette korpuset må ein byrje frå grunnen av, og det er nettopp det Maria Arvidsson ved Vadstena-prosjektet i Oslo har gjort. Det siste året har ho lemmatisert heile den svenske (rett nok noko forkorta) versjonen av Barlaam-legenda ved hjelp av MLA. Dette vil gje eit tilsvarande grunnlag for vidare lemmatisering av fornsvenske tekstar som den gammalnorske Barlaam-legenda har gjeve for norrøne tekstar.

Då MLA har vore testa av brukarar i Oslo og Reykjavík (og tentativt i København) var det nødvendig å byggje opp eit nett-grensesnitt. Dette inneber at eksterne brukarar kan laste opp ein tekst i MLA, oppdatere og utvide basen over lemmatiserte tekstar, og deretter laste ned den same teksten med Menota-kompatibel koding. Det er ein vinn-vinn-situasjon, der Menota-prosjektet kan nyte godt av dei som brukar lemmatiserings-tenesta, og der dei som er brukarar, får koda tekstar tilbake – som dei kan bruke i andre samanhengar. Ei slik wikipedia-liknande form for lemmatisering kunne vi ikkje på nokon måte sjå for oss då vi skreiv søknaden i 2002. Det har gjeve eit anna – og mykje meir dynamisk – svar på planane om å byggje opp ein ordbank. Men det seier seg sjølv at ei slik teneste krev tett oppfølging og ikkje minst kvalitetskontroll.

### **Vurdering av resultatata**

Måla i prosjektsøknaden er oppfylte, med unntak av at vi ikkje har utarbeidd nokon evalueringsrapport for MLA. Vi meiner det bør vente til vi har drifta tilbodet over noko tid og dermed får vite meir konkret korleis det fungerer for fleire brukarar enn dei som har arbeidd på MLA. Det uløyste problemet ved avslutninga av prosjektet er at vi ikkje har finansiering for vidare drift av tilbodet. Resultatet av lemmatiseringa, m.a. av eddadikta i GKS 2365 4to, er likevel eit permanent og høgst konkret resultat. Menota TVB var konsipert som eit utviklingsprosjekt, og vi må seie oss godt fornøgde med å ha bygd opp eit så sterkt verktøy som MLA. Dersom det lykkast oss å finne finansiering for ei vidare drift av denne tenesta, vil det bli eit internasjonalt tilbod, hittil utan konkurranse. Det er verd å merke seg at MLA har vore brukt for både vestnordiske (norske og islandske) og austnordiske tekstar, i samsvar med den breie, nordiske profilen til Menota.

## Delprosjekt 3: elektronisk formatering av *Heimskringla*-tekstar

### Målformulering

Utarbeiding av ein tekst av *Heimskringla* I (y-klassa) i samsvar med den nye Menota-standard, og konvertering av *Heimskringla* I, II og III (x-klassa) frå prosjektet i Oslo etter same standard, inkl. korrektur og kontroll av *Heimskringla* I.

### Resultat i prosjektperioden

Arbeidet med ei ny utgåve av *Heimskringla* går tilbake til slutten av 1980-talet, og vart på 1990-talet drive – med fleire opphald og manglande finansiering – ved Universitetet i Oslo. Det vart vidareført i det før nemnde prosjektet *Editing medieval manuscripts* (Centre for Advanced Study, Oslo, 2000–2001). I utgangspunktet var *Heimskringla*-utgåva tenkt som ei *kritisk* utgåve, til avløyning av dei tidlegare utgåvene til Finnur Jónsson og Bjarni Aðalbjarnarson. Under drøftingane på prosjektet *Editing medieval manuscripts* vart det stadig klarare at det ikkje var tenleg å utarbeide ei kritisk (i praksis eklektisk) utgåve, men at ein heller skulle gje ut dei individuelle handskriftene av *Heimskringla*, både Hkr. I i y-klassa og Hkr. I, II og III i x-klassa.

Rune Kyrkjebø har koordinert arbeidet med utgjevinga av Hkr. I i y-klassa, men då han tok til i ei anna stilling ved oppstarten av prosjektet, seier det seg sjølv at han ikkje har hatt tid til å fullføre det planlagde arbeidet med Hkr. I. Ved avslutninga av prosjektet har han likevel utarbeidd ein del av teksten i Hkr. I, *Haralds saga hárfagra*, til Menota-kompatibel XML, og denne er no publisert på nettsidene til Menota. Som det går fram av delprosjekt 4 nedanfor har Rune Kyrkjebø også ytt eit viktig bidrag til utviklinga av katalogstrukturen for Menota, forutan at han i perioden har leia to andre prosjekt, Diplom-prosjektet og MPF-prosjektet (omtalte i samanfatninga på s. 10 nedanfor).

Jon Gunnar Jørgensen har koordinert arbeidet med Hkr. I, II og III i x-klassa, og ved avslutninga av Menota TVB er Hkr. I og II ferdig korrekturlesne og vil bli publiserte innan midten av februar 2007. Kjartan Ottosson har også vore med i *Heimskringla*-prosjektet, og har teke eit særleg ansvar for Hkr. II. Dei tre *Heimskringla*-tekstane ligg føre i eit gammalt, men velstrukturert WordPerfect-format, og Florian Grammel (København) har gjort ein stor innsats i å konvertere denne kodinga til XML, m.a. ved å omstrukturere og oppløyse ei rekkje tilfelle av overlappende koding. For Hkr. II er situasjonen at det står tilbake om lag ei veker arbeid med korrektur i København og dertil ein tilsvarende innsats for å konvertere teksten til valid XML. Kjartan Ottosson har teke på seg å gjennomføre korrekturarbeidet, og det er no realistisk å rekne med at korrekturen og konverteringa kan fullførast i løpet av mars 2007.

### Vurdering av resultata

Den elektroniske utgjevinga av Hkr. I, II og III i x-klassa er nesten fullført i samsvar med planane, men som nemnt står det noko arbeid att på Hkr. II. Vi reknar finansieringa av det attståande arbeidet som sikra, og veit at det er høgt prioritert. Forseinkinga med Hkr. II er eit lite skør i gleda, men det avgjerande er at *Heimskringla*-prosjektet no har nådd så langt at tekstane om kort tid vil vere publiserte på nettsidene til Menota. Dét er ein milestolpe.

## Delprosjekt 4: utvikling av elektronisk arkivstruktur

### Målformulering

Utarbeiding av arkiv-infrastruktur for tekstar, katalogdata og informasjon om elektroniske versjonar for Menota, på grunnlag av EAD (Encoded Archival Descriptions) og Text Encoding Initiative (TEI).

### Resultat i prosjektperioden

Dette arbeidet har vore utført med stor energi gjennom det siste halvåret av prosjektet. Som planlagt har Rune Kyrkjebø på grunnlag av sin bibliotekfaglege ekspertise utgreidd spørsmålet om korleis arkivstrukturen for Menota bør vere. Dette har han gjort gjennom ei vurdering av EAD og TEI, i samsvar med planane, men også av dei bibliografiske FRBR- og MARC-standardane. Denne utgreinga er publisert i eit notat på nettsidene til Menota TVB: <http://www.menota.org/TVB/rapportar/Menotainformasjon.pdf>

Dei første tre og eit halvt åra av prosjektet har vore prega av det ein kan kalla standard-utvikling innanfor fontar (*MUFI character recommendation*) og tekstkoding (*The Menota Handbook*), forutan oppbygging av lemmatiseringstenestene i MLA. På dette fundamentet må sjølv arkivet byggjast opp, og i det siste halvåret har prosjektet kome så langt at valide XML-filar av nordiske mellomaldertekstar er publiserte eller under publisering.

Katalogmodulen for Menota har vore utvikla i samarbeid mellom prosjektleiar på den eine sida og Paul Meurer og Tone Merete Bruvik på den andre. Katalogen blir automatisk generert på grunnlag av header-informasjon i dei enkelte tekstane, og har derfor ført til at kap. 10 i *The Menota handbook* er revidert fortløpande den siste tida. Katalogen opnar for både lesing og søking i tekstane, og er koordinert med bruken av den før nemnde lemmatiseringsmodulen MLA. I praksis vil det seie at kvar tekst berre ligg føre i *ein* XML-fil (sjølv om den kan ha tekstrepresentasjon på inntil tre nivå, forutan lemmatisering og grammatisk koding) og at den blir indeksert og tilgjengeleggjord gjennom berre *ein* modul, nemleg MLA. Dette gir godt oversyn over tekstane i arkivet og det gjer det mogleg å halde ein god versjonskontroll. Katalogen kjem til å vere under utvikling i tida framover (under føresetnad av finansiering), men har no nådd så langt at den etter vårt syn har ein brei og tilstrekkeleg funksjonalitet: <http://decentius.aksis.uib.no:8060/corpus/menota.xml>

Då søknaden vart send inn i 2002, såg vi for oss at dei følgjande tekstane skulle gjerast tilgjengelege (i tillegg til dei før nemnde *Heimskringla*-tekstane):

- *Barlaams ok Josaphats saga*, Holm perg 6 fol
- *Konungs skuggsjá*, AM 243 b α fol
- *Óláfs saga ins helga*, Upps DG 8
- *Strengleikar*, Upps DG 4–7
- *Piðriks saga af Bern*, Holm perg 4 fol
- *Thómass saga erkibiskups*, Holm perg 17 4to
- Alle norske diplom i gammalnorsk språkform frå ca. 1200 til 1308

Ved sluttrapporteringa av prosjektet kan vi slå fast at nokre av desse tekstane er publiserte, nokre står tilbake, og fleire andre har kome til. Lista ser slik ut (med atterhald om at dei tekstane som er merkte med stjerne vil bli publiserte innan 15. febr. 2007):

- *Barlaams ok Josaphats saga*, Holm perg 6 fol [norsk]
- *Óláfs saga ins helga*, Upps DG 8 [norsk]
- *Strengleikar*, Upps DG 4–7 [norsk]
- *Thómass saga erkibiskups*, Holm perg 17 4to [norsk]
- *Heimskringla I, y-klassa: Haralds saga hárfagra*, AM 37 fol [islandsk/norsk]
- *Heimskringla I, x-klassa*, AM 35 fol [islandsk] (\*)
- *Heimskringla III, x-klassa*, AM 63 fol [islandsk] (\*)
- *Codex Wormianus*, AM 242 fol [islandsk]
- *Codex Regius: Vølusþá*, GKS 2365 4to [islandsk]
- *Codex Verelianus*, Holm D 4 [svensk]
- *Religiösa texter*, Holm A 10 [svensk]
- *Legenden om Barlaam och Josaphat*, Holm A 49 [svensk]

Når det gjeld *Konungs skuggsjá* og *Piðriks saga af Bern* står det att noko korrekturarbeid ved Gammelnorsk Ordboksverk før dei kan publiserast, men vi held det for sannsynleg at dette vil skje i løpet av 2007. Som før nemnt kjem *Heimskringla II* i nær framtid, truleg i løpet av mars 2007.

I tillegg held vi det for sannsynleg at det vil kome inn fleire andre store tekstar i 2007, m.a. frå det svenske Vadstena-prosjektet, men også nokre større norske og islandske tekstar frå andre utgjevarar (m.a. i København, truleg også i Leiden). Når alle dei gammalnorske tekstane frå Oslo er komne på plass, vil også mengda med fullt lemmatiserte og morfologisk analyserte tekstar auke sterkt.

### **Vurdering av resultata**

Utgreiinga av arkivstrukturen er gjennomført i samsvar med planane, medan den praktiske oppbygginga av katalogstrukturen i MLA gjennom ein stilarkbasert ekstraksjon av header-informasjon er eit resultat som går utover det som vart planlagt. Når det gjeld dei publiserte tekstane, er utvalet litt annleis enn det vi såg for oss i 2002, men volumet er totalt høgare enn det vi planla og det er realistisk å få inn dei attståande tekstane i løpet av inneverande år. I tillegg kjem det at Menota har fått ein sterk felles-nordisk profil gjennom tilskotet av tre større svenske tekstar. Volumet av publiserte tekstar i arkivet må no seiast å vere så stort at arkivet har oppnådd ein kritisk masse.

### **Samanfattande vurdering**

Menota TVB har vore eit stort og komplekst prosjekt. Det har gått over fire år og har hatt mange medarbeidarar. Med nokre få og ikkje avgjerande unntak har prosjektet gitt større utteljing enn det vi såg for oss då vi skreiv søknaden i 2002. I ei samla vurdering må det seiast at prosjektet har oppfylt og til dels overoppfylt dei måla som var oppsette, og dei punkta der prosjektet ikkje har nådd heilt fram, er få, og det dreiar seg berre om mindre forseinkingar.

Det kan vere fleire grunnar til at Menota TVB har nådd så langt. Gjennom Dokumentasjonsprosjektet ved Universitetet i Oslo (leia av Christian Emil Ore) vart det på 1990-talet bygd opp omfattande tekstressursar; Menota TVB har i realiteten kome inn og teke desse ressursane det siste steget fram mot publisering og integrering i si lemmatiseringsteneste, MLA. Vidare vart det gjennom prosjektet *Editing medieval manuscripts* (Centre for Advanced Study, 2000–2001) bygd opp ein tekstkodingskompetanse som har gjort at arbeidet med *The Menota handbook* har falle langt lettare enn det elles ville ha gjort. Grunnlagsarbeidet for fontutviklinga skjedde også i det same prosjektet, slik at *Medieval Unicode Font Initiative* kunne ta til høgt oppe på banen. Vadstena-prosjektet, finansiert av den svenske Riksbankens Jubileumsfond og under leiing av Karl G. Johansson, har levert både kompetanse og tekstar til arkivet.

Fleire parallelle prosjekt har vore til god nytte for Menota TVB. Det er for det første Diplom-prosjektet (2004–2006) leia av Rune Kyrkjebø, som har arbeidd med etterreformatoriske diplomtekstar i Bergen, og som har gjort bruk av og gjeve tilbakemelding på ressursar frå Menota-prosjektet. Dernest er det prosjektet *Medieval Parchment Fragments* (MPF), også det leia av Rune Kyrkjebø, der Åslaug Ommundsen ved Senter for mellomalderstudiar har koda alle latinske pergamentfragment ved Universitetesbiblioteket i Bergen, igjen etter retningslinjene i Menota (jf. <http://gandalf.aksis.uib.no/mpf/>). Og endeleg skal doktorgradsprosjektet til Eugenia Beu-Dachin nemnast; ho har koda alle latinske innskrifter frå Dacia i dagens Romania. Kodinga av dei dakiske innskriftene har vore etter full menotisk standard, dvs. på dei tre representasjonsnivåa <facs>, <dipl> og <norm>, og har vist at Menotas retningslinjer for nordiske primærkjelder har fungert godt, og faktisk utan at det har vore behov for justeringar, på eit latinspråkleg, epigrafisk materiale.

Aksis i Bergen har gjennom heile perioden vore ein god stad for drifta av prosjektet, framfor alt fordi dette miljøet har konsulentar med høg kompetanse og røynsle frå fleire tilstøytande prosjekt, og dessutan god innsikt i filologiske problemstillingar. Alt dette er slikt som blir omtalt som synergieffektar, og som – når rekneskapen no skal gjerast opp – har ført til at prosjektet har nådd så langt som det har gjort.

Det seier seg sjølv at ei eigenvurdering er prega av liten avstand, og denne vurderinga er neppe noko unntak. Derfor kan det vere grunn til å nemne at Menota-prosjektet, i lag med to andre tekstteknologiske prosjekt ved Aksis, har vore vurderte av *European Strategy Forum on Research Infrastructures* (ESFRI) i den nyss publiserte “European Roadmap for Research Infrastructures” (<http://cordis.europa.eu/esfri/>). På grunnlag av ein brei gjennomgang av europeiske infrastrukturprosjekt har ESFRI identifisert 24 “emerging projects”, derav 6 reint humanistiske. Tre tekstteknologiske miljø ved Aksis (Det hanseatiske nettverket, Wittgensteinarkivet og Menota) er definerte som eitt av seks utvalde humanistiske prosjekt i Europa, (jf. omtalen på s. 69 i rapporten: [ftp://ftp.cordis.europa.eu/pub/esfri/docs/esfri-roadmap-report-26092006\\_en.pdf](ftp://ftp.cordis.europa.eu/pub/esfri/docs/esfri-roadmap-report-26092006_en.pdf)).

Det må vere tillate å ta dette som uttrykk for at Menota TVB har lykkast i å byggje opp ein godt fundert standard for mellomalderleg, nordisk tekstutgjeving, av interesse også på europeisk nivå.

## Formidling i prosjektperioden

Dei nyaste bidraga er oppførte først i kvar seksjon.

### A. Rapportar og bøker

- The Menota handbook*, v. 2.0 beta. Utg. Odd Einar Haugen, Matthew J. Driscoll, Rune Kyrkjebø, Karl G. Johansson og Tarrin Wills. Bergen: Medieval Nordic Text Archive. Under utgjeving i 2007. <http://www.menota.org/guidelines>
- MUFI character recommendation*, v. 2.0. Utg. Odd Einar Haugen. Bergen: Medieval Unicode Font Initiative, 22.12.2006. BORA: <http://hdl.handle.net/1956/2003>.
- The Menota handbook*, v. 1.1. Utg. Odd Einar Haugen, Matthew J. Driscoll, Rune Kyrkjebø og Karl G. Johansson. Bergen: Medieval Nordic Text Archive, 05.05.2004. BORA: <http://hdl.handle.net/1956/440>.
- MUFI character recommendation*, v. 1.0. Utg. Odd Einar Haugen. Bergen: Medieval Unicode Font Initiative, 08.12.2003. BORA: <http://hdl.handle.net/1956/2004>.

### B. Tidsskriftsartiklar og kapittel i bøker

- Alois Pichler og Odd Einar Haugen. "Fra kombinerte utgaver til dynamisk utgivelse. Erfaringer fra edisjonsfilologisk arbeid med Wittgensteins filosofiske skrifter og nordiske middelaldertekster." I *Læsemåder*, utg. Per Dahl, Johnny Konderup og Karsten Kynde, s. 178–224 og 241–249. København: Reitzel, 2005.
- Odd Einar Haugen. "The Unicodification of Runic script." *NOWELE* 46/47 (2005): 87–102.
- Jon Gunnar Jørgensen. "Håndskrift- og arkivkunnskap." *Handbok i norrøn filologi*, utg. Odd Einar Haugen, Bergen: Fagbokforlaget, 2004, 27–79.
- Odd Einar Haugen. "Tekstkritikk og tekstfilologi." *Handbok i norrøn filologi*, utg. Odd Einar Haugen, Bergen: Fagbokforlaget, 2004, 81–118.
- — —. "Paleografi." *Handbok i norrøn filologi*, utg. Odd Einar Haugen, Bergen: Fagbokforlaget, 2004, 175–214.
- Johansson, Karl G. "Computing Medieval primary sources from the Vadstena Monastery. Arguments for the primary source text." *Literary and Linguistic Computing* 19/1 (2004): 93–104.
- Jon Gunnar Jørgensen, Stine Brenna Taugbøl og Hilde Bøe. "Philology meets text encoding in the new scholarly edition of Henrik Ibsen's writings." *Literary and Linguistic Computing* 19/1 (2004): 55–71.
- Odd Einar Haugen. "Parallel views. Multi-level encoding of Medieval Nordic primary sources." *Literary and Linguistic Computing* 19/1 (2004): 73–91.

### C. Konferansebidrag, presentasjonar og gjesteførelingar

- Odd Einar Haugen. "Paleography meets graphonomy: On the analysis of writing in Medieval Nordic texts." Invitert gjesteføreling. Universitetet i Zürich, 18.01.2007.
- — —. "Electronic editing of Medieval Nordic texts: Developing tools for representation, display and lemmatisation." Invitert gjesteføreling. Universitetet i Basel, 30.11.2006.
- — —. "On the diplomatic turn in editorial philology." Foredrag. *The 13th International Saga Conference*, Durham, 06.–12.08.2006.

- Odd Einar Haugen. “Mediating between manuscript appearance and logical structure: Editorial philology and XML encoding.” Invitert foredrag. *History, Archives and Digital Publishing*, Bergen, 11.–14.05.2006.
- — —. “Does appearance matter? On the usage of Unicode fonts in XML encoding.” Foredrag. *Publication et lecture numériques. Défis et processus engagés par l'édition critique et les activités de lecture*, Lyon, 05.–06.09.2005.
- — —. “Making medieval texts readable: manuscript and work fidelity in electronic text encoding.” Invitert foredrag. *Symposium on literacy studies*, Stavanger, 22.–24.04.2005.
- — —. “What happened to the critical edition?” Invitert foredrag. *Creating the medieval saga – versions, variability and editorial approaches*, Bergen, 13.–15.04.2005.
- — —. “Meeting the needs of the academic community: Unicode in medieval studies.” Foredrag. *Unicode 27*, Berlin, 06.–08.04.2005.
- — —. “Dealing with difficult characters: XML encoding and the Private Use Area of Unicode.” Foredrag. *ALLC / ACH conference*, Göteborg, 11.–16.06.2004.
- — —. “Medieval Nordic primary sources: Encoding, display and analysis.” Invitert foredrag. *Deutsch. Diachron. Digital: Antragskonferenz der Initiative Historisches Referenzkorpus des Deutschen*. Berlin, 10.–11.12.2003.
- Christian Emil Ore. “The Meta Dictionary and Menota.” Invitert foredrag. *Deutsch. Diachron. Digital: Antragskonferenz der Initiative Historisches Referenzkorpus des Deutschen*. Berlin, 10.–11.12.2003.
- Odd Einar Haugen. ”Koding av nordiske middelaldertekster i XML/TEI.” *Det digitala källmaterialet och forskningen*. Invitert foredrag. Vitterhetsakademien, Stockholm, 27.10.2003.
- — —. “Encoding strategies for Medieval Nordic primary sources.” Invitert foredrag. *The Trondheim Workshop on Electronic Corpora of the Old Languages*. NTNU, Trondheim, 19.–20.09.2003.
- — —. “Utgjeving av norrøne tekster: Elektroniske og trykte løysingar.” Invitert foredrag. *Udgivelsesprinsipper og tekstformidling*. Háskóli Íslands, Reykjavík, 05.–06.09.2003.
- — —. “Introducing the Medieval Nordic Text Archive (Menota)”. Foredrag. *The 12th International Saga Conference*. Bonn, Tyskland, 28.07–02.08.2003.
- Alec McAllister, Tarrin Wills, Odd Einar Haugen. “Medieval Unicode Font Initiative (MUFI) – a round table discussion.” Rundebordspresentasjon. *International Medieval Congress*. Weetwood, Leeds, 14.–17.06.2003.
- Odd Einar Haugen. “Abbreviations in Old Norse manuscripts: Analytical and synthetical interpretations.” Foredrag. *International Medieval Congress*. Weetwood, Leeds, 14.–17.06.2003.
- — —. “Encoding medieval characters in the Private Use Area (PUA) of the Basic Multilingual Plane (BMP) of Unicode.” Presentasjon. *MUFI group meeting*. Linacre College, Oxford University, 11.–12.06.2003.
- — —. “Koding av nordiske middelaldertekster i XML.” Invitert foredrag. *Vadstenaklosteret som text och handskriftsproducerande miljö*. Stockholm, 25.–27.04.2003.