

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

Revidert versjon, 24. september 2005

Prosjektskildring

Dette er eit treårig prosjekt [*no: forlenga med eitt år til 31.12.2006*] som skal leiast av prof. Odd Einar Haugen, Nordisk institutt, Universitetet i Bergen, i tilknytning til eit nyoppretta elektronisk arkiv, Medieval Nordic Text Archive (Menota). Administrativt vil prosjektet bli lagt til HIT-senteret [*no: Aksis*] i Bergen, og i tillegg vil fire av dei faglege medarbeidarane ved senteret delta i prosjektet; Tone Merete Bruvik, Paul Meurer, Vemund Olstad og Johan Poppe. Det vil knyte til seg tre aktive samarbeidspartnarar; Christian-Emil Ore ved DOK-gruppa (tidl. Dokumentasjonsprosjektet) og Jon Gunnar Jørgensen ved *Heimskringla*-prosjektet, begge Universitetet i Oslo, og Karl G. Johansson i styret for Menota, Växjö universitet [*no: Universitetet i Oslo*].

Det er sett av ei heil forskarstilling gjennom tre år på prosjektet, med tanke på dr.art. Rune Kyrkjebø, p.t. ved Universitetsbiblioteket i Bergen. Kvar av dei fire medarbeidarane ved HIT-senteret er innrekna med 1–2 månadsverk per år gjennom heile prosjektperioden. Vidare vil ein del av stillingsressursen til prosjektleiaren, Odd Einar Haugen, bli lagt inn i prosjektet, delvis i form av eigeninnsats og delvis i form av frikjøp. For delar av arbeidet med *Heimskringla* er det også innarbeidd eit frikjøp av Jon Gunnar Jørgensen. Utover dette vil dei tre aktive samarbeidspartnarane bidra til prosjektet gjennom eigeninnsats. [*Merknad: Då Rune Kyrkjebø vart tilsett som førstebibliotekar ved Universitetsbiblioteket i Bergen etter at søknaden vart sendt inn, er det gjort eit makeskifte slik at Odd Einar Haugen går inn i prosjektet i heil stilling, og Rune Kyrkjebø innanfor ramma av forskingsdelen av si stilling som førstebibliotekar ved UBB.*]

Menota er eit fellesnordisk prosjekt som samlar alle dei tunge utgjevingsinstitusjonane innanfor mellomalderfilologien i Norden: dei arnamagnæanske institutta i Danmark og Island, riksarkiva i Noreg og Sverige, og fleire utgjevingselskap i Sverige og Danmark. Det vart oppretta 10.09.2001 etter felles initiativ frå HIT-senteret i Bergen og DOK-gruppa i Oslo. Leiar av interimstyret for Menota er Odd Einar Haugen.

Hovudmålet for prosjektet er å bygge opp ein brei tekst- og verktøybase for Menota. Det er avgjerande at arkivet ikkje blir eit “filhotell”, men at dei tekstane som går inn i arkivet dannar eit koordinert korpus, som gjer at brukarane kan søkje i arkivet etter eit breitt utval av kriterier, samtidig som brukarane også kan få tilgang til dei enkelte tekstane som litterære verk; kort sagt slik at arkivet kan dekkje typisk behov blant både litteraturforskarar, språkforskarar og historikarar. For å oppnå dette målet må arkivet raskt etablere ein kritisk masse. Tidlegare forsøk har stranda nettopp på grunn av dette, t.d. den “Computer bank” for norrøne tekstar som vart oppretta i København på slutten av 1980-talet og som seinare vart nedlagd på grunn av manglande standardisering, tekstbidrag og teknisk vedlikehald. Det store materialet som er bygd opp i form av eit tradisjonelt setelarkiv ved Gamalnorsk ordboksverk i Oslo, kan gje Menota nettopp

denne kritiske massen. I regi av det tidlegare Dokumentasjonsprosjektet (no DOK-gruppa) har det vore gjort eit stort arbeid med å overføre desse orbokssetlane til eit elektronisk format, men dette er enno ikkje gjort tilgjengeleg innanfor rammene av ein omforeint standard. Det er dette som no kan skje i regi av Menota.

For å oppnå dette hovudmålet har prosjektet formulert tre [*no: fire*] delmål. Det første inneber ei vidareutvikling av kodingsstandarden med referanse til Text Encoding Initiative, og dessutan nødvendige fontløysingar for diplomatarisk tekstattgjeving. Det andre delmålet knyter seg til oppbygging av ein ordbank for arkivet, dvs. eit komputasjonelt leksikon, som gjer at tekstane vil danne eit verdifullt materiale for leksikografisk og annan språkvitskapleg forskning.

Det tredje delmålet knyter seg til ei oppfølging av *Heimskringla*-prosjektet, som inneber at ein sentral tekst kan gjerast tilgjengeleg i det nye arkivformatet, forutan at teksten også skal brukast som testing av ordbankprosjektet. Til dette delprosjektet er det knytt spesifikke forskingsoppgåver i tilknytning til analyse av skriftprofilar i *Heimskringla*. [*no: delprosjektet om skriftprofilar er erstatta av eitt om utvikling av elektronisk arkivstruktur*].

Det fjerde delmålet [*opprett i 2005*] er opprettinga av ein elektronisk arkivstruktur for Menota, bygd på dei nyaste standardane og i samarbeid med seinare oppretta kodingsprosjekt, m.a. Diplomprosjektet ved Universitetet i Bergen.

Samla sett omfattar prosjektet både utviklingsoppgåver og forskingsoppgåver. Ettersom utviklingsdelane av prosjektet delvis har preg av å vere pionéroppgåver, og i alle tilfelle må byggje på konkret filologisk ekspertise, er dette skiljet ikkje skarpt, og alle tre [*no: fire*] delprosjekta har derfor klar forskingskarakter.

Delprosjekt 1: utvikling av kodingsstandard og tekstvisning

(a) Handboksutvikling og visningsproblematikk

Dette delprosjektet byggjer på den kodingshandboka som vart utvikla på prosjektet *Editing medieval manuscripts* (Centre for Advanced Study, Oslo, 2000–2001) under leiing av Odd Einar Haugen. Handboka ligg no føre i ein preliminær versjon på engelsk under tittelen *Menota handbook*, redigert av Matthew Driscoll, Odd Einar Haugen, Karl G. Johansson og Rune Kyrkjebø. Dette arbeidet har knytt seg nært til Text Encoding Initiative (TEI), som er ein internasjonal standard utvikla særleg for koding av tekstar innanfor humaniora. TEI er utforma som eit sett av retningslinjer der ein legg vekt på at det er strukturen og innhaldet av tekstane, ikkje den ytre utforminga (layout, typografi), som skal kodast. Dette er implementert i kodingspråket XML (Extensible Markup Language), som er tilrådd av W3C (World Wide Web Consortium) og som no er blitt ein etablert standard på verdsveven. Det er eit rikt format, dvs. at det er eit stort sett av kodar som er tilgjengelege, og det finst alternative kodeforslag, slik at svært ulike tekstar og ulike tekstsyn skal kunne presenterast tilfredsstillande.

Sentralt for utgjeving av nordiske mellomaldertekstar er det at tekstane skal kunne presenterast på ulike nivå, i samsvar med tradisjonen for ulike typar av utgåver. I *Menota handbook* har vi valt å identifisere tre fokale nivå, nemleg (a) eit faksimilert nivå, der teksten blir attgjeven ord for ord, teikn for teikn, inkl. alle forkortingsteikn, og med relativt stort inventar av teiknvariantar; (b) eit fortolka diplomatarisk nivå, der teksten blir attgjeven teikn for teikn, men med fortolking og ekspansjon av forkortingar (typisk vist i kursiv), og ei viss forenkling av teikninventaret; (c) eit normalisert nivå, der forkortingar er oppløyste utan kommentar, og ortografi og teiknsetjing er normalisert i samsvar med gjeldande grammatikkar og ordbøker (jf. Haugen 1995). *Menota handbook* spesifiserer korleis desse tre nivåa kan presenterast innanfor ein og same fil.

Sjølv om det no ligg føre ei preliminær TEI-basert kodingshandbok for mellomalderleg tekst i form av *Menota handbook*, er det framleis mykje som står att når det gjeld verktøy for innskriving og framvising av tekst. Utfordringa ligg her i at vi ved hjelp av eigne standardar for stilark til XML skal kunne utvikle eit system for framvising i fleire ulike former. Det vil i praksis seie at ein ut frå ein enkelt XML-koda tekst skal kunne generere både visningar på skjerm (HTML) og eigne utskriftsversjonar (f. eks. Adobes PDF). I motsetnad til XML, som har vore ein offentleg standard i nokre år, er dei tilhøyrande stilarkstandardane så nye at det ikkje er utvikla nokon programvareløysingar som tilfredsstillar dei krava prosjektet har sett til visning av slike tekstar. Denne delen av prosjektet vil derfor i høg grad vere eit pionérprosjekt.

Vidare skjer det no ei rask utvikling av TEI-standarden. Versjon P4 er nettopp utkomen, og på trappene for P5 står m.a. ein gjennomgripande revisjon av kapitla om handskrifttranskripsjon og behandling av abbreviaturar. Dette gjer det nødvendig med oppdatering av retningslinjene i handboka, og det er ikkje urealistisk å rekne med at dei erfaringane som er gjort i prosjektet vil kunne påverke den vidare utviklinga av TEI-standarden på visse felt. Dette er eit arbeide som ikkje berre kan løysast på teknisk nivå, men som også krev høg filologisk kompetanse.

Text Encoding Initiative er eit konsortium som blir drive av fire samarbeidande institusjonar, nemleg University of Oxford, Brown University, Universitetet i Bergen og University of Virginia. HIT-senteret i Bergen er konsortiets hovudkontor, og Tone Merete Bruvik ved HIT-senteret er no executive director i TEI-konsortiet.

Mål

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.

Medarbeidarar

Vemund Olstad: stilarkutvikling og visningsproblematikk

Rune Kyrkjebø og Odd Einar Haugen: oppdatering av *Menota handbook*

Tone Merete Bruvik: kvalitetssikring i forhold til TEI-standarden

(b) Unicode-tilpassing

Presentasjon av diplomatarisk tekst frå norrøne handskrifter – anten det skjer på skjerm eller i utskrift – krev særskilde fontar. Gjennom den standarden som er utvikla av Unicode har ein kome langt når det gjeld visning og publisering av normaliserte og fortolka diplomatariske utgåver, ettersom mange av dei teikna som krevst for dette ligg i nytilkomne kodesider i Unicode (Latin Extended-A, Latin Extended-B, Latin Extended Additional). Men det står mykje att når det gjeld presentasjon av fullt diplomatarisk og faksimilert tekst, der ein ikkje kjem utanom bruk av det såkalla private området i Unicode.

Medieval Unicode Font Initiative, oppretta av Odd Einar Haugen, Alec McAllister (Leeds) og Tarrin Wills (Sydney), arbeider for å opprette ein felles standard på dette feltet, som vil kome både Menota og andre prosjekt til gode. Prosjektet har også knytt kontakt med fontutviklingsmiljø innanfor gammalengelsk, representert ved Peter S. Baker (Virginia), som har utvikla den eksperimentelle fonten *Junicode*, og iberisk, representert ved António H.A. Emiliano (Lisboa), som har utvikla fontar for visigotisk.

Odd Einar Haugen har utarbeidd eit første utkast til bruk av Private Use Area i Basic Multilingual Plane i Unicode. Dette er publisert på nettet (Haugen 2002), og vil bli framlagt på International Medieval Congress i Leeds i juli 2002. Prosjektet krev oppfølging både på praktisk nivå i form av koordinering av innkomne utkast, og på teoretisk nivå i form av analyse av subsystema, særleg for abbreviaturar og metrisk transkripsjon.

Mål

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.

Medarbeidarar

Odd Einar Haugen: koordinering og vidareutvikling av utkastet til Unicode-tilpassing
Vemund Olstad: testing og oppdatering av Menotas DTD

Referansegruppe

Alec McAllister, University of Leeds
Tarrin Wills, University of Sydney
Peter S. Baker, University of Virginia

Delprosjekt 2: oppbygging av ordbank og utvikling av lemmatiseringsstrategiar

Hos DOK-gruppa i Oslo ligg det no føre ei stor samling av gammalnorske tekstar som er fullstendig lemmatiserte, dvs. at dei har informasjon om oppslagsform, ordklasse og grammatisk form til kvart einaste ord. Dette er eit materiale som er konvertert frå

ordbokssetlar laga ved Gamalnorsk ordboksverk dei siste 50 ára. Materialet omfattar ein vesentleg del av den gammalnorske litteraturen:

- *Barlaams ok Josaphats saga*, Holm perg fol. 6, ca. 1275
- *Konungs skuggsjá*, AM 243 b α fol., ca. 1275
- *Óláfs saga ins helga*, Upps DG 8, ca. 1225–1250
- *Strengleikar*, Upps DG 4–7, ca. 1270
- *Ǽðreks saga af Bern*, Holm perg 4 fol, ca. 1275–1300
- *Thómass saga erkibiskups*, Holm perg 17 4to, ca. 1300
- Alle norske diplom i gammalnorsk språkform frå ca. 1200 til 1308

Etter planen skal dette materialet konverterast til Menota-standard og gjerast tilgjengeleg innanfor rammene av dette arkivet. Det er fastsett ein standardisert TEI-header for kvar tekst som opplyser om datering, proveniens, sjanger o.l. På grunnlag av denne kan det tildelast eit sett med attributt til kvart løpande ord i teksten. Dermed kan brukarane formulere presise søk i det samla tekstarkivet, ikkje berre i enkelståande tekstar.

I dette delprosjektet er siktemålet å indeksere dei TEI-koda dokumenta på grunnlag av informasjonen i TEI-header og den leksikografiske informasjonen som ligg i teksten (oppslagsform, ordklasse, grammatisk form) og leggje dette inn i eigna korpus-programvare. Det er mest aktuelt å nytte programvarepakka “Corpus Workbench”, som har vore i bruk over lengre tid både i DOK-gruppa i Oslo og ved HIT-senteret i Bergen.

Fullt implementert inneber dette at tekstane i arkivet ikkje berre kan takast ut og studerast som sjølvstendige og uavhengige verk – altså som om det var tradisjonelle tekstutgåver, men at ein også kan søkje gjennom heile tekstmaterialet og hente ut førekomstar på kvart enkelt ord, sortert etter dei kriteria som er spesifiserte i den standardiserte TEI-headeren i kvar tekstfil, ordna i KWIC-konkordansar og liknande format. Dette vil gjere tekstarkivet til eit reelt arkiv, og føre til at det store og fagleg svært solide lemmatiseringsarbeidet som er utført ved Gamalnorsk ordboksverk gjennom fleire tiår kan gjerast tilgjengeleg på ein heilt annan måte enn tidlegare.

Det neste steget i delprosjektet er å utvikle halvautomatiske lemmatiseringsstrategiar, basert på den modellen som er utvikla i “Taggerprosjektet” ved DOK-gruppa og Tekstlaboratoriet i Oslo. Dette inneber at vi vil byggje opp eit komputasjonelt leksikon på grunnlag av materialet frå DOK-gruppa og utforme dette etter mønster av den tilsvarande basen for moderne norsk, Norsk ordbank. Paul Meurer, som har reimplementert programvara for Oslo–Bergen-taggeren, vil gje nødvendig assistanse for dette prosjektet, i samarbeid med Christian-Emil Ore ved DOK-gruppa i Oslo. Dette delprosjektet vil opprette kontakt med LOGIN-prosjektet, som nettopp har fått tildelt midlar får Noregs forskingsråd.

Etappar

1. Ekstraksjon av lemma, bøyingsformer og grammatiske trekk i det gammalnorske tekstmaterialet frå Oslo og konvertering til eit format som svarar til Norsk ordbank. Utviding for å gje rom til ein rikare bøyingsmorfologi enn i moderne norsk.

2. Utvikling av ein halvautomatisk prosess for å komplettere materialet som er skapt i den første etappen. Vi ser for oss at alle attesterte lemmata blir utbygde til fullstendige paradigme, og at systemet gjer framlegg om manglende ordformer i analogi til ord i same bøyingsklasse. Tilsvarende rutinar er utarbeidde for Norsk ordbank for både bokmål og nynorsk.
3. Utvikling av halvautomatiske lemmatiseringsrutinar for norrøn, i første omgang gammalnorsk tekst, på grunnlag av den nyutvikla ordbanken.

Til den halvautomatiske lemmatiseringsrutinen må det utviklast eit enkelt brukargrensesnitt, og det må testast ut på nye, hittil ikkje lemmatiserte tekstar. I denne prosessen må nye lemmaformer og paradigme føyast til leksikonet. Til dette føremålet vil utvalde *Heimskringla*-tekstar i delprosjekt (3) bli brukt, forutan delar av handskriftet Codex Wormianus (AM 242 fol), transkribert av Karl G. Johansson.

Mål

Oppbygging av ein norrøn ordbank som spesifisert ovanfor, 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.

Medarbeidarar

Paul Meurer: ekstraksjon, konvertering og programmering
Johan Poppe: databaseoppbygging og utvikling av søkjeverktøy
Rune Kyrkjebø: testing av lemmatiseringsrutinar og grensesnitt

Samarbeidspartnar

Christian-Emil Ore, DOK-gruppa, Oslo
Karl G. Johansson, Universitetet i Växjö [no: *Universitetet i Oslo*]

Vidare perspektiv

Dette delprosjektet bør følgjast opp i seinare prosjekt, der det blir utarbeidd disambigueringsrutinar basert på ein høveleg grammatikk, mest sannsynleg av typen *Constraint grammar*.

Delprosjekt 3: elektronisk formatering av *Heimskringla*-tekstar

Heimskringla er eit eineståande bidrag til den vestnordiske historieskrivinga i mellomalderen. Verket representerer det litterære høgdepunktet i sin sjanger, og har også ei veldig resepsjonshistorie her til lands. Samtidig har det ei problematisk tekstoverlevering, ettersom fleire av dei sentrale handskriftene no er tapte, og – med unntak av nokre få blad – berre kjende gjennom postreformatoriske avskrifter; dette gjeld både *Kringla*- og *Jöfraskinna*-handskriftet. Valet av *Heimskringla* er såleis bestemt både av allmenne, nasjonale omsyn, og dei faglege utfordringane som ligg nettopp i desse tekstane.

Heimskringla blir til vanleg delt i tre hovuddelar; Hkr. I, som går fram til saga om Olav den heilage, Hkr. II, som omfattar saga om Olav den heilage, og Hkr. III, som inneheld resten av verket. *Heimskringla* er overlevert i fleire handskrifter, og blir delt i ei såkalla x-klasse (herunder Kringla-handskriftet) og ei y-klasse (herunder Jöfraskinna-handskriftet).

Hovudhandskriftet Kringla (frå ca. 1270) brann i København i 1728, med unntak av eitt enkelt blad, som då var i Sverige. Vi har det i dag overlevert i seinare avskrifter av islendingar, m.a. Ásgeir Jónsson; denne delen av *Heimskringla*-tradisjonen er seinast behandla av Jon Gunnar Jørgensen i hans doktoravhandling (2000). Jöfraskinna (ca. 1325) er også tapt, med unntak av fire heile blad, men er overlevert i ei avskrift frå 1567 av nordmannen, biskop Jens Nilssøn. Studiet av dette handskriftet, AM 37 fol, danna grunnlaget for doktoravhandlinga til Rune Kyrkjebø (2001).

Dei sentrale handskriftene i x-klassa av *Heimskringla* ligg no føre innanfor rammene av *Heimskringla*-prosjektet i Oslo, dels i form av direkte transkripsjon av Kringla-teksten og dels i form av variantopplysningar innlagde i denne transkripsjonen. Det vil dermed vere av stor verdi å utvikle ein felles kodingsstandard for desse tekstane og det tekstmaterialet som no er under oppbygging i regi av Menota.

Framstilling av høgkvalitets koda primærttekst av *Heimskringla* gjer det mogleg å prøve ut ei rekkje interessante hypotesar som gjeld den ortografiske variasjonen i materialet relatert til spørsmåla om skrivarhender og språkleg variasjon i desse manuskripta. Både i Jöfraskinna-materialet aleine, og ikkje minst når ein dreg inn Kringla-versjonen, står forskinga i dag med gode hypotesar bygde på metodiske diskusjonar og manuelle granskningar. Med utgangspunkt i dei seinaste avhandlingane på feltet (Jørgensen 2000, Kyrkjebø 2001), vil ein med den koda teksten få tilgang til empiriske data som kan spele tilbake mot dei problemstillingar som er reiste og diskuterte.

For Jöfraskinna ligg det i dag føre elektroniske transkripsjonar i koda form. Desse er grunnlaget for Kyrkjebøs avhandling (2001). Transkripsjonane vart til dels framstilte før det vart semje i fagmiljøet om ein standardisert framgangsmåte. I dag ligg standarden føre i *Menota handbook*. For Jöfraskinna-transkripsjonane ligg arbeidet derfor i å konvertere den eksisterande kodinga til Menota-form. Det er ei oppgåve som er teknisk enkel, i og med at tekstane allereie er i SGML, men det er ei oppgåve som krev tid fordi volumet på den koda teksten er nokså stort (250 sider i tekstutgåva i Kyrkjebø 2001).

For Kringla-teksten sitt vedkommande er det nødvendig å konvertere frå eit ikkje-SGML format. Dei data som er koda i Kringla-transkripsjonane er likevel av same type som i Jöfraskinna-transkripsjonane, og tekstane er på alle måtar samanliknbare når det gjeld språk og ortografi. Transkripsjonsnivået er også tilsvarande. Overføring av Kringla-filane til Menota-format er derfor ei utfordring som Kyrkjebø kan løyse i samarbeid med Jørgensen. I første omgang vil dette gjelde *Heimskringla* II og III, medan det står tilbake ein korrektur- og kontrollrunde før *Heimskringla* I kan konverterast til Menota-standard. Det sistnemnde arbeidet må utførast av høgt kvalifisert personale, og skal etter planen gjerast av prof. Jon Gunnar Jørgensen, Oslo.

Etappar

1. Utarbeiding av ein fullt validert tekst i Menota-standard av første delen av *Heimskringla* (frå *Ynglinga saga* til *Óláfs saga ins helga*) etter Jöfraskinna-teksten hos Jens Nilssøn (AM 37 fol.). Denne kodinga skal omfatte både faksimilert og diplomatarisk nivå.
2. Utarbeiding av rutinar for konvertering av den standard som hittil er brukt i *Heimskringla*-prosjektet i Oslo til den som er utvikla i Menota-prosjektet, og dokumentert i *Menota handbook*. Siktemålet er å konvertere *Heimskringla* II og III i x-klassa.
3. Korrekturlesing og kontroll av *Heimskringla* I i x-klassa, og deretter konvertering av også denne til Menota-standard.

Mål

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.

Medarbeidarar

Rune Kyrkjebø: edisjonsfilologisk arbeid og tekstkoding

Odd Einar Haugen: edisjonsfilologisk assistanse

Samarbeidspartnar

Jon Gunnar Jørgensen, *Heimskringla*-prosjektet, Universitetet i Oslo

Delprosjekt 4: utvikling av elektronisk arkivstruktur

Menota-prosjektet bygger eit elektronisk arkiv som består av digitale tekster. Tekstene er produserte gjennom avskrivning av fysiske tekstberarar som typisk har sin plass i ei større eller mindre nordisk manuskriptsamling. Frå manuskriptsamlingane følgjer det med kvart manuskript eit meir eller mindre standardisert sett av katalogdata. Dette er opplysningar som identifiserer manuskriptet med signatur, alder, opphav og historie, og ofte grovt beskriver strukturen i det. Slike kjerneopplysningar er avgjerande å la følgje med den elektroniske manuskriptteksta som blir lagd inn i Menota sitt arkiv.

Den elektroniske teksta, ein gong komen inn i Menota, blir ikkje arkivert der på ein statisk måte slik som originalobjektet i manuskriptsamlinga, men går inn i ein open prosess av oppriking med elektronisk tekstmerking. Med det startar ein ny fase i teksta si forskingshistorie. Det oppstår eit arkivfagleg behov for dokumentering av denne fasen. Informasjon som må takast vare på, er opplysningar om tekstkodinga og den elektroniske handsaminga elles av teksta, det er nye bitar av filologisk kunnskap som oppstår i Menota-arkivet, og det er eventuelt nytt tilfang av bibliografiske referansar.

Det er såleis relevant for Menota-prosjektet å utarbeide seg ein arkiv-infrastruktur som gjev plass for, og knyter saman, både informasjon frå manuskriptinstitusjonane sine katalogar, og informasjon som oppstår frå og med at teksta er komen inn i Menota. Det er målet for dette delprosjektet.

For elektronisk koding av tradisjonell filologisk kataloginformasjon om manuskript er det i dag vanleg å nytte kodenstandarden TEI, med det særleg tilpassa Header-elementet. Dette er ei tekstkoding som godt fangar opp kortare og lengre arkivfaglege og filologiske beskrivelsar av eit manuskript. TEI er den standarden Menota arbeider etter, og den ligg til grunn for prosjektet sine publiserte retningslinjer. TEI handterer i Header-elementet også informasjon om skapinga av og oppbygginga av den elektroniske manuskriptteksta. Menota har spesifiert utfyllinga av eit minimalt sett av opplysningar i TEI Header.

Menota treng også ei tekstkoding og ein digital infrastruktur som ikkje fokuserer på enkeltokument, men på mengder av digitale dokument inndelt i undergrupper og kategoriar. Ein slik standard, internasjonalt brukt og veldokumentert, er EAD (Encoded Archival Descriptions). EAD inneber eit arkivperspektiv, meir enn eit dokumentperspektiv. Det er grunn til å tru at denne standarden kan tilpassast Menota for å ta vare på arkivbehova nemnde ovanfor.

Dette delprosjektet vil dra nytte av utviklingsarbeid med standarden EAD som skjer i det fagleg tilgrensande og samlokaliserte prosjektet Diplom (Digital infrastruktur for diplomsamlinga ved Universitetsbiblioteket i Bergen). Den digitale infrastrukturen som Diplom utviklar, tek hand om kataloginformasjon, tekst og bilde av dokument. Den vil vere eit konkret utgangspunkt for utviklinga av arkivinfrastrukturen for Menota.

Lenker:

<http://www.loc.gov/ead/>

<http://gandalf.aksis.uib.no/diplom/>

Etappar

1. Utarbeiding av meir presise retningslinjer for opptak til Menota av katalogdata frå manuskriptinstitusjonar, med vekt på bruken av TEI Header til kvar tekst.
2. Utarbeiding av EAD-struktur som ser heile Menota som eitt hierarkisk strukturert arkiv, med fokus på fordeling av data mellom EAD-postar og TEI Header.

Mål

Utarbeiding av arkiv-infrastruktur for tekster, katalogdata og informasjon om elektroniske versjonar for Menota, på grunnlag av EAD og TEI.

Medarbeidar

Rune Kyrkjebø: arkivfilologisk ekspertise

Referansar og sitert litteratur

Text Encoding Initiative:	www.tei-c.org
Medieval Nordic Text Archive:	www.menota.org
Editing medieval manuscripts:	www.shs.uio.no/Groups/EdMa2000/
Menota handbook:	www.aksis.uib.no/menota/guidelines/
Medieval Unicode Font Initiative:	www.aksis.uib.no/mufi/
Oslo–Bergen-taggeren:	http://decentius.hit.uib.no/cl/cgp/test.html

- Haugen, Odd Einar. 1995. “Constitutio textus. Intervensjonisme og konservatisme i utgjevinga av norrøne tekster.” *Nordica Bergensia* 7: 69–99.
- — —. 2002. “A proposal for subranges in the Private Use Area of Unicode: Supplements to the Latin alphabet for Medieval texts.” Bergen: Medieval Unicode Font Initiative (www.hit.uib.no/mufi/proposal/proposal-v1.html).
- Johansson, Karl G. 1997. *Studier i Codex Wormianus. Skrifttradition och avskriftsverksamhet vid ett isländskt skriptorium under 1300-talet*. Nordistica Gothoburgensia 20. Göteborg: Universitetet i Göteborg.
- Jørgensen, Jon Gunnar. 1999. *Ynglinga saga etter Kringla (AM 35 fol)*. Oslo: Universitetet i Oslo.
- — —. 2000. *Det tapte håndskriftet Kringla*. Acta humaniora 80. Oslo: Det historisk-filosofiske fakultet, Universitetet i Oslo.
- Kyrkjebø, Rune. 2001. *Heimskringla I etter Jöfraskinna. Karakteristikk av tekstvitna samt tekstkritisk utgåve av Jens Nilssøns avskrift i AM 37 folio*. [Upublisert. Under utgjeving i *Editiones Arnarnæanae*, København]. Bergen: Nordisk institutt, Universitetet i Bergen.
- Menota handbook*. 2002. Eds. Matthew J. Driscoll, Odd Einar Haugen, Karl G. Johansson and Rune Kyrkjebø. Bergen etc.: Medieval Nordic Text Archive.
- Ólafur Halldórsson. 2001. *Text by Snorri Sturluson in Óláfs saga Tryggvasonar en mesta*. London: Viking Society for Northern Research, University College London.