# Clarin.dk WP 5

## Status pr. 25/11-2010

Anders Conrad

Version 1 af infrastrukturen frigives til intern brug den 1/12-2010, og denne frigivelse anses som WP 5.1’s aflevering ”T28: Beta version of infrastructure”. Denne release omfatter deponering af tekst-resurser i det såkaldte *TEIP5DKCLARIN* format, som defineret af WP 2. Deponering foregår via en web-grænseflade på <http://infra.clarin.dk/deposit/web/>, og denne er fra og med den 1/12 åben for projektets parter for deponering. **Bemærk venligst at der er tale om en midlertidig grænseflade som alene er til intern brug i projektperioden!** Version 1 af infrastrukturen omfatter desuden design af Clarin.dk datacentret og implementering af dette i rammerne af hum.ku.dk’s datacenter hvor vi har lejet os ind. Der er derfor etableret backup osv., og datacenteret er fra og med 1/12-2010 i driftstilstand og administreres af Tomas Groth Christensen fra DSL ([tgc@dsl.dk](mailto:tgc@dsl.dk)). Infrastrukturen er med andre ord klar til at modtage og håndtere data fra arbejdspakkerne.

Der mangler dog endnu muligheden for at kunne deponere alle de øvrige typer af resurser i projektet. Deponering af disse planlægges implementeret i januar 2011 til frigivelse i version 1.1 den 1/2-2011. For at kunne implementere deponering af en given datatype kræves at vi har modtaget testdata af den pågældende type. Vi har for indeværende modtaget testdata af typen TEI XML (non-WP 2 standard), TEIP5DKCLARIN +billeder (WP 2.3) og leksikalske resurser (WP 4). For WP3 har vi til dels modtaget testdata, men der resterer endnu at få etableret en validator og validere de leverede data. Alle arbejdspakker bedes sikre at I har leveret de fornødne testdata til Lene Offersgaard **senest den 31/12-2010.** Hvis denne dato overskrides, kan vi ikke love at de eventuelle manglende resursetyper vil kunne deponeres før ved projektets afslutning i sommeren 2011. Tag gerne kontakt til Lene med det samme for at aftale detaljerne omkring dette vigtige punkt!

Web-udviklingen er i gang, i rammerne af WP 5.2. Der er udviklet skitser, nedsat en følgegruppe og aftalt møder med denne. Webprocessen er tæt koblet med udviklingen af den bagvedliggende søge- og visningsfunktionalitet i WP 5.1. Dette arbejde omfatter, ud over selve søgningen og visningen af resurserne, processen med at implementere kravene til adgangskontrol og hele den juridiske håndtering. Søgning kommer til at omfatte metadata- og fritekstsøgning med visning af resultatlister. Webudviklingen forventes afsluttet med en intern aflevering den 1/4 -2011 hvorefter både Tom Juul Andersen og Jacob Larsen fra Det Kongelige Bibliotek overgår til andre opgaver end Clarin. Frem til sommeren 2011 vil WP 5.1 herefter færdiggøre den bagvedliggende funktionalitet og gøre klar til den endelige aflevering af infrastrukturen.

Med hensyn til tools foregår der en del udviklingsaktivitet. Tools-udviklingen har fra starten foregået asynkront og relativt uafhængigt af resten af infrastrukturarbejdet. Det er derfor vanskeligt for mig at give en præcis status, men det forventes at det ved projektets afslutning vil være muligt at integrere tekst-baserede værktøjer. Der foreligger for indeværende stadig ikke nogen generel arkitektur for integrationen af værktøjer, og der er derfor en risiko for at der i enkelte arbejdspakker i Clarin.dk kan blive udviklet værktøjer eller værktøjslignende funktioner som ved projektets afslutning ikke vil kunne integreres i den tekniske infrastruktur. Jeg ser som arbejdspakkeleder desværre umiddelbart ingen mulighed for at ændre dette forhold.

Den afsluttende aflevering af den tekniske infrastruktur vil ske i form af version 2 som er planlagt til projektets afslutning den 30/6-2011. Version 2 vil ud over deponering af samtlige resursetyper til repositoriet (såfremt vi har modtaget testdata i tide!) – indeholde den endelige webgrænseflade med søge- og visningsfunktioner. Der er i WP 5.1 forventning om at der vil kunne leveres en basal visningsfunktion af alle resurse- og datatyper. Derimod tager vi forbehold over for levering af mere avancerede visninger som næppe kan forventes implementeret inden for tidsrammen. Det kan betyde at nogle af parterne i projektet kan blive skuffet over hvordan lige netop deres resurse præsenterer sig. Jeg ser umiddelbart kun begrænsede muligheder for at kunne imødekomme mere specifikke ønsker fra de forskellige forskningsmiljøer inden for den givne projektramme. Som jeg tidligere har nævnt, er infrastrukturen rent teknisk designet på en måde der giver mulighed for at der senere kan tilføjes nye og mere avancerede brugergrænseflader som i højere grad tilgodeser bestemte forskergrupper.

En eventuel skærpelse fra styregruppens side af kravene til den endelige levering i forhold til den ovenstående status vil kræve en stillingtagen til finansiering, bemanding, tidsfrister og ledelse for det videre arbejde. Jeg står i givet fald gerne til rådighed for en diskussion omkring dette.

Jeg vil endvidere gøre opmærksom på at WP 5.1 gerne skal have at vide senest omkring 1/4-2011 hvor og hvordan platformen skal sættes i drift ved projektperiodens udløb, for at kunne forberede sammenpakning og overflytning af data og programmer til driftsleverandøren ved projektets afslutning.

Desuden skal styregruppen være opmærksom på at Mark Logic supporten udløber den 31/3-2011 da vi ikke på tegningstidspunktet var vidende om at projektet ville løbe længere end denne dato. Vi har ikke i WP 5.1 midler til at forny supportaftalen, men kan selvfølgelig fortsætte driften af den installerede software efter denne dato.

Venlig hilsen

Anders Conrad

Arbejdspakkeleder WP 5