



## Nyhedsbrev december 2009

Velkommen til Netarkivets nyhedsbrev.

### Status på bitarkiverne

De nye bitarkiv-maskiner på Det Kongelige Bibliotek er nu på plads. Der er indkøbt 12 nye, kraftige HP maskiner og de gamle arkivdata er overført til disse maskiner. Den sidste af de 56 gamle maskiner forventes slukket allersnarest. De nye maskiner kører helt stabilt, modsat de gamle som gav mange driftsforstyrrelser. Desuden bruger de nye maskiner kun halvt så meget strøm - per maskine.

Netarkivets bitarkiv fylder i øjeblikket ca. 130 TB. Vi har i alt 180 TB på de nye maskiner, men med den hastighed nettet vokser på, forventer vi kun at have ledig plads til medio 2010.

Statsbibliotekets bitarkiv, der består af blot to kraftige maskiner, kører upåklageligt.

### Anden hardware udskiftning

Den centrale maskine i Netarkivet med det flotte navn "kb-prod-adm" er blevet forældet i forhold til dens væsentlige funktion: At binde hele systemet sammen og blandt andet afvikle batchjobs. Typisk begyndte standard batchjobs at trække hele Netarkivet ned. Fx var det ikke længere muligt at køre checksumsjobbet via NetarchiveSuite GUI'en. Derfor er der sat et hastearbejde i gang med at finde en ny kraftig efterfølger. Maskinen er fundet og installeret. I starten af januar skal den indvies.

Det viste sig, at der er andre emner, der skal tages i betragtning for at have en ny løsning, der er fremtidssikret. For eksempel bliver det allerede i 2010 nødvendigt at køre Netarkivet som 64-bit applikation og ikke som 32-bit applikation, som vi plejer.

### Ottende tværsnitshøstning

Denne tværsnitshøstning er netop afsluttet (d. 20. dec. 2009).

Den var planlagt til at køre over 3 måneder evt. lidt mere. Men sådan skulle det ikke gå.. En fejl i indeksgenereringen, nedsatte høstningskapacitet på netværket på Det kongelige Bibliotek og nogle andre mindre problemer forsinkede planen 2 måneder selvom vi imødegik disse problemer så godt som muligt. Og problemerne er løst således at de ikke vil optræde ved næste tværsnitshøstning. Den gode nyhed er dog at vi fik en ekstra step-1 høstning i 2009. Step-1 er den første af de to dele af tværsnitshøstningen som indeholder de mange små websites i Danmark.

I 2010 burde vi nå op på 3 tværsnitshøstninger pr år - teoretisk med mulighed for en fjerde tværsnitshøstning.

Dette sikres med nyeste version 3.10. af NetarchiveSuite og fordi der nu er givet grønt lys til at skruer 50 % op for høstningskapaciteten på Det kongelige Bibliotek, og det vil vi selvfølgelig benytte os af fra januar.

Samlingsafdelingen på Det Kongelige Bibliotek står for arbejdet med tværsnitshøstninger.

## Selektive høstninger

De selektive høstninger af udvalgte netsteder kører planmæssigt, men der er efterhånden meget indhold som er svært høstbart. Derfor assisterer udviklergruppen nu samlingsfolkene med at høste de komplekse websites. Det er typisk steder med login adgang eller fx med video indhold. Dette såkaldte webcrawler-engineer arbejde er nu prioriteret ind på linje med de øvrige udviklingsopgaver, så dette arbejde ikke overses.

Samlingsafdelingen på Statsbiblioteket står for arbejdet med de selektive høstninger.

## Begivenhedshøstninger

Begivenhedshøstninger i 2009 blev:

- EU-parlamentsvalg i juni
- Kommunalvalg2009 i november
- Klimakonference i København i december.
- Influenza-pandemien
- Begivenheden omkring irakere i Vor Frue kirke og Brorsons Kirke
- Jægersoldatbogen (enkelt URL høstning)

Det er relative små datamængder, der høstes.

Samlingsafdelingerne på Det Kongelige Bibliotek og Statsbiblioteket koordinerer sammen arbejdet med begivenhedshøstningerne.

## Internationalt samarbejde og konferencer

Netarkivet deltog i IIPC (International Internet Preservation Consortium) i San Francisco i oktober.

## Open source NetarchiveSuite

Det internationale samarbejde omkring udvikling af NetarchiveSuite er nu blevet hverdag. Der er regelmæssigt telefonmøder med deltagelse af udviklere/kuratorer fra ONB (Østrigs nationalbibliotek) og BnF (Frankrigs nationalbibliotek). Vores udenlandske deltagere producerer nu nyttig kode, test og review.

Dette samarbejde forventes udbygget i 2010. Der er aftalt en workshop i Wien i Østrig til september. Det er foreløbig afsat 2½ dage.

Open source NetarchiveSuite har frigivet stable release 3.10. Testmiljøet er netop opgraderet til denne, og der er lavet formel accepttest her. Næste trin er at opgradere produktionsmiljøet også.

## Redaktionsgruppemøde

Kulturministeriet har beskikket nye medlemmer til redaktionsgruppen, der træder sammen til første møde d. 11. januar.

Glædelig jul og godt nytår  
Claus Lomborg, Netarkivet