



Nyhedsbrev november 2010

Velkommen til Netarkivets efterårs nyhedsbrev.

Status på bitarkiverne

Driften af de 12 HP bitarkivmaskiner på Det Kongelige Bibliotek er som forventet fuldstændig stabil. Driften af tilsvarende bitarkivmaskiner på Statsbiblioteket er som vanligt problemfri.

Checksum kørsler kører igen problemfrit på arkivkopierne på Det Kongelige Bibliotek og på Statsbiblioteket. De to arkivkopier holdes regelmæssigt op mod hinanden og evt små forskelle rettes op så de er 100 % identiske.

Datamængden i Netarkivet er nu nået op på samlet ca. 160 TB. En tværsnitshøstning fylder ca. 28 TB og dette tal er stigende med ca. 15 % per år. Derudover høstes hvert år ca. 5-6 TB for selektive høstninger og begivenhedshøstninger tilsammen, hvor de selektive høstninger fylder langt størstedelen. Der arbejdes stadig på at konvertere en række gamle samlinger fra før år 2005 ind i Netarkivet.

Hardware status

Proceduren for udskiftning af den centrale administrationsmaskine i Netarkivet, KB-PROD-ADM, skal afprøves "live". Da denne maskine er single-point-of-failure, er vi nødt til at være sikre på, at vi kan håndtere et katastrofesenarie. Derfor vælger vi at lave en udskiftning af maskinen helt fra scratch ud fra en båndkopi af det software, der ligger på den. Dette er planlagt til november.

Netarkivets hardware er pænt opdateret på bitarkiv delen og på den administrative maskine. Men de ældste høstermaskiner er over 5 år gamle og begyndt at give forskellige fejl og heller ikke længere dimensionerede til høstningsopgaverne. Derfor har Det kongelige Bibliotek besluttet at købe 2 nye høstermaskiner og en ny accessmaskine, der også har nået pensionsalderen. På Statsbiblioteket giver sparerunder ikke luft til hardware udskiftning i øjeblikket.

Tværsnitshøstning

Den 10. tværsnitshøstning blev startet d. 15. juli 2010 efter nogle drillerier med opgradering af NetarchiveSuite til version 3.12.2. Tværsnitshøstningen var planlagt til at køre over cirka 4 måneder, men en fejl, der hober beskeder op i systemet og bruger maskinhukommelsen op, har været svær at få fjernet helt. Af denne grund er tværsnitshøstningen forlænget med ca en måned.

Selektive høstninger

De selektive høstninger af udvalgte 80 - 100 netsteder kører planmæssigt. Kvalitetskontrol er en udfordring, da webstederne ændrer sig hele tiden og høstningerne ikke nødvendigvis følger med.

Da mængden af svært høstbart webindhold fylder mere og mere, er det blevet en fast opgave for udviklerne sammen med samlingsmedarbejderne. Fra nytår vil der være faste regelmæssige møder mellem samlingsfolk og udviklere. De komplekse websites indeholder for eksempel login adgang, video, lyd eller helt nye teknologier.

Begivenhedshøstninger

Mærkeligt nok er der ikke indtruffet begivenheder i 2010, der har været tilstrækkeligt interessante og relevante til begivenhedshøstninger. Faktisk har der ikke været begivenhedshøstninger siden klimatområde i København, Vinter OL og jordskælvet i Haiti – bortset fra en mindre høstning af Plummer sagen.

Af tidligere erfaring vides det dog, at begivenheder sagtens kan komme i klumper. Når en begivenhed indtræffer, tager samlingsmedarbejderne på Det kongelige Bibliotek og Statsbiblioteket sammen initiativ til at finde relevante URL'er og iværksætter en høstning af dem.

Internationalt samarbejde og konferencer

Netarkivet deltog i Wien, Østrig d. 22 – 24. september i

- 2010 IIPC (International Internet Preservation Consortium) Working Group Meetings
- iPres 2010 (International Conference on Preservation of Digital Objects)
- IAWAW 2010 (International Web Archiving Workshop)

Umiddelbart forud herfor afviklede NetarchiveSuite samme sted sin årlige workshop som omtalt i næste afsnit.

Open source NetarchiveSuite

20 - 22. september afvikledes en 2½-dages workshop omkring NetarchiveSuite i Wien med deltagelse af de aktive bidragere – det vil sige BnF, ONB og Netarkivet. Netarkivet havde i alt 6 personer af sted (plus yderligere 2 til de øvrige konferencer). Workshopen indeholdt en teknisk del, en mere samlingsorienteret del samt en ½ dags tutorial. Tutorial var ment som en appetitvækker til måske kommende brugere og aktive deltagere. Mest repræsenteret var nationalbibliotekerne fra Tjekkiet og Ungarn.

Den tekniske og mindre tekniske del opsamlede med succes hver en række ønsker til fremtidig udvikling af NetarchiveSuite. Tutorial blev også en succes med mange deltagere.

Det er 2. gang, der har været holdt en NetarchiveSuite workshop, og første gang der har været holdt en tutorial. Vi kender hinanden fra regelmæssige telefonmøder med udviklere fra Österreichischen Nationalbibliothek (ONB) og La Bibliothèque nationale de France (BnF). Men på de fysiske workshops har vi lært hinanden bedre at kende. Dette gør samarbejdet mere levende og tilfredsstillende.

Det hører med til historien at vores udenlandske samarbejdspartnere igennem et stykke tid har skabt nyttig kode, test og reviews. Og at dette samarbejde fremover skal udbygges. Specielt BnF lægger en stor del timer i udviklingsarbejdet.

NetarchiveSuite har netop færdigtestet stable release 3.14. Netarkivets produktionsmiljø er planlagt opgraderet til denne til januar.

Redaktionsgruppemøde

Den nye redaktionsgruppe har haft sit andet møde d. 8. juni og skal have sit tredje møde d. 22. november. Netarkivet tager gruppens råd og vurderinger op i styregruppen og gør dem så vidt muligt operationelle.

Claus Lomborg, Netarkivet
clo@netarkivet.dk