

## Stor interesse for Open Source software: NetarchiveSuite

En fremsynet lov har resulteret i et dansk Open Source produkt, NetarchiveSuite til arkivering af websider, der allerede har vakt international interesse. Udviklingen af NetarchiveSuite blev påbegyndt i 2004 af Det Kongelige Bibliotek og Statsbiblioteket i Århus således at de to biblioteker kunne varetage den webarkivering, som var en del af den nye elektroniske pligtafleveringslov, der trådte i kraft 1. juli 2005.

Sidst i denne uge mødes 17 udviklere fra 10 forskellige lande til en workshop på Det Kongelige Bibliotek, hvor NetarchiveSuite gennemgås. De repræsenterer flere af de store betydende biblioteker i Europa, samt Internet Archive fra USA og afspejler den interesse, der har været omkring den danske udvikling siden koden blev frigivet som open source i juli. Den private amerikanske, non-profit organisation Internet Archive har siden 1996 opbygget verdens største webarkiv. Før frigivelsen blev NetarchiveSuite testet af Internet Archive og af det norske nationalbibliotek og fik ved den lejlighed meget rosende omtale. Tilbagemeldingen fra Det norske nationalbibliotek indeholdt ord som "potentiale for at blive defacto værktøj til administration af webcrawler".

Arbejdet bygger ovenpå en aktivitet, som nationalbibliotekerne i de fem nordiske lande har haft sammen med det amerikanske firma Internet Archive og som resulterede i crawlersoftware Heritrix, Open Source produktet. Heritrix er i dag de facto standard i hele verden for indhøstning af websider.

Målgruppen for NetarchiveSuite er større institutioner, som ønsker at indsamle websider. NetarchiveSuite understøtter de tre indsamlingsstrategier, som der arbejdes med i Danmark i forbindelse med pligtafleveringen af internettet, nemlig indsamling af alt fra en række prædefinerede domæner (tværsnitshøstning), periodevis (eksempelvis daglig) indsamling fra udvalgte domæner samt definition og efterfølgende indsamling i forbindelse med begivenheder, der skønnes at være særligt relevante for eftertiden.

NetarchiveSuite er en komplet software pakke til webarkivering. NetarchiveSuite's væsentligste funktioner er planlægning og afvikling af arkiveringer af dele af internettet. NetarchiveSuite adskiller sig fra andre webarkiveringspakker ved, at den ikke blot er beregnet til at håndtere indsamling fra et mindre antal websites, men også kan anvendes til de store nationale indsamlinger på top domæne niveau, som f.eks. .dk, hvor der nu er mere end 800.000 domæner. NetarchiveSuite har desuden en indbygget bitbevaringsfunktion, der gør det muligt løbende at replikere det indsamlede materiale på flere geografiske lokaliteter.

Læs mere om projektet og den frigivne software på [www.netarkivet.dk/suite](http://www.netarkivet.dk/suite).

Nærmere oplysning kan fås hos projektets leder Christen Heedegaard , [chh@kb.dk](mailto:chh@kb.dk) eller hos Netarkivets driftsleder Bjarne Andersen, [bja@statsbiblioteket.dk](mailto:bja@statsbiblioteket.dk)