



Strategi for indsamling af materiale til Netarkivet

Introduktion

Dette dokument udgør strategien for indsamling af materiale til Netarkivet ved Det Kongelige Bibliotek og Statsbiblioteket. Dokumentet indgår i et hierarki af dokumenter, der også omfatter:

- *Politik for indsamling af materiale til Netarkivet*
- *Høstning i praksis*
- *Årsplan for indsamling af materiale til Netarkivet*

Den teknologiske udvikling siden Netarkivets start i 2005 har gjort, at udfordringerne med indsamling af materiale, der falder ind under loven, er øget betragteligt. Samtidig er der større krav fra brugerne af arkivet om, at materialet bliver indsamlet på en sådan måde, at det forbedrer mulighederne for at arbejde med materialet.

Netarkivet ønsker i denne strategiperiode fra 2015-2017 at arbejde med at forbedre indsamlingen i forhold til:

- nye webteknologier
- sociale medier
- betalingsbelagt materiale

Dette vil medføre aktiviteter som

- afprøvning og evt. idriftsættelse af nye værktøjer samt opdatering af eksisterende indsamlingssystem
- analyse og opdatering af indsamlingsmetoder med henblik på optimal indsamling
- udforskning og prioritering af mulighederne inden for området betalingsbelagte elektroniske kommunikationsnet, evt. pilotprojekter

Dokumentet opdateres og godkendes af direktionserne på Det Kongelige Bibliotek og Statsbiblioteket) hvert tredje år.

Dokumentet er offentligt og publiceres på Netarkivets hjemmeside (www.netarkivet.dk) i en dansk og en engelsk udgave.

Formål

Formålet med dette dokument er at opregne de strategiske mål for perioden 2014-16 og herigennem at udmønte den politiske målsætning i operationaliserbare punkter til implementering i Netarkivets handlingsplaner.

Dokumentet skal dels tjene til at informere offentligheden om, hvad Det Kongelige Bibliotek og Statsbiblioteket vil satse på for at løse opgaven med indsamling af det danske internet, dels fungere som intern rettesnor for prioritering af opgaverne.

Rammer for indsamling

Ansvar og roller

Det Kongelige Biblioteks og Statsbibliotekets samarbejde om Netarkivet er nedfældet i en samarbejdsaftale, som revideres hvert tredje år.

Netarkivet skal løbende evaluere de tekniske værktøjer og metoder i forhold til det aktuelle indsamlingsfokus for at sikre, at udviklingsarbejdet understøtter løsningen af aktuelle udfordringer i indsamlingen.

Videndeling og kompetenceudvikling

Netarkivet vil sikre videndeling og kompetenceudvikling gennem såvel interne som eksterne aktiviteter.

Internt afholdes månedlige videomøder mellem kuratorerne på Det Kongelige Bibliotek og Statsbiblioteket samt jævnlige fysiske og virtuelle møder mellem kuratorer og udviklere om prioriteringen af mulige tekniske løsninger på udfordringerne i forhold til høstning-, og visning/adgang.

Eksternt vil kompetenceudvikling og videndeling ske ved aktiv deltagelse i internationale konferencer og workshops, samt ved at bidrage til globale indsamlingsinitiativer i regi af IIPC (International Internet Preservation Consortium).

Netarkivet vil tilstræbe samarbejde med relevante forskningsmiljøer på indsamlingsområdet for at sikre relevans og kvalitet af det indhøstede. Der er desuden særlig fokus på at understøtte forskningsprojekter, som adresserer forbedring og effektivisering af indsamlingen ved automatiserede metoder og udvikling af alternative høstningsmetoder.

Anvendelse af standarder og åben software

Netarkivet vil bidrage til udviklingen af grundlæggende indsamlingsværktøjer med henblik på forbedring af den generelle indsamling. Dette vil i den kommende strategiperiode ske ved at indarbejde de nyeste produktionsmodne open source værktøjer i vores infrastruktur samt aktivt bidrage til vedligeholdet og videreudviklingen af de grundlæggende open source værktøjer.

Netarkivet bidrager desuden til analyse, udvikling og evaluering af standarder, metoder og værktøjer i international sammenhæng gennem IIPCs arbejdsgrupper til løsning af aktuelle tekniske problemstillinger ved indsamlingen af web-materiale.

Indsamling

Pligtafleveringsloven foreskriver, at Netarkivets indsamling kombinerer tre grundlæggende indsamlingsstrategier, som komplementerer hinanden og derved sikrer, at indsamlingen bliver så komplet som mulig.

Disse tre indsamlingsstrategier er:

Tværsnitshøstning: Indsamling fire gange årligt af alt materiale fra domænet .dk samt fra andre domæner (fx .net, .com), der retter sig mod et publikum i Danmark.

Der vil desuden være fokus på at finde metoder til at udvide indsamlingen af materiale, som publiceres uden for det danske toplevel domæne .dk i takt med, at en stigende del af det materiale, som er omfattet af pligtafleveringsloven, befinder sig der.

Selektiv høstning: Hyppig indsamling af materiale, som offentliggøres på ca. 80 udvalgte netsteder, dækkende nyhedsmedier, et repræsentativt udsnit af typiske og meget besøgte netsteder, samt netsteder, der er særligt karakteristiske, eksperimentelle eller unikke.

Netarkivet vil løbende analysere og revidere de 80 udvalgte netsteder med særligt henblik på sammensætningen af de forskellige typer af netsteder. Den manuelle indsamling af teknisk og indholdsmæssigt udfordrende netsteder skal bidrage direkte til at give viden og erfaring til alle Netarkivets indsamlinger.

Begivenhedshøstning: Indsamling af materiale, som offentliggøres i tilknytning til begivenheder – planlagte så vel som spontane.

Netarkivet vil bl.a. anvende begivenhedshøstninger til at få erfaring med indsamling af nye medietyper, formater og indhold til gavn for alle Netarkivets indsamlinger.

Herudover vil Netarkivet indsamle materiale, der ikke er omfattet af ovennævnte strategier. Det drejer sig bl.a. om understøttelse af forskningsprojekter el. lign. indsamlinger, og om vanskeligt tilgængelige webteknologier, der kræver særskilt opmærksomhed på teknologiske udfordringer eller websteder.

Netarkivet vil i denne strategiperiode analysere konsekvenserne af de eksisterende indsamlingsfrekvenser og høstningsstrategier med henblik på at kunne optimere indsamlingens dækningsgrad og økonomi.

Collection management (samlingshåndtering)

Kassation

Netarkivet vil gennem analyse og udvikling arbejde med at mindske indsamlingen af redundant materiale

Netarkivet vil løbende raffinere metoderne til at sikre, at kun relevant materiale indsamles.

Kvalitetssikring

Netarkivet udfører løbende en overvejende manuel kvalitetskontrol af de dele af materialet, som kommer ind via den selektive og begivenhedsorienterede høstning.

Det er målsætningen, at væsentlige dele af indsamlingen kan kvalitetssikres gennem automatiserede processer. Et første mål vil dog være at få løst opgaven i forhold til den selektive indsamling, da det er her vi forbruger manuelle ressourcer, som ønskes frigjort til andre processer, som ikke kan automatiseres.

I strategiperioden 2015-2017 vil vi evaluere de automatiserede metoder til kvalitetssikring, som anvendes ved andre webarkiver for at vurdere, hvilken metode der vil egne sig bedst i forhold til Netarkivets selektive indsamling.

I forbindelse med afprøvning af nye indsamlingsmetoder og -værktøjer udfører Netarkivet særskilt kvalitetskontrol som en del af evalueringen.

Dokumentation

Indsamling af materiale til Netarkivet dokumenteres gennem automatiserede processer og registreringer i indsamlingssystemerne.

Desuden foretages en manuel dokumentation af indsamlingen af udvalgte netsteder, særligt i forbindelse med *selektive høstninger* og *begivenhedshøstninger*.

Netarkivet vil opdatere sine dokumentationsværktøjer og -processer, således at de i højere grad afspejler brugernes og kuratorernes behov og letter arbejdsgange.

Internationalt samarbejde

Netarkivet vil indtage en central placering i *International Internet Preservation Consortium* (IIPC), både i styregrupperegi og i relevante undergrupper og udvalg.

Netarkivet vil være aktiv deltager i relevante indsamlings- og udviklingsprojekter for at fremme vidensdeling og erfaringsudveksling.

Litteratur

Netarkivet vil ajourføre litteraturlisten på <http://netarkivet.dk/om-netarkivet/publikationer/>