

DeiC Nationale Kulturarvscluster, Det Kgl. Bibliotek

## Kulturarvsclusteret

Forskere kan søge i og analysere data fra nationale samlinger af digitaliseret kulturarv på Det Kgl. Bibliotek.

DeiC Nationale Kulturarvscluster, Det Kgl. Bibliotek, er en supercomputer med adgang til Det Kgl. Biblioteks digitale kulturarvsmaterialer. Det omfatter blandt andet radio-tv-samlingen og den digitale avis-samling, som begge kan browses på Mediestream, samt Netarkivet. Forskere kan også anvende systemet til at arbejde med egne data.

Radio-/tv-samlingen indeholder mere end en million timers tv-udsendelser og mere end halvanden million timers radioudsendelser udsendt på danske kanaler fra 1980'erne til i dag. Samlingens data er tilgængelige som video- og audiofiler. Samlingen indeholder også store mængder metadata såsom programtitler, udsendelsestidspunkter og undertekster.

Netarkivet indeholder mere end 800 TB data svarende til mere end 25 milliarder objekter hentet fra den danske del af internettet fra 2005 til i dag. Arkivet indeholder både data og metadata. Netarkivet er et fælles nationalt projekt ved Det Kongelige Bibliotek.

Den digitale avissamling indeholder foreløbig 29 millioner avissider fra 1700-tallet indtil i dag. Når det nuværende digitaliseringsprojekt er afsluttet, vil der være 32 millioner sider i samlingen. Alle disse sider ligger som billedfiler sammen med en stor mængde metadata og optisk tegngenkendelsesdata (OCR), der gør det muligt at søge i samlingen.

Ud over disse store samlinger har Det Kgl. Bibliotek også andre mindre specialsamlinger.

Alt i alt står mere end 4 PB data til rådighed for forskningsprojekter.

Brugervenlig software gør det muligt at foretage søgninger og analyser ved hjælp af webbaserede adgange til værktøjer. Dog er det også muligt at benytte avancerede programmeringsmetoder såsom Spark og R til at foretage analyser.



### Specifikationer

- 10 noder med hver to Intel Xeon-CPU'er forbundet i et højhastighedsnetværk.
- I alt 360 CPU-kerner.

### Datalagring

- EMC Isilon med en kapacitet på 3 PB
- EMC VNX med plads til 200 TB.

### Software

- Open Data Platform med Spark, Hadoop, YARN, Hive, HBase, m.m.

### Partnere

Kulturarvsclusteret er etableret i et samarbejde mellem DeiC og Det Kgl. Bibliotek.

### Få adgang

Alle danske forskere kan mod betaling anvende DeiC Nationale Kulturarvscluster, Det Kgl. Bibliotek. Supercomputeren er dog primært tiltænkt forskere inden for humaniora og samfundsvidenskab.

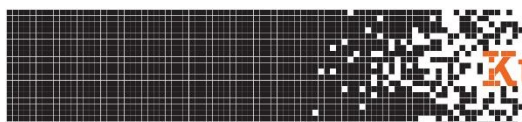
### Kontakt

Katrine Gasser, Det Kgl. Bibliotek,  
[khg@kb.dk](mailto:khg@kb.dk)

[mediestream.dk](http://mediestream.dk)

[netarkivet.dk](http://netarkivet.dk)

[www.deic.dk/kulturarvscluster](http://www.deic.dk/kulturarvscluster)



**Kulturarvscluster**