

Fagpanelet for e-Science

Kortlægningsrapport om nationale behov for forskningsinfrastrukturer inden for e-Science

12. oktober 2010

Fagpanelets medlemmer:

Formand: Institutleder John Renner Hansen, Niels Bohr Institutet, Københavns Universitet
Professor Henrik Madsen, Institut for Informatik og Matematisk Modellering, Danmarks Tekniske Universitet
Professor Brian Vinter, Datalogisk Institut, Københavns Universitet
Professor Ole Sigmund, Institut for Mekanisk Teknologi, Danmarks Tekniske Universitet
Lektor Gerth Stølting Brodal, Datalogisk Institut, Aarhus Universitet
Lektor Josva Kleist, Institut for datalogi, Aalborg Universitet
Lektor Daniel Merkle, Institut for Datalogi og Matematik, Syddansk Universitet.

eScience dækker alle former for brug af computerteknologier; dataindsamling, datatransport, datalagring, databehandling, datapræsentation (visualisering), simulering og syntese. eScience er det hurtigst voksende område inden for moderne videnskab. Udover at være en grunddisciplin inden for informations- og kommunikationsteknologi (IKT), er eScience for alle andre videnskaber et uundværlig supplement til de to klassiske metoder, eksperiment/observation og teori.

Inden for få år vil det ikke være muligt at gennemføre videnskabelige studier uden på en eller anden måde at inddrage computere til beregninger eller søgninger i databaser på tværs af fag- og landegrænser. Antallet af brugere vil blive betydeligt udvidet, og deres behov for ressourcer vil overstige det niveau, man i dag kan tilbyde gennem universiteternes lokale systemer og gennem det nationale Dansk Center for Scientific Computing. Med de mange nye forskere som traditionelt ikke har haft adgang til eScience-baserede værktøjer, bliver der et øget behov for brugerstøtte og uddannelse. Nye støttefunktioner må derfor oprettes med en central organisering, som skal sikre synergi mellem forskellige fagområder. Kendskabet til mulighederne i eScience skal styrkes.

Udover at være hurtigt voksende er eScience og den tilhørende eInfrastruktur dybt afhængig af udviklingen i verden omkring os. Det er essentielt, at den danske eInfrastruktur til stadighed er på et højt internationalt niveau, og at udviklingen sker i takt med den, der planlægges i de øvrige europæiske lande. Der er behov for en stram styring, hvis ressourcerne skal udnyttes bedst muligt i et felt, hvor ekspansion og udvikling sker sideløbende.

Opsummering og anbefalinger

Den nationale eInfrastruktur kan groft deles i tre områder;

- 1) et landsdækkende højhastighedsdatanetværk,
- 2) systemer til beregning
- 3) enheder til sikker opbevaring af data.

De tre dele er tæt forbundet og må hele tiden holdes indbyrdes i balance og på højde med de internationale standarder på området.

Det danske forskningsdatanetværk, som forbinder forskningsinstitutionerne, Forskningsnettet, har en standard, der i kvalitet er på højde med andre nationale netværk, og som udnytter de internationale forbindelser forbilledligt. Der er på kort sigt ikke behov for yderligere investeringer. På længere sigt skal Forskningsnettet tilføres tilstrækkelige ressourcer for altid at kunne udbyde netværksforbindelser med en service på højde med dem, som danske forskeres samarbejdspartnere og konkurrenter nyder godt af.

Hvad angår **High Performance Computing (HPC)**, dvs. computersystemer og sikker lagerplads til store data-samlinger foreslår panelet, at der etableres en infrastruktur med flere elementer, som kan ind- og udfases i takt med, at behovene udvikler sig og ressourcerne bliver tilgængelige. En bestyrelse bestående af fagligt kompetente medlemmer skal tilse, at midlerne fordeles efter hensigten - med kvalitet og konkurrencedygtighed som de primære parametre.

Forslaget består af følgende fire HPC relaterede elementer organiseret under en fælles bestyrelse.

En bestyrelse for danske HPC aktiviteter: Udviklingen sker med en hastighed, som kræver konstant overvågning og mulighed for at flytte ressourcer mellem forskellige elementer eller til en ny aktivitet, som udvikler sig lovende. For at sikre en fornuftig koordinering af de forskellige investeringer og en udvikling i takt med de internationale partnere anbefales det, at nye og eksisterende aktiviteter, som involverer regne- og

lagerenheder, samles under en fælles bestyrelse. Bestyrelsen sammensættes af højt estimerede videnskabsfolk med eScience-erfaring, indstillet efter samme principper som medlemmerne af Det Frie Forskningsråds faglige råd.

En National Aktivitet for Videnbaseret Udnyttelse af eInfrastrukturer (Ny eInfrastruktur)

Der etableres en gruppe af eScience-eksperter med indgående kendskab til de ressourcer, som de danske og internationale eInfrastrukturer tilbyder. Gruppen skal opsamle, formidle og danne en national platform for viden om metoder brugt inden for videnskabelige beregninger og datalagring. Det er en central opgave at indsamle, dokumentere og på anden måde formidle viden om den internationale udvikling og dermed sikre, at danske forskere kan udnytte tilgængelige internationale eInfrastrukturer, fx det europæiske PRACE-projekt, som giver adgang til nogle af de største HPC-anlæg i verden. En væsentlig opgave for gruppen bliver at evaluere behovet for et større centralt nationalt regnecenter på linje med dem, som allerede findes i andre europæiske lande. Det foreslås, at gruppen etableres over en toårig periode med hhv. 2 MDKK og 5 MDKK fra infrastrukturpuljen.

Dansk Center for Scientific Computing – DCSC (Eksisterende eInfrastruktur)

DCSC uddeler midler efter forskningsrådslignende principper til danske forskere med henblik på indkøb af HPC udstyr. Derudover varetager DCSC kontakten til internationale HPC eInfrastrukturer. Denne opgave bør udvides med et partnerskab i PRACE projektet. Det foreslås, at DCSC's nuværende budget på 16 MDKK årligt fra finansloven suppleres med 14 MDKK årligt fra Infrastrukturpuljen for at opretholde det niveau, som blev opnået ved tilskud fra den første infrastrukturpulje. Dertil kommer budgettet for de internationale aktiviteter på ca. 9 MDKK årligt.

DCSC's aktiviteter videreføres under en ny bredere bestyrelse.

Sikker Lagring af Videnskabelige Data (Ny eInfrastruktur)

IT-baserede kollaborative projekter indenfor alle videnskaber er karakteriseret ved behovet for at dele data. Lagring af data handler ikke kun om at gemme data, men også om effektivt og sikkert at håndtere forskellige datatyper fra forskellige kilder, som lagres, behandles og genkaldes på forskellige måder. Der er ofte tale om følsomme data, som kræver særlig sikkerhed. Det foreslås at etablere en national datalagringsenhed med en stab på tre fuldtidsansatte eksperter til at varetage brugerstøtte og udvikling. Som udgangspunkt anses det tilstrækkeligt at opbygge en facilitet med 1 PByte, som derefter udvides efter behov. Samlet udgift til etablering for de første to år er 5 MDKK. Enheden har en klar tilknytning til aktiviteterne i eInfrastruktur ekspertgruppen.

Cloud Computing (Ny eInfrastruktur)

Området, hvor cloud computing passer ind som forskningsværktøj, dækker adgang til mellemstore computere, der i funktionalitet ikke adskiller sig markant fra en almindelig PC. Brugere har store datasæt og har behov for store beregninger, men er ikke traditionelle brugere af supercomputere. Cloud computing giver en mulighed for på en simpel måde at mangedoble beregningskapaciteten. Det foreslås at bruge 5 MDKK årligt på cloud-installationer fordelt på flere universiteter. Enheden har en klar tilknytning til aktiviteterne i eInfrastruktur ekspertgruppen. Da der er flere initiativer inden for dette område, bl.a. fra en arbejdsgruppe under VTU, er det overordentlig vigtigt med en national koordinering.

eScience, eInfrastruktur og panelrapportens afgrænsning

Indsamling og behandling af videnskabelig information på digital form samt numerisk simulering og syntese, eScience, har de sidste tre årtier udviklet sig fra at være et værktøj for de få, oftest inden for de tekniske og naturvidenskabelige fag, til at være det tredje grundelement i den videnskabelige metodesamling - et supplement til teori og eksperiment/observation. eScience dækker alle former for brug af computerteknologier; dataindsamling, datatransport, datalagring, databehandling, datapræsentation (visualisering), simulering og syntese. Netværk, computere og datalagre udgør tilsammen eInfrastrukturen, som gør det muligt at benytte værktøjer fra eScience.

I det efterfølgende skelnes mellem den eInfrastruktur, som universiteter og andre forskningsinstitutioner naturligt stiller til rådighed for medarbejderne i form af lokale netværk, arbejdsstationer og mindre klynger af computere og så den nationale eInfrastruktur. Sidstnævnte forbinder forskningsinstitutionernes lokale ressourcer gennem et højhastighedsnetværk på tværs af landet og stiller store regne- og lagerfaciliteter til rådighed for forskning og uddannelse.

I sin analyse af de fremtidige behov har panelet brugt de indkomne interesselikendegivelser og sammenholdt disse med erfaringer høstet gennem brug af den eksisterende eInfrastruktur. De fremsendte infrastrukturforslag kan groft opdeles i fire typer:

- a) Infrastrukturer, som tilbyder at dække andre infrastrukturens eScience-aktiviteter og andre videnskabelige projekter med behov for eInfrastruktur
- b) Infrastrukturer, som har behov for adgang til eInfrastrukturen, men hvis primære formål retter sig mod at støtte videnskabelige projekter
- c) Infrastrukturer, som har et større behov for en eInfrastruktur, men som i deres udformning er så store, at de udgør en selvstændig eInfrastruktur
- d) Aktiviteter omkring udvikling af komponenter til ny eInfrastruktur

Forslaget til en fremtidig national eInfrastruktur tager udgangspunkt i a). Forslagene under b) og c) betragtes som brugere af eInfrastruktur. Punkt c) dækker blandt andet store databaser eller samlinger af national relevans. Det er panelets opfattelse, at disse skal kunne rummes i den nationale eInfrastruktur, men at opbygningen skal ske som selvstændige projekter.

Forslag under d) er meget vigtige for den fremtidige udvikling af infrastrukturer, men hører ikke til som en del af den nationale eInfrastruktur.

Brugerne

De danske brugere af den nationale forsknings-eInfrastruktur kan løst deles op i to klasser:

- i. Brugere med behov for stor regnekraft og måske adgang til store datasæt, men uden specielle krav til de computere, de regner på. Det kan være en Linux- eller Windows-baseret PC, hvis bare den er stor nok.
- ii. Brugere med behov for stor regnekraft, hvor det samtidigt er nødvendigt at konfigurere regneenheder på en særlig måde. Der kan være tale om specielle styresystemer, behov for mange computere, som intensivt kommunikerer resultater til hinanden under afviklingen af programmerne, eller regneenheder som kræver hurtig adgang til massive datamængder enten i nærheden af computeren eller fordelt på flere geografisk afskilte lagerenheder.

Traditionelt er eScience-metoder blevet brugt indenfor de teknisk- og naturvidenskabelige fag, men humaniora, sundhed og samfundsvidenskaberne har gennem opbygning af store databaser behov for at benytte de samme metoder. Der er fx tale om digitalisering af de danske museumssamlinger og biblioteker eller om at gøre større registre tilgængelige for komparativ forskning, herunder internationale projekter, hvor data skal udveksles på tværs af landegrænser. Der skal derfor skabes sikre, vedvarende datalagre, som garanterer adgang for autoriserede personer - og kun disse, da der ofte er tale om personfølsomme eller andre sikkerhedsfølsomme data.

De brugere, som denne rapport hovedsageligt adresserer, er videnskabsfolk og studerende med behov for regnekraft, der går ud over nogle få PC'er, eller med datasæt, som i omfang overstiger nogle få TByte. Det er dog meget vigtigt at medtage de øvrige brugere i den fremtidige planlægning, da mange med tiden kommer til at udnytte de større nationale ressourcer. Overgangen fra at være lokal til national bruger er i dag ikke velorganiseret, og en national strategi for eScience må derfor indeholde en plan for brugersupport.

Samtlige projekter beskrevet i ESFRI's roadmap for en europæisk forskningsinfrastruktur er dybt afhængige af en stærk eInfrastruktur. Danske forskere er allerede stærkt engageret i flere projekter. Skal Danmark deltage på lige fod med de øvrige EU-lande, må eInfrastrukturen og brugernes kompetencer leve op til internationale standarder. Dette betyder en betydelig investering i selve infrastrukturen og et øget fokus på uddannelse i brug af ressourcerne.

Den nuværende nationale eInfrastruktur

Den nationale eInfrastruktur har siden 2001 været baseret på to organisationer, Forskningsnettet, som varetager netværksforbindelserne mellem forskningsinstitutionerne og forbindelserne til udlandet og Dansk Center for Scientific Computing, som forsyner forskerne med regnekraft og store datalagre.

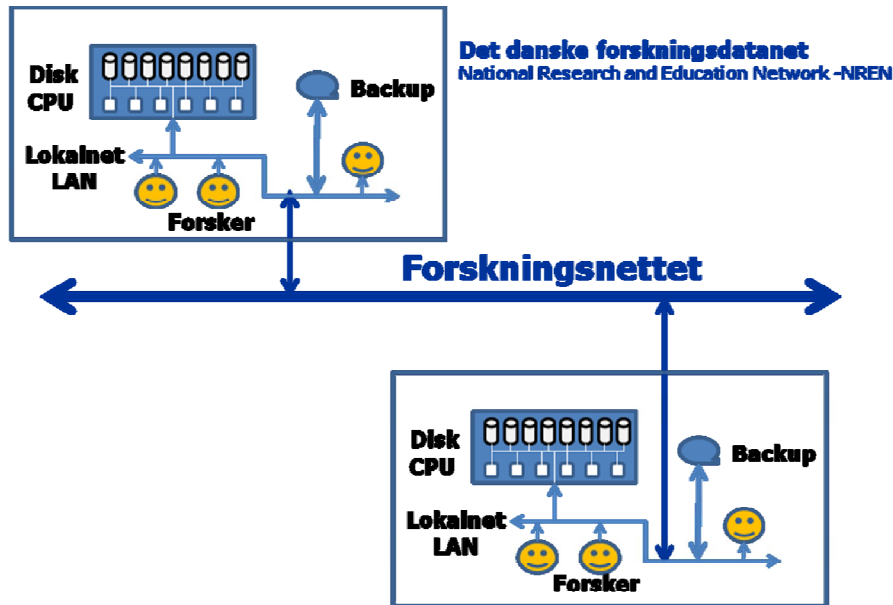
Forskningsnettet

Datanetværk muliggør deling af ressourcer og data og er derfor rygraden af en eInfrastruktur. Forskningsnettet er det højhastighedsnetværk, der forbinder de danske universiteter og forskningsinstitutioner, figur 1. Nettet har eksisteret siden 1987 og har over 170.000 brugere fordelt på 100 institutioner. Forskningsnettet er placeret under Ministeriet for Videnskab, Teknologi og Udvikling, VTU. Organisationen har egen styregruppe sammensat af højt estimerede forskere og forskningsledere og har sit eget sekretariat.

Efter en bevilling fra den første infrastrukturpulje på 40 mio. kr. opgraderede Forskningsnettet i 2008 sit netværk og leverer nu en service på højde med de bedste nationale netværk i verden. Forskningsnettet leverer derudover udlandsforbindelser via det nordiske NORDUnet og det europæiske net GÉANT.

Det er panelets helt klare opfattelse, at Forskningsnettet leverer dataforbindelser af meget høj kvalitet. Panelet skønner, at der ikke på kort sigt er behov for at opgradere netværkerne yderligere.

På mellemlangt sigt skal udviklingen følges og nettet opgraderes, når behovene melder sig.



Figur 1. Danske forskere har adgang til computere og datalagre gennem lokale datanetværk. På tværs af landet forbinder **Forskningsnettet** forskerne med hinanden og med ressourcer uden for den lokale forsknings-eInfrastruktur.

Dansk Center for Scientific Computing – DCSC

Dansk Center for Scientific Computing er en national forskningsinfrastruktur under Ministeriet for Videnskab, Teknologi og Udvikling, med det formål at levere High Performance Computing, HPC, til stærke danske forskningsmiljøer. DCSC varetager også de danske HPC-interesser internationalt, både på nordisk plan i Nordic Data Grid Facility (NDGF) og i en række EU-projekter.

DCSC ledes af en bestyrelse sammensat af højt estimerede forskere og forskningsledere, en fra hvert af de pt. fem værtsuniversiteter og fem udpeget af forskningsråd og VTU, heriblandt to fra erhvervslivet.

DCSC's strategi er baseret på decentrale driftscentre ved de danske universiteter, pt. AU, AAU, DTU, KU og SDU. Bestyrelsen uddeler hvert år, efter foregående opslag og ansøgning fra forskere, midler til indkøb af udstyr, som i samarbejde med et af de fem driftscentre indkøbes og drives efter bevillingshaverens anvisninger. Bevillingerne uddeles på basis af projekternes videnskabelige højde, relevans og ansøgernes CV.

DCSC drives som et konsortium bestående af værtsuniversiteter og VTU. Siden starten i 2001 har aktiviteterne været finansieret med et tilskud på finansloven på ca. 16 MDKK årligt. Universiteterne har gennem tilskud til drift, dvs. lokaler, strøm, køling og lønninger til systemadministratorer, medfinansieret i et omfang, som typisk har været af samme størrelse som bevillingerne fra DCSC. Udgifterne til strøm og køling er de sidste år steget betydeligt. Flere universiteter har givet udtryk for, at dette på sigt ikke kan fortsætte. Løsningen kunne findes i en øget fokus på Grøn-IT samt udflytning af visse funktioner til steder i verden hvor energipriserne er lavere og køling en naturlig del af miljøet. Udgifterne til denne omstilling kunne indgå som en del af den centrale finansiering af den nationale eInfrastruktur.

I forbindelse med den første uddeling af midler til forskningsinfrastruktur i 2007, modtog DCSC en bevilling på 37 MDKK, hvoraf 7 MDKK var afsat til betaling af det danske bidrag til NDGF i 2009 og 2010. Særbevillingen har muliggjort en næsten fordobling af de udelte midler. Den gennemsnitlige bevillingsrate fra DCSC var i perioden 2008-2010 på ca. 35 % af de ansøgte midler. DCSC's bestyrelse har ret til at tildele 10 % af enhver bevilling til andre brugere, fx forskere med akutte behov eller nye brugere.

DCSC har også til opgave at styrke dansk deltagelse i udviklingen af internationale eInfrastrukturer og sikre danske forskere adgang til udenlandske faciliteter. Dette sker dels gennem det nordiske samarbejde NDGF og dels gennem direkte deltagelse i projekter under EU's 7. rammeprogram. Det danske bidrag til NDGF og medlemskaber af de europæiske projekter udgør årligt 3.5 MDKK, som betales ud af DCSC's ramme.

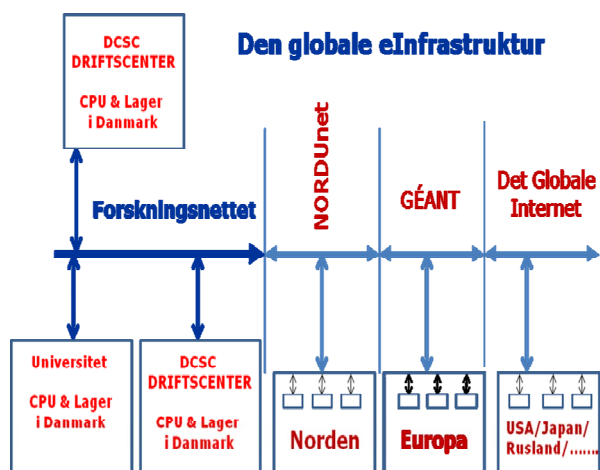
Den internationale eInfrastruktur

Dansk forskning er dybt afhængig af internationale forskningsinfrastrukturer, herunder eInfrastrukturer, figur 2. Ligesom den danske eInfrastruktur har to dele, er den internationale bygget op om netværksorganisationerne, NORDUnet og GÉANT og beregnings- og lagerprojektet NDGF.

NORDUnet er et datakommunikationsnetværk, der forbinder de nationale forsknings- og uddannelsesnet i de nordiske lande. NORDUnet forbinder omkring en million brugere og 600 forskningsinstitutioner i Norden. Det er via NORDUnet, at de nordiske nationale forskningsnetværk (NREN's) har adgang til andre forskningsnetværk over hele verden.

NORDUnet er organiseret som et aktieselskab, der ejes af de nordiske undervisnings- og forskningsministerier. Det danske medlemskab af NORDUnet varetages af Forskningsnettet.

NORDUnet varetager det nordiske medlemskab af det europæiske datanetværk for forskning og uddannelse, GÉANT. Sammen med de nationale forsknings- og uddannelsesnetværk forbinder GÉANT 40 millioner brugere på over 8.000 institutioner i 40 lande. GÉANT er i fællesskab finansieret af EU's 7. rammeprogram og de nationale partnere.



Figur 2. Forskningsnettet forbinder danske forskere til kolleger i udlandet og til internationale ressourcer gennem det nordiske NORDUnet og det europæiske GÉANT. Derfra er der forbindelse til alle andre forskningsnetværker i verden.

Nordic DataGrid Facility - NDGF

Den nordiske samarbejdsorganisation NDGF sikrer forskerne fra de nordiske lande adgang til internationale forskningsprojekter med behov for computer- og lagerressourcer i et omfang, som ikke er tilgængelig gennem de nationale eInfrastrukturer alene. NDGF er en produktionsfacilitet, som knytter eksisterende nationale IT-ressourcer sammen i en grid infrastruktur. NDGF har de sidste fire år serviceret en række projekter inden for fysik, kemi og bio-informatik. Den største operationelle bruger er de nordiske partikelfysikere, som deltager i CERN-eksperimenterne ved Large Hadron Collider.

Det årlige danske bidrag til NDGF har siden starten i 2005 været 3.5 MDKK. De første år blev kontingentet betalt af Forskningsrådet for Natur og Univers. 2009 og 2010 bidraget blev hentet fra den bevilling på 37 MDKK, som DCSC modtog fra den første infrastruktur uddeling i 2007.

I løbet af foråret 2010 har NDGF gennemgået en international evaluering, idet det nuværende projekt stopper den 31. december 2010. Den efterfølgende rapport anbefaler en fortsættelse af samarbejdet med visse organisatoriske ændringer. Panelet anbefaler derfor, at Danmark fortsætter sit medlemskab på sammen niveau som tidligere, og at DCSC eller dens efterfølger fortsat administrerer de danske ressourcer til projektet.

Databehandlings- og softwarecentret for den europæiske neutronkilde - ESS

Af helt speciel interesse for Danmark er etableringen af det nye europæiske center for neutronspredning i Lund, kaldet European Spallation Source – ESS. Som en del af aftalen om at vælge Lund som hovedsæde for ESS indgår, at Data Management og Software Centret, DMSC, placeres på Nørre Fælled i København - tæt integreret med den nye Nørre Campus ved Københavns Universitet.

ESS/DMSC skal fra 2010 servicere alle videnskabelige og tekniske aktiviteter i forbindelse med behov for regne- og lagerkapacitet. Det er planen, at der langsomt opbygges en stab bestående af IT- og simuleringseksperter på ca. 70 medarbejdere. Indtil 2013 skal centret støtte processen hen mod et endeligt design af ESS, herunder også udviklingen af planer for ESS/DMSC. Efter 2012 skal centret udvikles til et produktionscenter, som kan varetage alle eScience funktioner for ESS.

For dansk forskning vil tilstedeværelse af ESS/DMSC have enorm betydning. En stab på ca. 70 personer vil dække alle aspekter relateret til dataindsamling, databehandling, simulering og visualisering. ESS/DMSC vil uden tvivl virke som en dynamo for udviklingen af eScience i Danmark og Sydsverige. Kombinationen af ESS/DMSC og den viden som allerede er opbygget ved de danske universiteter, gør det meget attraktivt at tilbyde datahåndteringsfaciliteter til andre store projekter. Der har således allerede været kontakter i retning af, at centret kunne have samme rolle for synkrotronstrålingscentret MAX-IV i Lund og XFEL i Hamborg, som det får for ESS. Der har også været kontakt fra danske astronomer, som overvejer at bringe et databehandlingscenter i spil i forbindelse med deltagelse i et kommende NASA satellitprojekt. Sådanne centre kan have stor glæde af de erfaringer med internationale samarbejder, som samles i umiddelbar nærhed af ESS/DMSC og hele platformen vil kunne bruges som drivkraft til at trække andre store IT-projekter til Danmark.

For at få det fulde udbytte af den danske investering er det essentielt at universiteterne udbyder uddannelser, som rummer emner inden for eScience i bred forstand. Kandidater med speciale i billedbehandling, dataindsamling, sikkerhed og datalagring vil være i høj kurs.

Mulige nye veje i den fremtidige eInfrastruktur til forskning

Den nuværende nationale eInfrastruktur med Forskningsnettet og DCSC som de centrale danske aktører og med NORDUnet og NDGF som de umiddelbare internationale partnere bør udbygges med bedre brugerstøtte, state-of-the-art lagerfaciliteter og udvidet adgang til regnekraft. I det følgende beskrives en række mulige veje til at fremme dette. Formålet er at sikre fortsat konkurrencedygtighed på den internationale forskningsscene.

En National Aktivitet for Videnbaseret Udnyttelse af eInfrastrukturer

Ressourcer til eScience bliver med tiden mere kostbare og samtidigt mere komplekse. For at udnytte eInfrastrukturen bedst muligt, bør der oprettes en national enhed bestående af eScience-eksperter med indgående kendskab til de ressourcer, som de danske og internationale eInfrastrukturer tilbyder. Ekspertgruppen skal opsamle, formidle og danne en national platform for viden om videnskabelige beregninger og datalagring. Det er en central opgave at indsamle, dokumentere og på anden måde formidle viden om den internationale udvikling og dermed sikre, at danske forskere kan udnytte tilgængelige internationale eInfrastrukturer.

Effektivitet er ikke alene et spørgsmål om optimering af de programmer, som afvikles på en given computer, men om alle de dele, som indgår i et eScience-projekt - god udnyttelse af hardware, priser af hardware og software, tilgængelighed, adgang til den nyeste teknologi, energiforbrug og kendskab til de enkelte forskeres arbejdsmetoder. Det er ekspertgruppens opgave at hjælpe såvel nye som erfarne brugere med at få mest ud af investeringerne.

Betydningsfulde opgaver for ekspertgruppen er dels at deltage i uddannelsesaktiviteter på eScience-området, dvs. sommerskoler, kandidatkurser og ph.d. skoler, dels at arrangere møder og workshops indenfor forskellige HPC-områder med henblik på at opnå lokale og nationale synergieffekter. Gruppen indkalder flere gange årligt potentielle brugere til informationsmøder, hvor mulighederne for det kommende år præsenteres. En anden væsentlig opgave for gruppen bliver at evaluere behovet for et større centralt, nationalt regnecenter på linje med dem, som allerede findes i mange europæiske lande.

For at styrke gruppens gennemslagskraft og effektivitet foreslås det, at lederen overtager administrationen af de frie midler, som DCSC's bestyrelse i dag bruger til mindre projekter, begyndere og akutte behov. De frie midler udgør i dag 10 % af bevillingerne fra DCSC. Det foreslås, at kvoten hæves til 15 %. Gruppen holder løbende regnskab med, hvordan ressourcerne bruges.

Gruppen bør bestå af 8-10 eksperter placeret i en central enhed og 2-3 eksperter knyttet til hver af de lokale eScience driftscentre. Med en gradvis opbygning til 18 årsværk vil omkostningen for de første to år være 16 MDKK, hvoraf ca. halvdelen foreslås finansieret fra infrastrukturpuljen.

Sikker Lagring af Videnskabelige Data

IT-baserede kollaborative projekter indenfor alle videnskaber er karakteriseret ved behovet for at dele data og sikre, at disse også er tilgængelige på lang sigt. Lagring af data kræver viden om effektiv og sikker håndtering af mange forskellige datatyper fra forskellige kilder. Introduktionen af internationale standarder for datahåndtering bliver derfor en fremtrædende opgave for enheden.

En særlig udfordring bliver at assistere projekter, hvis data betragtes som følsomme og hvor adgang er underlagt særlige restriktioner. Som eksempler kan gives nationale registre med personfølsomme oplysninger, hvor svar på forespørgsler ikke må være personificerende eller data, der af konkurrencehensyn kun skal være tilgængelige for en mindre gruppe af personer.

Det foreslås at etablere en national datalagringsenhed, med en stab på 3 fuldtidsansatte eksperter til at varetage brugerstøtte og udvikling. Som udgangspunkt anses det tilstrækkeligt at opbygge en facilitet med 1 PByte, som derefter udvides efter behov. Samlet udgift til etablering for de første to år skønnes til at være 5 MDKK. Enheden har en klar tilknytning til aktiviteterne i den nationale ekspertgruppe for videnbaseret udnyttelse af eInfrastrukturer.

En række tilsvarende enheder findes eller er ved at blive opbygget i de øvrige nordiske lande. Det vil være naturligt, at disse med tiden kommer til at indgå i en større enhed, som udnytter den fælles viden til effektiv og sikker datahåndtering.

Cloud Computing

Området, hvor cloud computing passer ind som forskningsværktøj, er groft beskrevet som adgang til mellemstore computere, der ikke adskiller sig markant fra en almindelig PC. Ressourcerne i en sådan cloud kan med rimelighed beskrives som PC-agtige computere, der er 50-100 gange hurtigere end en almindelig PC.

Brugerne er forskere fra alle områder, der langsomt er ved at anvende større beregninger i deres videnskab. De adskiller sig fra deres kolleger ved at have større datasæt og/eller behov for flere beregninger. Omvendt adskiller de sig fra de etablerede DCSC-brugere ved typisk at have meget svært ved at anvende en klassisk supercomputer, som ikke tilbyder samme interaktive værktøjer som brugerne kender fra deres egen PC. Det skal bemærkes, at cloud-initiativet ikke kan stå alene og nødvendigvis må parres med lagerfaciliteter.

Computerne i det foreslåede cloud-system skal altid kun tilgås af én bruger ad gangen og maskinerne skal nulstilles mellem to brugere. Nul-stilling er nødvendigt for at kunne anvende sensitive data, der fx er omfattet af persondataloven eller på anden måde kræves isoleret fra andre brugere.

Det foreslås at bruge 10 MDKK på cloud-installationer fordelt på flere universiteter. På sigt anses det for sandsynligt, at denne form for beregningsfacilitet bliver så billig, at den enten bliver kommercialiseret eller bliver overtaget af universiteternes egne driftsorganisationer.

En større National High Performance Computing Facilitet

På linje med de øvrige europæiske lande kan man med tiden etablere et større nationalt regnecenter til supplement af de regionale DCSC-understøttede centre. Centret skal råde over ressourcer i form af hardware og personale i et omfang, som er sammenligneligt med andre tilsvarende centre rundt om i verden. En central enhed skal kunne afvikle projekter af en størrelse, som ikke er mulig på de regionale centre. Centret kan samtidig indgå som et fremtidigt dansk bidrag til det europæiske samarbejdsprojekt PRACE, hvor pt. 14 lande tilbyder forskere adgang til nogle af de største regnemaskiner i verden.

Panelet finder, at et sådant center med tiden kan spille en vigtig rolle i den danske eInfrastruktur, men at behovet pt. ikke er tilstrækkeligt dokumenteret til, at en egentlig opbygning kan anbefales.

Panelet erkender dog, at der indenfor ingeniørvidenskaberne kan være særlige behov for en stor installation. Erfaringer med internationale partnerskaber og brugernes behov for en så kostbar installation må undersøges og danne grundlag for en fremtidig beslutning. Det foreslås, at den Nationale Aktivitet for Videnbaseret Udnyttelse af eInfrastrukturer inden for en treårig periode tilvejebringer det nødvendige grundlag for en beslutning om oprettelse af en centralt placeret større HPC-facilitet.

Deltagelse i projekter fra Nordic eScience Action plan

NordForsk under Nordisk Ministerråd har udarbejdet en plan for udviklingen af eScience i de nordiske lande. Den såkaldte 'Action Plan' indeholder ti anbefalinger inden for hovedområderne:

- A) Uddannelse i brug af eScience metoder
- B) Udvikling af metoder
- C) Ressourcedeling

Som en opfølgning på planen har en arbejdsgruppe under NordForsk (eNoria) foreslået tre indsatsområder, som umiddelbart bør støttes fra fælles nordisk side. Én anbefaling omhandler videreførelsen af NDGF. Som tidligere omtalt er det panelets opfattelse, at NDGF spiller en central rolle i udviklingen af den nordiske eInfrastruktur og anbefaler derfor, at NDGF fortsat vil være en del af de internationale eInfrastruktur aktiviteter som Danmark deltager i.

De to øvrige anbefalinger fra eNoria omhandler projekter inden for områderne:

'Research in response to grand challenges on climate change, with a focus on climate modeling' og 'Research in response to grand challenges related to health and social preconditions for health'.

Deltagelse i disse projekter må afvente anbefalinger fra de relevante fagområder.

The Partnership for Advanced Computing in Europe – PRACE

PRACE er en paneuropæisk HPC-tjeneste bestående af op til seks af verdens førende computersystemer, dybt integreret i den europæiske eInfrastruktur. Hvert system vil på 2-5 års sigt tilbyde computerkraft i petaflop/sek klassen. På længere sigt (2019) er målet exaflop /sek. Brugerne vil blive støttet af eksperter i forbindelse med installation, opskalering og optimering af programmer til nye computer-arkitekturer. PRACE tilbyder undervisning af forskere og studerende i den bedste udnyttelse af systemerne.

En videnskabelig styringskomite vil rådgive PRACE sideløbende med en peer review proces, hvorigennem adgang til centrale ressourcer vil blive ydet på baggrund af videnskabelige meritter.

PRACE's første produktionssystem, én petaflop/sek IBM Blue Gene, er installeret på Forschungszentrum Jülich, Tyskland. Indkaldelse af forslag om adgang til ressourcer på dette første PRACE-system blev åbnet den 10. maj 2010. Forskere fra alle EU-lande kan i denne omgang søge om regnetid på systemet. Senere vil ansøgninger blive begrænset til forskere fra partnerlande.

Panelet anbefaler, at Danmark indtræder som partner i PRACE for derigennem at få adgang til de allerstørste regneenheder i verden. Den årlige medlemsgebyr er på 0,5 MDKK og brug af ressourcer afregnes efter forbrug. Det anbefales derfor tillige, at der afsættes 5 MDKK årligt fra infrastrukturpuljen til at købe regnetid på PRACE. Hvis det viser sig hensigtsmæssigt, bør det også være muligt at købe tid på andre internationale faciliteter, fx udenfor Europa. Beløbet uddeles af bestyrelsen efter samme kriterier som de øvrige bevillinger.

Den nationale indsats på kort og mellemlang sigt

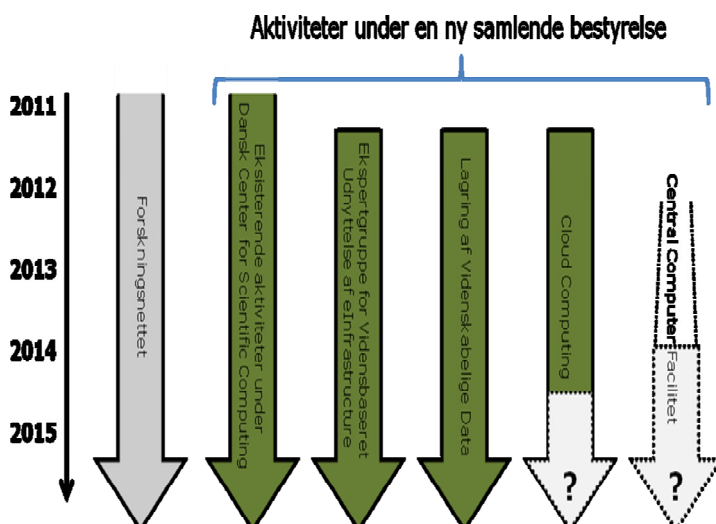
For at sikre en fornuftig koordinering af de forskellige investeringer og en udvikling i takt med de internationale partnere anbefales det, at nye og eksisterende aktiviteter samles under en fælles bestyrelse. Panelet finder dog, at Forskningsnettet, som er en højteknologisk service- og produktionsfacilitet med selvstændig teknologiudvikling, men uden projektansøgninger og direkte brugerdeltagelse, bevares som en selvstændig enhed med egen bestyrelse, figur 3.

Bestyrelsen sammensættes af højt estimerede videnskabsfolk med eScience erfaring, indstillet efter samme principper som medlemmerne af Det Frie Forskningsråds faglige råd. Bestyrelsen overtager det nuværende DCSC sekretariat.

Opgaveporteføljen dækker:

- **Uddeling af midler** efter ansøgning til eInfrastruktur understøttelse af forskningsprojekter, dvs. nyt hardware til de lokale driftscentre pt. under DCSC, cloud computing, plads på det nationale datalager og adgang til internationale faciliteter herunder PRACE.
- **Sikring af professionel brugerstøtte** såvel for nye som for mere erfarne brugere.
- **Organisering af adgang** til eInfrastrukturen for forskere med myndighedsbetjeningsopgaver, som ikke kan konkurrere om ressourcerne på lige vilkår med ordinære forskningsprojekter.
- **Løbende at følge udviklingen** og anbefale nye tiltag til styrkelse af eInfrastrukturen, herunder at tilvejebringe beslutningsgrundlaget for evt. at oprette et større centralt regnecenter.
- **Ansvar for udviklingen** af den danske deltagelse i internationale projekter på eInfrastruktur-området, bortset fra netværk. Herunder også uddannelsesprogrammer indenfor eScience.
- **Koordinering** med Forskningsnettet på relevante områder.
- **At balancere aktiviteterne** ved at flytte op til 10 % af den samlede bevilling mellem de enkelte indsatsområder. Øgede behov eller støtte til nye lovende aktiviteter vil kunne imødekommes gennem en overordnet prioritering og dermed ikke indenfor de snævrere rammer for et enkelt eInfrastruktur element.

Den nuværende bestyrelse for DCSC afslutter sit virke, når der oprettes en ny samlende bestyrelse på HPC området.



Figur 3

Forskningsnettet videreføres som en selvstændig organisation. DCSC's aktiviteter videreføres og organiseres sammen med nye virkemidler, som inkluderer en ekspertgruppe for viden om brug og udvikling af eInfrastrukturer, aktiviteter som håndterer lagring af videnskabelige data, cloud computing og videreudvikling af en større central computerinstallation.

Konkrete indsatsområder

Som udgangspunkt anbefales det, at Forskningsnettet videreføres på det nuværende høje serviceniveau og tilføres nye ressourcer i takt med udviklingen af de internationale netværker. Adgang til de internationale linjer er på fremragende vis håndteret gennem NORDUnet og GÉANT, og det anbefales at fortsætte samarbejdet på uændret niveau.

Panelet anbefaler derudover at:

- **Aktiviteterne under det nuværende Dansk Center for Scientific Computing** videreføres med et årligt budget til indkøb af nyt udstyr på 30 MDKK, hvoraf 16 MDKK kommer fra den faste bevilling på finansloven. Driften varetages fortsat af værtsinstitutionerne.
- **NDGF** og de relaterede EU-FP7 projekter fortsættes med en årlig bevilling på 3.5 MDKK.
- **Danmark indtræder som partner i PRACE.** Det årlige medlemsbidrag anslås til 0,5 MDKK. Derudover bør der afsættes ca. 5 MDKK til køb af ressourcer fra PRACE
- **Der oprettes en national datalagringsenhed,** omkostning for de første to år 5 MDKK.
- **Udviklingen af cloud computing** stimuleres gennem oprettelsen af et distribueret center for cloud computing med betydelige ressourcer, 10 MDKK. Centret vil få stor betydning for introduktionen af nye brugere til eScience. Aktiviteten evalueres efter to år og det besluttet om den videreføres som en national aktivitet.
- **Der etableres en Nationalt ekspertgruppe for Videnbaseret Udnyttelse af eInfrastrukturer.** Udgiften på ca. 16 MDKK for de første to år dækkes med 8 MDKK fra infrastrukturmidlerne resten fra universiteterne.
- **Det undersøges,** om der på mellemlangt sigt er behov for en større central computerinstallation, der evt. kan indgå som et dansk bidrag til PRACE.

Afsluttende kommentar

Nationale og internationale infrastrukturer for forskning og innovation vil i fremtiden blive en central bestanddel af det videnskabende samfund. Det er derfor vigtigt, at der oprettes et nationalt forum, hvor forskningsinfrastrukturer er i centrum. Et organ hvor det politiske system fremadrettet kan søge råd om investeringer og de centrale aktører, forskerne, kan påvirke udviklingen af næste generations forskningsinfrastrukturer. Panelet støtter derfor tanken om at oprette et forskningsråd for forskningsinfrastrukturer. Med selvstændige midler til udvikling, opbygning og udnyttelse af infrastrukturer vil et sådan råd sikre en stabil og gennemskuelig evolution af dette meget komplekse felt. De to eInfrastrukturbestyrelser for hhv. netværk og HPC passer perfekt ind som underudvalg til et forskningsråd for forskningsinfrastrukturer.