

e-infrastruktur til forskning i verdensklasse i 2025

- en analyse og udredning vedrørende dansk e-infrastruktur samarbejde

Bilag 1:

Kommissorium for analyse og udredning vedr. dansk e-infrastruktur samarbejde



Modtager(e):

Bestyrelsesformand Børge Obel, og Steen Pedersen, CEO, Danish e-Infrastructure Cooperation

Kopi:

Kommissorium for analyse og udredning vedr. dansk e-infrastruktur samarbejde

Indledning

Danske Universiteters Rektorkollegium (RK) og Styrelsen for Forskning og Uddannelse (SFU) har besluttet at igangsætte en strategiproces med det formål at udarbejde en fælles strategi for det nationale e-infrastruktur-samarbejde for perioden 2019-2029

Formål

Udgangspunktet for strategiarbejdet er følgende udkast for vision for dansk e-infrastruktur i 2025:

”Forskere ved de danske universiteter skal have adgang til en e-infrastruktur, der muliggør forskning og uddannelse i verdensklasse.”

De grundlæggende principper for den nationale strategi forventes at være følgende:

- At forskere inden for alle hovedområder har adgang til e-infrastruktur, og at adgangen til national e-infrastruktur sker på transparente vilkår.
- At der etableres en klar og tydelig rolle- og arbejdsfordeling mellem aktiviteter på det nationale niveau og på institutionsniveau.
- At investeringer i kostbar e-infrastruktur til forskning skal koordineres samt udnyttes og drives mest effektivt og optimalt.
- At der er stabilitet omkring økonomi for e-infrastruktur til at sikre både kontinuerlig modernisering af teknologi og kompetenceudvikling.

Som led i strategiprocesen gennemføres et analyse- og udredningsarbejde med det formål at tilvejebringe et videngrundlag, der kan danne baggrund for drøftelserne omkring det fremtidige behov.

RK og SFU er enige om at anmode Danish e-Infrastructure Cooperation (DeIC) om at gennemføre udredningsarbejdet. På den baggrund skitseres opgaven for DeIC's udrednings- og analysearbejde i forbindelse med strategiprocesen.

RK og SFU skal desuden gøre opmærksom på, at det i udgangspunktet er forventningen, at strategien kan udfoldes inden for de kendte økonomiske rammer.

Opgaven

Analyse- og udredningsarbejdet skal bestå af tre dele: En behovskortlægning, en benchmarking af Danmark i forhold til andre lande samt en analyse af den europæiske udvikling på området.

12. september 2017

Styrelsen for Forskning og
Uddannelse
Forskningsinfrastruktur

Bredgade 40
1260 København K
Tel. 3544 6200
Fax 3544 6201
Mail sfu@ufm.dk
Web www.ufm.dk

CVR-nr. 1991 8440

Sagsbehandler
Troels Rasmussen
Tel. 72 31 82 96
Mail trra@ufm.dk

Ref.-nr. 17/016907-07



Behovskortlægning

Der foretages en kortlægning af danske forskeres nuværende behov samt behovene i et fem-ti-årigt perspektiv. Det er centralt, at behovsanalysen er baseret på bidrag fra både ledelsesrepræsentanter og aktive forskere fra alle hovedområder og på tværs af forskningsinstitutioner.

Kortlægningen skal for hvert hovedområde beskrive forskernes behov for HPC, net, data management og lagring samt kompetencebehov og services, herunder særlige støttefunktioner. Den skal beskrive, hvilke dele der bør håndteres på nationalt niveau, og hvilke dele der bør håndteres på lokalt/institutionelt niveau.

Formålet er at få tilvejebragt en vurdering af de fremtidige behov fordelt på hovedområder, samt de implikationer dette har for udviklingen af en national strategi.

Konkret skal følgende spørgsmål besvares:

- Hvad er behovene inden for de videnskabelige hovedområder i et tiårigt perspektiv?
- Hvad stiller dette af krav til løsninger i både teknisk og økonomisk henseende?
- Hvilke løsninger håndteres teknisk og økonomisk mest optimalt på nationalt niveau og hvilke på lokalt/institutionelt niveau?

Styrelsen for Forskning og
Uddannelse

Benchmark i forhold til andre lande

Der ønskes en sammenligning af den nuværende danske model for e-infrastruktursamarbejde med de nationale løsninger, der er valgt i en række andre lande med lignende karakteristika som Danmark f.eks. de øvrige nordiske lande og Holland.

Formålet er at få et sammenligningsgrundlag for, hvorledes det nationale e-infrastruktursamarbejde håndteres i andre lande. Benchmarket skal omfatte HPC, net og datamanagement, kompetenceudvikling og andre services samt håndtering af internationalt samarbejde.

Konkret ønskes der en besvarelse af følgende spørgsmål:

- Hvordan er det nationale e-infrastruktursamarbejde organiseret?
- Hvordan er opgaver og ansvar fordelt mellem det nationale og lokale/institutionelle niveau og mellem forskellige aktører?
- Hvilken governance-model er valgt, og hvilken selskabsform er brugt?
- Hvad er den samlede økonomi for det nationale e-infrastruktursamarbejde, og hvordan er dette finansieret?
- Hvordan håndteres deltagelse i europæiske e-infrastrukturer – hvilket niveau (nationalt/lokalt) og med hvilke ressourcer?

Beskrivelse af den europæiske udvikling

Dansk e-infrastruktur er i mange tilfælde integreret i det europæiske samarbejde f.eks. via etablering af fælles e-infrastruktur, navnlig fælleseuropæiske beregningsanlæg og databaseinfrastruktur, samt net- og sikkerhedstjenester. Samtidig er EU-Kommissionen i gang med at formulere en strategi for den såkaldte European Open Science Cloud (EOSC), som er tænkt som et europæisk økosystem for digital dataanvendelse baseret på FAIR-principper (Findable, Accessable, Interoperable, Reuseable) for open access.

Der ønskes en analyse, som forholder sig til den europæiske udvikling inden for e-infrastruktur, som den pt. er under udvikling og afklaring, og en analyse af de muligheder og udfordringer, som dette giver Danmark. Det gælder både for net, HPC



og datamanagement samt andre væsentlige europæiske e-infrastrukturaktiviteter, som Danmark er involveret i.

Formålet er at få beskrevet udviklingen af de enkelte e-infrastrukturaktiviteter og det danske engagement heri.

Konkret ønskes en kort beskrivelse af de enkelte e-infrastrukturer og en besvarelse af følgende spørgsmål:

- Hvad er formål og de væsentligste aktiviteter af de europæiske e-infrastrukturer?
- Hvordan er Danmark involveret økonomisk og ressourcemæssigt?
- På hvilket niveau (nationalt eller institutionelt/lokalt) vurderes den danske deltagelse mest hensigtsmæssigt varetaget – teknisk og økonomisk?
- Hvordan forventes aktiviteten at udvikle sig fremadrettet, og hvilke implikationer forventes dette af få for dansk deltagelse?
- Er der europæiske e-infrastrukturer, som Danmark pt. ikke deltager i, men som vurderes som relevante for Danmark?

Styrelsen for Forskning og
Uddannelse

Udviklingen af EOSC beskrives ligeledes efter ovenstående skabelon. Dog skal der tages forbehold for, at EOSC er et relativt nyt begreb i europæisk sammenhæng, og der derfor er mange uafklarede forhold, som gør det vanskeligt præcist at vurdere, hvorledes det vil påvirke danske forhold på længere sigt.

Output

RK og SFU forventer en udredning, der er egnet til offentliggørelse, på ca. 40-60 sider.

Tidsplan

Udrednings- og analysearbejdet skal gennemføres i perioden juni – november 2017. Endelig rapport skal afleveres til RK og SFU senest med udgangen af november 2017.

e-infrastruktur til forskning i verdensklasse i 2025

- en analyse og udredning vedrørende dansk e-infrastruktur samarbejde

Bilag 2:

Talere i eScience status præsentation på DeiC konferencen den 26. september og deltagere ved rundbordssamtale den 5. oktober 2017

Liste over inviterede dekaner og deltagere 26.09.2017 og 05.10.2017

| Universitet | Område | Inviteret | Deltager 26. september | Deltager 5. oktober |
|------------------------|--------------------------|----------------------------|------------------------|------------------------|
| Københavns Universitet | Humaniora | Ulf Hedetoft | Julie Sommerlund | |
| | Teologi | Kirsten Busch Nielsen | | |
| | Naturvidenskab | John Renner Hansen | Mads Nielsen | John Renner Hansen |
| | Samfundsvidenskab | Troels Østergaard Sørensen | Mette Wier | |
| | Juridisk fakultet | Jacob Graff Nielsen | Henrik Palmer Olsen | |
| Aarhus Universitet | Sundhedsvidenskab | Ulla Wewer | Peter Steffen Loft | Søren Brunak |
| | Humaniora (ARTS) | Johnny Laursen | Johnny Laursen | |
| | Naturvidenskab (SCIENCE) | Niels Chr. Nielsen | Søren Rud Keiding | Søren Rud Keiding |
| | Samfundsvidenskab (BSS) | Thomas Pallesen | Jakob Kjær Eskildsen | |
| | Sundhedsvidenskab (SUND) | Lars Bo Nielsen | Anders Børglum | Lars Bo Nielsen |
| Aalborg Universitet | Humaniora | Henrik Halkier | Henrik Halkier | |
| | Naturvidenskab | Mogens Rysholt Poulsen | | |
| | Teknologi | Eskild Holm Nielsen | Henrik Pedersen | |
| | Samfundsvidenskab | Rasmus Antoft | | Søren Kristiansen |
| | Sundhedsvidenskab | Lars Hvilsted Rasmussen | | |
| Syddansk Universitet | Humaniora | Simon Møberg Torp | Simon Møberg Torp | Simon Møberg Torp |
| | Naturvidenskab | Martin Zachariasen | Martin Zachariasen | Martin Zachariasen |
| | Teknologi | Henrik Bindslev | | Peder Thusgaard Ruhoff |
| | Samfundsvidenskab | Nikolaj Malchow-Møller | Oliver Baumann | |
| | Sundhedsvidenskab | Ole Skøtt | Torben A. Kruse | Ole Skøtt |
| Roskilde Universitet | Humaniora | Hanne Løngreen | | |
| | Naturvidenskab | Lars Nørsgaard | | |
| | Teknologi | Anders Siig Andersen | | |
| | Samfundsvidenskab | Peter Kragelund | | |
| DTU | Teknologi | Katrine Krogh Andersen | Katrine Krogh Andersen | Katrine Krogh Andersen |
| IT-Universitetet | Teknologi | Jens Chr. Godskesen | | |
| CBS | Samfundsvidenskab | Peter Møllgaard | Søren Hvidkjær | Søren Hvidkjær |
| | | | | |

Øvrige deltagere 5. oktober 2017:

- Direktør Hans Müller Pedersen, SFU
- Rektor Anders Bjarklev, DTU DKUNI,
- Nikolaj Helm-Petersen, DKUNI
- Lars Christensen, SFU
- Troels Rasmussen, SFU
- Børge Obel, DeIC
- Steen Pedersen, DeIC
- Gitte Kudsk DeIC

e-infrastruktur til forskning i verdensklasse i 2025

- en analyse og udredning vedrørende dansk e-infrastruktur samarbejde

Bilag 3a:

**eScience status præsentationer på DeiC konferencen 25.
september 2017**



DeiC konference 2017
September 26, 2017

Status for e-science Syddansk Universitet / Odense Universitetshospital

Professor Torben A. Kruse

**Department of Clinical Genetics
Clinical Genome Center
Odense University Hospital
and
Institute of Clinical Research
University of Southern Denmark**





Behovskortlægning:

behov for e-infrastruktur som HPC/computing, net, data management og lagring, samt kompetencebehov og services, herunder særlige støttefunktioner.

Vigtigste område:

Genomforskning – BigData i Biomedicin





>gi|6552298|ref|NM_007294.1| Homo sapiens breast cancer 1, early onset (BRCA1), transcript variant BRCA1a, mRNA

AAAAC TGC GACTGCGCGGCGT GAGCTCGCTGAGACTTCTGGACCCCGCACCAGGCTGTGGGGTTTCTCAGATAACTGGGCCCTGCGCTCAGGAGGCCTCACCCCTGCTCTGGGT
AAAGTTCATTGGAACAGAAAGAAATGGATTTATCTGTCTTCCGCTTGAAGAAGTACAAAATGTCATTAATGCTATGCAGAAAAATCTTAGAGTGTCCCATCTGTCTGGAGTGTATCAA
GGAACCTGTCTCCACAAAGTGTGACCACATATTTTGCAAAATTTGTCATGTGAAAACCTCAACAGAAAGAAAGGGCCTTACAGTGTCCCTTATGTAAGAATGATATAACCAAAAGG
AGCTACAAGAAAAGTACGAGATTTAGTCAACTTGTGAAAGACTATTGAAAATCATTGTGCTTTTCAGCTTGACACAGGTTTGGAGTATGCAAAACAGCTATAAATTTGCAAAAAAGG
AAAATAACTCTCTGAACATCTAAAAGATGAAGTTTCTATCATCCAAAGTATGGGCTACAGAAACCGTGCCAAAAGACTTCTACAGAGTGAACCCGAAAATCCTTCTTGCAGGAAA
CCAGTCTCAGTGTCCAACCTCTAACCTTGGAACTGTGAGAACTCTGAGGACAAAAGCAGCGGATACACCTCAAAAGACGCTGTGTCTACATTGAATGGGATCTGATTCTTCTGAAGA
TACCCTTAATAAGGCAACTTATTGCAGTGTGGGAGATCAAGAATTGTTACAAATCACCCCTCAAGGAACCAGGGATGAAATCAGTTTGGATTCTGCAAAAAAGGCTGCTTGTGAATTT
TCTGAGACGGATGAACAAATAGCTGAACATCATCAACCCAGTAATAATGATTTGAAACACCAGTGAAGCGTGCAGCTGAGAGGCATCCAGAAAAGTATCAGGGTAGTCTGTGTTCA
AACTTGCATGTGGAGCCATGTGGCACAAATACTCATGCCAGCTATTACAGCATGAGAACAGCAGTATTACTACTAAAGACAGAATGAATGTAGAAAAAGGCTGAATTTCTGTAAT
AAAAGCAAACAGCCTGGCTTAGCAAGGAGCCAACATAACAGATGGGCTGGAAGTAAAGAAAACATGTAATGATAGGCGGACTCCCAGCACAGAAAAAAGGTAGATCTGAATGCTG
ATCCCTGTGTGAGAGAAAAAGAATGGAATAAGCAGAAACTGCCATGCTCAGAGAATCCTAGAGATACTGAAGATGTTCCCTGGATAACACTAAAATAGCAGCATTAGAAAAGTTAATG
AGTGGTTTTCCAGAAGTGATGAACTGTTAGGTTCTGATGACTCACATGATGGGGAGTCTGAATCAAATGCCAAAGTAGCTGATGTATTGGACGTTCTAAATGAGGTAGATGAATATTC
TGGTTCTCAGAGAAAATAGACTTACTGGCCAGTGCCTCATGAGGCTTTAATATGTAAGTGAAGAGTTCCTCAAAATCAGTAGAGAGTAATATTGAAGACAAAATATTTGGG
AAAACCTATCGGAAGAAAGGCAAGCCTCCCAACTTAAGCCATGTAAGTGA AAAATCTAATTATAGGAGCATTGTTACTGAGCCACAGATAATAACAAGAGCGTCCCTCACAAAATAAA
TTAAAGCGTAAAAGGAGACCTACATCAGGCCTTACCTGAGGATTTTATCAAGAAAAGCAGATTTGGCAGTTCAAAAGACTCCTGAAATGATAAATCAGGGAATAACCAACGGAG
CAGAATGGTCAAGTGATGAATATTACTAATAGTGGTCAATGAGAAATAAAACAAAAGGTGATTCTATTAGAATGAGAAAAATCTAACCCAATAGAATCACTCGAAAAAGAATCTGCT
TTCAAACGAAAAGCTGAACCTATAAGCAGCAGTATAAGCAATATGGAACCTCGAATTAATATCCACAATTCAAAAGCACCTAAAAAGAATAGGCTGAGGAGGAAGTCTTCTACCAG
GCATATTCATGCGCTTGAACCTAGTAGTCACTAGTGAATCTAAGCCACCTAATTGTACTGAATTGCAAATGATAGTTGTTCTAGCAGTGAAGAGATAAAGAAAAAAGTACAACCA
AATGCCAGTCAGGCACAGCAGAACCTACAACCTATGGAAGGTAAGAACCTGCAACTGGAGCCAAGAAGAGTAACAAGCCAAATGAACAGACAAGTAAAAAGACATGACAGCGAT
ACTTTCCAGAGCTGAAGTTAACAATGCACCTGGTCTTTTACTAAGTGTCAAAATACCAGTGAACCTAAAAGAAATTTGTCAATCCTAGCCTTCCAAGAGAAGAAAAAGAAGAGAAAC
TAGAAACAGTTAAAGTGTCTAATAATGCTGAAGACCCCAAGATCTCATGTTAAGTGGAGAAAAGGGTTTTGCAAACTGAAAGATCTGTAGAGAGTAGCAGTATTTCAATTGTTACCTG
GTACTGATTATGGCACTAGGAAAAGTATCTCGTTACTGGAAGTTAGCACTCTAGGGAAGGCAAAAACAGAACCAATAAATGTGTGAGTCACTGTGCAGCATTTGAAAACCCCAAGG
GACTAATTCATGGTTGTTCCAAAGATAATAGAAATGACACAGAAGGCTTTAAGTATCCATTGGGACATGAAGTTAACACAGTGGGAAACAGCATAGAAAATGGAAGAAAGTGA
CTTGATGCTCAGTATTTGCAGAATAACATTCAAGGTTTTCAAAGCGCCAGTCAATTTGCTCCGTTTTCAAATCCAGGAAATGCAGAAAGAGGAATGTGCAACATTCTGCCACTTGGGT
CTTAAAGAAACAAAGTCCAAAAGTCACTTTTGAATGTGAACAAAAGGAAGAAAATCAAGGAAAGAATGAGTCTAATATCAAGCCTGTACAGACAGTTAATATCACTGCAGGCTTTCC
TGTGGTTGGTCAAGAAAGATAAGCCAGTTGATAATGCCAAATGTAGTATCAAAGGAGGCTCTAGGTTTTGTCTATCATCTCAGTTCAGAGGCAACGAAACTGGACTCATTACTCAAAT
AAACATGGACTTTTACAAAACCCATATCGTATACCACCACTTTTCCCATCAAGTCATTTGTTAAAACTAAATGTAAGAAAAATCTGCTAGAGGAAAACTTTGAGGAAACATTCATGTA
CACCTGAAAGAGAAAATGGGAAATGAGAACTTCCAAGTACAGTGCAGCACAATTAGCCGTAATAACATTAGAGAAAATGTTTTTAAAGAAGCCAGCTCAAGCAATTAATGAAGTA
GGTTCAGTACTAATGAAGTGGGCTCCAGTATTAATGAAATAGGTTCCAGTGTGAAAAACATTC AAGCAGAACTAGGTAGAAAACAGAGGGCCAAAATTTGAATGCTATGCTTAGATTA
GGGGTTTTGCAACCTGAGGTCTATAAACAAAGTCTTCTGGAAGTAATTGTAAGCATCTGAAATAAAAAAGCAAGAATATGAAGAAGTAGTTCAGACTGTTAATACAGATTTCTCTC
CATATCTGATTTAGATAAAGTATGAAACAGCCTATGGGAAGTATGTCATGTCATCTCAGGTTTGTCTGAGACACCTGATGACCTGTTAGATGATGGTGAATAAAGGAAGATACTAGTTT
TGCTGAAAATGACATTAAGGAAGTTCTGCTGTTTTTGAAGAAAAGCGTCCAGAAAAGGAGAGCTTAGCAGGAGTCTAGCCCTTTTACCATAACATTTGGCTCAGGGTTACCGAAGA
GGGCGCAAGAAATAGAGTCTCAGAAGAGAACTTATCTAGTGAAGGATGAAGAGCTTCCCTGCTTCCAACACTTGTTATTTGGTAAAGTAAACAATATACCTTCTCAGTCTACTAGGC
ATAGCACCGTTGCTACCAGTGTCTGTTAAGAACACAGAGGAGAATTTATTTATCATTGAAGAATAGCTTAAATGACTGCAGTAACCCAGGTAATATTGGCAAAAGGCAATCTCAGGAAC

AT





GENOME SIZE AND VARIATION

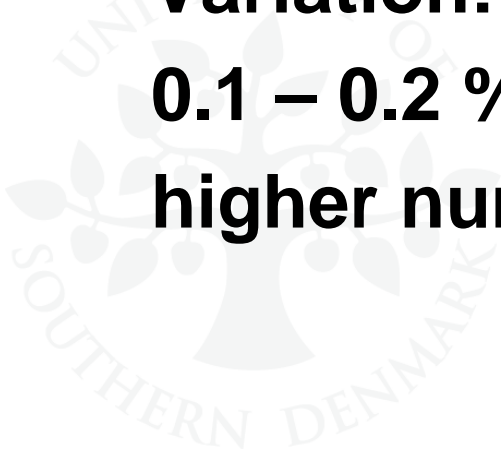
Size:

3×10^9 base pairs

Variation:

0.1 – 0.2 % of basepairs polymorphic

higher number of rare variants

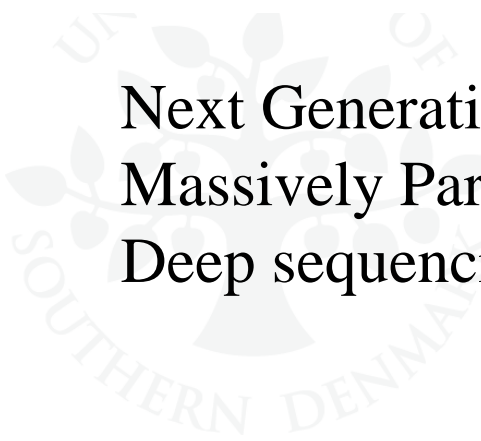


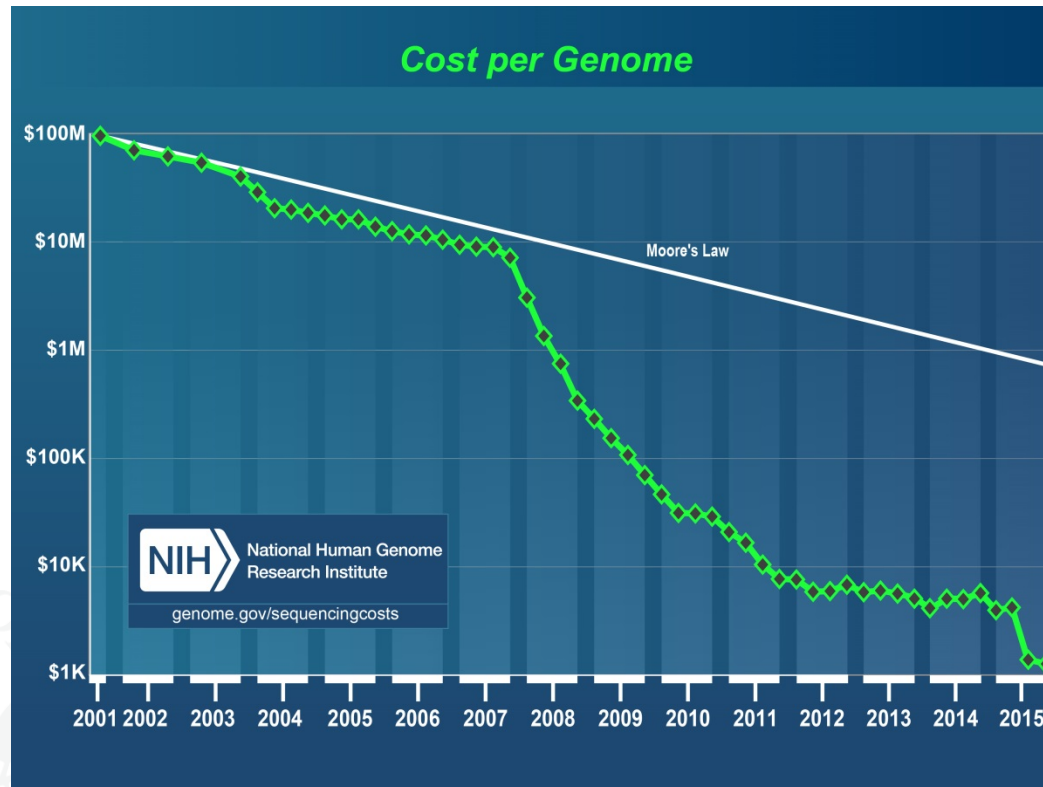
Sequencing



Sanger sequencing

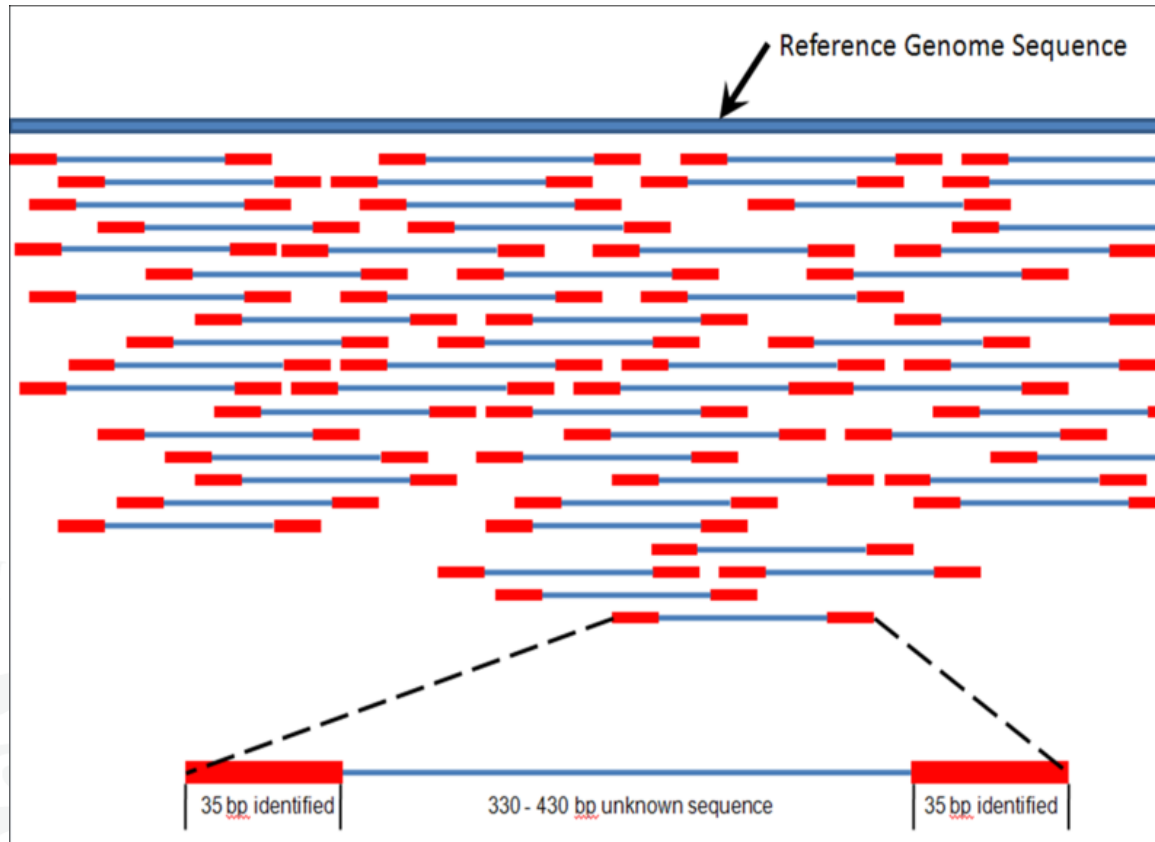
Next Generation Sequencing (NGS)
 Massively Parallel Sequencing
 Deep sequencing





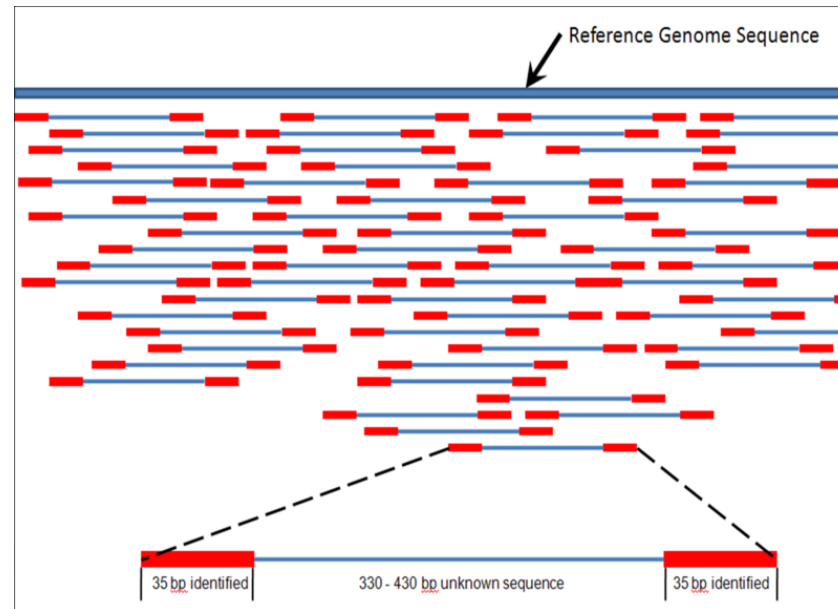


Next-generation-sequencing (NGS) – paired end sequencing





Sequence alignment – need for high performance computing



100 bp fragment afsøger 3×10^9 positioner
1 genom er 30,000,000 fragmenter
30X genom er ca. 9×10^8 fragmenter



>gi|6552298|ref|NM_007294.1| Homo sapiens breast cancer 1, early onset (BRCA1), transcript variant BRCA1a, mRNA

AAAAC TGC GACTG C GCG GCG CTG GAG CT C GCT GAG ACT T C T G G A C C C C G C A C C A G G C T G T G G G G T T T C T C A G A T A A C T G G G C C C C T G C G C T C A G G A G G C C T T C A C C C T T G C T C T G G G T
AAA G T T C A T T G G A A C A G A A A G A A A T G G A T T T A T C T G C T T T C G C G T T G A A G A A G T A C A A A A T G T C A T T A A T G C T A T G C A G A A A A T C T T A G A G T G T C C C A T C T G T C T G G A G T T G A T C A A
G G A C C T G T C C C A A A G T G T G A C C A C A T A T T T T G C A A T T T T G C A A T T T C T C A A C C A G A A G A A A G G G C C T T C A C A G T G C C T T A T G T A A G A A T G A T A T A C C A A A A A G G
A G C T T A C A A G A A A G T A C G A G A T T T A G T C A A C T T G T T G A A G A G C T A T T G A A A A T C A T T T T G C T T T T C A G C T T G A C A C A G G T T T G G A G T A T G C A A A C A G C T A T A A T T T T G C A A A A A G G
A A A A T A A C T C T C T G A A C A T C T A A A A G A T G A A G T T T C T A T C A T C C A A A G T A T G G G C T A C A G A A A C C G T G C C A A A A G A C T T C T A C A G A G T G A A C C C G A A A A T C C T T C C T T G C A G G A A
C C A G T C T C A G T G T C C A A C T C T A A C C T T G G A A C T G T G A G A A C T C T G A G G A C A A A G C A G C G G A T A C A A C C T C A A A A G A C G T C T G T C T A C A T T G A A T T G G G A T C T G A T T C T T C T G A A G A
T A C C G T T A A T A A G G C A A C T T A T T G C A G T G T G G G A G A T C A A G A A T T G T T A C A A A T C A C C C C T C A A G G A A C C A G G G A T G A A A T C A G T T T G G A T T C T G C A A A A A A G G C T G C T T G T G A A T T T
T C T G A G A C G G A T G A A C A A A T A G C T G A A C A T C A C A C C C A G T A A T A A T G A T T T G A A C A C C A C T G A G A A G C G T G C A G C T G A G A G G C A T C C A G A A A A G T A T C A G G G T A G T T C T G T T T C A
A A C T T G C A T G T G G A G C C A T G T G G C A C A A A T A C T C A T G C C A G C T A T T A C A G C A T G A G A A C A G C A G T T T A T T A C T A C T A A A G A C A G A A T G A A T G T A G A A A A G G C T G A A T T C T G T A A T
A A A A G C A A A C A G C C T G G C T T A G C A A G G A G C C A A C A T A A C A G A T G G G C T G G A A G T A A G G A A A C A T G T A A T G A T A G G C G G A C T C C C A G C A C A G A A A A A A A G G T A G A T C T G A A T G C T G
A T C C C C T G T G T G A G A G A A A A G A A T G G A A T A A G C A G A A A C T G C C A T G C T C A G A G A A T C C T A G A G A T A C T G A A G A T G T T C C T T G G A T A A C A C T A A A T A G C A G C A T T C A G A A A G T T A A T G
A G T G G T T T T C C A G A A G T G A T G A A C T G T T A G G T T C T G A T G A C T C A C A T G A T G G G G A G T C T G A A T C A A A T G C C A A A G T A G C T G A T G T A T T G G A C G T T C T A A A T G A G G T A G A T G A A T T C
T G G T T C T C A G A A A A T A G A C T T A C T G G C C A G T G A T C C T A T G A G G C T T T A A T A T G T A A A A G T G A A A G A G T T C A C T C C A A A T C A G T A G A G A T A A T T T G A A G A C A A A A T A T T T G G G
A A A A C C T A T C G G A A G A A G G C A A G C C T C C C C A A C T T A A G C C A T G T A A C T G A A A A A T C T A A T T A T A G G A G C A T T T G T T A C T G A G C C A C A G A T A A T A C A A G A G C G T C C C C T C A C A A A T A A A
T T A A A G C G T A A A A G G A G A C C T A C A T C A G G C C T T C A T C C T G A G G A T T T T A T C A A G A A A G C A G A T T T G G C A G T T C A A A A G A C T C C T G A A A T G A T A A A T C A G G G A A C T A A C C A A C G G A G
C A G A A T G G T C A A G T A T G A A T A T T A C T A A T A G T G G T C A T G A G A A T A A A A C A A A A G G T G A T T C T A T T C A G A A T G A G A A A A T C C T A A C C C A A T A G A A T C A C T C G A A A A A G A A T C T G C T
T T C A A A C G A A A G C T G A A C C T A T A A G C A G C A G T A T A A G C A A T A T G G A A C T C G A A T T A A A T A T C C A C A A T T C A A A A G C A C C T A A A A A G A A T A G G C T G A G G A G G A A G T C T T C T A C C A G
G C A T A T T C A T G C G C T T G A A C T A G T A G T C A G T A G A A A T C T A A G C C A C C T A A T T G T A C T G A A T T G C A A A T T G A T A G T T G T T C T A G C A G T A A G A G A T A A A G A A A A A A A A A G T A C A A C C A
A A T G C C A G T C A G G C A C A G C A G A A A C C T A C A A C T A T G G A A G G T A A A G A A C C T G C A A C T G G A G C C A A G A A G A G T A A C A A G C C A A T G A A C A G A C A A G T A A A A A G A C A A T G A C A G C G A T
A C T T T C C C A G A G C T G A A G T T A C A A A T G C A C C T G G T T C T T T T A C T A A G T G T T C A A A T A C C A G T G A A C T T A A A G A A T T T G T C A A T C C T A G C C T T C C A A G A G A A G A A A A A G A A G A G A A A C
T A G A A A C A G T T A A A G T G T C T A A A T A A T G C T G A A G A C C C C A A A G A T C T C A T G T T A A G T G G A G A A A G G G T T T T G C A A A C T G A A A G A T C T G T A G A G A G T A G C A G T A T T C A T T G T G T A C C T G
G T A C T G A T T A T G G C A C T A G G A A A G T A T C T C G T T A C T G G A A G T T A G C A C T C T A G G G A A G G C A A A A C A G A A C C A A A T A A A T G T G T G A G T C A G T G T G C A G C A T T T G A A A A C C C C A A G G
G A C T A A T T C A T G G T T G T T C C A A A G A T A A T A G A A A T G A C A C A G A A G C T T T A A G T A T C C A T T G G G C A T G A A G T T A A C C A C A G T C G G G A A A C A A G C A T A G A A A T G G A A G A A A A G T G A A
C T T G A T G C T A G T A T T T G C A G A A T A C A T T C A A A G G T T T C A A A G C G C C A G T C A T T T G C C G T T T T C A A A T C C A G G A A A T G C A G A A G A G G A A T G T G C A A C A T T C T G C C C A C T C T G G G T C
C T T A A G A A A C A A A G T C C A A A A G T C A C T T T T G A A T G T G A A C A A A G G A A G A A A A T C A A G G A A A G A A T G A G T C T A A T A T C A A G C C T G T A C A G A C A G T T A A T A T C A C T G C A G G C T T T C C
T G T G G T T G G T C A G A A A G A T A A G C C A G T T G A T A A T G C C A A A T G T A G T A T C A A A G G A G G C T C T A G G T T T T G T C T A T C A T C T C A G T T C A G A G G C A A C G A A A C T G G A C T A T T A C T C C A A A T
A A A C A T G A C T T T T A C A A A A C C C A T A T C G T A T A C C A C C A C T T T T T C C C A T C A A G T C A T T T G T T A A A A C T A A A T G T A A G A A A A A A C T G C T A G A G A A A A C T T T G A G G A A C A T T C A A T G T
C A C C T G A A A G A G A A A T G G G A A A T G A G A A C A T T C C A A G T A C A G T G A G C A C A A T T A G C C G T A A T A A C A T T A G A G A A A A T G T T T T A A A G A A G C C A G C T C A A G C A A T A T T A A T G A A G T A
G G T T C C A G T A A T G A A G T G G G C C C A G T A T T A A T G A A A T A G G T T C C A G T G A T G A A A A C A T T C A A G C A G A A C T A G G T A G A A A C A G A G G G C C A A A A T T G A A T G C T A T G C T T A G A T T A
G G G G T T T T G C A A C C T G A G G T C T A T A A A C A A A G T C T T C C T G G A A G T A A T T G T A A G C A T C C T G A A A T A A A A A A G C A A G A A T A T G A A G A A G T A G T T C A G A C T G T T A A T A C A G A T T T C T C T C
C A T A T C T G A T T T C A G A T A A C T T A G A A C A G C C T A T G G G A A G T A G T C A T G C A T C T C A G G T T T G T T C T G A G A C A C C T G A T G A C C T G T T A G A T G A T G G T G A A A T A A A G G A A G A T A C T A G T T T
T G C T G A A A A T G A C A T T A A G G A A A G T T C T G C T G T T T T A G C A A A A G C G T C C A G A A A G G A G A G C T T A G C A G G A G T C C T A G C C C T T T C A C C C A T A C A C A T T T G G C T C A G G G T T A C C G A A G A
G G G C C A A G A A A T T A G A T C C T C A G A A G A A C T T A T C A G T G A G G A T G A A G A G C T T C C C T G C T T C C A A C A C T T G T T A T T T G G T A A A G T A A A C A A T A T A C C T T C T A G T C T A C T A G G C
A T A G C A C C G T T G C T A C C A G T G T C T G T C T A A G A A C A C A G A G G A G A A T T T A T T A T C A T T G A A G A A T A G C T T A A A T G A C T C A G T A A C C A G G T A A T A T T G G C A A A G G C A T T C A G G A A C

AT



Computer time:

LINUX-server 32 cores:

One 30x genome alignment 2-3 days

Abacus2.0 one node 24 cores

One 30x genome alignment 2-3 days

Storage:

One 30x genome incl backup 400GB



Next generation sequencing (computer time)

Clinical Genome Center, OUH/SDU

2017:

30 genomes/month

alignment on 3 nodes on Abacus2.0

2018:

600 genomes/month

alignment on 60 nodes on Abacus2.0 (10% of capacity)



Next generation sequencing (storage)

Clinical Genome Center, OUH/SDU

2017:

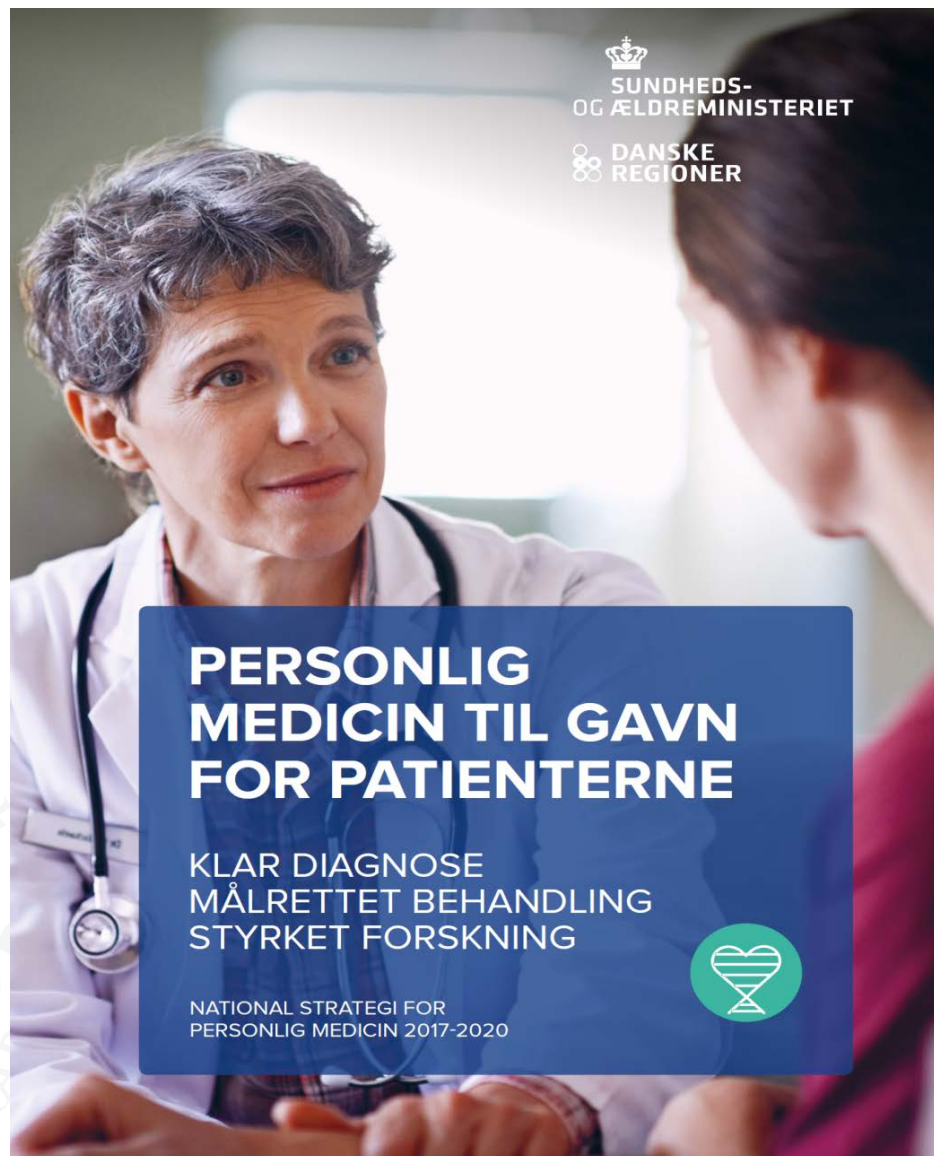
30 genomes/month

15 TB/month

2018:

600 genomes/month

300 TB/month i.e. 3.5 PB/year




SUNDHEDS-
OG ÆLDREMINISTERIET

 DANSKE
REGIONER

PERSONLIG MEDICIN TIL GAVN FOR PATIENTERNE

KLAR DIAGNOSE
MÅLRETTET BEHANDLING
STYRKET FORSKNING

NATIONAL STRATEGI FOR
PERSONLIG MEDICIN 2017-2020



 Region
Syddanmark





Areas with future high e-Science needs:

- **Data analysis of NGS raw data**
- **Machine learning of Big Data – Omics data**
- **Analysis of real-time data from wearable devices**
- **Image analysis**

Konklusion:

- **Kraftigt øget behov for HPC**
- **Kraftigt øget behov for storage**
- **Øget behov for hurtigt net mellem HPC centre**
- **Kraftigt øget behov for dataloger/bioinformatikere som støtte for biomedicinske og kliniske forskere**

Da forskere, klinikere og patienter findes decentralt er det vigtigt at ovennævnte udbygning bygger på de allerede eksisterende relativt decentrale strukturer og kompetencer



Computational Biomedicine



Region
Syddanmark





eScience at The Faculty of Health and Medical Sciences, UCPH (SUND)

Steffen Loft, SUND
Head of Department of Public Health

DeIC conference, Sep 26, 2017

UNIVERSITY OF COPENHAGEN



eScience at SUND

- Facts about SUND
- Strategy
- Status and road map
- Future needs



SUND – a large faculty

SUND is among the largest faculties of its kind in Europe.



4,400
employees
headcount



2,800
scientific staff
headcount



7,800
students



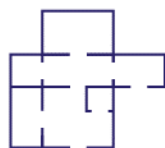
1,700
PhD students



No. 2
university among the
Nordic countries within
Life Science and
Medicine in 2015



EUR **415** million
annual revenue



302,000 m²
building floor space



11
departments



3
teaching hospitals

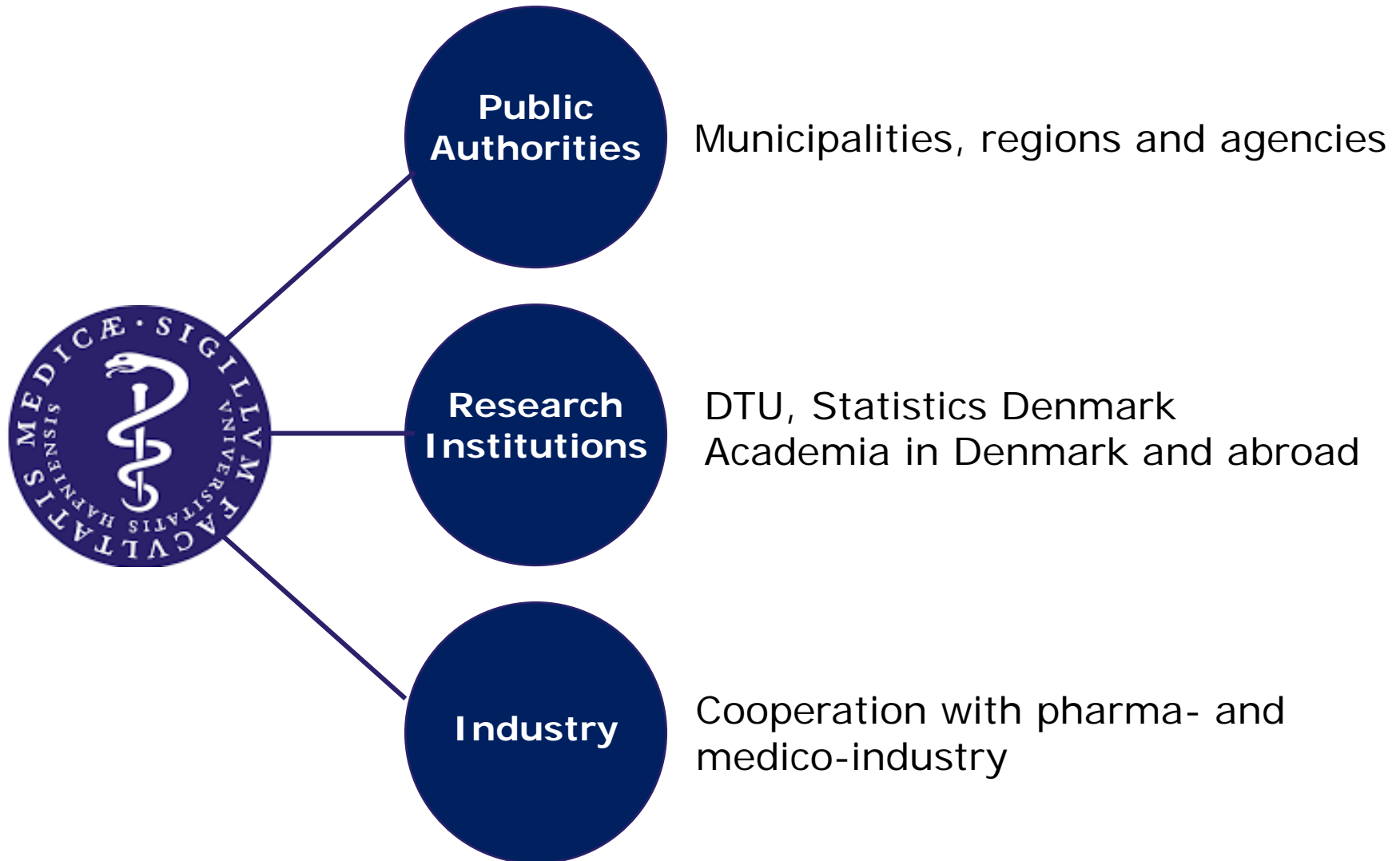


40,000
patients treated
each year

SUND – a diverse faculty



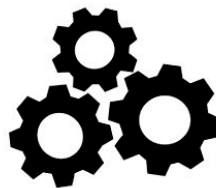
SUND – cooperation with other actors



eScience at SUND – strategy

Make it simple to get

F_{indable} A_{ccessible} I_{nteroperable} R_{eusable}

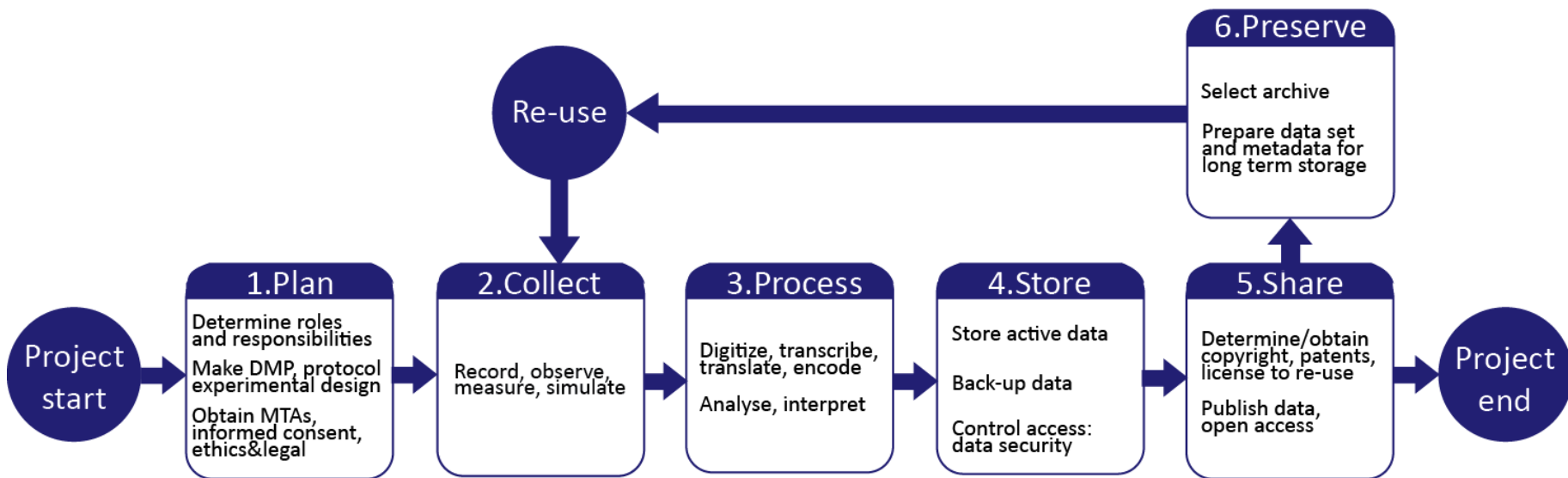


Implementation of FAIR principles to achieve OpenScience

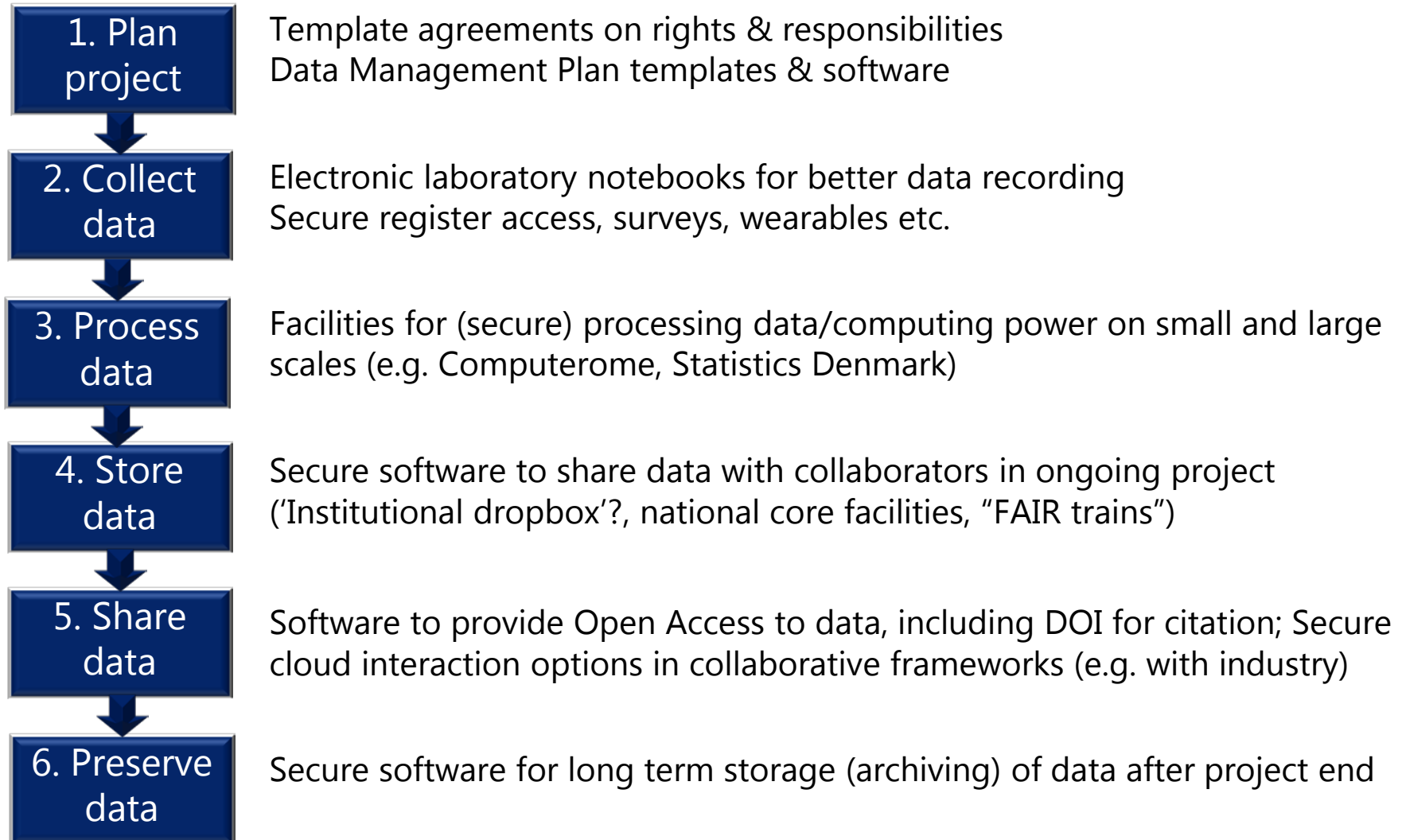
- excellent infrastructure
- expert support staff
- acceptance & compliance of researchers

What do researchers need?

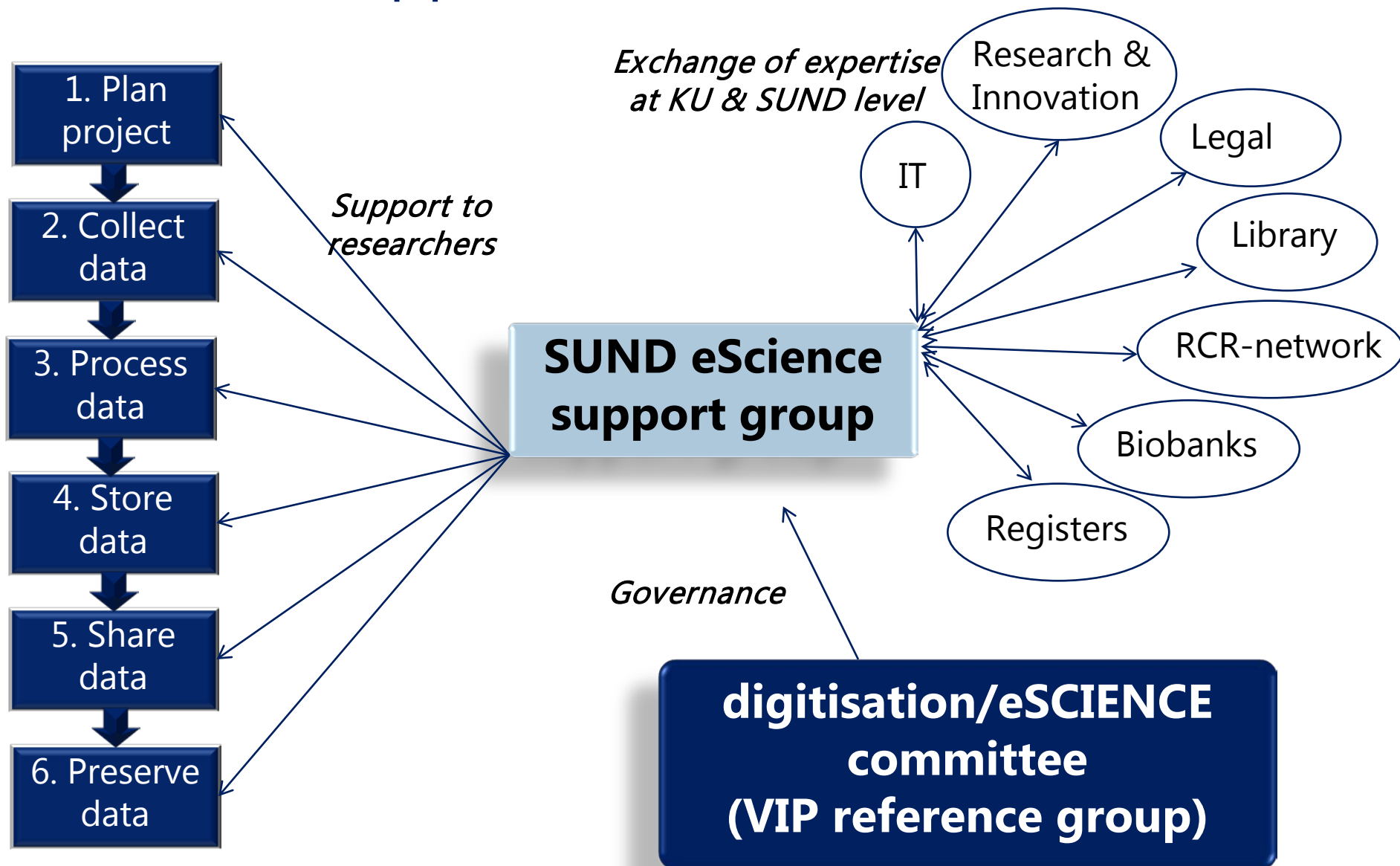
- excellent infrastructure and expert support to deal with all aspects of the data life cycle



What do researchers need?



An ideal support scenario...



eScience at SUND – where are we now?

Support

- Policy for Research Data Management (published April 2016)
- Roadmap and implementation plan for RDM/eScience infrastructure and support at SUND
- Network activities in and coordination with UCPH and national eScience committees (e.g. Det Nationale Forum under DeIC)

Infrastructure

- DMPonline
- Data processing and storage at SUND-IT
- Data processing at Computerome, a DTU/UCPH/DeiC initiative
- Data processing and storage at Statistics Denmark and Danish Health Data Authority

eScience at SUND – what does it take?

Support

- Further implementation of eScience support
- Training of researchers, e.g. development of e-learning modules
- Capacity building, e.g. train-the-trainer activities
- Advice on compliance with legislation

Infrastructure

- Electronic lab book (ELN) solutions
- Storage solutions with annotated metadata (e.g. Dataverse)
- Increasing local capacity for secure processing, storage and archiving
- Increasing capacity for secure exchange, processing and storage of (sensitive) data in national networks and core facilities as well as internationally

A case...Protein Imaging Platform

at the Novo Nordisk Foundation Center for Protein Research



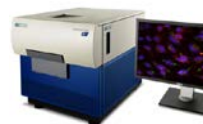
High-end Research Microscopes at CPR and CFIM



*Zeiss LSM880 Airy Scan
Confocal Laser Scanning Microscope*



*Perkin Elmer Spinning Disk
Confocal Microscope*



*Molecular Devices
High-content microscope*



*2x Olympus ScanR
High-content microscope*

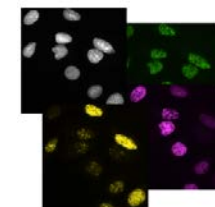


*Planned for 2019: Perkin Elmer
Confocal High-content microscope*



Large Image Data sets:

High-content microscope time-lapse imaging
High-content microscope screening
Large 3D high-resolution confocal imagery



Projected Image Data Volume:

100 TB for 2018-2019
300 TB for 2020-2024

Requirement for fast, local access of image data for image analysis and data processing

eScience at SUND – future needs

- Local support and infrastructure, adaptable to increasing data volume and demand for secure data collection, processing, exchange and storage
- National coordination and infrastructure initiatives to maximize synergies

Thank you.



DeiC konference 2017

Status for eScience - Sundhedsvidenskab

Anders Børglum, MD, PhD
Professor, centerleder

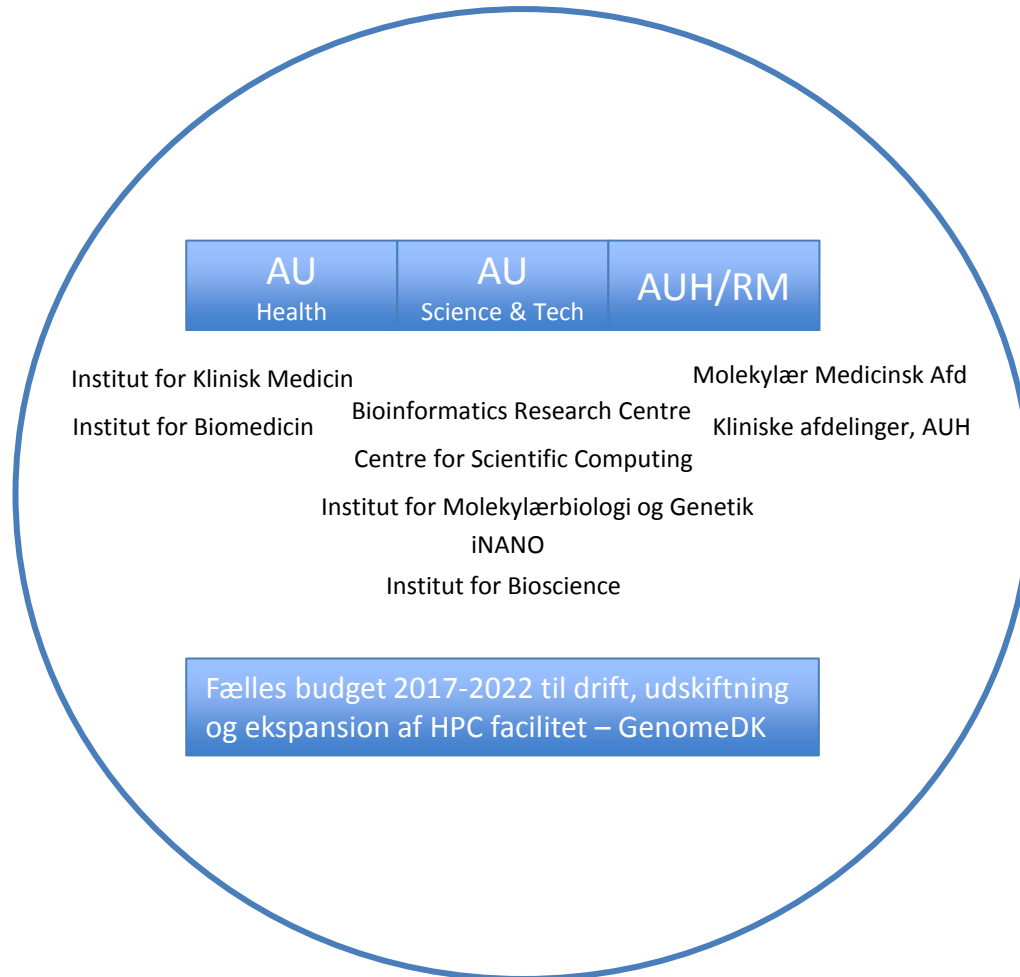
AU eScience, sundhedsvidenskab

- i tæt samarbejde med AU teknologi- og naturvidenskab (ST) og AUH/RM -
+ samfundsvidenskab (BSS)

Personlig Medicin og HPC

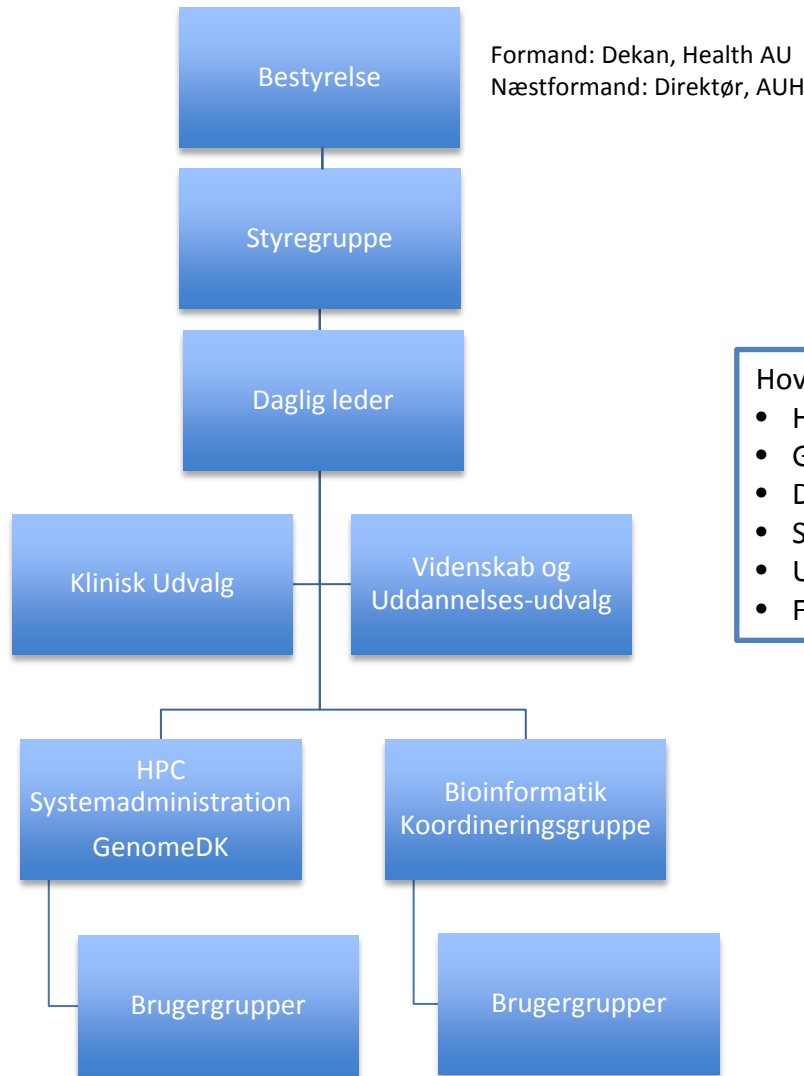
- Håndtering og analyse af massive genomiske data
- Integration med andre sundhedsrelaterede data-kilder, herunder data fra klinik (inkl EPJ), registre og anden life science (OMICS) data

AU-RM genomcenter/iSEQ



Mange nationale og internationale samarbejdspartnere og HPC brugere

AU-RM genomcenter/iSEQ



Hovedopgaver:

- HPC og medicinsk bioinformatiske pipelines
- Genomdatabase
- Differentieret brugerflade (bioinf → klinik)
- Service og support (bioinf → klinik)
- Uddannelse
- Forskning

GenomeDK supercomputer (HPC)



Key system designers and administrators:



Estimeret kapacitetsbehov indenfor 5 år

Håndtering, analyse og lagring af sekvensdata på 200.000 hele genomer:

Lager: 40 PB

Beregningskerner: 60.000

GenomeDK etableret 2012, løbende udvidet

Designet og dedikeret til genomics og personlig medicin

Integration med anden sundhedsrelateret Big Data

- klinisk og register data
- anden life science / OMICs data

HPC for al klinisk PM i Region Midtjylland

Fortrinsret og tids-garanti for kliniske analyser

Høj sikkerhed (bl.a. fysisk adgangskontrol, logging, monitorering og vagtordning)

Betydelig og skalerbar kapacitet

- ~3.600 beregningsenheder (~4.800 Dec-2017)
- Paralleliseret filsystem med samlet I/O på 75
- Lagerkapacitet 4.5 PB (~7 PB Dec-2017).
- Expansion X 10 mulig i nuværende rammer
- Backup ved AU-IT
- 260 software pakker installeret
- >65 millioner beregningsjobs siden 2014

350+ brugere ved AU/AUH og nationalt samt for

International HPC hub for
PGC data analysis supported by
NIH/NIMH



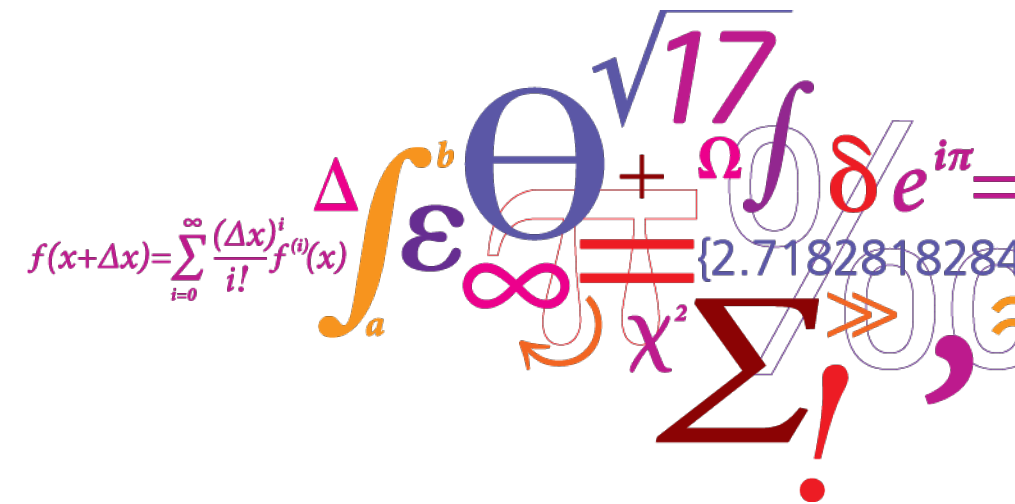
Psychiatric Genomics Consortium



Status og planer for eScience og eInfrastruktur

Forskningsdekan Katrine Krogh Andersen

September 2017



Status for eScience og digitalisering

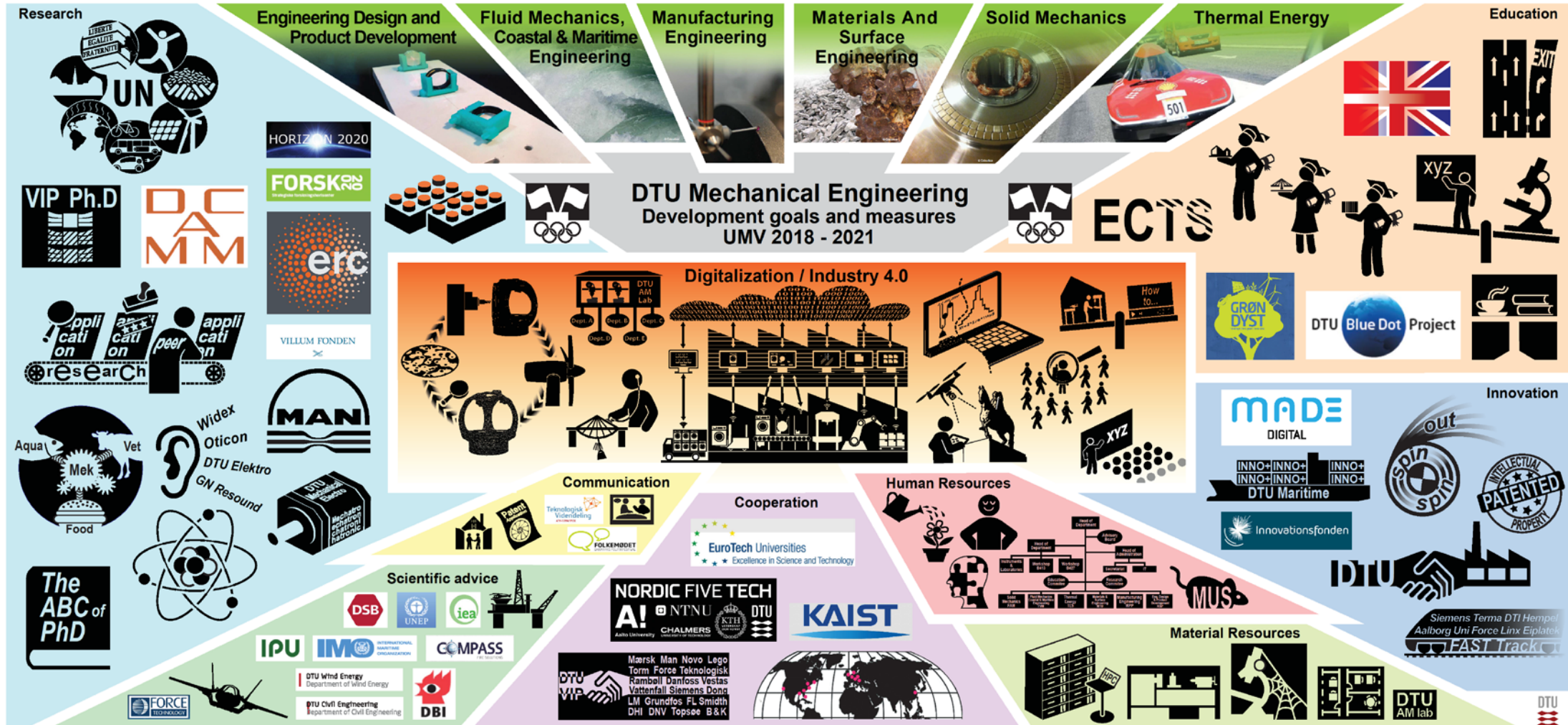
- eScience og digitalisering er på dagsordenen i DTU's strategi og handleplan for 2017
 - Digitalisering 4.0 som særligt fokusområde
 - Illustration af digitalisering som centralt omdrejningspunkt for DTU Mekanik
 - Etablering af Computerome center på DTU
- Generelt eScience Center i samarbejde med DTU Compute
 - *Centralt finansieret indsats på 10 ÅV*
- Forsknings datamanagement forum etableret og politik udarbejdet i 2016:
 - *Politik for opbevaring af primære materialer og data*
- Informationssikkerhed og persondataforordning
 - *Fornyet og øget fokus på håndtering af persondata ifb. digitalisering*

Fra DTU's handleplan for 2017

1.1 Digitalisering 4.0 som særligt fokusområde

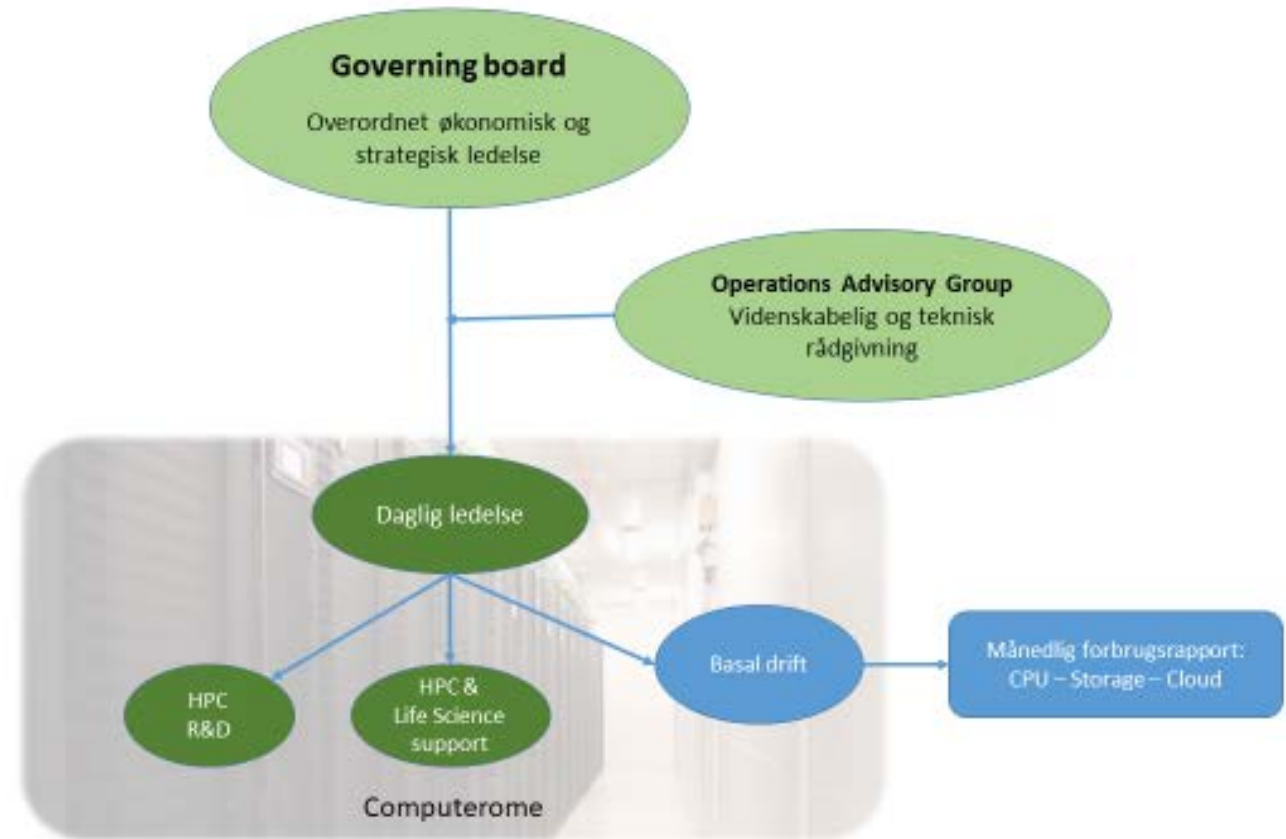
Digitaliseringen skabte mulighederne for den tredje industrielle revolution og åbnede for automatisering og brug af robotter i industriproduktion. Den 4. industrielle revolution bygger videre på dette, dog med den store forskel, at der er koblet kommunikation ind mellem devices; sensorer, aktuatorer og computere er i stadig kommunikation med hinanden. Dette vil dramatisk ændre på ingeniørens arbejdsrum og -opgaver. Internet of Things (IoT) giver mekatronikken en renæssance, og maskin-, elektronik- og IT/softwareingeniører er helt centrale stakeholders i denne transformation. DTU's særlige fokus på digitalisering 4.0 favner denne udvikling.

Digitalisering som omdrejningspunkt



Computerome - National Life Science Supercomputing center, DTU

- Etableret som DTU center 1. juli 2017
- Fortsætte og udbygge det eksisterende samarbejdet med KU og DeiC
- Formålet er at skabe de bedste rammer for et nationalt samarbejde på Life Science området og mellem eksterne parter og institutterne på DTU



Status for eScience og digitalisering

- eScience og digitalisering er på dagsordenen i DTU's strategi og handleplan for 2017
 - Digitalisering 4.0 som særligt fokusområde
 - Illustration af digitalisering som centralt omdrejningspunkt for DTU Mekanik
 - Etablering af Computerome center på DTU
- Generelt eScience Center i samarbejde med DTU Compute
 - *Centralt finansieret indsats på 10 ÅV*
- Forsknings datamanagement forum etableret og politik udarbejdet i 2016
 - *"Politik for opbevaring af primære materialer og data"*
- Informationssikkerhed og persondataforordning
 - *Fornyset og øget fokus på håndtering af persondata ifb. digitalisering*

Status for eInfrastruktur

- Netværk på campus og mellem campusser
 - *Opgradering til nyeste netværksteknologi med fleksibel og sikker netadgang*
- Central storage til generelle data
 - *4 PB i et centralt og redundantt storagemiljø med intern og ekstern adgang*
- Generelle supercomputere til flere brugere
 - *8.500 CPU kerner – heraf 3.400 CPU kerner som også studerende kan anvende*
- Dedikerede supercomputere til specifikke forskningsformål,
 - *DTU Fysik – 11.800 CPU kerner*
 - *DTU Wind – 6.400 CPU kerner*
- Der er begge steder en stigende grad af samarbejde med forskergrupper på andre institutter

Status for eInfrastruktur

- DeIC Nationale LifeScience Supercomputer, DTU
 - *Computerome med 15.000 CPU kerner og 8 PB hurtig disk storage – etableret 2014 og skal derfor snart fornyes*



- Opbygning af avanceret dataplatform til automatiseret til håndtering af analyse af store og forskelligartede datamængder
 - *Proof of Concept med tre pilot projekter fra forskellige områder – samlet indsats 9 MDKK*

Fremtidige behov

- Fastholde løbende udbygning og vedligeholdelse af eInfrastrukturens komponenter, såvel lokalt som i fællesskaber med andre.
- Især fokus på koordinering og samarbejde mellem centrale og decentrale kompetenceenheder og med nationale aktiviteter
- Udvikling af eScience/eInfrastruktur som en service
- Der skal være plads til store behov – såvel beregningsmæssigt som storagemæssigt
- Øget satsning på anvendelse af modellering og analyse indenfor alle områder
- Øget behov for adgang til modellerings-, dataanalyse- og programmerings- kompetencer i et tæt samarbejde med den faglige forskning
- Open Data og Open Science skal håndteres
- DTU bidrager gerne aktivt til en national strategi for udvikling og anvendelse af eScience

DIGITALISERING PÅ AAU

26. SEPTEMBER 2017
DEKAN HENRIK PEDERSEN
DET TEKNISKE FAKULTET FOR IT OG DESIGN



AGENDA

- Status
- IT-plattform til forskningsområdet
- Samarbejde med SDU og AU
- Forslag til national indsats
- Anbefaling



STATUS

- Der er udarbejdet en bred digitaliseringsstrategi på AAU: Uddannelse, PBL, Forskning, Viden-samarbejde, Administration og Service.
- Fokus på både Datamanagement inkl. GDPR og på eInfrastruktur.
- Styres af det Strategiske Digitaliseringsråd; ledes af et direktionsmedlem. Repræsentanter fra fakulteter, IT- og Økonomiafdeling.
- Engagement i Abacus og opbakning til foreløbig rapport om Personlig medicin inkl. roller for Abacus, Computerome og AU-computer

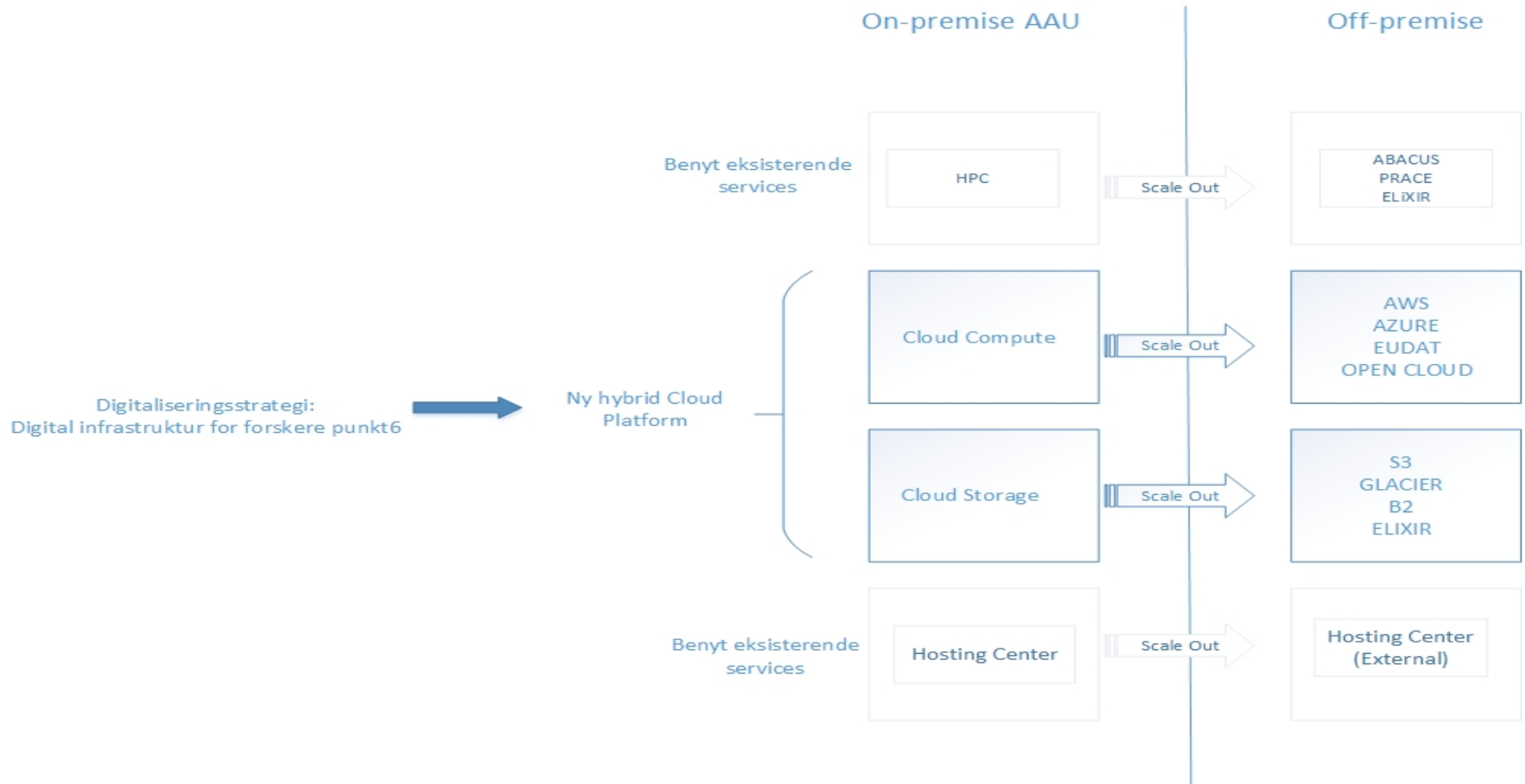


STATUS

- Stærkt fokus på kompetenceopbygning, samarbejde og rådgivningsfunktion, men ikke en tydelig satsning på eScience i ansættelsesstrategi
- Få mindre "HPC-anlæg"
- Storage efterlyses af flere forskere
- En tydelig implementeringsplan mangler, og tydelig stillingtagen til økonomi og ambition fra direktionen udestår
- Samarbejde med SDU om Storage-system, men uklarhed om konvergens i samarbejdet



IT-PLATFORM TIL FORSKNINGSSOMRÅDET



SAMARBEJDE MED SDU OG AU...

- AAU deltager på Abacus
- AAU har bedt om at blive optaget i styregruppen for Abacus og "B2SHARE" – og SDU har indvilget
- AAU og SDU samarbejder om at bygge data-storage: AAU har især bidraget til basisplatform – SDU arbejder videre ovenpå dette med iRODS i lighed med EUDAT B2SHARE
- AAU, AU og SDU har et fælles DelC projekt National Science App Store



FORSLAG TIL NATIONAL INDSATS

- BUDGET: Årligt 120 mio. kr. fra DU og UFM
- HPC: 2-3 nationale Tier2-anlæg
- Gratis adgang for alle forskere
 - op til et basisniveau af timer
- Adgang via Peer Review til yderligere timer
- Adgang for forskere og virksomheder via transparent betalingsmodel
- Lokale eScience-kompetencecentre og Lokale DM Fora
 - begge med mulighed for faciliteret national vidensdeling
- Et Nationalt Digitalt Bibliotek med metadata på alle danske forskningsdata



FORSLAG TIL NATIONAL INDSATS

- National koordinering af internationalt samarbejde (PRACE, RDA, EUDAT, EOSC,)
- Bestyrelse med betydelig ejerskab fra DU og UFM (og DEFF) – bred repræsentation fra DU: HUM, SAMF, NAT, TEK OG SUND...
- Direktør ansat af bestyrelsen
- En smal arbejdende direktion med stærke kompetencer på HPC og Storage – suppleret med kompetencer fra DEFF til DM
- Netværk med betaling som hidtil og med fokus på hovedopgaven vedr. forskningsnettet: Sikker, hurtig og billig netværksforbindelse på tværs af landet og ud af landet via NORDUnet



ANBEFALING

- De gamle dage med DCSC var en god start – men nu skal vi videre sammen med resten af verden og i sync med EU
- Nogle forskere på danske universiteter føler ikke medejerskab til DeICs dispositioner – lederne skal tage medejerskab nu!
- EOSC, OPEN SCIENCE, FAIR DATA, GDPR OG PERSONLIG MEDICIN tilsiger, at der skal rykkes nu – både lokalt OG nationalt

DERFOR: TAG STYRINGEN, DISPONER BUDGET OG LØFT BLIKKET OG BIDRAG TIL EN NATIONAL VISION FOR eSCIENCE



TAK FOR OPMÆRKSOMHEDEN!

**DEKAN HENRIK PEDERSEN
DET TEKNISKE FAKULTET FOR IT OG DESIGN**



AALBORG UNIVERSITY
DENMARK

Status for initiativer og udfordringer omkring anvendelsen af eScience

Det Tekniske Fakultet, SDU

Det Naturvidenskabelige Fakultet, SDU

DeiC konference
26. september 2017

Martin Zachariasen, dekan, NAT, SDU



Status for organisering – SDU eScience Center

SDU eScience Center blev etableret i foråret 2017

Centeret er tværfakultært og akademisk ledet (med professor Claudio Pica som centerleder) og har en styregruppe med repræsentanter fra alle fakulteter

Visionen for SDU eScience Center er at skabe én fælles indgang på SDU til eScience og eScience infrastruktur

Centeret er en kompetence- og forskningsbaseret organisation, som har systemejerskabet for fælles eScience infrastrukturer på SDU

Nuværende eScience infrastruktur på SDU

Beregningskapacitet leveres fortrinsvis igennem DeiC nationale HPC center på SDU (supercomputeren Abacus 2.0 med 14.016 kerner fordelt på 3 arkitekturer samt 1 PB centralt datalager)

Abacus 2.0 har aktuelt 400 brugere fordelt på alle danske universiteter og en bred vifte af fagområder; dog udgør SDU og naturvidenskab den største brugerskare

På TEK/NAT findes lokalt nogen beregnings- og datalagringskapacitet, som er knyttet til bestemte forskergrupper, men meget få forskere har brug for fysisk adgang til servere

Udviklingen går mod at konsolidere brugerne på færre og større systemer

Status for behov og adgang til eScience infrastruktur (TEK/NAT)

Beregningskapacitet

Behovene er aktuelt godt dækket ved brug af Abacus 2.0; finansieringsmodellen og usikkerhed i forhold til videreførelse volder dog bekymring

Datalagringskapacitet

Stor efterspørgsel efter bedre datalagringsystemer og yderligere kapacitet; håndtering af personfølsomme oplysninger en særlig udfordring

Kompetencebehov

Behov for yderligere understøttelse og kompetenceudvikling på TEK/NAT, og et endnu større behov på andre fagområder

Fagområder med stort behov for beregnings- og datalagringskapacitet (TEK/NAT)

Følgende fagområder har aktuelt et stort behov for **beregningskapacitet**:

- materialeforskning
- biofysik
- kvantekemi
- teoretisk højenergifysik
- machine learning / signal processing

Følgende fagområder har aktuelt et stort behov for **datalagringskapacitet**:

- biokemi og molekylærbiologi (proteomics, genomics m.v.)

Eksempler på fagområder i vækst (TEK/NAT)

Følgende fagområder udviser en særlig stor vækst i forhold beregnings- og datalagringskapacitet:

- bioinformatik
- imaging
- computer vision / augmented reality (AR)
- big data analytics / machine learning

Estimeret behov frem mod 2022 (TEK/NAT)

Beregningskapacitet

Behov for op mod 25.000 kerner med en særlig vækst inden for dataanalyse og life science; et mindre antal arkitekturer (3-4) vurderes at kunne dække behovet for beregningskapacitet på tværs af alle fagområder

Datalagringskapacitet

Behov for op mod 5 PB med god understøttelse af data management og håndtering af personfølsomme data; igen forventes den største vækst inden for life science (proteomics, genomics m.v.)

Kompetencebehov

Behov for forskningsopbygning og kompetenceopbygning på tværs af TEK/NAT

Adgang til eScience infrastruktur fremadrettet

Tre typer af adgang bør understøttes i et fremtidigt nationalt samarbejde:

A. Fri adgang til basisydelser

Gratis adgang til basisydelser i en national eScience infrastruktur, fx adgang til beregnings- og datalagringskapacitet op til et vist niveau

B. Adgang via merit

Ansøgning om adgang til beregningskapacitet/datalagringskapacitet i konkurrence med andre forskere; ansøgere og projekter vurderes fagligt

C. Adgang via betaling

Adgang på vilkår der ligner dem de kommercielle udbydere anvender

Et bud på en national vision for eScience: *forskerperspektiv*

1. Alle forskere ved de danske forskningsinstitutioner har (sikker) adgang til tidssvarende national eScience infrastruktur via en fælles indgang/hub
2. Alle forskere har gratis og uhindret adgang til basisydelser
3. Adgang til ekstreme ressourcer sker med udgangspunkt i faglig merit eller betaling
4. Infrastrukturen opbygges centralt, mens kompetencer og support opbygges lokalt
5. Forskere inddrages i udviklingen af den nationale infrastruktur

Et bud på en national vision for eScience: *institutionsperspektiv*

1. De danske forskningsinstitutioner ser det som et fælles ansvar at få opbygget en tidssvarende national eScience infrastruktur
2. Forskningsinstitutionerne prioriterer eScience infrastruktur ledelsesmæssigt og økonomisk
3. Adgang til eScience infrastruktur skal ikke være en intern konkurrenceparameter i Danmark
4. Økonomiske mellemværender klares fortrinsvis mellem institutionerne, dvs. udenom forskerne
5. Samarbejdet forankres stærkt ledelsesmæssigt med transparent/gennemskuelig ledelse og langsigtet økonomi

A U

AARHUS
UNIVERSITET

STATUS FOR E-SCIENCE

SCIENCE AND TECHNOLOGY AARHUS UNIVERSITY

A U

AARHUS
UNIVERSITET
SCIENCE AND TECHNOLOGY

SØREN RUD KEDING
PRODEKAN

25. SEPTEMBER 2017

eScience at Aarhus University 1963-2017→



GIER-RECAU



CSCAA- (ST/ HE/ BSS/ ARTS)

Common themes during 50 years: National planning and local machines, nearness to research environments, technology moves faster than politics, financing,.....

Research Highlights

Quantum Chemistry, Surface Catalysis, Bioinformatics, Quantum Computing, Climate Change, Big Data, Circular Bio Economy, Algorithms, Structural Engineering, Cryptography, Environmental Science, Materials, Image Analysis, Stellar Astronomy, Structural Biology, ESS/ MAX IV, Molecular Dynamics.

National Research Council Centers (DG), Villum Centers, ERC, Lundbeck ST Centers: iMAT, DIGIT, WATERTEC, CBIO,



22 September 1995

Chemical Physics Letters 243 (1995) 409–418

CHEMICAL
PHYSICS
LETTERS

The second-order approximate coupled cluster singles and doubles model CC2

Ove Christiansen, Henrik Koch, Poul Jørgensen

Department of Chemistry, Aarhus University DK-8000 Aarhus C, Denmark

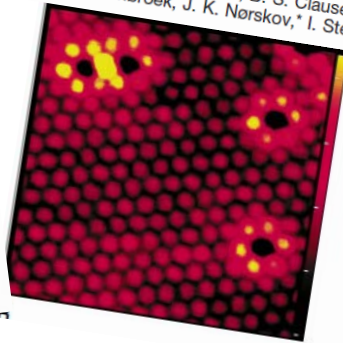
Received 22 May 1995; in final form 14 July 1995

Abstract

An approximate coupled cluster singles and doubles model is presented, denoted CC2. The CC2 total energy is of second-order Møller-Plesset perturbation theory (MP2) quality. The CC2 linear response function is derived. Unlike MP2, second-order Møller-Plesset perturbation theory (MP2) quality. The CC2 linear response function is derived. Unlike MP2, excitation energies and transition moments can be obtained in CC2. A hierarchy of coupled cluster models, CC2, CC3, CCSD, CC3, CCSDT etc., is presented where CC2 and CC3 are approximate coupled cluster models defined by similar approximations. Higher levels give increased accuracy at increased computational effort. The scaling of CC2, CC3, CCSD, CC3 and CCSDT is N^4 , N^5 , N^6 , N^7 and N^8 , respectively where N is the number of orbitals. Calculations on Be, N_2 and C_2H_4 are performed and results compared with those obtained in the second-order perturbation propagator approach SOPPA.

Design of a Surface Alloy Catalyst for Steam Reforming

F. Besenbacher, I. Chorkendorff, B. S. Clausen, B. Hamme, A. M. Molenbroek, J. K. Nørskov, * I. Stensgaard



RNA Exosome Depletion Reveals Transcription Upstream of Active Human Promoters

Pascal Preker,¹ Jesper Nielsen,² Susanne Kammler,^{1*} Søren Lykke-Andersen,¹ Marianne S. Christensen,² Christophe K. Mapendano,² Mikkel H. Schierup,² Torben Heick Jensen

1 MARCH 1999

PHYSICAL REVIEW LETTERS

Quantum Computation with Ions in Thermal Motion

Anders Sørensen and Klaus Molmer
Institute of Physics and Astronomy, University of Aarhus, DK-8000 Aarhus C, Denmark
(Received 26 June 1998; revised manuscript received 25 November 1998)

We propose an implementation of quantum logic gates via virtual vibrational excitations in an ion trap quantum computer. Transition paths involving unpopulated vibrational states interfere destructively to eliminate the dependence of rates and revolution frequencies on vibrational quantum numbers. As a consequence, quantum computation becomes feasible with ions whose vibrations are strongly coupled to a thermal reservoir. [S0031-9007(99)08589-0]

The Influence of Late Quaternary Climate-Change Velocity on Species Endemism

B. Sandel,^{1,2*} L. Arge,² B. Dalsgaard,² R. G. Davies,⁴ K. J. Gaston,² W. J. Sutherland,⁴ J.-C. Svenning
The effects of climate change on biodiversity should depend in part on climate displacement rate (climate-change velocity) and its interaction with species' capacity to migrate. We estimated Late Quaternary glacial-interglacial climate-change velocity by integrating macroclimatic shifts since the Last Glacial Maximum with topoclimatic gradients. Globally, areas with high velocities were associated with marked absences of small-ranged amphibians, mammals, and birds. The association between endemism and velocity was weakest in the highly vagile birds and strongest in the weakly dispersing amphibians, linking dispersal ability to extinction risk due to climate change. High velocity was also associated with low endemism at regional scales, especially in wet and seasonal regions. Overall, we show that low-velocity areas are essential refuges for Earth's many small-ranged species.

SCIENCE & TECHNOLOGY

- 12 departments, 2 national centers and 3 schools
- 4 interdisciplinary centers
- 1,700 academic staff
- 1,400 technical-administrative staff
- 750 PhD students
- 7,200 bachelor and master students
- 2,900 MDKK in turnover in 2015
- External funding: approx. 63%(2013)



ST LOCATIONS



STRONG INFRASTRUCTURE



Illustration of the European Spallation Source (ESS) © ESS



AU 2016: DIGITALISATION STRATEGY

DIGITALI- SERINGS- STRATEGI

Digitisation in the area of research: Aarhus University must establish, develop and maintain digital facilities and expertise in order to maintain and consolidate its position as an internationally leading research university and its strong engagement in the development of society.

Digitisation in the area of education: Aarhus University must establish, develop and maintain digital facilities and competencies in order to increase use of Educational IT in its efforts to offer degree programmes of the highest international standard.

The university's digital study and work environment: Aarhus University will offer a digitally supported, attractive and inspiring environment for work and study which attracts and retains the best students and employees.

Digitisation of the university's administration: Aarhus University's digital solutions must contribute to efficient, cohesive and professional administration, in order for students and employees to maintain their focus on the university's core tasks.

AU 2016: DIGITALISATION STRATEGY

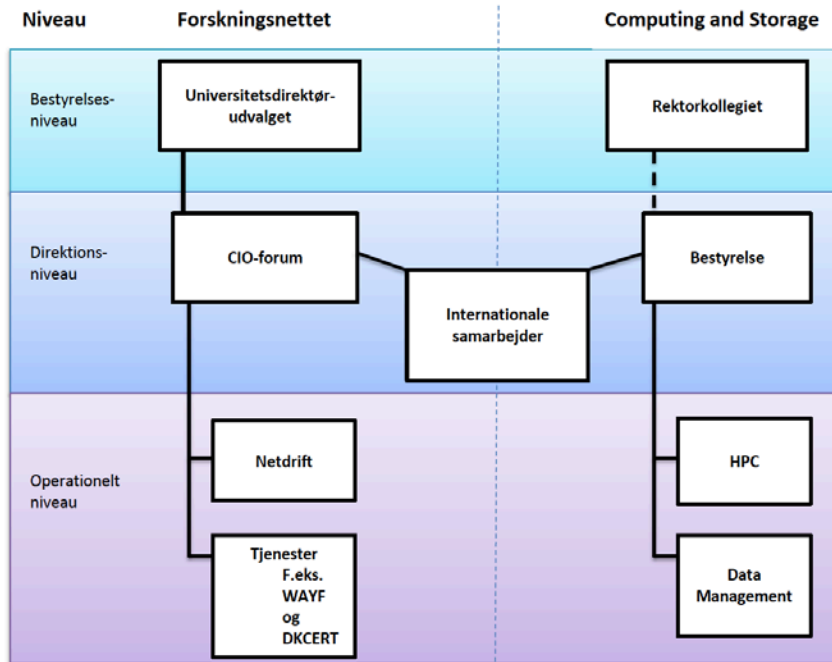
DIGITALI- SERINGS- STRATEGI

Digitisation in the area of research:

- HPC
- Data Management

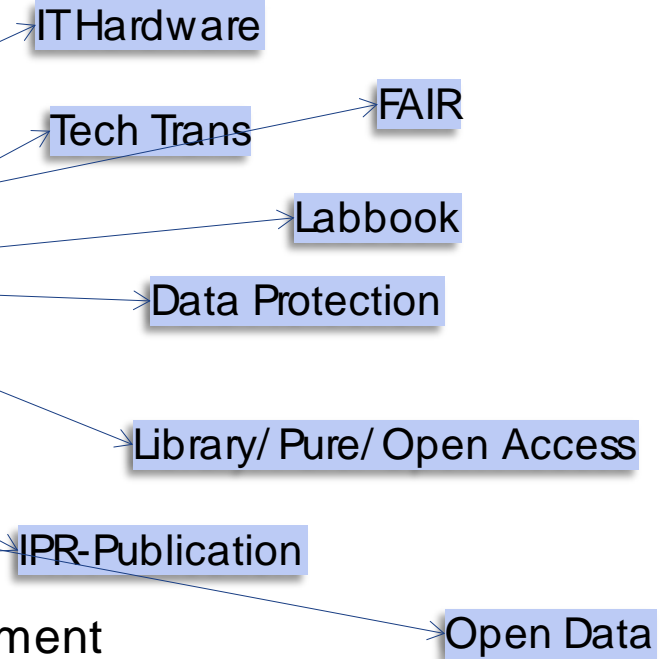
Aligned with Universities in Denmark

UNIVERSITIES IN DENMARK AND AU



- **Researchers** must have access to **e-infrastructure** to promote **excellent research**;
- If possible, **parallel investments** in costly e-infrastructure at Danish universities **should be avoided**
- **Transparency in governance**
- **Stable financing** to ensure a basis for **continued upgrading** and a stable and competent staff
- **Computing and Storage:**
 - In future, collaboration on **Computing and Storage** should be **managed by a board of directors** with representatives from the **management of the Danish universities**
 - The funding of Computing and Storage should derive from the **National budget**. A **main task of the new board of directors must be to raise more funding** –both external funding and from the universities themselves

AU Data Research Management:



Many organisations contribute to data management

AU Data Research Management:



Data Management

IT Hardware

FAIR

Tech Trans

Labbook

Data Protection

Library/ Pure/ Open Access

IPR-Publication

Open Data

.....etc

For the researcher Data Management is a singular task

AU-HPC STATUS AND OVERVIEW

1985-->2001

UNI-C

Mostly DTU, smaller machines in Aarhus

2001-->2013

DCSC

National coordination and local DCSC centers at DTU, KU, AAU, SDU
15 Mkr/year to hardware.

2002-->

CSCAA

AU's DCSC hub

External hardware funding and 3Mkr ST contribution in 2016

2013-->2018

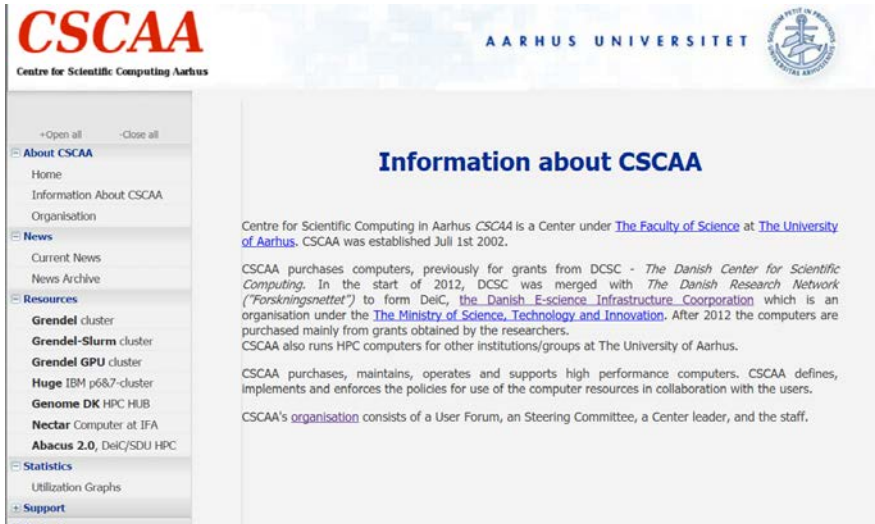
DeiC

National HPC Center at

ABACUS2.0 @ SDU

Computerome @ DTU/Risø

Kulturarvsclu. @ Det Kongelige Bibl./Aarhus



The screenshot shows the CSCAA website interface. At the top, there is the CSCAA logo (Centre for Scientific Computing Aarhus) and the Aarhus University logo. Below the logos, there is a navigation menu with categories: About CSCAA (Home, Information About CSCAA, Organisation), News (Current News, News Archive), Resources (Grendel cluster, Grendel-Slurm cluster, Grendel GPU cluster, Huge IBM p687-cluster, Genome DK HPC HUB, Nectar Computer at IFA, Abacus 2.0, DeiC/SDU HPC), Statistics (Utilization Graphs), and Support. The main content area is titled "Information about CSCAA" and contains text describing the center's role, funding, and organization.

CSCAA
Centre for Scientific Computing Aarhus

AARHUS UNIVERSITET

Information about CSCAA

Centre for Scientific Computing in Aarhus. CSCAA is a Center under [The Faculty of Science](#) at [The University of Aarhus](#). CSCAA was established July 1st 2002.

CSCAA purchases computers, previously for grants from DCSC - *The Danish Center for Scientific Computing*. In the start of 2012, DCSC was merged with *The Danish Research Network ("Forskningsnettet")* to form DeiC, the [Danish E-science Infrastructure Cooperation](#) which is an organisation under the [The Ministry of Science, Technology and Innovation](#). After 2012 the computers are purchased mainly from grants obtained by the researchers. CSCAA also runs HPC computers for other institutions/groups at The University of Aarhus.

CSCAA purchases, maintains, operates and supports high performance computers. CSCAA defines, implements and enforces the policies for use of the computer resources in collaboration with the users.

CSCAA's [organisation](#) consists of a User Forum, an Steering Committee, a Center leader, and the staff.

AU-HPC STATUS AND OVERVIEW

DATA DRIVEN:

- iSEQ(HE)
- GenomeDK
- BIRC
- MADALGO
- Regional County (RM) and AUH
- Industry and IP (Apple)

CPU DRIVEN

- >250 users
- GRENDel and ABACUS
- Accessibility
- Research training
- Software development



AU HPC:

- HPC is a key element in AU's research strategy
- HPC pervades all areas of research at Science and Technology
- Accessibility is priority no. 1
 - *Data-CPU-Software-Rules and regulations-Teaching*
- HPC technology moves fast and requires agile organizations
- Aligned strategies: AU and Universities in Denmark

A U

AARHUS
UNIVERSITET

eSCIENCE ved SCIENCE/KU

Det Natur- og Biovidenskabelige Fakultet

KØBENHAVNS UNIVERSITET



Key figures for SCIENCE

- 4,500 full-time equivalents (academic staff + technical and administrative staff)
- 9,500 full-time BSc and MSc students
- 21 BSc programmes and 35 MSc programmes
- 1,150 PhD students
- 16 active Centers of Excellence, 24 ERC grants with scientists from SCIENCE as the Principal Investigator (PI) in FP7 and up to now 7 ERC grants in Horizon2020 with researchers from SCIENCE as the PI.

Department of Biology

Department of Chemistry

The Niels Bohr Institute

Department of Computer Science

Department of Mathematical Sciences

Department of Geosciences and Natural
Resource Management

Department of Food Science

Department of Nutrition, Exercise
and Sports

Department of Plant and Environmental
Sciences

Department of Food and
Resource Economics

Department of Science Education

Natural History Museum of Denmark



Det Natur- og Biovidenskabelige Fakultet – SCIENCE Eksempler hvor eSCIENCE er centralt for udviklingen

- Genetik og Bioinformatik
- Generelt billedbehandling
- Machine learning
- Højenergifysik (CERN)
- Materialeforskning – Røntgen, neutroner, e- og lysmikroskopi, NMR) (ESS, MAX-IV, XFEL, ILL, ESRF, PSI, ...)
- Jordobservation
- Klimaforskning
- Astronomi, astrofysik og kosmologi
- Kvantekemi
- Fødevareforskning
- Økonomi- og ressourcevidenskab
- Ernæring, idræt og sport
- Museale samlinger
- Interaktion mellem mennesker og maskine
- Biodiversitet
- eLæring

Status og fremtidige behov

Eksterne netværksforbindelser

- Dedikerede forbindelser med høj kapacitet
- Dedikerede forbindelser med høj sikkerhed

Status og fremtidige behov Datalagerkapacitet

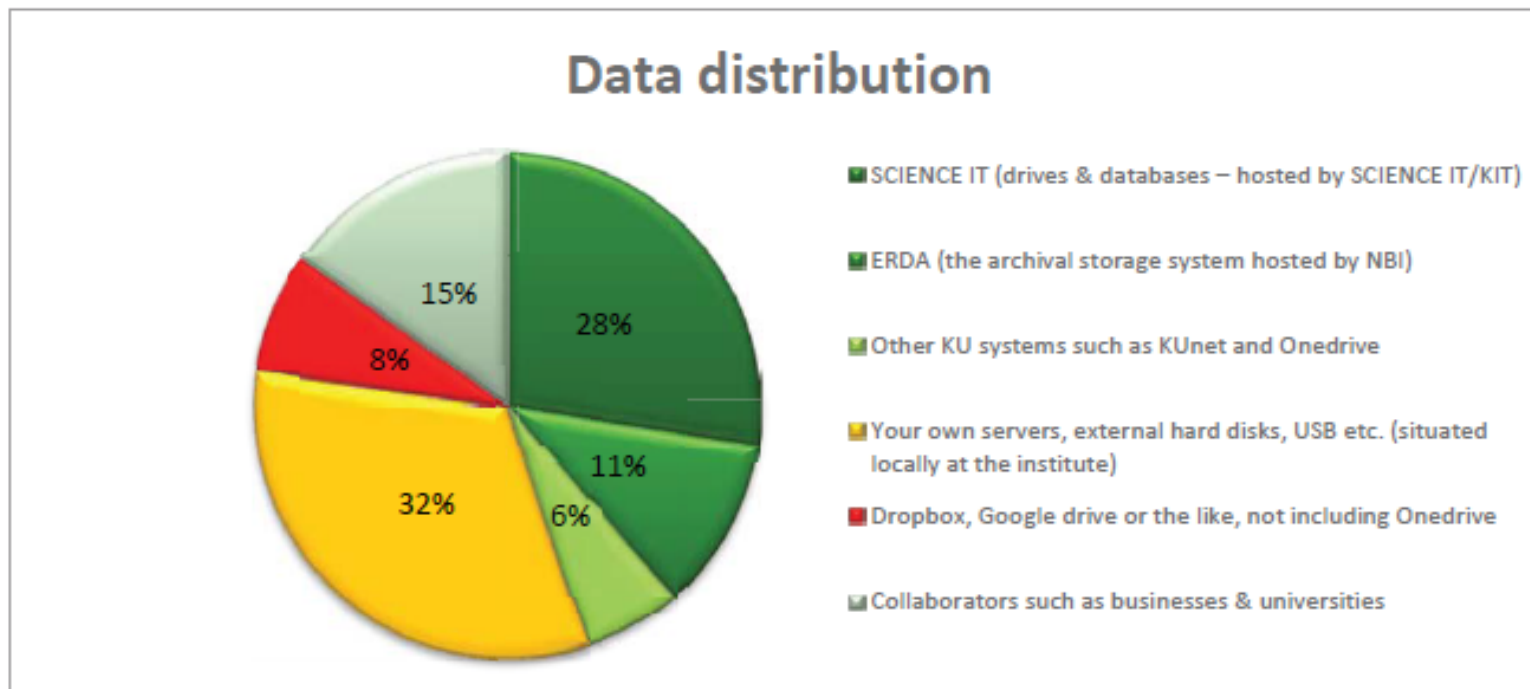
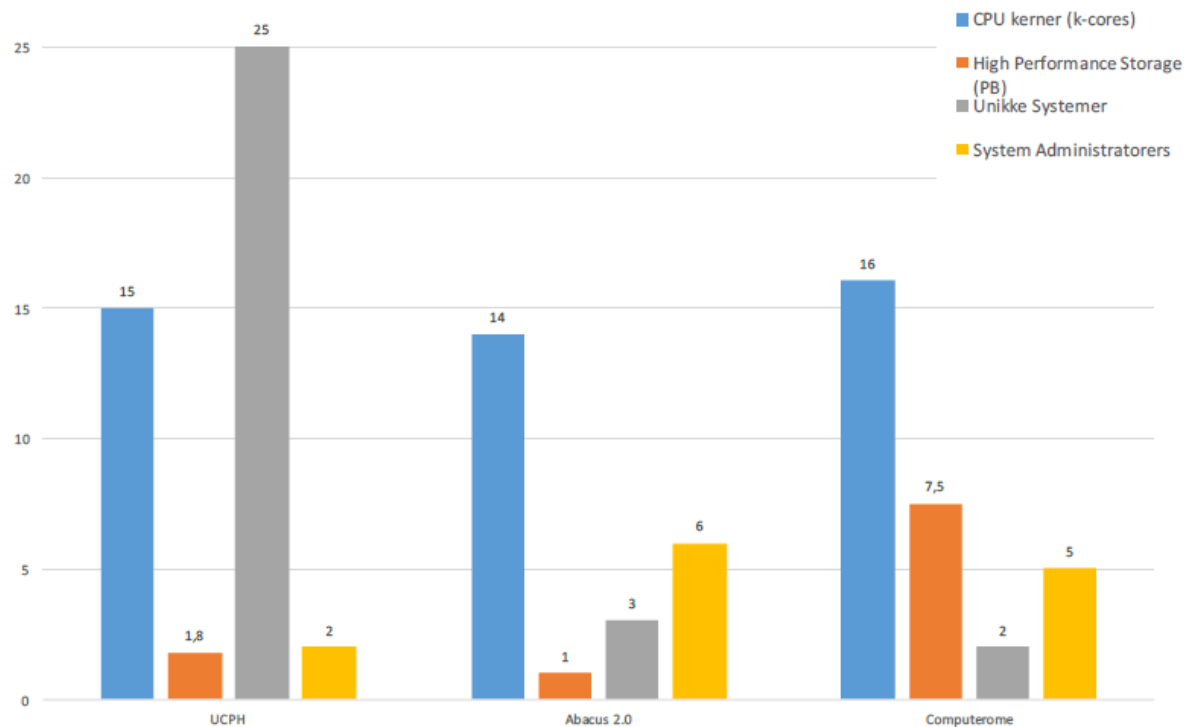


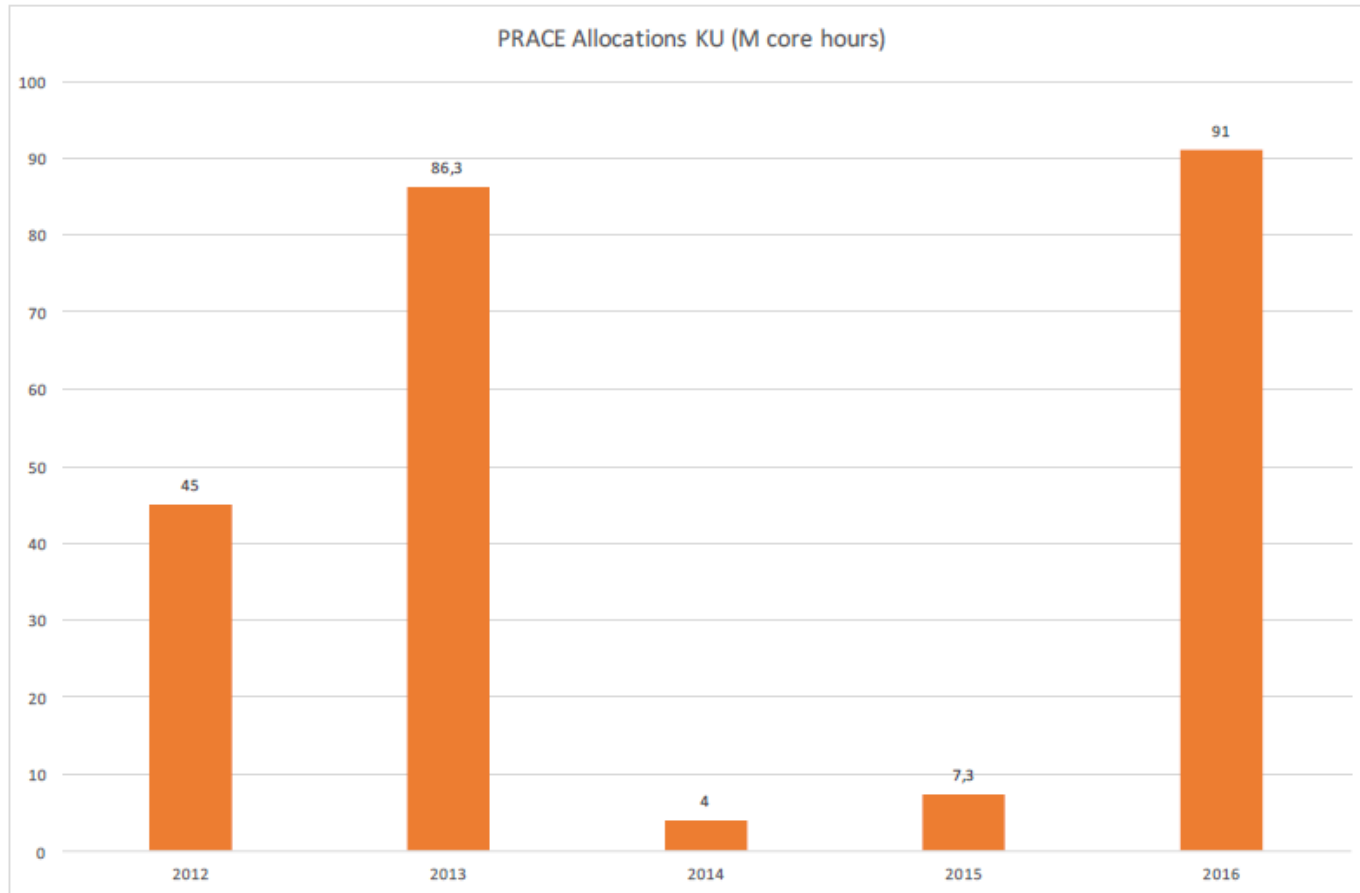
Figure 2. Distribution of approximate amount of data on different classes of data storage systems. Total amount ca 7.3 PB.

Status og fremtidige behov HPC behov



Figur 1. HPC Centerets størrelse sammenlignet med de to nationale centre. UCPH er sammenligneligt i størrelse, men drives af meget færre folk og driver langt flere uafhængige systemer.

KU PRACE allokeringer 2013-2016



Figur 2 KU PRACE allokeringer over de sidste 5 år.

Styring af eSCIENCE processen ved SCIENCE

- Udvalg for dataopvevring
Formand inst. leder prof. Mads Nielsen
- SCIENCE Digitaliseringsråd
Formand dekan prof. John Renner Hansen

SCIENCE Digitaliseringsråd

Dekan John Renner Hansen er formand for Rådet, som tillige har deltagelse af de øvrige medlemmer af fakultetsdirektionen.

Herudover består Rådet af en sammensætning af centrale aktører fra erhvervslivet og den offentlige sektor samt ledelses-, medarbejder- og studenterrepræsentanter fra SCIENCE.

Otte eksterne medlemmer

- Morten Bo Christiansen, Head of Strategy, Transport and Logistics Division, Mærsk
- Torben Dalgaard, direktør for Center for It, Medico og Telefoni, Region Hovedstaden
- Torben Fabrin, IT-direktør, Arla Foods
- Adam Lebech, branchedirektør, DI Digital
- Morten Petersen, bestyrelsesformand i Worksome
- Peter Scharff, Head of Engagement & Digital Natives, Google Danmark
- Hanne Shapiro, tidl. innovationschef ved Teknologisk Institut
- Marianne Sørensen, vicedirektør, Digitaliseringsstyrelsen

Seks interne medlemmer fra SCIENCE

- Hanne Andersen, institutleder på Institut for Naturfagenes Didaktik
- Jette Fugl, informationsspecialist på Københavns Universitetsbibliotek Frederiksberg
- Jeppe Hagedorn Hansen, studerende
- Peter C. Kjærgaard, museumsdirektør på Statens Naturhistoriske Museum
- Mads Nielsen, institutleder på Datalogisk Institut
- Brian Vinter, professor ved Niels Bohr Institutet

Formål med de strategiske satsninger vedr. digitalisering på SCIENCE

At sikre, at SCIENCE på et velfunderet grundlag... :

- igangsætter initiativer, som kan understøtte visionen for digitalisering på SCIENCE.
- har fokus på såvel kortsigtede udviklingsaktiviteter, der hurtigt kan skabe resultater, som større satsninger, der kræver mere langsigtede investeringer.
- bidrager til at understøtte og imødekomme de behov, som opstår i samfundet som følge af den digitale udvikling.
- bidrager til at gøre Danmark til et internationalt digitalt hot spot.

**SCIENCE Digitaliseringsråd fungerer som rådgivende panel for direktionen
vedr. den strategiske vision om digitalisering på SCIENCE.**

Fire større satsninger igangsat i 2017

På baggrund af anbefalinger fra SCIENCE Digitaliseringsråd er følgende fire strategiske satsninger vedr. digitalisering på SCIENCE igangsat i 2017:

1. Flere kandidater med digitale kompetencer

Styrkelse af de fagfaglige digitale kompetencer for alle SCIENCE's uddannelser

Kontakt: Grete Bertelsen, prodekan for Uddannelse (overordnet ansvarlig)
Martin Lillholm, viceinstituttleder på DIKU (tovholder med ansvar for satsningens gennemførelse)

2. Digitalisering af Danmarks Nationale Naturhistoriske Samlinger

En unik, national naturhistorisk samling i dag. En åben, global forskningsinfrastruktur i fremtiden

Kontakt: Morten Pejrup, prodekan for forskning (overordnet ansvarlig)
Peter C. Kjærgaard, museumsdirektør, SNM (tovholder med ansvar for satsningens gennemførelse)

3. Bedre mulighed for datahåndtering på SCIENCE


Simpel og sømløs adgang til dataanalyse for alle

Kontakt: John Renner Hansen, dekan (overordnet ansvarlig)
Brian Vinter, professor på NBI (tovholder med ansvar for satsningens gennemførelse)

4. Etablering af København som International Digital Hub inden for Kunstig Intelligens

Styrkelse af brugen af kunstig intelligens i alle SCIENCE's faglige områder

Kontakt: Erik Bisgaard Madsen, prodekan for erhvervs- og myndighedssamarbejde (overordnet ansvarlig)
Mads Nielsen, instituttleder på DIKU (tovholder med ansvar for satsningens gennemførelse)



AAU HUM
STATUS. BEHOV. BEGRÆNSNINGER
DEKAN HENRIK HALKIER



AALBORG UNIVERSITET

Status – E-Science på HUM-AAU

- E-Science bredt begreb – her særligt fokus på ‘labs’ og ‘data’ (video, sensor-data, eye-tracking m.m.)
- Labs til studier og udvikling af ‘menneskelig praksis’ (læring, oplevelse, teknologi-brug)
- Mange labs med avanceret udstyr – VR, 3D print, 360 kameraer, Eye-tracking, Robotics, volumetric scanning
 - VILA, X-lab, RELATE, Design Lab
- Store, komplekse, blandede typer af data (fx sensor data + video data + location data)
- Eksperimenter med cutting-edge teknologier til indsamling af kvalitative data
- Internationale samarbejdspartnere og samarbejdsprojekter



BEHOV

Data-management/Netværk: Transport, opbevaring og behandling af data

- Fx Video Data kræver sikker langtidsopbevaring, hurtig transport, analyse og behandling
- Modeller og processer for forskeres opbevaring, genfindning og deling af data – (DeIC pilotprojekt: **Video Life Cycle Data Management**)
- Lagring af multi-data sæt Video + location + bio-data

HPC

- Rendering af video fra flere kilder, batch-opgaver, video-konvertering
- Arbejde med store mængder af fx video er processor-krævende
- Algoritmer til gennemgang af data (fx udvælgelse af sekvenser i stort korpus)

Analyseværktøjer

- Analyseværktøjer er rettet mod mono-data – behov for værktøjer til multimodale analyser af forskelligartet data



BEGRÆNSNINGER

- Der eksisterer ikke nu en infrastruktur, der nemt kan håndtere transport, opbevaring og behandling af data – pilotprojekter, diversitet
- Kompetenceudvikling – svært at ændre eksisterende praksis
 - Brug for lokale ‘boundary brokers’ – ‘tekniske forskere’
 - Formidler til andre, hvordan nye teknologier kan give nye indsigter
- Forskere travle. Vil forske!
 - Ikke håndtere og administrere data og data-management planer
 - Er svære at træne/uddanne – vil først når de står med udfordringen
- Open Science udfordring – mennesker og ikke klima-data
 - HUM-data ofte sensitive, personhenførbare – kan ikke bare deles offentligt
 - Stor interesse for data-delning og åbne data – men der kan være begrænsninger





Empowering eHumanities

DeiC konference 2017

26. september

Dekan Simon Møberg Torp

Det Humanistiske Fakultet, Syddansk Universitet

Udgangspunkt

Det ”nye” er mængden af digitale data
– *ikke computere og digitalisering*

Teknologien på SDU er i stor udstrækning tilstrækkelig
– *det er ikke manglende soft- og hardware, der bremser udviklingen*

Viden om eScience blandt forskerne er lille
– *løsningen ligger dog næppe i bred efteruddannelse*

Den vigtigste udfordring fremover, hvis vi vil være førende, bliver bekæmpelsen af siloer
– *datasiloer såvel som videnssiloer.*

→ e-infrastruktur skal nedbryde virtuelle mure og understøtte samarbejdskultur



Vision for eHumanities

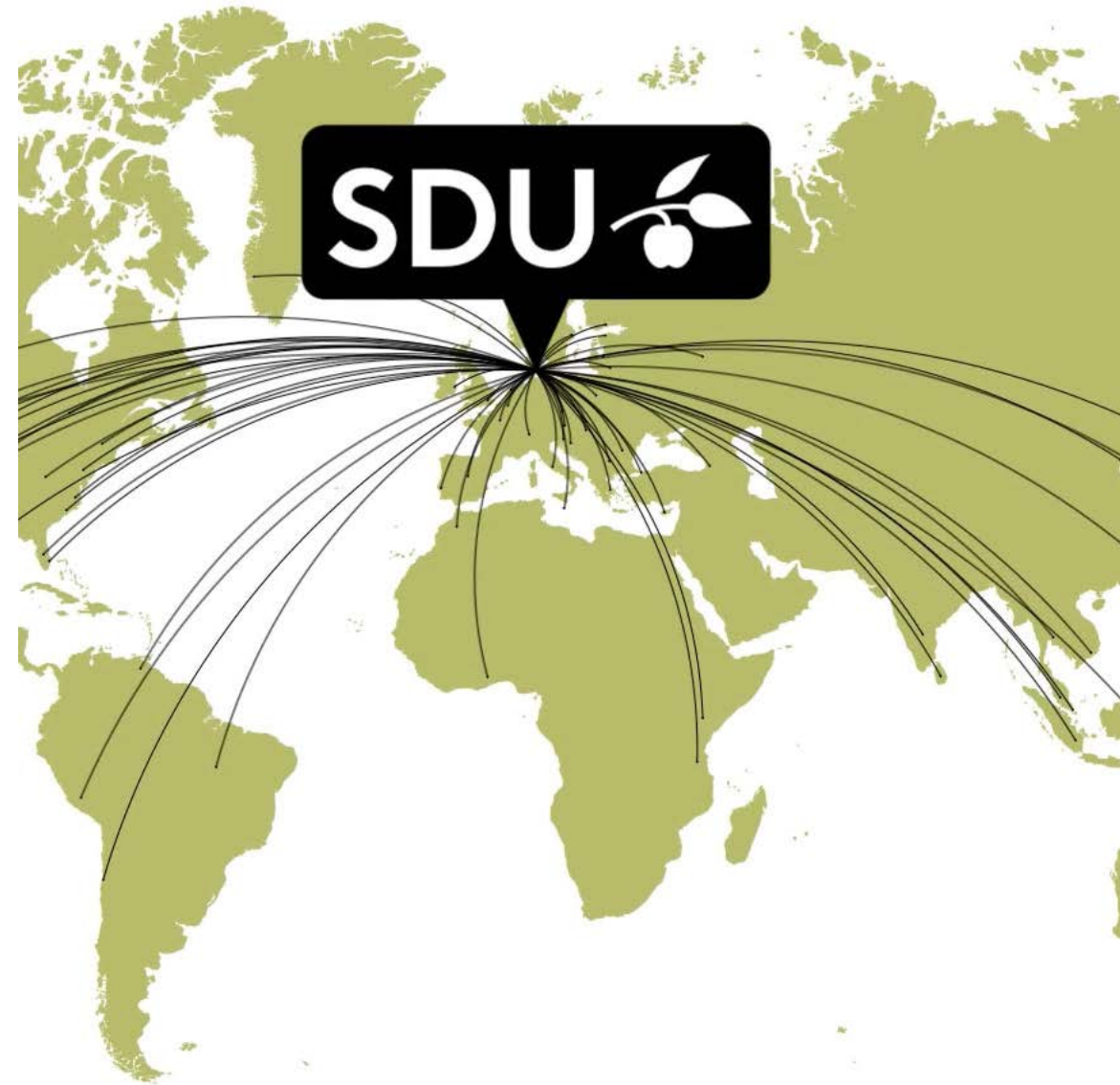
Stærkt lokalt eScience setup

- eScience-center på universitetsniveau og stærkt eHumanities-lab på fakultetsniveau

Koordineret national infrastruktur med servicecentre

- Adgang til ressourcer på tværs af institutioner, fagområder og rettighedshavere, understøttet af specialiserede enheder (alle skal ikke kunne alt)

→ vi skal være førende i Skandinavien



eHumanities på SDU

Empower domain experts

- Opbygning af serviceenhed, der understøtter hurtige prototyper, pilotprojekter på rigtige data og skaber et rum, hvor forskere, der tænker i eHumanities, kan lufte idéer og blive hjulpet videre.

”guerilla”-strategi omkring udbredelse

- Sæt ind, hvor der er vilje og muligheder, og lad resultaterne tale.

Tværinstitutionelle uddannelsesprogrammer

- Sikring af kritisk masse og optimal udnyttelse af ressourcerne ved at løfte uddannelsesopgaverne sammen – på den måde fremmes også samarbejde, netværk og vidensdeling.



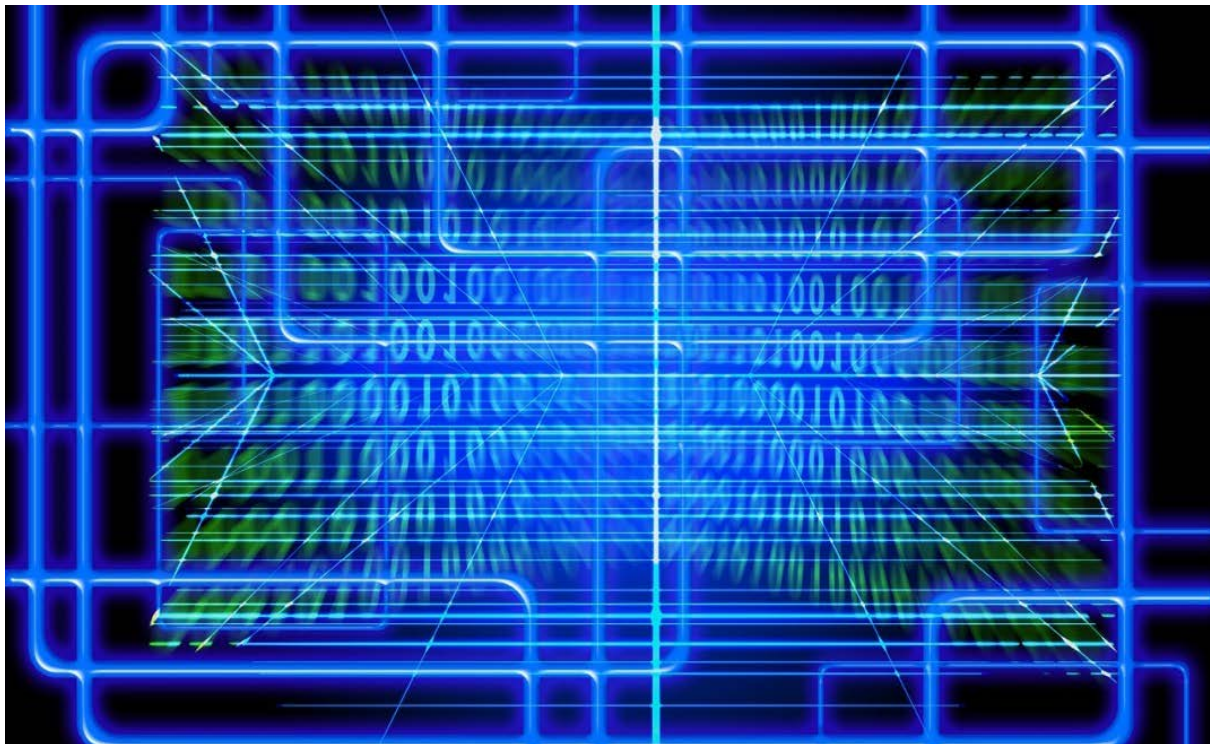


Status for e-science på det samfundsvidenskabelige område (retsvidenskab)

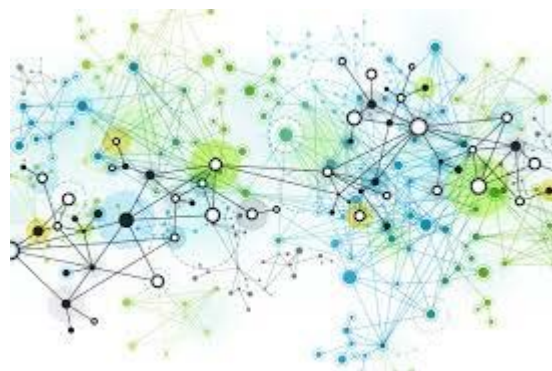
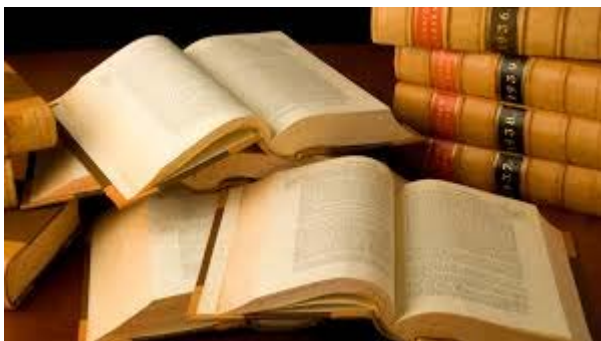
Prodekan for forskning, Henrik Palmer Olsen



e-infrastruktur i retsvidenskaben



Modernisering og kompetenceudvikling af den retsvidenskabelige forskning





Eksisterende adgang til e- forskningsinfrastruktur



Hvordan er retsvidenskabelig e-infrastruktur finansieret





Kan god e-infrastruktur skabe kompetitive fordele for dansk retsvidenskabelig forskning?



E-infrastruktur på CBS: Status og behov

Søren Hvidkjær
Professor, Institutleder
Institut for Finansiering

- Sine qua non datagrundlag
 - registerdata (DST): individer og virksomheder
 - internationale kommercielle databaser
 - i stigende grad: big data
- Hidtil lokale (højst på institutniveau) løsninger på storage, serverkapacitet, adgang til HPC, fildeling etc.
- Research Data Management politik vedtaget 2017
- RDM styrekomite oprettet 2017

- Storage: Behov for systemer som på tilfredsstillende vis leverer på lagerplads, backup, fildeling og adgangskontrol.
- HPC: Abacus fra CBS perspektiv
 - The good: Højtydende system med stort softwareudvalg
 - The bad: Kørselsrunderne kræver meget planlægning eller lang ventetid ift typisk projektkadence
 - The ugly: Lagerplads og behandling af fortrolig data
- Behov for HPC med hurtig adgang, behandling af fortrolig/personfølsom data

- HPC løsning i DST
- Kobling af sundhedsdata og socioøkonomiske registerdata
- Videreudvikling af software til sikker deling af data

eScience

Det Samfundsvidenskabelige Fakultet, SDU

Professor mso Oliver Baumann
Strategic Organization Design og Danish IAS
Institut for Marketing & Management
oliv@sod.dias.sdu.dk / 6550-4433

DeiC konference, Kolding, 26.09.2017

Agenda

Organisation og tilgang

Fokusområder

Udfordringer og behov

Den lokale eScience gruppe på fakultetet er bindeleddet mellem SDUs eScience Center og den enkelte forsker

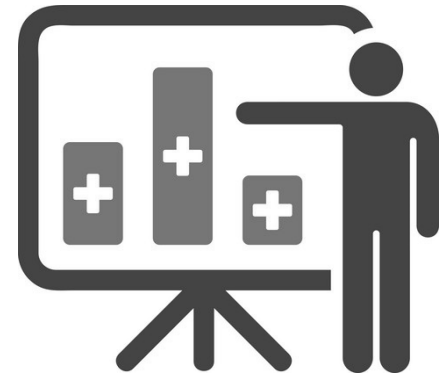
SDU eScience Center

- Har systemejerskab over eScience infrastruktur og services
- Yder brugersupport til forskere på tværs af fakulteter

SDU eScience Center
Abacus 2.0 – DelC National HPC Centre, SDU

eScience gruppen på SAMF

- Består af lederen af IT afdelingen og en repræsentant fra hvert institut
- Medlemmerne fungerer som lokale videnskabelige kontaktpersoner



eScience gruppen på fakultetet fokuserer på at fremme peer-to-peer vidensdeling ved at:

1.

Skabe bevidsthed for eScience bland forskerne

2.

Forstå forskernes behov og udfordringer

3.

Hjælpe forskerne komme i gang med at bruge eScience services

Fakultetets eScience fokusområderne er datalagring og high performance computing

- 1.** Datalagring
- 2.** High Performance Computing
- 3.** Understøttende tjenester
- 4.** Netforbindelse

Effektiv datalagring er relevant for alle, men især for dem, der arbejder med personfølsomme data

Baggrund

- Fakultetet tilstræber skift til non-lokal backup løsning (fx MS OneDrive)
- Persondataforordning kræver særlig lagring og behandling af personfølsomme data

Aktiviteter

- eScience Center er ved at finde en understøttende løsning, som vil sikre overholdelse af datamanagement kravene
- Fakultetet og eScience Center er ved at udarbejde en løsning for at tilgå disse data
- Der er planer om at finde en løsning for både at lagre "almindelige" og personfølsomme data



Antallet af brugere af supercomputeren på fakultetet ligger p.t. på ca. 10, men flere forskere har vist interesse

Anvendelsesområder

- Diskret matematik
- Numerisk analyse og simulation
- Statistisk analyse ("big data")

Fordele

- Højere stikprøvestørrelser
- Flere robusthedstjeks
- Hurtigere resultater og prototypefremstilling
- Højere hastighed og kvalitet



eScience på fakultetet står over for både operative og strategiske udfordringer og behov

1.

Mere præcis afklaring af det faktiske behov for HPC: Nogle beregningsopgaver er svære at parallelisere, og der kan også være nemmere alternativer.

2.

Mange kunne være interesseret i at arbejde med store mængder af data, men for forskere uden en "teknisk" baggrund er hurdlerne høje.

3.

Grundet høj efterspørgsel fra de "tekniske" fakulteter er der ikke nogen incitament til at gøre HPC services mere brugervenlige.

4.

Rent lavpraktisk er det endnu ikke klart, hvilken betydning persondataforordningen kommer til at have og hvordan den bedst skal implementeres.

5.

Større infrastrukturinitiativer (fx DRDS, BICLabs) skulle helst tænkes ind i eScience agendaen for at skabe optimale rammer for forskning på højeste niveau.

DRDS: Danish Research Data for the Social Sciences; BICLabs: Behavior, Interaction and Cognition Labs



Det Samfundsvidenskabelige Fakultet

E-science og databehov

DeIC konference 26/9 2017, Comwell, Kolding

Mette Wier, prodekan for forskning og omverdensrelationer

KØBENHAVNS UNIVERSITET



Ofte anvendte datatyper og særlige udfordringer

Big data, herunder data fra sociale medier

Behov for digital ekspertise ifm. scraping og behandling af big data, (herunder machine learning)

Computerbaserede eksperimentelle opstillinger (fx EEG, øjenbevægelsesmålinger)

Behov for særlige eksperter her.

Ofte anvendte datatyper og særlige udfordringer

Hybriddata: Når egne indsamlede data kombineres med registerdata

Her stilles særlige krav til opbevaring og sikkerhed. Dyrt for de enkelte projekter, mangler fælles infrastruktur

Sensitive/sikkerheds-godkendte data (fx. politidata):

Behov for værktøjer/best practice mht. kryptering.

Registerforskning og andre personhenførbare data

- Persondatalovgivning og persondataforordning administrativt krævende
- Bedre e-infrastruktur i DST:
 - ✓ Fra individuelle servere ➡ samlet større lager- og computerkraft (HPC)
 - ✓ Videreudvikle software ➡ nemmere og mere sikker datadeling.

Generelle ønsker

- Behov for regnekraft og datafaciliteter
- Nem deling af data med andre forskere globalt, fx via Dropbox-løsning til forskere
- Regionernes Genomcenter på DeIC's supercomputer ved Risø: Interesse for kobling af data med socio-økonomiske data i DST.

AARHUS BSS E-SCIENCE BEHOV



INSTITUT FOR VIRKSOMHEDSLEDELSE
AARHUS UNIVERSITET

26. SEPTEMBER 2017

JACOB KJÆR ESKILDSEN
INSTITUTLEDER



DATATYPER MED SÆRLIGE BEHOV

Data der "scrapes" – sociale medier, blogs, intranet etc.

Data fra eksperimenter – højfrekvente psykofysiologiske/sensoriske målinger

Finansielle data – højfrekvente data fra børser og lign.

Data der genereres – Monte Carlo studier

Data der flettes fra flere kilder – registerdata, følsomme virksomhedsdata mm.

Data der er personfølsomme – offentlige registre, surveys, lokale registre, observationsstudier

PERSONFØLSOMME/REGISTER DATA

Behov for et koordineret setup ifm med persondataforordningen

Behov for konsolidering af ressourcerne (lagring/beregning, HPC)

GENERELT

Ønske om regnekraft og lagring (cluster) dedikeret til højereordens programmeringssprog (f.eks. matlab, python, R)

Ønske om nem og sikker måde at dele data med forskere/virksomheder nationalt såvel som internationalt

Opmærksomhed på at de forskellige datatypers betydning vil ændre sig i fremtiden



INSTITUT FOR VIRKSOMHEDSLEDDELSE
AARHUS UNIVERSITET

e-infrastruktur til forskning i verdensklasse i 2025

- en analyse og udredning vedrørende dansk e-infrastruktur samarbejde

Bilag 3b:

Præsentationer ved Rundbordssamtale 5. oktober 2017

eScience status og e-infrastruktur behov fra Tek/Nat området

Møde med SFU 5. oktober 2017

Katrine Krogh Andersen

På baggrund af præsentationer ved DeiC konferencen 26. september



AALBORG UNIVERSITET



AARHUS UNIVERSITET



KØBENHAVNS
UNIVERSITET

Hovedudsagn

- E-science og digitalisering er prioriteret på alle universiteter
- Data management stiller krav til data lagring og data adgang
- Tiltagende behov for data lagring ifht computing, særligt blandt nyere brugergrupper
- Nærhed til forskere og fleksibel adgang væsentlig
- Behov for specific arkitektur og regnekraft på visse områder
- Samtidig tiltagende behov for som data lagring og computing som "service" med nem for adgang og evt tilkøb af kompetencer
- Mange berøringsflader til internationale dagsordener og sundhedsområdet, særligt personlig medicin

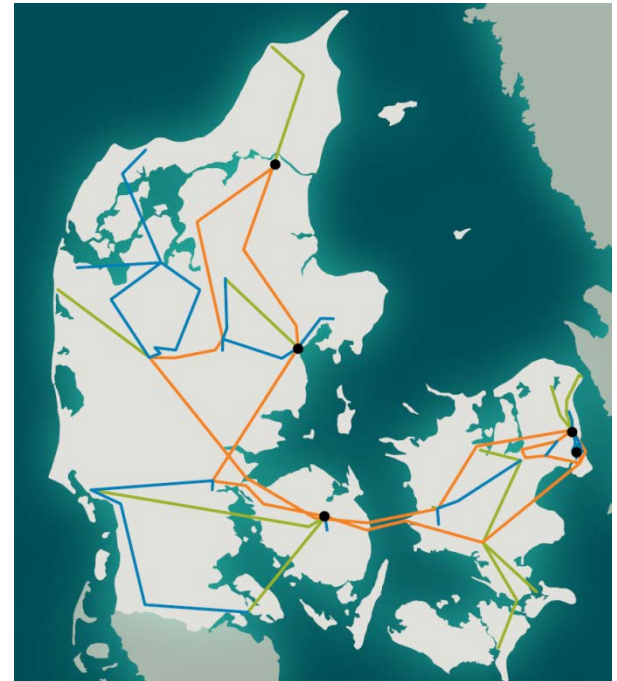
Væsentlige elementer

- Forskningsnet
- Data lagring
- Computing og HPC infrastruktur
- Governance og prioritering
- E-science kompetencer
- Sikkerhed
- Finansiering



Forskningsnettet

- Grundlæggende behov for alle
- Skal fortsat organiseres og betales fælles
- Høj kapacitet, hastighed og sikkerhed vigtige



Data lagring

- Tæt koblet til data management arbejdet
- Stærkt stigende behov for kapacitet
- Sammenhæng med Open Science og Open Data dagsordener

HPC infrastruktur

- Forskellige behov:
 - Dedikerede anlæg til særlige brugergrupper og behov
 - Adgang til tværgående - gerne nationale - fællesfaciliteter
 - Internationalt til meget store behov (PRACE)



Governance og prioritering

- Governance og økonomi for fælles faciliteter skal være transparent
- Prioritering del er af en governancemodel
- Adgang:
 - Grundlæggende adgang til basisydelser for alle
 - Yderlige kapacitet ansøges på baggrund af merit eller tilkøbes
 - Ansøgning om internationale ressourcer til meget store behov

E-science kompetencer

- Vigtigt at opbygge og bevare lokale kompetencer på alle universiteterne
- Behov for e-science kompetencer i eller tæt på en række forskningsmiljøer
- Flere universiteter/fakulteter har etableret tværgående e-science kompetence centre
- Uddannelser med fokus på e-science og digitalisering meget vigtige

Sikkerhed

- Fokus på sikkerhed ved data adgang og data håndtering er stigende
- Personfølsomme data og GDPR stiller særlige krav

Finansiering

- Finansieringsmodel der sikrer løbende re-investeringer
- Finansiering gerne på universitets/fakultets niveau for basisydelse, men med muligheder for udvidede behov
- Hvem?
 - UFM
 - Universiteter
 - Eksterne midler

Rundbordssamtale DeiC/Styrelsen for Forskning og Kvalitet. 5. oktober 2017

Humaniora (AU, KU AAU og SDU)

Jeg vil gerne takke for muligheden for her at give udtryk for nogle tanker om DeiC's samarbejde med det humanistiske område.

Ikke helt korrekt, når DeiC i sin evaluering skriver, at det humanistiske område er novicer inden for e-science. Vi er meget langt inden for:

- 1: Digitalisering af kulturarv
- 2: Studier af digitalt fødte data (Fx web-arkæologi, radikaliserings)
- 3: Udvikling og konstruktion af IT-værktøjer: (Fx sprogteknologi. At få computere til at genkende, forstå, tolke, producere og efterligne sprog – såvel talt som skrevet)
4. Udvikling af digitalt understøttede forskningsmetoder (fx arkæologi).

OG vi i dag samarbejder med en række store fuldmodne europæiske infrastrukturer, som DARIAH og CLARIN.

DeiC's anbefalinger:

Det er rigtigt, at *"kompetencerne og adgangen til den nødvendige infrastruktur"* – overordnet set – ikke er på plads, men det er **ikke** rigtigt, at *"... der er primært behov for support og kompetenceudvikling indenfor e-Scienceområdet, i alle faser af forskningen."* Der er derimod brug for et fagligt ledet udviklingsarbejde med fokus på at udvikle digitale kapaciteter og værktøjer udviklet med sigte på de humanistiske videnskaber. Det betyder, at ansvaret må være fagligt og institutionelt, og ikke nationalt og generisk/generelt/ på tværs af områder.

Der er behov for **to typer**:

- 1) **Tæt på computerne:** Teknik, forskningsnet, HPC, viden om optimering og tilretning af koder) (dvs. DeiC),
- 2) **Tæt på forskerne og fagmiljøerne:** Kompetenceudvikling: Udvikling af og kendskab til digitale værktøjer. Fx viden om hvilke metoder man bør anvende, og hvilke programmer som kan bruges til forskellige former for analyse. Samarbejde mellem universiteter, også internationalt. Det er vigtigt at have en teknisk supportfacilitet tæt på miljøerne (fx DigHumLab, Clarin, Dariah).
Juridisk rådgivning om fx personfølsomme data og copyright-spørgsmål vurderes af et flertal af ovenstående repræsentanter for humaniora at skulle løses tæt på fagmiljøerne.

DigHumLab

DigHumLab: to generationer:

DigHumLab1: aktuelt i drift inden for en række fagområder baseret på en infrastrukturbevilling fra ministeriet.

DigHumLab2: Ansøgning inde i forhold til en række nye faglige samarbejder og udviklingsområder.

I DigHumLab arbejder vi med communities, dvs. forskningsgrupper, som er tæt på forskningen. Vi har pt. 5 communities:

Clarin, som er Danmarks repræsentant i det Europæiske Clarin, arbejder på to fronter:

- 1) en databank med dansksproget materiale og
- 2) forskellige værktøjer som kan bruges i sproglige analyser. Værktøjerne bruges bredt af bl.a. litteraturforskere, historikere og **selvfølgelig** lingvister.

Netlab, som har fokus på at støtte forskning, som benytter webben som datakilde for at studere webudvikling, kontroverser (eksempelvis holdning til vaccination) og sprogudvikling.

LARM, som giver adgang til at søge blandt alle DR's (nu udbudt via KB) digitaliserede radioudsendelser og arbejde med materialet (eksempelvis annotere udsendelser).

VILA og Mobile Labs, som stiller ekspertise til rådighed omkring brug af video - både produktion, brug og efterfølgende analyse.

DigHumLab sikrer en dialog mellem de forskellige communities og, at værktøjer bliver brugt på tværs.

DigHumLab bidrager fx også til standardisering i forhold til opmærkning af tekst. Samt fælles publicering af digitaliserede værker af Kierkegaard, Brandes og alle de andre store danske forfattere.

Generelt: Der er en uløst økonomisk udfordring i overgangen fra udvikling til drift og den efterfølgende indlejring og vedligeholdelse eller nyrehvervelser for at holde niveauet i forhold til den teknologiske udvikling (Matthæusprincippet kan blive konsekvensen).

"Humaniora og samfundsvidenskab er typisk set som novicerne inden for e-Science. Der eksisterer dog store e-Science forskningsprojekter, og det fremtidige behov og potentiale er anerkendt, men kompetencerne og adgangen til den nødvendige infrastruktur er ikke på plads.

*Mulighederne for anvendelse af e-Science inden for humaniora og samfundsvidenskab udspringer af den store mængde af data, der er opstået blandt andet efter digitalisering af skriftlige kilder og billedmateriale og indsamling af data fra internettet. For at kunne udnytte de muligheder digitale data giver, er der primært behov for support og kompetenceudvikling indenfor e-Scienceområdet, i alle faser af forskningen."*¹

¹ Anbefalinger til en sammenhængende e-infrastruktur til fremme af dansk forskning for perioden 2018-2028, set i lyset af forskningens digitalisering og globalisering, s. 8.

E-infrastruktur for samfundsvidenskab

Søren Hvidkjær
Professor, Institutleder
Institut for Finansiering

Datatyper med særlige behov

Data der "scrapes" – sociale medier, blogs, intranet etc.

Data fra eksperimenter – højfrequente psykofysiologiske/sensoriske målinger

Finansielle data – højfrequente (milisekund) data fra børser og lign.

Data der genereres – Monte Carlo studier

Data der flettes fra flere kilder – registerdata, følsomme virksomhedsdata mm.

Data der er personfølsomme – offentlige registre, surveys, lokale registre, observationsstudier

Sensitive/sikkerhedsgodkendte data (fx politidata)

Registerforskning og andre personhenførbare data

Persondatalovgivning og persondataforordning administrativt krævende

Bedre e-infrastruktur i DST:

- Fra individuelle servere samlet større lager- og computerkraft (HPC)
- Videreudvikle software nemmere og mere sikker datadeling.

Regionernes Genomcenter på DeIC's supercomputer ved Risø: Interesse for kobling af data med socio-økonomiske data i DST.

Kommercielle løsninger?

Identificer hvad der bedst løses af kommercielle providers:

- Dropbox bliver fx i fuld compliance med persondataforordningen inden denne træder i kraft
- Kommercielle HPC cloud løsninger

Kommercielle løsninger dominerer ofte for data providers og udgør essentiel e-infrastruktur fx i økonomi og finansiering (DRDS).

Finansieringen af e-infrastruktur bør ikke afhænge af om man vælger en kommerciel løsning (skab incitament til at vælge optimale løsninger).

Behov for regnekraft og datafaciliteter

Fokus på nem (brugervenlig, hurtig og billig/gratis) adgang til HPC for alle forskere

Behov for nem og sikker måde at dele data med forskere/virksomheder nationalt såvel som internationalt



eScience status e-Infrastrukturbehov Sundhedsvidenskab

Ole Skøtt

Dekan, Sundhedsvidenskabelige Fakultet, SDU



Megatrends SUND

- Bevægelse fra fokus på store befolkningsgrupper til individer
 - Patienten i centrum, Patienten er alt ..etc.
 - Den enkelte patients personlige præferencer og forudsætninger
 - Præcisionsmedicin og individualiseret behandling
 - Langt større krav til at kunne kombinere oplysninger fra mange forskellige datakilder, også på individniveau, og bagud i tid
 - Langt større datamængder (genomer, billeder, sundhedsdata fra "wearables" smartphones/smart watches)
 - Integration i fx elektroniske patientjournaler

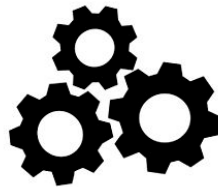
- Nye lovmæssige rammer med øgede krav til håndtering af personfølsomme data
 - EUs persondatafordning
 - Institutionelle løsninger, nationale løsninger til afløsning af individuelle løsninger
- Teknologiske nybrud
 - Billiggørelse af genomanalyser
 - Nye effektive metoder til datahåndtering, analyse og opbevaring

- Mange områder på vej over i BigData
 - Beslutningsstøttereds kab er, hvor lægens beslutningsgrundlag flytter sig fra erfaring fra nogle tusind patienter til millioner af patienter (ex Deep Blue/Watson).
 - Omics af mange typer
 - Mikroskopisnit i patologien på vej.
 - Uddannelsesområdet, virtuelle laboratorier, anatomi etc

- I fremtiden
 - Større fokus på fælles nationale løsninger
 - Sikring af nationale højhastighedsnet
 - Sikring af analysekapacitet med High Performance Computing og lagringskapacitet
 - Fokus på nationale "secure cloud" løsninger til datahåndtering og dataadgang
 - Fokus på national standardisering af datahåndtering, prøvetagning etc.
 - Og selvfølgelig i overensstemmelse med internationale forhold

Make it simple to get

F_{indable} A_{ccessible} I_{nteroperable} R_{eusable}



Implementation of FAIR principles to achieve OpenScience

- excellent infrastructure
- expert support staff
- acceptance & compliance of researchers

e-infrastruktur til forskning i verdensklasse i 2025

- en analyse og udredning vedrørende dansk e-infrastruktur samarbejde

Bilag 4:

Interview med forskere, deltagere og uddrag af interviews

Forskerinterviews på universiteterne

Fremgangsmåde

Med udgangspunkt i "Kommissorium for analyse og udredning vedr. dansk e-infrastruktur samarbejde" dateret 12. september 2017, er der foretaget en interviewrunde med 17 danske forskere (Tabel 1) inden for de fire videnskabelige fagområder hhv. Humaniora (HUM), Samfundsvidenskab (SAMF), Natur-Teknik (TEKNAT) samt Sundhedsområdet (SUND). Der er under interviewene spurgt ind til forskernes behov for regnekraft (HPC, High Performance Computing), net, data management og lagring samt kompetencebehov og services, herunder særlige støttefunktioner. Der forefindes lydoptagelser fra interviewene.

Formål

Formålet har været at kortlægge eksisterende brug af e-infrastruktur, flaskehalse, hvad er behovet på nationalt niveau, og hvilke dele der bør håndteres på lokalt/institutionelt niveau i et fem-til-tiårigt perspektiv. Udgangspunktet har været at forskerne frit har kunnet anskueliggøre hvad der skal til for at fremme deres specifikke forskning med fokus på digitale infrastrukturer og ikke mindst at: "Forskere ved de danske universiteter skal have adgang til en e-infrastruktur, der muliggør forskning og uddannelse i verdensklasse."

Dette bilags indhold

Med udgangspunkt i mere end ti timers interview er her sammenfattet

- Angivelse af behov for regnekraft (Størrelse i antal kerner hvis muligt, Tabel 2)
- Hovedsynspunkter fra hver af de 16 danske forskere (En halv A4 side)

Tabel 1. Forskere der er blevet interviewet i forbindelse med denne rapport.

ID kan henføres til Tabel 2.

| Fagområde | Forsker | ID | Universitet |
|-----------|----------------------------|-----|-------------|
| HUM | Costanza Navarretta | 5. | KU |
| HUM | Mads Rosendahl Thomsen | 8B. | AU |
| HUM | Kristoffer Laigaard Nielbo | 9B. | SDU |
| HUM/SAMF | Mads Kæmsgaard Eberholst | 11. | RUC |
| SAMF | Fane N. Groes | 3. | CBS |
| SAMF | Martin M. Andreasen | 8C. | AU |
| SAMF/TEK | Sebastian Büttrich | 2. | ITU |
| SUND | Samuel Schmidt | 7B. | AAU |
| SUND | Thomas Damm Als | 8D. | AU |
| TEKNAT | Troels Haugbølle | 1. | KU |
| TEKNAT | Niels Aage | 4. | DTU |
| TEKNAT | Karsten W. Jacobsen | 6. | DTU |
| TEKNAT | Kåre L. Nielsen | 7A. | AAU |
| TEKNAT | Lars B. Madsen | 8A. | AU |
| TEK | Jost Adam | 9A. | SDU |
| NAT | Claudio Pica | 9C. | SDU |
| TEKNAT | Jens H. Walther | 10. | DTU |

HUM – Costanza Navarretta, Seniorforsker, Afdeling for Nordiske Studier og Sprogvidenskab, Center for Sprogteknologi, Københavns Universitet (5.)

HPC/Datatyper/Behov: HPC benyttes i samarbejde med tyske samarbejdspartnere; machine learning på multimodale data, sprog, gestus. Studerende på vores tværgaglige uddannelse har i dag problemer med regnekraft. Øgede datamængder og større grad af tværfaglige samarbejder i fremtiden. Mere behov for regnekraft. Mange og komplekse datatyper. Modellering – mønstergenkendelse – automatiske opmærkning - semantik, morfologi, syntaks, pragmatik, billeder i tekster, video af menneske-menneske, menneske maskine kommunikation mv.

Forskningsnet/Sikkerhed: Udfordringer med sikkerhed. Kryptering på diske – gemmes personligt på computer og eksterne diske- ikke redundans – kun 1 backup. Google doc bruges ofte til fildeling. Behov for klare linjer mht. datasikkerhed.

Datamanagement/Lagring: Vil helst ikke arbejde med data der ikke kan deles i fremtiden. Problematik omkring personfølsomme data; tekst, talesprog, billeder, video. Eksempelvis kan det foretrækkes at bruge nyheder fra USA til forskning i stedet for danske. Sætter forskningsmæssige begrænsninger. Behov for fælles (evt. national) løsning mht. data lagring/storage – håndtering af forskellige grad af personfølsomme data. Ikke-kompatibelt med FAIR på visse områder.

Kompetencebehov: Undervisere skal uddannes – digitalisering er ved at revolutionere vores liv. Stor kompetencebehov bl.a. i digital humanities . Vi kan lokke forskere med user-cases indenfor deres fag. Behov for at vise tværfaglige muligheder til inspiration. Workshops. Ekspertter der kan fortælle hvordan tingene udvikler sig – måske via DeiC. Tilgængelige landskabsbilleder mht. fagrelevante værktøjer der kan deles tværfagligt. Behov for at implementere HPC regnekraft i undervisningen og forskning.

HUM – Mads Rosendahl Thomsen, Professor MSO, Kommunikation og Kultur, Aarhus Universitet (8B.)

HPC/Datatyper/Behov: Tekst – globalisering af litteratur – kræver umiddelbart ikke de store regnekrafter. Men regnekraft bliver udfordret ved fx Google-books – Hatchitrust – uni og Notre dame – geotagged litteratur – fx 10mill bøger – Samlet korpus kan blive stort – topic modelling – Nordisk litteratur – mindre korpus – billedbehandling kræver mere regnekraft. Brug af video og billeder og højfrekvente data fra sociale medier er i stigning i et fremtidsperspektiv.

Forskningsnet/Sikkerhed: Netadgang og hastighed fungerer umiddelbart i eksisterende form.

Datamanagement/Lagring: Storage af data – evt. fagdirektorer - behov for synlige generiske værktøjer — Behov for håndtering af personfølsomme data og nem adgang til kulturelle data (Afklare hvordan man håndterer rettigheder eller evt. ok at søge på features?). Copyright issues er et problem – Den norske model er god – digitalisering af kulturarv ønskes.

Kompetencebehov: Generelt kompetenceløft – digital literacy – vigtigt med brugervenlighed for nye brugere/nye fagområder - sænke kompleksiteten mht. hvordan folk involveres og introduceres til digitale værktøjer. Digitale kompetencer mangler til nye studerende. Brug for flexibel, brugervenlig tilgang til HPC og support.

HUM – Kristoffer Laigaard Nielbo, Lektor, Institut for Kultur og Samfund, Aarhus Universitet (9B.)

HPC/Datatyper/Behov: Natural language processing (NLP) – stigende behov for Graphics Processing Units (GPUer) mht HPC. Image analysis - video analysis. Heterogene og ustrukturerede data! Behov for teknisk løsning der gør det muligt at arbejde med personfølsomme data i lukket system på større/ nationale HPCer.

Forskningsnet/Sikkerhed: Behov for beskyttede netværkslinjer til data transport når det gælder personfølsomme data – Data mængden går op – ”remote access” til alle fagdiscipliner nyttigt.

Datamanagement/Lagring: FAIR – Hvor skal data lagres? – Behov for guidelines – HUM data kan ofte ikke deles – behov for acceptabel model hvor copy right data © kan deles – massiv trend der arbejder væk fra Open Data – men kan stadig dele ”features”; eks psykiatriske data (Århus). Hvordan håndteres HPC beregninger på 14 millioner artikler – kan ikke spørge hver enkelt mht © men kan evt. være ok at dele features → Behov for national ”Research Cloud” med indbygget håndtering af ©-issues og personfølsomme data. Behov for let data adgang og data transport. Paradoks: Lettere at få adgang til patient journal end copy-right data. Holland håndterer © elegant mht HPC – data ansvarlig på regneprojekter.

Kompetencebehov: Ser massiv overlap af de kompetencer der er brug for mellem forskningsområder tværfagligt. Brug kompetencerne! Kan f.eks. gruppere humaniora med sundhedsområdet hvor der er overlap mht. personfølsomme data og uensartet datastruktur. Vigtigt at kombinere folk med humaniora fagviden og dem med tekniske digitale kompetencer. Nyttigt med nationale HUM kompetence ”hubs” - Stort overlap mht. programmings værktøjer – signal processing – Python og R typisk – ligger en vigtig opgave i at interface dette elegant og brugervenligt på HPC anlægene.

SAMF/TEK – Sebastian Büttrich, Teknisk ansvarlig, Pervasive Interaction Technology Laboratory (pIT Lab), IT Universitetet (2.).

HPC/Datatyper/Behov: Meget diverse tværfaglige datatyper. Voksende behov for computing, GPU, CPU – voksende data science retning – billedbehandling som datatungt element. Video, audio, multimedia, satellitdata – copernicus (multispektrale sensor, geo-data/GIS) mm. Regnekraft behov kan vokse eksponentielt lige pludselig.

Forskningsnet/Sikkerhed: Transport 10 Gbyte fungerer fint i dag dog forespørgelser om 100 GB net allerede. Kan tage en uge at nedhente data/ indtil videre ikke kritisk men forudser det kan blive en flaskehals. Netværksoptimerings fokus i internationale samarbejdsprojekter (Blive på forskningsnet – ej hoppe mellem kommercielle netværk).

Datamanagement/Lagring: Udfordringer – hvad kan være i skyen og hvad kan være remote? Økonomi afgørende; dyrt at have 24 hrs regnekraft i skyen Amazon/ Azure versus lokale løsninger. Udfordring at finde den rigtige blanding. Kommercielle løsninger kan være ok i spidsbelastningsperioder. Vi kan ikke erstatte lokale ressourcer med løsninger i skyen (spidsbelastninger vs. kontinuerligt). Behov for national sky tjeneste. Både mht. lagring og at regne i skyen. Fildelings behov!! Nordisk initiativ velkomment! Ønsker om brugervenlighed i stil med Dropbox.

Kompetencebehov: Behov for kortlægning af fælles forståelse af infrastruktur på tværs af fagområder, DK, Nordisk, Internationalt! Opbygge fælles sprog mellem IT og fagkompetencer. Uddannelse i GDPR, Open Science, fildeling, storage feks via seminarer/workshops. Vigtig kompetence er kundeorientering (Hurtige, lette, brugervenlige løsninger også som front runnere).

HUM/SAMF – Mads Kæmsgaard Eberholst, Studielektor, Institut for Kommunikation og Humanistisk Videnskab, Roskilde Universitet (11.)

HPC/Datatyper/Behov: Medieforskning, journalistik, trends og retorik på sociale medier. Øget datamængde i fremtiden – heterogene ustrukturerede data! Forudser behov for HPC (feks afdække mønstre i 2,5 millioner Tweeds fra danske valg). Vil rigtig gerne bruge HPC! Modelling der kræver HPC regnekraft er realiteten i dag. Strander ved manglende hjælp, kan godt programmere, ved hvad der skal til. Brug for intro og sparring mht HPC. Tilpasser sin forskning efter den regnekraft der er til rådighed (Risiko for at data frasorteres og finder alternative løsninger).

Forskningsnet/Sikkerhed: Forskningsnet hastighed fungerer i nuværende form.

Datamanagement (DM)/Lagring: Hvordan høster man data smartest fra forskellige kilder/databaser? Screenshot af data benyttes til trods for andre alternativer findes men ikke haves. Savner fælles strategier på området feks lokalt eller nationalt. Behov for håndtering af personhenførbare data – arbejder primært med offentligt tilgængeligt medieindhold – savner tydelige juridiske rammer for håndtering af denne type data – ophavsret – opbevaring af data mere end 24 timer hvordan administreres dette reelt? Stort behov for opbevaring af denne type data og retningslinjer. Datamængde stiger. Koordinering og synliggørelse af hvilke repositories man kan bruge. Positiv overfor open data tilgangen – mere af det!

Kompetencebehov: Behov for indsigt og adgang til den infrastruktur der kan understøtte og rykke sin forskning og inkludere support! Vigtigt med tid til kompetenceudvikling, men lige så vigtigt at man understøtter kompetencerne så de kan komme i spil (efteruddannelse eller match med andre der har kompetencerne). Vigtigt at investere i datahåndtering og regnekraft. Det svarer til at sige til en der forsker: Du har dit laboratorium. Du har din tid. Men der mangler indhold/apparater i laboratoriet (HPC) og nøglen har forskeren fået men ikke fået af vide hvilken bygning den er til. Et laboratorium er dyrt. Det er også dyrt at håndtere data. Behov for brugervenligt, skematiseret forskningsunderstøttelse mht. IT, tilgængeligt (synligt) og serviceminded så man får hurtig assistance.

SAMF – Fane N Groes, Lektor, Økonomisk Institut, Copenhagen Business School (3.)

HPC/Datatyper/Behov: Mikrodata – for individer og virksomheder- registerdata – al analyse foregår hos Danmarks Statistik (DS) – instituttet har købt servere hos DS. Behovet vil stige for HPC. CBS skal kunne tilbyde capability servere (regne kraftigt). Sammenkobling af data fra feks. elever på erhvervsskoler som sammenkøbes med registerdata bla. personfølsomme data.

Forskningsnet/Sikkerhed: Rammerne sikre hos DS – giver tryghed at man ikke kan gøre noget forkert/ feks. kun udføre tilladte gennemsnit af databeregninger. Dog stor administrativ del. Behov for at de juridiske rammer også er på plads hos CBS. Drop box – fil deling. Stort arbejde at indsamle data fra fx erhvervsskolerne, hvor man skal ud på hver enkelt skole.

Datamanagement (DM)/Lagring: Økonomer der benytter registerdata – det skal kunne køre mere efficient – data ligger mange steder – hvert institut har adgang til DS data i sin egen silo/serveradgang – ej optimal udnyttelse af serverplads og regnekraft - økonomi – stort resourcespild. Har et behov ud over det som det Windows-baserede system kan håndtere. Begrænset af kapaciteten og brugerfladen. Hvor er mandat til langtidsopbevaring? Hvor? Hvordan? Nationalt?

Kompetencebehov: Research DM team (under biblioteket) – forståelse for at kunne opbevare - forskerstøttedfunktion storage. Behov for registerdata delen. Internt på CBS – lokal support og ekspertise (data standarder, forskningsdata – personfølsomme data). SAMF forskeren har alle 3 kasketter på: har fagområde viden til sin ansøgning og bygger som regel selv sin server (finder penge til den) og udvikler algoritmen hvis der er brug for modellering - enmandstrekant.

SAMF – Martin Møller Andreasen, Professor, Institut for økonomi, Århus Universitet (8C.)

HPC/Datatyper/Behov: Stabil finansiering af regnekraft er vigtig! – forskellige mængder HPC, statistisk modellering, Behov for flexibel adgang til HPC (let, når man har behov) og stabilitet omkring det cluster man bruger mht arkitektur og support. Tidsrække-analyser – aktiemarkedet, makro økonomi – ikke fortrolige offentlige data – memory tungt – HPC drevet ting – simulering og parallellisering. Center for registerforskning – Arbejdsmarkedsforskning – Danmarks Statistik – samkøring af registre – Memory begrænsning og bruger derfor kun et udsnit af data i stedet for alle data på én gang. Stigende behov forudsiges for regnekraft. Tilpasser sin forskning til den regnekraft der er til rådighed. Mangler middelstore HPC clustre (Haves i bankverdenen). Nationalt HPC setup ses betinget af tilstedeværelsen af lokale (lærings)miljøer.

Forskningsnet/Sikkerhed: Netadgang og hastighed i dag ok. Kunne give mening med lagring af personfølsomme data forankret nationalt for at varetage sikkerheden for den enkelte.

Datamanagement/Lagring: Bekymret for projektfinsierede storage ordninger hvor storage eller cluster kan ændre sig over tid/ på baggrund af prioritering og finansiering. Brug for stabilitet. Både mht. regnekraften (clustre) og datahåndtering. For usikker tilgang med projektfinsiering når datacyklus med publicering for øje kan tage mange år også efter afslutning af et projekt. Vigtigt med stabilitet.

Kompetencebehov: Begrænsede kompetencer – vigtigt at der er et lokalt support miljø – Kunne være nyttigt med MATLAB cluster og kan mangle fagligt miljø/ sparring når det gælder programmeringssprog og værktøjer. Behov for HPC forskningsmiljøer til de studerende. Uddannelsesmiljø omkring HPC – på tværs af feks. statistik/datalogi – Lavt hængende frugter. Phd-studerende.

SUND – Samuel Schmidt, Lektor, Institut for Medicin og Sundhedsteknologi, Aalborg Universitet (7B.)

HPC/Datatyper/Behov: Billeder, sensordata og algoritmer – diagnostisering af hjertesygdomme. Behov for selv at kunne installere/ drag-and-drop software moduler på nationale HPC anlæg. Ellers går det for langsomt med at komme i gang. Bruger fortrinsvis CPUer da GPUer ikke tilgængelige på den anvendte nationale HPC. Behov for forskellige HPC arkitekturer og styresystemer (Windows vs Linux mm).

Forskningsnet/Sikkerhed: Vigtigt at sikre rette sikkerhed feks ved gæstelogin men adminIT der styrer adgangen dertil skifter ofte login koder. Dette kan føre til oprettelse af Ghost accounts.

Datamanagement/Lagring: Mange samarbejdspartnere så datadeling er helt centralt. Behov for fildelings system der er brugervenligt i stil med OneDrive, Dropbox, GoogleDrive etc. Adgang til sundhedsdata på "remote stick", personfølsomme data. Guidelines? Stort behov for data deling med rette sikkerhed til at håndtere personfølsomme data – gerne på tværs af hele Europa. Nuværende infrastruktur (netværksdrev, servere) ikke optimal. Behov for lagring af alle rådata: feks. bemærkelsesværdigt at EKG indsamlet på sygehus feks gemmes på jpg-format; dermed er al rådata forsvundet og man har kun et billede af en EKG graf. Databaser konstrueret til at få data ind men ikke til at få det hele ud igen.

Kompetencebehov: Vigtigt at integrere de unge forskere. Nye Phd kurser og tværfaglige seminarer.

SUND – Thomas Damm Als, Lektor, Institut for Biomedicin, Aarhus Universitet (8D.)

HPC/Datatyper/Behov: Memory baseret HPC regnekraft til kortlægning af genomer, F.eks. 80.000 individer der analyseres 10.000.000 steder i deres genom. Øget HPC behov i fremtiden. Nogle typer af data analyser kan ikke udføres i dag da det er kombination af personfølsomme register data og genomer og data kan ikke sammenkøres. I fremtiden også machine learning i forhold til personlig medicin.

Forskningsnet/Sikkerhed: Hastighed kan være langsom. Endnu ikke en flaskehals.

Datamanagement/Lagring: Backup og lagring af data er begrænset af at der er store data mængder og for lille lagringskapacitet. Øgede datamængder i fremtiden i form af endnu flere sekvensdata for hele genomer stiller krav til en øget lagringskapacitet. Personfølsomme data. Specifikke regler for data håndtering via GenomeDK (Med en særlig sikret del altså ala Danmarks Statistik).

Kompetencebehov: Kendskab til relevante værktøjer og kendskab til opskalering på større HPC anlæg.

TEKNAT – Troels Haugbølle (TH), Lektor, Institut for stjerne og planetdannelse, Niels Bohr Institutet, Københavns Universitet (1.)

HPC/Datatyper/Behov: Massive strukturerede data – vejrudsigten bare i universet – kan ikke bruge national HPC setup/arkitektur. TH er Superbruger på PRACE. Numerisk astrofysik. Vigtigt med forskellige niveauer af HPC kapacitet/typer af hardware for skalering både nationalt og internationalt. Fornuftig betalingsmodel påkrævet – f.eks. baseret på videnskabelig kvalitet som fremmer videnskabelig konkurrence (Peer review) – F.eks. 3 kategorier: small – medium – large – dimensioneret efter behov. I dag betales for HPC tid – kvalitet ikke hovedparameter. Nationale HPC kapacitet i DK er 2-4 mindre end øvrige Nordiske lande og Holland. De nationale computere skal være større end de lokale. Nuværende nationale HPC konstruktion afskærer de største brugere (faciliteter ikke store nok) og betalingsmodel er uattraktiv og storbrugere investerer i stedet lokalt. Ser potentielt flere DK kandidater/større DK HPC masse til at tilgå PRACE. National strategi /finansiering vigtigt på den lange bane (Stabilt HPC landskab).

Forskningsnet/Sikkerhed: Sikkerhed anses som et HPC center ansvarsområde. Båndbredde vigtigt.

Datamanagement/Lagring: Øget behov for lagring af data. Adgang til data – analysen foregår der hvor data ligger. Kæmpe behov for at dele data FAIR. Som inspiration kigger PRACE på at lave data-portaler med support, så forskningsdata kan overleve i længere tid.

Kompetencebehov: Oplysningsopgave – der findes HPC ressourcer – find de gode user-cases Vise folk hvad det er blevet brugt til. Ingen uddannelse der producerer HPC eksperter. Behov for nationale landskaber over kurser, uddannelser også inkl. kurser i Europa via PRACE. Kompetencer skal være tæt på forskeren. EXAscale computere i Europa, hvor mange og hvem kommer til at køre på dem? Udvikling af de programmer der skal køre i fremtiden. Hardware og SW på plads.

TEKNAT – Niels Aage, Lektor, Institut for mekanisk teknologi, Sektion for faststofmekanik, Danmarks Tekniske Universitet (4.)

HPC/Datatyper/Behov: HPC – fysik – design – simuleringer - metodeudvikling – en masse beregningskraft – brug for god infrastruktur. Mangler HPC kapacitet x10 større end det til rådighed i dag. Løser kun det som nuværende HPC kapacitet kan bære. Nuværende HPC forretningsmodel fungerer ikke til kursus-afholdelse. Visualisering set-up – mens den regner kunne det være godt at visualisere sine resultater – Behov for Interaktiv data-simulering af BigData. Erfaring fra international HPC/PRACE: Curier – reglerne meget præcise – prioriteringsregler - struktureret og professionel support. Brug erfaringerne nationalt? Vigtigt at koder fra paletten af fagområder testes på HPC dvs. noget fra alle! Vind, fluid, mekanik – skalering – test suite til definition af løsningen – arkitekturen. Og synliggøre dette for brugeren. Ignorere at folk vil have deres eget HPC – tilgang til kapacitet er centralt – samarbejde gør livet sjovere...

Forskningsnet/Sikkerhed: Netværk fungerer godt. Bange for at miste data. Sårbart setup med serveren i kælderen – hvad hvis huset brænder? Dårlig sikring af datasikkerhed/ back-up.

Datamanagement/Lagring: Behov for central storage – gerne med med brugerregler – eksempler med study cases. Hvad med samarbejdsplatforme der er nemme at tilgå i projekter? Sharepoints? eLabbøger for track-and-trace af forsøg, kodeversioner mm. Vigtigt med synlig e-infrastruktur.

Kompetencebehov: Kurser e.g. hvordan kommer jeg i gang med HPC – der er muligheder. Brugervenlighed vigtigt. Infrastrukturen – der skal være nogen som styrer – administratorer – mere synligt - mere nemt. Behov for best practice – guidelines både mht HPC, storage, værktøjer mv.

TEKNAT – Karsten Wedel Jacobsen, Professor, Institut for fysik, Danmarks Tekniske Universitet (6.)

HPC/Datatyper/Behov: Materiale forskning på atomar niveau –modellering med afsæt i naturlovene. Design af nye egenskaber. Stigende behov for regnekraft - Anvender i dag den kraft der bliver stillet til rådighed - Niflheim – Arkitektur minder om Abacus - Kunne godt bruge mere HPC regnekraft. Tilpasser sin forskning til kapaciteten. Behov for at kunne lave mange beregninger, mange jobs. 3 x gange så mange jobs i kø, som computeren kan håndtere – behov for den lokale flexibilitet og brugervenlighed. Servicen og kontrollen over maskinen er afgørende. Niflheim: optimeret – er skræddersyet til det vi har brug for. Behov for at anerkende den store diversitet inden for HPC og lave optimale løsninger hvor de forskellige behov er. At få tilbudt samme ressourcer på national HPC sammenlignet med Niflheim er ikke en gulerod (blot større afstand) – mere regnekraft kapacitet og tid er en klar forudsætning for at national alternativ vil være attraktivt.

Forskningsnet/Sikkerhed: Netværkshastighed opfylder pt. behov. Stigende behov på sigt.

Datamanagement/Lagring: Inspiration: NOMAD: Novel materials discovery (Max Planck instituttet): alt data lagt ind fra beregninger - enestående indsats – gratis lagring i 10 år (uddeler DOI).

Kompetencebehov: Kurser – Computing. Kompetencer skal være tæt på forskningsmiljøet.

TEKNAT – Kåre Lehmann Nielsen, Professor, Institut for Kemi og Biovidenskab - Sektion for Bioteknologi, Aalborg Universitet (7A.)

HPC/Datatyper/Behov: Planteforædling – genomsekventering - Central barriere: man skal selv kunne installere software (feks i moduler) på HPC – ellers går det for langsomt. Flexibilitet – imødekommes ikke på centrale HPC anlæg. Fremtidigt behov: Server i serverne – sandboxes – Linux miljø – Forskellige HPC niveauer af regnekraft. Øget behov for regnekraft fremadrettet. Flexibel adgang og modellerbar allokering af HPC regnetid vigtig da processen indeholder flere forskellige faser af beregning: Optimering af analyser/koder – Reiterationer/genberegning på samme model – drift / Selve beregningen – tage højde for software komplikationer (kommercielle og hjemmelavet SW eller kode). To typer HPC brugere: dem der er glade for eksisterende (lokale) løsninger og "Emerging" HPC brugere med et behov der overstiger det HPC man har til rådighed.

Forskningsnet/Sikkerhed: Dataoverførsel/ transport fart kan være for langsom. Behov for højere net hastighed. Behov for fleksibel administration- datasikkerhedssystemer der er godkendt – guidelines, ordentlige transparente retningslinjer.

Datamanagement/Lagring: Open Science (OC) problematisk ved samarbejde med firmaer. Utopi i praksis. Interesse konflikter mht. patentsrettigheder og dataværdi. Som eksempel er der kendskab til amerikanske samarbejdspartnere/forskere som netop bevidst står udenfor OS området pga modsatrettede kommercielle interesser i projekter. Finansieringsmodellen for forskning i DK er linket up på public funding (50%) hvilket spænder ben for OS konceptet (F.eks. 50% medfinansiering fra virksomheder via Innovationsfonden). Sharing: brug for en fælles løsning – på tværs af universiteter. Dataformat: konvience mod cost. Et stort behov for fremtidsikring af data. Data skal ligge centralt – i sharable systemer der kan tilgås af eksterne samarbejdspartnere. Vigtigt med delbarhed mellem internationale, nationale og lokale data repositorier. Svært i forskning centralt at definere standarder "tidligt i processen" fordi man ofte på det tidspunkt ikke ved, hvad der er væsentligt information at standardisere.

Kompetencebehov: Oplysning – PhD kurser – interne kurser – mentorkorps ud i organisationen der udbreder kendskab til HPC. BEMÆRK: Forskere skaber eksterne midler via deres publikationer, citationer og Journal Impact Factor – skal derfor finde incitament/frikøb af forsker for at deltage i mentorkorps. Workshops om muligheder/inspiration på tværs af fagområder.

TEKNAT – Lars Bojer Madsen, Professor, Institut for Fysik og Astronomi, Aarhus Universitet (8A.)

HPC/Datatyper/Behov: Fysik – Teori for kvantesystemer – bølgefunktioner/algebra - store udregninger – meget CPU-kraft – Behov for flere noder/ parallelle jobs. Vigtigt med lokale HPC faciliteter der understøtter vækstlaget. Øget HPC behov i fremtiden, men svært at spå om præcise typer og behov over en 10 års periode. Det giver ingen mening at have nationale HPC anlæg hvis man ikke også har stærk lokal regnekraft. Vigtigt at have faciliteter lokalt og hvis der er midler til mere kan man derudover prioritere nationalt. I dag prismodel-barriere mht. nationale HPC faciliteter. anbefaler kraftigt at man nationalt afsætter en bestemt pulje penge til fri og stabil lokal HPC regnekraft. Ikke nemt at komme i gang på nationale clustre for nye fagbrugere der kommer andetstedsfra end TEKNAT. HPC arkitektur; Husk fokus på forskellige HPC behov. Feks. kan AU Kvantekemi godt bruge national HPC til at køre beregninger baseret på programmer, der er blevet udviklet over mange år, mens AU Fysik ikke generelt kan anvende Abacus 2.0 i eksisterende form (Natur af applikationer; hvis mere udviklingsorienteret ikke standard kode). Og AUs lokale HPC faciliteter er sammenlignelige med Abacus 2.0 og derfor er sidstnævnte ikke så attraktiv. Stort spring fra nationale HPC til PRACE. Tvingende nødvendigt med langtidsholdbare midler til HPC området (Stabilt landskab). Svarer til at en eksperimentel laserfysiker ikke har en laser, hvis fysik teoretikerne ikke har HPC!

Forskningsnet/Sikkerhed: Sandsynligt behov for øget hastighed af netværk på sigt.

Datamanagement (DM)/Lagring: Udarbejdning af tydelige retningslinjer der kan implementeres – 1) hvad skal der gøres 2) hvordan implementeres arbejdet. Venteposition pt. til DM området der er ved at blive defineret på alle danske universiteter. Forventer en løsning der er nem at tilgå.

Gemmer ikke al data i dag (Bølgefunktioner fylder meget) – har mulighed for elektronisk log (eLabBook) over ens arbejde så det kan reproduceres.

Kompetencebehov: Kompetence-løft ved hjælp af lokal support – uddannelse og kurser på nationalt niveau. Behov for at udbrede kendskab til storage muligheder, værktøjer (GitLab) og HPC faciliteter. Vigtigt at støtte de unge forskere og give dem kompetencer/ nem adgang til HPC.

TEKNAT – Jost Adam, Lektor, Mads Clausen Institutet, NanoSYD, Syddansk Universitet (9A.)

HPC/Datatyper/Behov: Nano photonics - computational modelling – micro electro mechanical system. Need of computational power (CPUs and RAM) – numerical models – technical disciplines – parallelism – use HPC and the interest is increasing – commercial Software– physics – signal processing – constantly moving due to the enhanced computational power. Local CPUs/silos for modelling keeps people from buying into HPC. Need for offering flexible virtual HPC machines! Important with a national long term stable HPC solution. Important with different HPC architecture for different scientific research areas.

Forskningsnet/Sikkerhed: Speed of net currently ok – always the faster the better.

Datamanagement (DM)/Lagring: FAIR - Challenge in where to put the data – how do we manage this – need guidelines. Simulation data makes sense to share –not common practice today! Huge requirement of storage (numerical data). Not yet a limitation regarding DM – using data.deic.dk good service for storage. Important that it is easy to store data and analyse. Recommend to use national Cloud computing service.

Kompetencebehov: Massive needs for local research support at HPC. E.g. building up national labs/hubs – bridging the gap to new technologies - Digital scholarship centers - centralised approach – training organised at national level – reuse of knowledge – Huge overlap in programming languages – signal processing – python – interfacing it elegantly on HPC. No need to develop new things – need to extend what is already out there and get awareness of storage possibilities and HPC through userfriendly communication/seminars etc.

NAT – Claudio Pica, Professor og Centerleder, Institut for Matematik og Datalogi, Anvendt Matematik, Syddansk Universitet (9C.)

HPC/Datatyper/Behov: High energy physics – theoretical – the field is among the largest user of HPC worldwide and developing new HPC architectures – biggest user at SDU – growing complexity for simulations – larger national computers needed – the simulation softwares used are among the standard benchmark for new HPC machines at large HPC centers worldwide – Important to have a national long-term, stable HPC solution – Growing interest for HPC in other fields. Connection speed to the supercomputer must not be a barrier. Diversity of users can be a challenge – new research areas are coming in - Traditional users are still the ones that use most resources. Important to have different HPC architectures best suited for different research questions.

Forskningsnet/Sikkerhed: PRACE: not hosting data. Need to transport back data to DK – fast network needed – GEANT is usually fast enough – capacity not filled up now, but in the future needs will increase – need to make services (HPC, storage) more accessible to everyone. A (national) research cloud should be developed. – Good infra structure in DK. National Network services: WAYF – good (identity provider). – Speed of the network – the faster the better.

Datamanagement/Lagring: Need for national research cloud. The same for research data and storage – coordinated effort needed – federation of capacity inside DK.

Kompetencebehov: Potential for more users on PRACE – skills are needed in order to increase the Danish users – larger national HPC facilities are needed – international collaborations are needed.

– A lot more training. Research support needed for newcomers on HPC.

TEKNAT – Jens Honore Walther, Professor, Institut for Mekanisk Teknologi, Sektion for Fluid Mekanik, Vandbygning og Maritim Teknik, Danmarks Tekniske Universitet (10.)

HPC/Datatyper/Behov: Strømnings mekanik – gasser og væsker - samarbejdsprojekter med industrien (80% m/ industrien) – tunge beregninger på lokal DTU Niflheim HPC (Fra 1 times varighed til en måned). Instituttet medbringer økonomiske midler til Niflheim (Brugervenlighed, Specialiseret i fysik, Linux, let tilgængelig, max 0-2 dag kø). Til tider begrænset af den HPC kapacitet der er til rådighed – tilpasser sin forskning. Stigende HPC behov i fremtiden. Klar til regnekraft/tid – beregning pr tid – kvota duer ikke - fleksibel tilgang for at kunne justere. Spild af penge til at sprede HPC ud – dele nationalt – tværfagligt. Windows – udfordringer fremfor Linux. Behov for forskellige HPC arkitekturer. anbefale forskningsrådene at man støtter store lokale regnekraft centre - under forudsætning af at pågældende HPC bliver drevet ordentligt. Brug for forskellige niveauer af HPC regnekraft. F.eks. betale med hardware for at få regnekraft adgang.

Forskningsnet/Sikkerhed: Login eller netværk kan sinke visualiseringsdelen under HPC beregninger hvilket kan være en flaskehals.

Datamanagement (DM)/Lagring: Behov for central storage server. Sikre back-up af data ! Backup er ikke systematisk sikret idag. Skrøbeligt. Svært at tage DM seriøst når det helt basale ikke er på plads. FAIR tilgangen: Sætter tvivl ved om arbejdet med at tage data/beskrivelse af data kan være brugbart for andre.

Kompetencebehov: Specialkurser feks. i programmering - masterservice – specialer. Nyttigt med HPC kompetencer i undervisning af studerende. Vigtigt med adgang til HPC til de studerende. Sårbare kompetencer hvis én person er HPC kyndig – også nyttigt med backup mandskabsmæssigt

UDKAST

Tabel 2. Oversigt over datatyper og behov for regnekraft.

| Datatyper | HUM | | | | TEK/NAT | | | | | | | | SUND | | SAMF | | | |
|--------------------------------------|-------|-----|-------|-------|---------|---------|---------|-----|-----|------|----------|---------|------|-------|-------|-----|------|----|
| | 5 | 8B | 9B | 11 | 1 | 4 | 6 | 7A | 8A | 9A | 9C | 10 | 7B | 8D | 2 | 3 | 8C | |
| Qualitative (Kategorier) | x | x | x | x | | x | x | x | | x | | x | x | x | x | x | x | |
| Kvantitative | x | x | x | x | x | | x | x | x | x | x | x | x | x | x | x | x | |
| Binære data | | | | | | x | x | | x | x | | x | | x | | | | |
| Tidsserier | | | x | x | | x | | | | x | x | | x | x | | x | x | |
| Enkelte observationer | | | | | x | | | | | | | | | | | | | |
| Højfrekvente/online data | | | | x | | x | | | | | | x | x | | | x | x | |
| Massive strukturerede data | | | | | x | x | x | x | x | x | x | | | x | x | | | |
| Heterogene ustrukturerede data | x | x | x | x | | x | | | | | | x | | | | | | |
| Databaser | x | x | x | x | | x | x | x | | | | | x | x | x | x | x | |
| Billeder | x | x | x | x | | x | | | | | | | x | | x | | | |
| Lyd/ Audio | x | | | x | | | | | | | | | | | | x | | |
| Film/ Video | x | x | x | x | | x | | | | | | | | | | x | x | |
| Personfølsomme data* | x (x) | (x) | x (x) | x (x) | | | | | | | | | x | x | x | x | x | |
| Geografiske data | | | x | | | | | | | | | | | x | x | x | | |
| Vejrudsigten bare i universet | | | | | x | | | | | | | | | | | | | |
| Satellitdata | | | | | | | | | | | | | | | | x | | |
| Spørgeskemaer | x | | x | | | | | | | | | | | | x | x | | |
| Tekst/ sprog | x | x | x | x | | | | | | | | | | | | x | | |
| Sociale medier | x | | x | x | | | | | | | | | | | | x | x | |
| IoT (Internet of Things) | x | | x | | | | | | | | | | | | | x | | |
| Sensordata (Temp, luft, eye, EKG...) | x | | x | | | x | | | | x | | | x | | | x | | |
| Økonomi data/ modellering | | | | | | | | | | | | | | | | | x | x |
| Kliniske data | | | | | | | | | | | | | x | x | | | | |
| Genomdata | | | | | | | | x | | | | | | | | x | | |
| Tekstfiler | | | x | x | | | | x | | | | | | | | x | | |
| Kombinerer tværfagligt | Ja | Ja | Ja | Ja | Nej | Ja | Nej | Nej | Nej | Nej | Nej | Ja | Nej | Ja | Ja | Ja | Nej | |
| Regnekraft | | | | | | | | | | | | | | | | | | |
| Bruger lokale/interne HPC idag | Ja | Nej | Ja | Nej | Ja | Ja | Ja | Ja | Ja | Ja | Ja | Ja | Ja | Ja | Ja | Ja | Nej | Ja |
| Bruger nationale HPC idag | Nej | Nej | Nej | Nej | Nej | Nej | Nej | Ja | Nej | Ja | Ja | Nej | Ja | Ja | Nej | Nej | Nej | |
| Behov/Kerner idag | Ja | Ja | Ja | Ja | 3.000 | 10.000 | 6.000 | 100 | 250 | <100 | 2.000 | 2.000 | <100 | 3.400 | 1-200 | Ja | <100 | |
| Behov/Kerner fremtid | Ja | Ja | Ja | Ja | 30.000 | 100.000 | >50.000 | Ja | Ja | Ja | >100.000 | >20.000 | Ja | Ja | Ja | Ja | Ja | |
| | KU | AU | SDU | RUC | KU | DTU | DTU | AAU | AU | SDU | SDU | DTU | AAU | AU | ITU | CBS | AU | |

* Registerdata, mikrodata (DS, Danmarks statistik), personhenførbare data; og for copyright data (x)

e-infrastruktur til forskning i verdensklasse i 2025

- en analyse og udredning vedrørende dansk e-infrastruktur samarbejde

Bilag 5:

HPC benchmark Top500 og BencHEIT

HPC

HPC – High Performance Computing – anvendes i analysen som et samlet begreb for regnekraft, der overstiger det en typisk desktop-computer kan levere, og som dermed kræver en større investering. HPC anvendes således gennemgående i rapporten som et samlende begreb for fx supercomputing, parallel- og serielcomputing og cloud computing. Når der i rapporten er behov for at skelne mellem de forskellige computingformer og deres anvendelse, vil de relevante begreber blive anvendt. Man har ønsket at benchmarke de nordiske lande samt Holland.

Metodeafsnit

Gennem en desk-top research har det vist sig, at der ikke findes én international standard til at kunne sammenligne HPC på tværs af lande og installationer. Det skyldes dels at genstandsområdet – HPC – ikke lader sig entydigt definere og dels forskellige opgørelsesmetoder.

Der er i denne analyse valgt 2 primære kilder til at kunne sammenligne de 5 landes HPC: Dels den internationalt anerkendte TOP500 liste over HPC¹, og dels omkostningsanalysen BencHEIT.

TOP500

I TOP500 opregnes computere sorteret efter deres præstationer på LINPACK Benchmark. LINPACK er et program som bliver kørt på den enkelte HPC installation af den lokale systemadministrator og resultatet bliver uploadet til TOP500. Præstationerne forsøges at blive verificeret fra brugere og leverandører, men der vil være fejl og der er forskel på hvor godt den enkelte HPC installation kan præsentere. Listen bliver opdateret halvårligt og på den måde vil man løbende få de nyeste installationer med

Hovedmålet med TOP500-listen er således at give en rangeret liste over HPC installationer hvor der kan køre avancerede applikationer til generelle formål. Ved generelle formål mener de at computersystemet skal kunne bruges til at løse en række videnskabelige problemer. Ethvert system, der er designet specielt til at løse LINPACK benchmarkproblemet eller har som hovedformål at opnå en høj TOP500 ranking bliver diskvalificeret fra benmarken.

TOP500 listen er således god indikator for hvor højt det enkelte anlæg yder og man er i stand til at sammenligne på tværs af mange forskellige teknologier og opsætninger. Selvom TOP500 listen indeholder HPC anlæg der skal kunne løse en række videnskabelig problemer, indeholder listen også anlæg som kun bruges af industrien. For Danmark kan man fx se at den hurtigste HPC i landet er et anlæg som kun bruges til industrielt formål.

Bencheit

BencHEIT² er en europæisk benchmarking af 50 europæiske universiteter hvor der sker en undersøgelse af IT-omkostninger, organisering og størrelse. BencHEIT arbejder på grundlag af frivilligt arbejde og total åbenhed - hvert deltagende universitet eller videregående uddannelsesinstitution - kan se alle data fra alle andre deltagere.

BencHEIT-undersøgelsen hjælper til at forstå universitetets it-omkostningsstruktur i tre dimensioner: Organisationsdimension: omkostninger hos centraliseret IT, andre centraliserede enheder (fx bibliotek, økonomi) og hos decentrale enheder på fakulteter og forskningscentre.

¹ Listen kan ses her, hvor det henvises til den senest opdateret pr. 13/11 2017: <https://www.top500.org/list/2017/11/>

² <http://www.eunis.org/task-forces/benchmarking/>

Servicedimension: f.eks. netværk, undervisning, audiovisuelt og HPC

Artskontodimension: løn, drift, konsulenter m.m

Universiteterne indsender oplysninger til BenchHEIT sekretariatet 1 gang om året på baggrund af regnskabstal fra forrige år. Dette øger validiteten af de indberettede tal, men derudover bliver tallene verificeret at det fælles sekretariat på Helsinki Universitet og af de andre brugere.

I forhold til denne sammenligning, skal det nævnes at alle de norske, svenske og finske universiteter er med i BenchHEIT. For Danmark og Holland gælder det dog at hvert land kun er repræsenteret med 2 universiteter. For disse lande er validiteten således lavere. Der er endvidere den begrænsning at HPC er en relativ ny omkostningstype i BenchHEIT som første gang blev indhentet i 2015.

Baggrunden for at vælge en kombination af 2 kilder skyldes den erfaring at der hos alle lande er både små og store HPC installationer. TOP500 styrke er at den med stor sandsynlighed afdækker de største HPC-anlæg, men inkluderer ikke de små, idet de er for små til at gøre sig gældende i TOP500. Modsat afdækker BenchHEIT de mindre og lokale anlæg som findes på universiteterne, men afdækker ikke nødvendigvis de største nationale anlæg, idet disse ikke nødvendigvis driftes af et universitet.

Benchmarking

Fra de 5 lande er der fra BenchHEIT identificeret 19 HPC anlæg og fra TOP500 er der identificeret 17 HPC anlæg.

| HPC | | | | | | |
|--|--------------------------|------------------|------------------|--------------------------|-----------------|-----------------|
| HPC in the Nordic countries and the Netherlands | | | | | | |
| All values are for 2016/2017 | | | | | | |
| HPC - High Performance Computing | | | | | | |
| | Total (all countries) | Denmark TOTAL | Finland TOTAL | The Netherlands TOTAL | Norway TOTAL | Sweden TOTAL |
| 61-TFlop/s (peak) CPU | 6.113 | 1.472 | 303 | 0 | 1.230 | 3.108 |
| 62-TFlop/s (peak) accelerators | 5.715 | 239 | 172 | 0 | 5.246 | 58 |
| 63-HPC related storage volume (in TB) | 32.497 | 10.373 | 3.087 | 200 | 4.179 | 14.658 |
| 64-Number of cores | 253.730 | 54.516 | 19.518 | 4.200 | 71.636 | 103.860 |
| 65-Total Power Consumption for HPC (kW) | 5.140 | 1.030 | 256 | 0 | 1.367 | 2.487 |
| Costs allocated to HPC | | | | | | |
| Investment in 2016/2017 or yearly depreciation (EURO) | 22.691.019,01 | 5.820.954,72 | 1.971.182,78 | 55.133,74 | 8.437.527,08 | 7.449.093,82 |
| Distributed computing | | | | | | |
| 66-Number of workstations for distributed computing on average | 587 | 0 | 245 | 0 | 71 | 271 |
| 67-How many CPU cores can be used for distributed computing on average | 2.231 | 0 | 650 | 0 | 332 | 1.249 |

For BenchHEIT er der hvert HPC anlæg indsendt oplysninger om teoretisk peak performance for CPU og GPU, hvor meget storage der er knyttet til anlæggene, antal af cores, total strømforbrug samt direkte omkostninger knyttet til anlægget. I opgørelsesmetoden har man enten opgjort omkostninger som den årlige afskrivning (der er typisk en afskrivningsprofil på 4 år for et HPC anlæg), eller den omkostning man har haft det pågældende år som en engangsinvestering.

Udover HPC er der også indsendt oplysninger om distributed computing. Hvis man har det, hvor mange computere der indgår i dette i gennemsnit og hvor mange CPU-cores som er anvendt.

For TOP500 er stamoplysninger som operatør, land, by og etableringsår oplyst. Derudover er resultat fra LINPACK Benchmark og bestykning.

Nedenfor er de enkelte lande opgjort

Holland

Hollands største HPC er SURFsaras Carteius 2 med pt. en 142 plads på TOP500.

SURFsara

| | |
|---------|---|
| URL | https://www.surfsara.nl/ |
| Segment | Academic |
| City | Amsterdam |
| Country | Netherlands |

SYSTEMS HISTORY

| System | Year | Vendor | Cores | Rmax (GFlop/s) | Rpeak (GFlop/s) |
|--|------|--------|--------|----------------|-----------------|
| Cartesius 2 - Bullx DLC B710/B720 Blades, Xeon E5-2695 v2 2.4GHz E5-2690v3 2.6GHz, Infiniband FDR | 2015 | | 38,880 | 1,088,510 | 1,327,104 |
| Cartesius - Bullx DLC B710/B720 Blades, Xeon E5-2695 v2 2.4GHz E5-2690v3 2.6GHz, Infiniband FDR | 2013 | | 38,880 | 1,052,510 | 1,327,104 |
| Cartesius Accelerator Island - Bullx B515 cluster, Intel Xeon E5-2450v2 8C 2.5GHz, InfiniBand 4x FDR, Nvidia K40m | 2014 | | 3,036 | 153,600 | 209,880 |

Det nationale niveau er organiseret i 3 selskaber under SURFsara: 1 for HPC, 1 for eScience og et for net/services. Det nationale anlæg er finansieret af ministeriet via en roadmap, og adgang og politik for forskerne til de nationale anlæg er defineret centralt via en videnskabelig og en teknisk styregruppe.

HPC

HPC in the Nordic countries and the Netherlands

All values are for 2016/2017

| HPC - High Performance Computing | Total | The Netherlands | |
|--|-----------------|-----------------|-----------|
| | (all countries) | TOTAL | NL_TUe |
| 61-TFlop/s (peak) CPU | 6.113 | 0 | |
| 62-TFlop/s (peak) accelerators | 5.715 | 0 | |
| 63-HPC related storage volume (in TB) | 32.497 | 200 | 200 |
| 64-Number of cores | 253.730 | 4.200 | 4.200 |
| 65-Total Power Consumption for HPC (kW) | 5.140 | 0 | |
| Costs allocated to HPC | | | |
| Investment in 2016/2017 or yearly depreciation (EURO) | 22.691.019,01 | 55.133,74 | 55.133,74 |
| Distributed computing | | | |
| 66-Number of workstations for distributed computing on average | 587 | 0 | |
| 67-How many CPU cores can be used for distributed computing on average | 2.231 | 0 | |

Adgang til det lokale anlæg sker via lokal adgang og finansieret lokalt.

Sammenligner man dette ene anlæg med de 2 nationale anlæg på i alt ca. 74.000 cores, er det lokale anlæg ubetydeligt.

Sverige

Den største installation i Sverige er KTHs Bekow som pt. er nr 69 på TOP500. Derudover har de 1 installation mere med på TOP500, Kepnekaise, med en placering 490.

[Home](#) / [KTH - Royal Institute of Technology](#)

KTH - Royal Institute of Technology

| | |
|----------------|-----------------------|
| URL | http://www.pdc.kth.se |
| Segment | Academic |
| City | Stockholm |
| Country | Sweden |

SYSTEMS

HISTORY

| System | Year | Vendor | Cores | Rmax (GFlop/s) | Rpeak (GFlop/s) |
|--|------|--------|--------|----------------|-----------------|
| Beskow - Cray XC40, Xeon E5-2695v4/E5-2698v3 16C 2.3GHz, Aries interconnect | 2017 | | 67,456 | 1,802,510 | 2,438,140 |
| Lindgren - Cray XE6, Opteron 12 Core 2.10 GHz, Custom | 2011 | | 36,384 | 237,200 | 305,625.6 |
| Cray XT6m 12-Core 2.1 GHz | 2010 | | 11,016 | 75,120 | 92,534.4 |
| Ekman - PowerEdge SC1435 Dual core Opteron 2.2 GHz, Infiniband | 2010 | | 9,800 | 61,760 | 86,024.4 |

HPC2N - Umea University

| | |
|----------------|--------------------------|
| URL | http://www.hpc2n.umu.se/ |
| Segment | Academic |
| City | Umea |
| Country | Sweden |

SYSTEMS

HISTORY

| System | Year | Vendor | Cores | Rmax (GFlop/s) | Rpeak (GFlop/s) |
|--|------|--------|--------|----------------|-----------------|
| Kebnekaise - Lenovo NeXtScale nx360M5, Xeon E5-2690v4 14C 2.6GHz/Xeon Phi 7250, Infiniband FDR, Tesla K80 | 2015 | | 17,632 | 438,841 | 749,588 |

Begge er finansieret af Sverige nationale HPC – SNIC. I Sverige er man i gang med at reorganisere området. På nuværende tidspunkt har man afklaret finansiering af de fremtidige anlæg med 50% finansiering fra ministeriet og 50% fra universiteterne.

HPC

HPC in the Nordic countries and the Netherlands
All values are for 2016/2017

| HPC - High Performance Computing | Total | Sweden | | | | | | |
|--|-----------------|--------------|-------|--------------|--------------|--------|------------|----------|
| | (all countries) | TOTAL | SE_GU | SE_LIU | SE_KTH | SE_ORU | SE_UU | SE_UMU |
| 61-TFlop/s (peak) CPU | 6.113 | 3.108 | | 945 | 2.039 | 4 | 120 | |
| 62-TFlop/s (peak) accelerators | 5.715 | 58 | | 11 | 45 | | 2 | |
| 63-HPC related storage volume (in TB) | 32.497 | 14.658 | 428 | 2.000 | 5.000 | 30 | 7.200 | |
| 64-Number of cores | 253.730 | 103.860 | 748 | 40.000 | 55.432 | 80 | 7.600 | |
| 65-Total Power Consumption for HPC (kW) | 5.140 | 2.487 | | 1.100 | 763 | 374 | 250 | |
| Costs allocated to HPC | | | | | | | | |
| Investment in 2016/2017 or yearly depreciation (EURO) | 22.691.019,01 | 7.449.093,82 | | 4.389.045,26 | 2.771.174,56 | | 288.874,00 | 1.238,49 |
| Distributed computing | | | | | | | | |
| 66-Number of workstations for distributed computing on average | | 587 | 271 | 271 | | | | |
| 67-How many CPU cores can be used for distributed computing on average | | 2.231 | 1.249 | 1.249 | | | | |

Det viser sig her at de lokale de nationale anlæg SE_KTH og SE_LIU er væsentlig større end de øvrige anlæg.

Finland

Finlands største HPC anlæg er CSCs Sisu med pt. en plads 107 på TOP500

Home / CSC (Center for Scientific Computing)

CSC (Center for Scientific Computing)

| | |
|---------|----------------------------|
| URL | http://www.csc.fi/english/ |
| Segment | Academic |
| City | Espoo |
| Country | Finland |

SYSTEMS

HISTORY

| System | Year | Vendor | Cores | Rmax (GFlop/s) | Rpeak (GFlop/s) |
|---|------|--------|--------|----------------|-----------------|
| Sisu - Cray XC40, Xeon E5-2690v3 12C 2.6GHz, Aries interconnect | 2014 | | 40,608 | 1,250,000 | 1,689,292.8 |
| Apollo 6000 XL230a, Xeon E5-2690v3 12C 2.6GHz, Infiniband FDR | 2014 | | 12,960 | 456,648 | 539,136 |
| Cluster Platform SL230s Gen8, Xeon E5-2670 8C 2.600GHz, Infiniband FDR | 2012 | | 9,200 | 163,900 | 191,360 |
| Cray XT5/XT4 QC 2.3 GHz | 2009 | | 10,864 | 76,510 | 101,996.8 |
| Cray XT5/XT4 QC 2.3 GHz | 2008 | | 10,816 | 76,510 | 99,507 |
| Vuori - Cluster Platform 4000 BL465c, Opteron 6C 2.6GHz, Infiniband DDR | 2010 | | 3,264 | 27,741 | 33,946 |

Nedenstående er en detaljeret visning af de lokale anlæg i Finland.

| HPC | | | | | | | |
|--|-----------------|--------------|--------------|------------|-----------|------------|-----------|
| HPC in the Nordic countries and the Netherlands | | | | | | | |
| All values are for 2016/2017 | | | | | | | |
| HPC - High Performance Computing | Total | Finland | | | | | |
| | (all countries) | TOTAL | FI_Aalto | FI_Centria | FI_LUT | FI_TUT | FI_UH |
| 61-TFlop/s (peak) CPU | 6.113 | 303 | 137 | 0 | 12 | 93 | 61 |
| 62-TFlop/s (peak) accelerators | 5.715 | 172 | 57 | | 9 | 55 | 51 |
| 63-HPC related storage volume (in TB) | 32.497 | 3.087 | | 61 | 126 | 500 | 2.400 |
| 64-Number of cores | 253.730 | 19.518 | 6.200 | 60 | 672 | 3.600 | 8.986 |
| 65-Total Power Consumption for HPC (kW) | 5.140 | 256 | 150 | 1 | 18 | 80 | 7 |
| Costs allocated to HPC | | | | | | | |
| Investment in 2016/2017 or yearly depreciation (EURO) | 22.691.019,01 | 1.971.182,78 | 1.393.740,94 | | 61.895,92 | 429.007,56 | 86.538,36 |
| Distributed computing | | | | | | | |
| 66-Number of workstations for distributed computing on average | 587 | 245 | | | 25 | 220 | |
| 67-How many CPU cores can be used for distributed computing on average | 2.231 | 650 | | | 100 | 550 | |

Sammenligner man disse anlæg med de finske nationale anlæg (se afsnit om TOP500) med samlet ca. 60.000 cores, vil man se at disse anlæg er relativ små, men samlet har de 19.518 cores.

Norge

Norges største er Simga2s FRAM med en 202 plads på TOP500

Home / UNINETT Sigma2 AS

UNINETT Sigma2 AS

| | |
|---------|------------------------|
| URL | https://www.sigma2.no/ |
| Segment | Academic |
| City | Trondheim |
| Country | Norway |

SYSTEMS

HISTORY

| System | Year | Vendor | Cores | Rmax (GFlop/s) | Rpeak (GFlop/s) |
|---|------|--------|--------|----------------|-----------------|
| Fram - Lenovo NeXtScale nx360M5, Xeon E5-2683 v4 16C 2.1GHz, Infiniband EDR | 2017 | | 32,192 | 953,571 | 1,081,651 |

Det nationale anlæg er finansieret af ministeriet og forskernes adgang sker gennem en central styregruppe. Der er planlagt 2 nationale anlæg.

Nedenstående er en sammenligning af de lokale norske HPC anlæg

HPC

HPC in the Nordic countries and the Netherlands
All values are for 2016/2017

| HPC - High Performance Computing | Total | Norway | | | | | |
|--|-----------------|--------------|--------------|--------------|--------------|--------|--------------|
| | (all countries) | TOTAL | NO_NTNU | NO_UIB | NO_UIO | NO_UIS | NO_UIT |
| 61-TFlop/s (peak) CPU | 6.113 | 1.230 | 467 | 205 | 258 | | 300 |
| 62-TFlop/s (peak) accelerators | 5.715 | 5.246 | 5.196 | 0 | 50 | | |
| 63-HPC related storage volume (in TB) | 32.497 | 4.179 | 1.000 | 579 | 600 | | 2.000 |
| 64-Number of cores | 253.730 | 71.636 | 22.464 | 22.272 | 12.900 | | 14.000 |
| 65-Total Power Consumption for HPC (kW) | 5.140 | 1.367 | 500 | 342 | 275 | | 250 |
| Costs allocated to HPC | | | | | | | |
| Investment in 2016/2017 or yearly depreciation (EURO) | 22.691.019,01 | 8.437.527,08 | 2.890.046,34 | 1.838.411,68 | 2.519.816,34 | | 1.189.252,72 |
| Distributed computing | | | | | | | |
| 66-Number of workstations for distributed computing on average | 587 | 71 | 60 | | | | 11 |
| 67-How many CPU cores can be used for distributed computing on average | 2.231 | 332 | 32 | | | | 300 |

Det viser sig her at det norske lokale anlæg har samlet set 71.636 cores mod det nye nationale anlæg FRAM på 32.000 cores. De lokale anlæg udgør her en betydelig del af det samlede antal cores.

(ps redegørelse om at de norske anlæg er ved at blive konsolideret til 2 anlæg.

Danmark

Danmarks har ikke længere HPC anlæg på TOP500. På forrige liste var det største anlæg Abacus 2.0

Home / University of Southern Denmark

University of Southern Denmark

| | |
|---------|---------------------|
| URL | http://deic.sdu.dk/ |
| Segment | Academic |
| City | Odense |
| Country | Denmark |

SYSTEMS

HISTORY

| System | Year | Vendor | Cores | Rmax (GFlop/s) | Rpeak (GFlop/s) |
|--|------|--------|--------|----------------|-----------------|
| Abacus 2.0 - Lenovo NeXtScale nx360M5, Xeon E5-2680v3 12C 2.5GHz, Infiniband FDR, NVIDIA Tesla K40 | 2015 | | 17,928 | 462,355 | 836,640 |

For Deic Nationale anlæg er finansieringen sket mellem DeIC og flere universiteter. Adgang for forskere til de nationale anlæg sker gennem lokale styregrupper.

HPC

HPC in the Nordic countries and the Netherlands

All values are for 2016/2017

HPC - High Performance Computing

| | Total | Denmark | | | |
|---|-----------------|---------|--------|--------|--|
| | (all countries) | TOTAL | DK_DTU | DK_SDU | |
| 61-TFlop/s (peak) CPU | 6.113 | 1.472 | 900 | 572 | |
| 62-TFlop/s (peak) accelerators | 5.715 | 239 | | 239 | |
| 63-HPC related storage volume (in TB) | 32.497 | 10.373 | 9.000 | 1.373 | |
| 64-Number of cores | 253.730 | 54.516 | 40.000 | 14.516 | |
| 65-Total Power Consumption for HPC (kW) | 5.140 | 1.030 | 700 | 330 | |

Costs allocated to HPC

| | | | | | |
|---|---------------|--------------|--------------|--------------|--|
| Investment in 2016/2017 or yearly depreciation (EURO) | 22.691.019,01 | 5.820.954,72 | 4.155.068,50 | 1.665.886,22 | |
|---|---------------|--------------|--------------|--------------|--|

Distributed computing

| | | | | | |
|--|-------|---|--|--|--|
| 66-Number of workstations for distributed computing on average | 587 | 0 | | | |
| 67-How many CPU cores can be used for distributed computing on average | 2.231 | 0 | | | |

Vi bør bruge de samme tal her som er præsenteret for bestyrelsen

Organisering

For drift af de største anlæg har Norge, Holland og Finland valgt at etablere selvstændige nationale organisationer som er operatører på de nationale anlæg. Sverige og Danmark har på nuværende tidspunkt driftsaftaler som udvalgte universiteter til at stå for driften.

e-infrastruktur til forskning i verdensklasse i 2025

- en analyse og udredning vedrørende dansk e-infrastruktur samarbejde

Bilag 6:

Beskrivelse af relevante internationale/global e-infrastrukturer.

Internationale e-infrastrukturer

1.1 Forskningsnet

1.1.1 Global R&E Network CEO Forum

| | |
|--|---|
| Formål | Global R&E Network CEO Forum (i daglig tale: "CEO Forum") blev etableret i september 2012 som fællesskab for direktørerne for verdens ledende forskningsnet. Som en del af de stiftende medlemmer tager NORDUnet aktiv del i udviklingen og ledelsen af dette fællesskab, der består af 17 direktører/ledere af forskningsnetsorganisationer. CEO forum er et udmærket eksempel på hvordan nordisk samarbejde gennem NORDUnet sikrer at de nordiske forskningsnet, incl. DeiC, opnår indflydelse på den globale udvikling. Hvis ikke de fem nordiske lande havde arbejdet sammen ville der ikke have været noget nordisk medlemskab af CEO forum. |
| Dansk involvering | NORDUnets administrerende direktør rapporterer om udviklingen i CEO forum på alle NORDUnet bestyrelsesmøder og tager input med derfra. Selvom CEO Forum er en lukket gruppe, så er dens arbejdsgrupper åbne for personale fra alle forskningsnet i verden. En af de mest aktive grupper under CEO forum står for "In the Field Stories" (https://www.inthefieldstories.net/), der er en blog om de resultater som forskere og undervisere på hele kloden opnår ved hjælp af forskningsnettene. DeiC og Danmark optræder indtil nu i 10 historier på In the Field Stories. |
| Vurdering af fremtidig involvering både teknisk og økonomisk | Selvom CEO Forum drives gennem en frivillig indsats fra dets medlemmer er det tydeligt at aktivitetsniveauet enten vil fortsætte på det nuværende høje niveau eller sågar øges. Grunden skal ses i at forskning i større og større grad er baseret på verdensomspændende samarbejder, der kræver global koordinering forskningsnettene imellem. |
| Forventet udvikling og implikationer | Medlemmerne af CEO Forum samarbejder om at identificere områder, hvor det giver global værdi at igangsætte udviklingsarbejde eller aktiviteter. CEO Forum vil: Rådspørge og samarbejde med den globale forskningsnet fællesskab, arbejdsgrupper og brugere; Accelerere idriftsættelse og brug af løsninger ved at understrege betydningen for samarbejdet; overføre resultaterne - hvis nødvendigt - til de relevante organisationer og arbejdsgrupper i det globale forskningsnetfællesskab for at de globalt kan tage implementere løsningen til fordel for alle. |

1.2 Datamanagement og -lagring

1.2.1 Zenodo

| | |
|--|--|
| Formål | Open source generisk lagerinfrastruktur (repositorie) udviklet og drevet af CERN og støttet af EU (OpenAIRE), der faciliterer langtidsbevaring af og open access til forskningsdata, publikationer oa. forskningsresultater herunder computerkode. Alle forskere (også non-EU) kan benytte gratis servicen (op til 50GB pr. datasæt). Understøtter alle dele (og næsten alle facetter) af FAIR (http://about.zenodo.org/principles/) og specifikt anbefalet i EUs H2020 data management guide, samt e-IRGs best practise guide til europæiske forskning (http://e-irg.eu/documents/10920/363494/2017-Supportdocument.pdf) |
| DKs nuværende involvering og økonomisk og ressourcemæssigt | DeiC har et teknisk samarbejde med Zenodo ifm. udviklingen data.deic.dk - en lagringsinfrastruktur til aktive forskningsdata, hvor forskeren i den kommende version kan eksportere fx resultatdata (data der ligger til grund for en publikation) herfra og til Zenodo mhp. langtidsbevaring og indeksering. |

| | |
|---|---|
| Vurdering af hensigtsmæssig varetagelse af deltagelsesniveau teknisk og økonomisk | Zenodo er ikke et projekt, som lande kan melde sig ind i, men det er muligt at bidrage økonomisk og med kode via Github. |
| Forventet udvikling og implikationer | Oprindeligt var Zenodo beregnet til at løse behovet for de forskere, der ikke har adgang til egnet lagerstruktur enten på egen institution eller inden for deres forskningsdomæne. I dag er servicen - bl.a. pga. den høje grad af FAIR-understøttelse og samspil med andre væsentlige infrastrukturer som fx Crossref - et attraktivt bill-board for open data og open access publikationer, dvs. forskere lægger deres forskningsresultater her også selvom de allerede har dem lagret andetsteds. Værd at bemærke: Zenodo er allerede ledende i verden mht. udstedelse af DOI'er til computerkode. Servicen linker EU-støttede forskningsresultater med de associerede forskningsbevillinger - denne service udvides til også at omfatte non-EU bevillingsgivere (National Science Foundation (USA), Wellcome Trust (UK) m.fl.), så scope er absolut internationalt. Zenodo som infrastruktur er helt uomgængelig i datamanagement sammenhænge. |

1.2.2 DataCite

| | |
|---|--|
| Formål | International non-profit organisation, der driver en netbaseret infrastruktur der leverer globalt unikke, persistente og anerkendte identifikatorer (digital object identifiers/DOI'er) til datasæt oa. dataobjekter, således at de er unikt identificerbare og citerbare. Persistente identifikatorer såsom DataCite DOI'er er en absolut forudsætning for FAIR (meta)data (FAIR princip #1). |
| DKs nuværende involvering og økonomisk og ressourcemæssigt | Danmark er medlem i DataCite , idet DTU er 'dansk allocator' (siden 2009), dvs. bindeled mellem den internationale organisation DataCite og de danske institutioner. 1.12.2017 overgår denne rolle til DeIC. Udgiften til medlemskab og administration er 140.000 DKK/år. |
| Vurdering af hensigtsmæssig varetagelse af deltagelsesniveau teknisk og økonomisk | Deltagelsesniveauet vil være hensigtsmæssigt varetaget med de nuværende 140.000 DKK/år og DeICs funktion som allocator. |
| Forventet udvikling og implikationer | Anvendelsen af DataCite DOI'er i Danmark har udviklet sig roligt, men i takt med at fokuset på data management er steget, er også interessen for DataCite øget tilsvarende. Det er rettidig omhu, at servicen nu udbydes nationalt som et essentielt bidrag til at sikre danske forskningsdatas genfindelighed. |

1.2.3 DMPonline

| | |
|--|--|
| Formål | Webbaseret værktøj, der med udgangspunkt i skabeloner og eventuelt bevillingsgiveres krav til data management planer hjælper forskere med at skrive en datamanagementplan. Data management planer er med til facilitere FAIR data og open science ved at sikre, at et forskningsprojekt har taget stilling til, hvilke data der indsamles/genereres/bruges hvordan samt den senere adgang til dem. |
| DKs nuværende involvering og økonomisk og ressourcemæssigt | DMPonline er open source udviklet i UK af Digital Curation Centre (mangeårigt europæisk fyrtårn inden for data management). DMPonline drives i UK af University of Edinburgh for de britiske universiteter og forskningsinstitutioner. I Danmark er en pilotinstallation af DMPonline, som blev startet af Data management i praksis-projektet, pr. 1.1.2017 blevet flyttet over på DeICs infrastruktur (dmponline.deic.dk) og sat i egentlig drift som en national service. Herunder har lokale administratorer på universiteterne mulighed for at tilføre lokale skabeloner og hjælpetekster. Driften er organiseret i et samarbejde mellem DeIC – som teknisk operatør – og DTU Bibliotek og Det Kgl. Bibliotek som service |

| | |
|---|---|
| | manager. DEFF har støttet etablering af drifts- og supportorganisation, med deltagelse af en række universitetsbiblioteker, som et konsolideringsprojekt der løber frem til 31.12.2018. |
| Vurdering af hensigtsmæssig varetagelse af deltagelsesniveau teknisk og økonomisk | Hvis der efter 1.1.2019 skal være en national service, skal der tilvejebringes driftsmidler. Det ovenfor nævnte konsolideringsprojekt skal bl.a. estimere omfanget heraf. |
| Forventet udvikling og implikationer | Flere bevillingsgivere kræver datamanagementplaner. Sådanne planer er bl.a. et krav for projekter der deltager under Open access to Research Data piloten (ORD pilot), der fra 2017 er udvidet til at dække alle tematiske områder under Horizon 2020. En infrastruktur til at facilitere udformning af datamanagementplaner er således svær at komme udenom, og DMPonline er et godt bud, da specifikt denne infrastruktur er nævnt i EC's Guidelines on FAIR Data Management in Horizon 2020 . Den danske installation af DMPonline, som er tilrettet danske forhold, bør forblive i regulær drift fra 31.12.2018. Der skal skaffes finansiering til at installationen løbende vedligeholdes og opgraderes i takt med at kravene fra bl.a. fundere til data management og FAIR data udvikler sig, også i Danmark. |

1.2.4 ORCID

| | |
|---|---|
| Formål | ORCID, Open Researcher and Contributor ID, er et internationalt udbredt og unikt forsker-id, som følger en forsker uanset skift af arbejdssted, stilling eller efternavn/stavemåde, så forsker, forskningsaktiviteter og -resultater er entydigt linkede og korrekte, hvilket er essentielt ift. at understøtte meriterende citering. Løsningen har støtte fra anerkendte universiteter og etablerede forlag og bevillingsgivere verden over. Pt. er der 4+ mio. aktive ORCID id'er, der sammenkæder 25 mio. forskningsaktiviteter. ORCID er således central infrastruktur. |
| DKs nuværende involvering og økonomisk og ressourcemæssigt | Syv ud af otte danske universiteter er med i det nationale ORCID konsortium, som har implementeret ORCID således at en forsker kan oprette sit ORCID id via sin institutions forskningsdatabase (PURE integration). Danmarks Elektroniske Fag- og Forskningsbibliotek, DEFF, er internationalt ORCID konsortiemedlem. |
| Vurdering af hensigtsmæssig varetagelse af deltagelsesniveau teknisk og økonomisk | DEFFs konsortiemedlemskab sikrer koordinering ift. det internationale, organisatoriske niveau. Derudover er det tekniske og økonomiske deltagelsesniveau i ORCID op til den enkelte institution at sikre, gennem formidlingsindsatser og fortsat integration i forskningsdatabasen. |
| Forventet udvikling og implikationer | Tilslutningen til ORCID blandt forskere i Danmark er støt stigende og to universiteter er allerede nu over 80% i VIP-tilslutning. Horizon 2020 vil automatisk drive tallet endnu højere op, idet fx "Guidelines on Open Access to Scientific Publications and Research Data in Horizon 2020" henstiller til at bidragydere anvender et persistent id og nævner specifikt ORCID som eksempel på egnet infrastruktur. Ligeledes efterspørger bevillingsgivere stadig oftere ORCID i forbindelse med indsendelsen af ansøgningen - dette gælder også nationalt, fx Novo Nordisk Fonden . |

1.3 Sikkerhed

1.3.1 FIRST

| | |
|--------|--|
| Formål | FIRST, the Forum of Incident Response and Security Teams, blev etableret i 1990, og er et samlingspunkt for sikkerhedsorganisationer, såvel offentlige som kommercielle fra hele verden. |
|--------|--|

| | |
|---|--|
| Dks nuværende involvering og økonomisk og ressourcemæssigt | DKCERT er medlem af FIRST og deltager i møder. |
| Vurdering af hensigtsmæssig varetagelse af deltagelsesniveau teknisk og økonomisk | Deltagelse bør fortsætte uændret for at sikre international koordinering af viden og hændelsehåndtering. |
| Forventet udvikling og implikationer | |

e-infrastruktur til forskning i verdensklasse i 2025

- en analyse og udredning vedrørende dansk e-infrastruktur samarbejde

Bilag 7:

EOSC ministerrådserklæring 27. maj 2015



Bruxelles, den 27. maj 2016
(OR. en)

9526/16

RECH 208
TELECOM 100

RESULTAT AF DRØFTELSENE

fra: Generalsekretariatet for Rådet

til: delegationerne

Tidl. dok. nr.: 8791/16 RECH 133 TELECOM 74

Vedr.: Rådets konklusioner om overgangen til et system for åben videnskab
(vedtaget 27/05/2016)

Vedlagt følger til delegationerne Rådets konklusioner om overgangen til et system for åben videnskab vedtaget af Rådet den 27. maj 2016 på 3470. samling.

.

**RÅDETS KONKLUSIONER OM OVERGANGEN TIL ET SYSTEM FOR ÅBEN
VIDENSKAB**

RÅDET FOR DEN EUROPÆISKE UNION,

SOM MINDER OM

- Kommissionens henstilling af 17. Juli 2012 om "Adgang til og bevaring af videnskabelig information"¹ og meddelelsen af 17. Juli 2012 om "Bedre adgang til videnskabelig information: Større udbytte af de offentlige investeringer i forskning"²
- de generelle principper i Horisont 2020³, der sikrer fri adgang til videnskabelige publikationer og fremmer fri adgang til forskningsdata fra offentligt finansieret forskning på EU-plan;
- Kommissionens rapport fra februar 2015: "Validation of the results of the public consultation on Science 2.0: Science in Transition"⁴
- Kommissionens meddelelse af 6. maj 2015 om "En strategi for et digitalt indre marked i EU"⁵, der anerkender betydningen af videnskab og forskning for stimulering af innovation, navnlig gennem sit mål om at forbedre rammebetingelserne for datadrevet forskning

¹ Dok. 12846/12.

² Dok. 12847/12.

³ Europa-Parlamentets og Rådets forordning (EU) nr. 1291/2013 af 11. december 2013 om Horisont 2020 - rammeprogram for forskning og innovation (2014-2020).

⁴ Rapporten findes på: http://ec.europa.eu/research/consultations/science-2.0/science_2_0_final_report.pdf#view=fit&pagemode=none

⁵ Dok. 8672/15.

- sine konklusioner af 29. maj 2015 om et "europæisk forskningsrum 2015-2020"⁶, som vedtager køreplanen for EFR, og understreger betydningen af at indlede gennemførelsen af de højt prioriterede tiltag, der er fastlagt i køreplanen inden medio 2016, herunder for prioritet 5 "optimal formidling af viden";
- sine konklusioner af 29. maj 2015 om "åben, dataintensiv og netværksbaseret forskning som drivkraft for hurtigere og bredere innovation"⁷, som udtrykker politisk støtte til at skabe de rette rammevilkår for forskning og innovation med fokus på big data og til at styrke hele dataværdikæden i Europa.

Åben videnskab

1. ANERKENDER, at den eksponentielle vækst i data, tilgængeligheden af stadig mere kraftfulde digitale teknologier, globaliseringen af det videnskabelige samfund samt den stigende efterspørgsel fra samfundet efter at løse vor tids samfundsmæssige udfordringer, danner grundlaget for en igangværende forvandling og åbning af videnskab og forskning, også omtalt som "åben forskning", som påvirker måden at forske og organisere videnskaben på.
2. ANERKENDER, at åben videnskab har potentiale til at øge kvaliteten, konsekvenserne og udbyttet af videnskab og til at fremskynde udviklingen af viden ved at gøre den mere pålidelig, mere effektiv og nøjagtig, mere forståelig for samfundet og lydhør over for samfundsmæssige udfordringer, og at den har potentiale til at skabe vækst og innovation, ved at alle interessenter på alle niveauer i samfundet kan genanvende videnskabelige resultater, hvilket i sidste ende kan bidrage til vækst og konkurrenceevne i Europa.

⁶ Dok. 8975/15.

⁷ Dok. 9360/15.

3. UNDERSTREGER, at åben videnskab bl.a. indebærer fri adgang til videnskabelige publikationer og optimal genanvendelse af forskningsdata, borgervidenskab og forskningsintegritet; NOTERER SIG, at fri adgang til videnskabelige publikationer og optimal genanvendelse af forskningsdata er af afgørende betydning for udviklingen af åben videnskab; NOTERER SIG ERAC's udtalelse om åbne forskningsdata⁸ og opfordringen til handling fra Amsterdam, som blev fremsat i anledning af det nederlandske formandskabs konference om åben videnskab⁹, og forskellige initiativer, der forfølges af medlemsstaterne, Kommissionen og interessenter¹⁰; OPFORDRER ERAC's stående arbejdsgruppe om åben videnskab og innovation til at vurdere de aktioner, der er foreslået som svar på opfordringen til handling fra Amsterdam, med hensyn til gennemførlighed, effektivitet og prioritering, og rapportere herom; UNDERSTREGER nødvendigheden af en koordineret indsats fra alle involverede parter: Kommissionen, medlemsstaterne, civilsamfundet og interessenterne¹¹

Platform for åben videnskabspolitik og europæisk dagsorden for åben videnskab

4. NOTERER SIG Kommissionens oprettelse af en platform for åben videnskabspolitik¹², som tager sigte på at støtte videreudviklingen af den europæiske politik for åben videnskab og fremme interessenternes anvendelse af god praksis, herunder spørgsmål som f.eks. tilpasning af belønnings- og evalueringssystemer, alternative modeller for publicering med fri adgang og forvaltning af forskningsdata (herunder arkivering), altmetrics, vejledende principper for optimal genanvendelse af forskningsdata, udvikling og anvendelse af standarder og andre aspekter af åben videnskab som f.eks. fremme af integritet i forskningen og udvikling af borgervidenskab; OPFORDRER Kommissionen til regelmæssigt at underrette medlemsstaterne og interessenterne om den igangværende udvikling og resultaterne af platformen for åben videnskabspolitik mindst to gange om året.

⁸ Dok. ERAC 1202/16.

⁹ Det nederlandske formandskabs konference om åben videnskab - Fra vision til handling, Amsterdam, den 4.-5. april 2016.

¹⁰ F.eks. udvikling af dagsordener og politikker for åben videnskab på nationalt plan.

¹¹ Interessenterne omfatter forskere (universiteter, forsknings- og teknologioorganisationer og virksomheder), finansieringsorganisationer, forlæggere og servicevirksomheder.

¹² <http://ec.europa.eu/research/openscience/index.cfm?pg=open-science-policy-platform>

5. OPFORDRER Kommissionen til i samarbejde med platformen for åben videnskabspolitik og i tæt samarbejde med medlemsstaterne og interessenterne og under hensyntagen til allerede eksisterende initiativer at videreudvikle den europæiske dagsorden for åben videnskab med henblik på yderligere at øge konsekvenserne af videnskaben for samfundet som helhed, ansøre til en effektiv mobilisering af enhver potentiel viden fra aktioner vedrørende åben videnskab og imødekomme kravene til uddannelse med hensyn til IKT og datafærdigheder, navnlig faglige dataforvaltningsfærdigheder, som kan bidrage til at give borgerne adgang til forskningsresultaterne.

Fjerne hindringer og skabe incitament

6. ER ENIGT I, at resultaterne af offentligt finansieret forskning bør gøres tilgængelige på en så åben måde som muligt, og ANERKENDER, at unødvendige juridiske, organisatoriske og finansielle hindringer for adgang til resultaterne af offentligt finansieret forskning, i den grad det er muligt og hensigtsmæssigt, bør fjernes for at opnå optimal vidensdeling, om nødvendigt under hensyntagen til behovet for udnyttelse af resultater; OPFORDRER Kommissionen og medlemsstaterne til at arbejde tættere sammen med tredjelande for at fremskynde overgangsprocessen mod åben videnskab og for at sikre gensidige fordele med hensyn til fri adgang til videnskabelige publikationer og optimal genanvendelse af forskningsdata i en global sammenhæng.

7. MENER, at vurdering af videnskabelig kvalitet bør bygge på selve arbejdet og udvides til at omfatte en vurdering af konsekvenserne af videnskab for samfundet som helhed, mens det nuværende fokus er på indikatorer, der er baseret på optællinger af citater fra tidsskrifter og publikationer; OPFORDERER Kommissionen, medlemsstaterne og interessenterne til på en koordineret måde at videreudvikle og gennemføre initiativer til bedre kvalitetssikring i deres kontrol- og evalueringssystemer; UNDERSTREGER, at der bør iværksættes incitamentsordninger for at belønne forskere (og forskningsinteressenter) for at udveksle resultaterne af deres forskning til genanvendelse; OPFORDERER Kommissionen, platformen for åben videnskabspolitik, medlemsstaterne og interessenterne til at undersøge mulige ordninger, herunder uddannelsesaktiviteter og bevidstgørende foranstaltninger, for at ændre de måder, som videnskab udføres på; OPFORDERER Kommissionen til gennem platformen for åben videnskabspolitik at samarbejde om navnlig fremme af et internationalt accepteret system til datacitation, der gør brug af allerede eksisterende initiativer og ekspertise.
8. GLÆDER SIG OVER Kommissionens meddelelse af 9. december 2015 med titlen ”På vej mod en tidssvarende, mere europæisk ramme for ophavsret”¹³ og SER FREM TIL det lovgivningsforslag fra Kommissionen, der blev bebudet heri; UNDERSTREGER betydningen af at styrke EU's konkurrenceevne og teknologiske og videnskabelige lederskab, hvilket kunne omfatte at tillade offentlige interesseforskningsorganisationer at udføre tekst- og dataudvinding af indhold, som de allerede har juridisk ret til, med henblik på videnskabelige forskningsformål; UNDERSTREGER behovet for fortsat støtte fra Kommissionen og medlemsstaterne for at gøre det muligt for alle organer og organisationer, herunder borgere, forskere, virksomheder og SMV'er, at udnytte resultaterne af offentligt finansieret forskning, som de allerede har lovlig adgang til.
9. MENER, at optimal adgang til og genanvendelse af videnskabeligt arbejde kan forbedres, hvis forskere eller deres arbejdsgivere bevarer ophavsretten til deres videnskabelige arbejde; OPFORDERER Kommissionen og medlemsstaterne til at undersøge de legale muligheder for foranstaltninger på dette område og fremme anvendelsen af licensmodeller som f.eks. Creative Commons til videnskabelige publikationer og forskningsdatasæt.

¹³ Dok. 15264/15.

Fri adgang til videnskabelige publikationer

10. BIFALDER fri adgang til videnskabelige publikationer¹⁴ som standardløsning i forbindelse med publicering af resultaterne af offentligt finansieret forskning; ANERKENDER, at en fuldstændig overgang til fri adgang bør baseres på fælles principper såsom gennemsigtighed, integritet i forskningen, bæredygtighed, fair prisfastsættelse og økonomisk gennemførlighed; og OPFORDRER medlemsstaterne, Kommissionen og interessenterne til at fjerne finansielle og juridiske hindringer og til at tage de nødvendige skridt til en vellykket gennemførelse inden for alle videnskabelige områder, herunder specifikke foranstaltninger på områder, hvor der er hindringer for denne udvikling.
11. OPFORDRER Kommissionen til sammen med interessenterne og medlemsstaterne at udvikle og fremme foranstaltninger til optimal opfyldelse af bestemmelserne om fri adgang til videnskabelige publikationer under Horisont 2020; OPFORDRER medlemsstaterne til at arbejde sammen med interessenterne om at gøre det samme på nationalt plan vedrørende offentligt finansieret forskning.

¹⁴ I disse konklusioner skal "fri adgang til videnskabelige publikationer" forstås således, "at disse stilles gratis til rådighed på det offentlige internet, således at alle brugere kan læse, downloade, kopiere, distribuere, udskrive, søge eller linke til den fulde tekst af disse artikler, gennemgå dem med henblik på indeksering, videregive dem som data til software eller anvende dem til andre lovlige formål, uden finansielle, juridiske eller tekniske hindringer. Den eneste begrænsning for reproduktion og distribution og ophavsrettens eneste funktion på dette område bør være at give ophavsmændene kontrol over integriteten af deres arbejde og ret til at blive behørigt anerkendt og citeret (primær kilde: Budapest Open Access Initiative, 2002, findes på: <http://www.budapestopenaccessinitiative.org/read>).

12. ER ENIGT I yderligere at fremme integrering af fri adgang til videnskabelige publikationer ved fortsat at støtte overgangen til omgående fri adgang som standard fra 2020 med anvendelse af de forskellige mulige modeller og på en omkostningseffektiv måde uden embargoer eller med kortest mulige embargoer og uden finansielle og juridiske hindringer, idet der tages hensyn til de mange forskellige forskningssystemer og -discipliner, samt at fri adgang til videnskabelige publikationer bør ske under fuld overholdelse af princippet om, at ingen forsker bør forhindres i at publicere; OPFORDRER Kommissionen, medlemsstaterne og relevante interessenter, herunder forskningsfinansierende organisationer, til at virke som katalysator for denne overgang; og UNDERSTREGER betydningen af klarhed i forbindelse med aftaler om videnskabelige publikationer.
13. OPFORDRER alle parter til at indgå i en synkroniseret indsats for at nå dette mål - med mellemliggende milepæle¹⁵ inden 2018; OPFORDRER Kommissionen og medlemsstaterne til regelmæssigt i løbet af denne periode at overvåge den nøje gennem de nationale kontaktpunkter¹⁶ og så vidt muligt tilpasse den til overvågningen af status for EFR; OPFORDRER Kommissionen til at bistå medlemsstaterne og interessenterne med at opbygge ekspertise og lette informationsudvekslingen, f.eks. gennem EU-finansierede projekter med henblik herpå.

Optimal genanvendelse af forskningsdata

14. UNDERSTREGER, at forskningsdata, som hidrører fra offentligt finansierede forskningsprojekter kan betragtes som et offentligt gode, og OPFORDRER medlemsstaterne, Kommissionen og interessenter til at fastsætte optimal genanvendelse af forskningsdata som udgangspunkt, men erkender samtidig behovet for forskellige adgangsordninger af hensyn til intellektuelle ejendomsrettigheder, beskyttelse af personoplysninger og fortrolighed, af hensyn til sikkerheden, samt den globale økonomiske konkurrenceevne og andre legitime interesser. Derfor bør det grundlæggende princip for optimal genanvendelse af forskningsdata være: "Så åbent som muligt, så lukket som nødvendigt".

¹⁵ Milepælen omfatter gylden, hybrid- og grøn fri adgang.

¹⁶ Som fastsat i Kommissionens henstilling af 17. juli 2012 om adgang til og bevaring af videnskabelig information (dok. 12846/12).

15. GLÆDER SIG OVER Kommissionens hensigt om at gøre forskningsdata frembragt under Horisont 2020-programmet frit tilgængelige pr. definition¹⁷, samtidig med at det anerkender retten til at fravælge af årsager baseret på intellektuelle ejendomsrettigheder, beskyttelse af personoplysninger og fortrolighed, af hensyn til sikkerheden og andre legitime interesser; OPFORDRER Kommissionen til at fremme datahåndtering, herunder uddannelsesaktiviteter og oplysningskampagner, og til at gennemføre dataforvaltningsplaner som en integrerende del af forskningsprocessen og til fortsat at gøre omkostninger forbundet med både dataforvaltning og forberedelse af forskningsdata støtteberettigede under Horisont 2020; OPFORDRER medlemsstaterne og interessenterne til at opstille strategier i overensstemmelse hermed og til at gennemføre anvendelsen af dataforvaltningsplaner som en videnskabelig standardpraksis i deres nationale forskningsprogrammer.
16. UNDERSTREGER, at mulighederne for at sikre optimal genanvendelse af forskningsdata kun kan realiseres, hvis data er i overensstemmelse med principperne om, at de skal være søgbare, tilgængelige, interoperable og genanvendelige inden for et sikkert og pålideligt miljø; MINDER OM betydningen af lagring, langtidsoptbevaring og forvaltning af forskningsdata under hensyntagen til forskningsgruppens eller organisationens kapacitet samt af at sikre tilstedeværelsen af metadata baseret på internationale standarder; TILSKYNDER medlemsstaterne, Kommissionen og interessenterne til at følge principperne om, at data skal være søgbare, tilgængelige, interoperable og genanvendelige i forskningsprogrammer og finansieringsordninger.

¹⁷ Kommissionens meddelelse af 19. april 2016 om "Det europæiske cloudinitiativ - Opbygning af en konkurrencedygtig data- og vidensøkonomi i Europa".

17. NOTERER SIG Kommissionens meddelelse af 19. april 2016 om "Det europæiske cloudinitiativ - Opbygning af en konkurrencedygtig data- og vidensøkonomi i Europa"¹⁸ og ERKENDER, at Europa kan drage fordel af en åben europæisk forskningscloud¹⁹, som bl.a. muliggør sikker og langsigtet lagring, effektiv analyse og brugervenlig (gen)anvendelse af forskningsdata på tværs af grænser og discipliner; OPFORDRER Kommissionen til i samarbejde med medlemsstaterne og interessenterne at udforske passende ledelses- og finansieringsmæssige rammer, idet der tages tilstrækkeligt hensyn til eksisterende initiativer og deres bæredygtighed og til lige vilkår i hele Europa.

Opfølgning

18. OPFORDRER Kommissionen, medlemsstaterne og interessenterne til at træffe de foranstaltninger, der er nødvendige for at gøre åben videnskab til virkelighed og slå til lyd for behovet for en samordnet indsats i relevante nationale, EU-baserede, multilaterale og internationale fora; OPFORDRER Kommissionen til at holde medlemsstaterne orienteret om og inddraget i arbejdet i platformen for åben videnskabspolitik og til at koordinere sine aktiviteter med medlemsstaterne og ER ENIGT OM sammen med Kommissionen og platformen med regelmæssige mellemrum at gøre status over fremskridtene, begyndende senest et år fra nu.

¹⁸ Dok. 8099/16.

¹⁹ Den åbne europæiske forskningscloud skal indledningsvis samle eksisterende videnskabelige datainfrastrukturer, som i dag er spredt på tværs af fagområder og medlemsstater.

e-infrastruktur til forskning i verdensklasse i 2025

- en analyse og udredning vedrørende dansk e-infrastruktur samarbejde

Bilag 8:

EOSC deklaration og underskrivere



EUROPEAN COMMISSION
DIRECTORATE-GENERAL FOR RESEARCH & INNOVATION

The Director-General

Brussels, 26 October 2017

EOSC Declaration

RECOGNISING the challenges of data driven research in pursuing excellent science;

GRANTING that the vision of European Open Science is that of a research data commons, widely inclusive of all disciplines and Member States, sustainable in the long-term,

CONFIRMING that the implementation of the EOSC is a process, not a project, by its nature iterative and based on constant learning and mutual alignment;

UPHOLDING that the EOSC Summit marked the beginning and not the end of this process, one based on continuous engagement with scientific stakeholders, the European Commission,

PROPOSES that all EOSC stakeholders consider sharing the following intents and will actively support their implementation in the respective capacities:

Data culture and FAIR data

- **[Data culture]** European science must be grounded in a common culture of data stewardship, so that research data is recognised as a significant output of research and is appropriately curated throughout and after the period conducting the research. Only a considerable cultural change will enable long-term reuse for science and for innovation of data created by research activities: no disciplines, institutions or countries must be left behind.
- **[Open access by-default]** All researchers in Europe must enjoy access to an open-by-default, efficient and cross-disciplinary research data environment supported by FAIR data principles. Open access must be the default setting for all results of publicly funded research in Europe, allowing for proportionate limitations only in duly justified cases of personal data protection, confidentiality, IPR concerns, national security or similar (e.g. 'as open as possible and as closed as necessary').
- **[Skills]** The necessary skills and education in research data management, data stewardship and data science should be provided throughout the EU as part of higher education, the training system and on-the-job best practice in the industry. University associations, research organisations, research libraries and other educational brokers play an important role but they need substantial support from the European Commission and the Member States.
- **[Data stewardship]** Researchers need the support of adequately trained data stewards. The European Commission and Member States should invest in the education of data stewards via career programmes delivered by universities, research institutions and other trans-European agents.
- **[Rewards and incentives]** Rewarding research data sharing is essential. Researchers who make research data open and FAIR for reuse and/or reuse and reproduce data should be rewarded, both

in their career assessment and in the evaluation of projects (initial funding, review of performance and impact). This should go hand in hand with other career policies in universities and research institutions (appointments, promotions etc.).

- **[FAIR principles]** Implementation of the FAIR principles must be pragmatic and technology-neutral, encompassing all four dimensions: findability, accessibility, interoperability and reusability. FAIR principles are neither standards nor practices. The disciplinary sectors must develop their specific notions of FAIR data in a coordinated fashion and determine the desired level of FAIR-ness. FAIR principles should apply not only to research data but also to data-related algorithms, tools, workflows, protocols, services and other kinds of digital research objects.
- **[Standards]** The EOSC must be underpinned by minimal and rigorous global standards for open research data, as well as standards for EOSC based services for collaboration through the EOSC (e.g. to facilitate inter-disciplinarity and avoid fragmentation). These standards (technical, semantic, legal and organisational) must combine long-term sustainability with optimal freedom of local implementation. They should be jointly defined by the research communities, taking into account existing instruments (e.g. EU Rolling Plan on ICT Standardisation). Cross-disciplinary agreements/protocols will lead to specific standards, inspired directly by relevant domain specific needs. Variations across scientific disciplines and their specific efforts of making research data open and FAIR should be respected.
- **[FAIR Data governance]** The design and implementation of FAIR principles must be built upon inclusive stakeholder participation (e.g. researchers from different scientific disciplines, EU Member States and the European Commission). Policy will go hand in hand with the implementation of technical and human resources, and a social infrastructure including education and training. To make FAIR data a reality, it is imperative to engage stakeholders and relevant multipliers, based on a solid stakeholder engagement strategy, on inter-institutional arrangements, well-established frameworks and decision making flows. Data governance needs to be agreed upon and the division of responsibilities be charted, ensuring transparency, representativity and accountability. European and national scientific research organisations, publishers and other actors must align their data-related business processes, responsibilities and expectations to achieve commonly agreed goals.
- **[Implementation & transition to FAIR]** Implementation of FAIR principles requires careful prioritisation and orchestration. The FAIR Data Action Plan 2018-2020 is an important collaborative instrument for the embedding of FAIR principles in the first phase of the EOSC. The plan will not necessarily suggest any specific technology, standard or implementation solution. For an even transition of data from different levels of maturity to FAIR, existing activities to make data FAIR (e.g. GO-FAIR) must be complemented by new initiatives that embed FAIR principles in all the phases of data life cycle.
- **[Research data repositories]** Trusted research data repositories play a fundamental role in modern science. Scientist must be able to find, re-use, deposit and share data via trusted data repositories that implement FAIR data principles and that ensure long-term sustainability of research data across all disciplines. Data repositories must be easy to find and identify, and provide to users full transparency about their services.
- **[Accreditation/certification]** Scientists must be assured that the European and national scientific research infrastructures where they deposit/ access data conform to clear rules and criteria (e.g. certified) and that their data is FAIR compliant. An accreditation or certification mechanism must be set in place based on agreed processes and an accreditation or certification body must maintain an up-to-date and accessible catalogue of certified repositories. Experience from existing accreditation processes must be taken into account.

- **[Data Management Plans]** A key element of good data management is a Data Management Plan (DMP); the use of DMPs should become obligatory in all research projects generating or collecting publicly funded research data, based on online tools conforming to common methodologies. Funder and institutional requirements must be aligned and minimum conditions for DMPs must be defined. Researchers' host institutions have a responsibility to oversee and complete the DMPs and hand them over to data repositories.
- **[Technical implementation]** While FAIR data must be implemented as part of good data governance at this highest possible level (e.g. certification, institutional implementation and support, as describe above), researchers also need handy tools to make data FAIR. These include:
 - [Citation system]** A data citation system should be put in place to reward the provision of excellent open data. This will assist both the assessment of researchers and their projects, and help implementing the findability, accessibility, interoperability and reusability of research data.
 - [Common catalogues]** There must be catalogues (e.g. for datasets, services, standards) based on machine readable metadata and identifiable by means of a common and persistent identification mechanism that will make research data findable via an 'EOSC Portal'..
 - [Semantic layer]** Research data must be both syntactically and semantically understandable, allowing meaningful data exchange and reuse among scientific disciplines and countries.
 - [FAIR tools and services]** Easy access must be available to a common set of FAIR tools and services, to guide the curation of FAIR data for re-use and to assess FAIR compliance.
- **[Data expert organisations]** The Research Data Alliance, CODATA, DDI Alliance and other organisations active in the research communities must be used as forums to reach consensus on practical implementation of FAIR data principles at European and global level.
- **[Legal aspects]** It is essential for the success of EOSC to clarify and address the legal uncertainty of Open Access to research data, as well as the correct legal implementation of the FAIR principles. Legal barriers to access and reusability of research data must be identified and overcome and the underpinning legal framework must be made simpler and more coherent. Conversely, issues of ownership must be addressed, particularly where institutions have created services and resources. All these measures should allow easier integration of research data across different legal frameworks, policy implementation plans and strategies.

Research data services and architecture

- **[EOSC architecture]** The EOSC will be developed as a data infrastructure commons serving the needs of scientists. It should provide both common functions and localised services delegated to community level. Indeed, the EOSC will federate existing resources across national data centres, European e-infrastructures and research infrastructures; service provision will be based on local-to-central subsidiarity (e.g. national and disciplinary nodes connected to nodes of pan-European level); it will top-up mature capacity through the acquisition of resources at pan-European level by EOSC operators, to serve a wider number of researchers in Europe. Users should contribute to define the main common functionalities needed by their own community. A continuous dialogue to build trust and agreements among funders, users and service providers is necessary for sustainability.

- **[Implementation]** Resources, components and initiatives of pan-European relevance will be federated on the basis of objective criteria, agreed by stakeholder-driven governance, such as organisational readiness and technical capacity to deliver EOSC main functionalities: provision of core common services, certification activities, joint-procurement initiatives, definition of minimum quality standards of service (based on clear Service Level Agreements SLAs), identity provisioning and management, common cataloguing data and computing/analytic services and tools.
- **[Legacy]** The EOSC should incentivise the re-use of existing building blocks, state-of-the-art services and solutions delivered by past and ongoing projects, local, national and European, as opposed to subsidizing actions aiming at reinventing the wheel. It should facilitate learning from the past, adopting best practices, tailoring scientific community needs through live use cases and leveraging the network effect.
- **[User needs]** Users should see the EOSC as a one-stop-shop to find, access, and use research data and services from multiple disciplines and platforms. Services and functionalities shall be user driven and determined by clear use cases. Intermediary users and other brokers of end-users' demand – IT departments, umbrella associations, community networks – should assist data scientists and ICT specialists in the identification of key requirements for EOSC services.
- **[Service provision]** Research Data Infrastructures, e-infrastructures and commercial operators will develop and provide services based on user needs, and discontinue provision when not justified by the level of adoption. Services will be offered at highest Technology Readiness Levels (TRLs) and kept future-proof based on a cutting-edge cloud based environment. In order to avoid lock-in by individual service providers, the EOSC should foster fair competition of public, PPP and private providers on clear value propositions of highly professional services.
- **[Service deployment]** The EOSC shall support different deployment models (e.g. Infrastructure as a Service, Platform as a Service, Software as a Service), to meet the needs of communities at different levels of maturity in the provision and use of research data service. The EOSC shall support the whole research lifecycle by strong development at platform level that facilitate the provision of a wide set of software, infrastructure, protocols, methods, incentives, training, services. Software sustainability should be treated on an equal footing as data stewardship.
- **[Thematic areas]** The EOSC shall promote the co-ordination and progressive federation of open data infrastructures developed in specific thematic areas (e.g. health, environment, food, marine, social sciences, transport). The EOSC will implement a common reference scheme to ensure FAIR data uptake and compliance by national and European data providers in all disciplines.
- **[Research infrastructures]** The role of ESFRI and EIROFORUM research infrastructures and organisations in the EOSC will be enhanced, Member States and the European Commission made significant investment; research infrastructures should be 'the steward of the community of standards' and provide scientists with a ramp-up for the utilisation of the EOSC.
- **[EU-added value and coordination]** The EOSC must implement policy hand in hand with technology. Condition of national and European measures is required to link the initiative to national strategies, to maximise the added value of inter-disciplinarity by making data FAIR, and to preventing duplication of efforts and investments. Over time, coordination will provide European added value by minimizing overlap and reducing fragmentation of infrastructures and services, helping long-term sustainability.
- **[High Performance Computing and the EOSC]** European commitment to HPC is clearly demonstrated by the signature of the EuroHPC Declaration by eight Member States since March 23, 2017. The Member States agreed to work together and with the European Commission in the context of a multi-government agreement called EuroHPC for acquiring and deploying by 2022/2023 a pan-European integrated exascale supercomputing infrastructure that will support

data-intensive advanced applications and services. It is a response to the surging demand from scientists, industry and the public sector for access to leading-edge computing capacity to cope with vast amounts of data produced in almost all scientific and engineering domains. This supercomputing and data infrastructure could support the European Open Science Cloud by providing data access and advanced computing and data management services. The EC plans to propose, by end of 2017, a legal instrument that provides a procurement framework for the exascale supercomputing and data infrastructure.

- **[Innovation]** The EOSC should create a level playing field for businesses and innovative SMEs to develop, and co-develop with publicly funded institutions, added-value services for researchers. Funding should support the migration of cutting-edge solutions to the EOSC, increasing European added value by fostering innovation.

Governance and funding

- **[Governance model]** A long-term, sustainable research infrastructure in Europe requires a strong and flexible governance model based on trust and increasing mutuality. As interdisciplinarity is one of the main objectives of the EOSC, the governance model should be based on representativity, proportionality, accountability, inclusiveness and transparency.
- **[Governance framework]** The EOSC governance framework will be co-designed, stakeholder driven and composed of three main layers: 1) institutional, including EU Member States and European Commission 2) operational, including a governance board and relevant working committees (e.g. thematic and functional) and 3) advisory, including a stakeholder forum.
- **[Governance board]** A governance board will coordinate the efforts of stakeholders endorsing the EOSC Declaration, with the broad mandate to reach practical agreements for the implementation of an EOSC Roadmap by 2020. The board will have an advisory role and an implementing role of the decisions by Member States and European Commission concerning the programming, financing and towards the setting up of a long-term governance and business model for the EOSC. It will make best use of the outcomes of past and current projects (e.g. EOSCpilot, eInfraCentral and EOSChub) and independent expert advice and studies.
- **[Coordination structure]** A coordination structure, funded by Horizon 2020, will help the governance board to manage the implementation, according to agreed rules and methods of stakeholder participation. The structure and its participating entities should be accountable for the responsibilities assumed, based on an objective assessment of their level of readiness in delivering the EOSC main functionalities.
- **[Long-term sustainability]** The European Commission, Member States and Research Funders will use existing and future resources strategically, to ensure long-term sustainability of open research data and research infrastructures, facilitating inter-disciplinarity.
- **[Funding]** Over time, a co-funding mechanism mixing different revenue streams for the EOSC will be set up, to increase the accountability of the governance, building trust, sharing resources and building long-term capacity for European research data. Early implementation of the EOSC will pilot innovative business models and support an integrated data and service platform for European research.
- **[Global aspects]** The EOSC will be European and open to the world, reaching out over time to relevant global research partners. It will increase the global value of open research data and support stakeholder engagement, including researchers and citizens. It will gradually widen the initiative to federated network of infrastructures and nodes from global research partners. The EOSC Stakeholder Forum will have an important role in this sense.



Brussels, 26 October 2017

List of institutions endorsing the EOSC Declaration

| Signatory | On behalf of | Date |
|--|--------------------------------|--------------|
| Arnaud Torres, Acting President and CEO | ANR | 19 September |
| BlueBridge | BlueBridge | 18 September |
| Fabiola Gianotti, Director-General | CERN | 20 September |
| Karel Luijben, President | CESAER | 06 September |
| Ron Dekker, Director | CESSDA | 12 July |
| Kimmo Koski, Managing Director | CSC-IT Center for Science Ltd. | 15 September |
| Peter Doorn, Director | DANS | 14 September |
| Jacques Demotes, Director-General Christian Ohmann, Chair of the Independent Certification Board | ECRIN | 15 September |
| Matthew Dovey, Council Chairman | EGI | 15 September |
| Gabriele von Voigt, Chair Executive Board | e-IRG | 07 September |
| Iain W. Mattaj, Chair of the EIROforum Council | EIROforum | 15 September |
| Iain W. Mattaj, Director-General | EMBL | 15 September |
| Tiziana Ferrari, Coordinator | EOSC HUB Consortium | 15 September |
| Juan Bicarregui, Coordinator | EOSCpilot | 21 September |
| Francesco Sette, Director-General | ESRF | 14 September |
| Lidia Borrell-Damian, Director Research & Innovation | EUA | 15 September |

| | | |
|---|--|--------------|
| Damien Lecarpentier, Acting HoU EUDAT CDI | EUDAT | 15 September |
| Tony Donn , Programme Manager | EUROFusion | 11 September |
| Ole F. Olesen, Director of North- North Cooperation | European & Developing Countries Clinical Trials Partnership (EDCTP) | 15 September |
| Rory Fitzgerald, Director | European Social Survey ERIC | 15 September |
| Xavier Barcons, Director-General | European Southern Observatory (ESO) | 06 October |
| John Womersley, Director-General | European Spallation Source | 11 September |
| Serguei Molodtsov, Scientific Director | European XFEL | 14 September |
| FAIRDOM Association e.V. Executive | FAIRDOM Association | 15 September |
| Klement Tockner, President | FWF | 09 September |
| Federico Ruggieri, Director | GARR | 14 September |
| Matthew Scott, Chief Programmes Officer | G ANT Association | 15 September |
| Werner Leo Kutsch, Director- General | ICOS RI | 18 September |
| Helmut Schober, Director | Institut Max von Laue - Paul Langevin (ILL) and FILL2030 | 15 September |
| Paul Feldman, Chief Executive | JISC | 15 September |
| Leibniz Executive Board | Leibniz Association | 14 September |
| Kristiina Hormia-Poutanen, President Wolfram Horstmann, Head of Scholarly Communication & Research Infrastructures Steering Committee | LIBER | 15 September |
| Gudmun Host, Director | NeIC | 13 September |
| Stan Gielen, President Executive Board | NWO | 12 September |
| Natalia Manola, Managing Director | OpenAire | 15 September |

| | | |
|---|-----------------------------------|--------------|
| Peter Wittenburg, RDA Europe 3 Coordinator Edit Herczog, RDA Council Member, Official RDA EU Liaison | RDA Europe | 14 September |
| John-Arne Røttingen, Chief Executive | Research Council of Norway (RCN) | 18 September |
| Philip Diamond, Director-General | SKA Organisation | 29 September |
| Neil Chue Hong, Director | Software Sustainability Institute | 15 September |
| Eva Méndez, Chair of Working Group on Open Science | YERUN | 12 September |