Resultatkontrakt

Massive data – processering, visualisering og fortolkning

14. december 2009

1.1 Massive data – processering, visualisering og fortolkning

Skema til beskrivelser af forsknings- og udviklingsaktiviteter			
Aktivitetsområde (navn):	Massive Data	**Aktivitetsområde nr.:**	3
Sammenfatning	Det overordnede formål med Massive data – processering, visualisering og fortolkninger er at udvikle nye ydelser og software-værktøjer så 1. Dansk erhvervsliv bliver i stand til at udnytte de massive mængder data de ligger inde med – som i dag kun i neddroslede versioner kan processeres, visualiseres og fortolkes. 2. At skabe nye anvendelser af data til processering, visualisering og fortolkning ved til fulde at udnytte multi-core processorer tilgængelige i moderne pc'ere. Disse to formål kommer af tilsvarende to aktuelle og alvorlige barrierer for den videre udvikling af pervasive computing og dets anvendelser indenfor næsten alle områder af det danske samfund og industri såsom; energi, finans, grøn-it, industri, Internet, new media, oplevelser, produktion, sundhed, service, telekommunikation, transport, o.a. De to barrierer er: a) Mængden af information og data stiger så kraftigt i takt med udbredelsen af pervasive computing at den ikke kan udnyttes med traditionel software udvikling. b) Den enkelte processor-kerne som størstedelen af softwaren i den danske industri er udviklet til, er ikke blevet hurtigere siden årtusindeskiftet. Alexandra instituttet vil udvikle relevante teknologiske services til afhjælpning af disse barrierer igennem konkrete anvendelses cases i tæt samspil med dansk erhvervsliv. Resultaterne af aktiviteterne vil blive formidlet bredt igennem Alexandra Instituttets store SMV kontaktnet, vores forskellige netværk, via nettet, fagpressen og konferencetidsskrifter. Formidlingsaktiviteterne vil bl.a. inkludere etablering af interessegrupper af bl.a. SMV'er, workshops, foredrag, pressemeddelelser, samt både popuplærvidenskabelige og akademiske artikler.		
Formål og målgruppe	Aktivitetsområdet dækker centrale dele af de IKT emner, der beskrives på www.bedreinnovation.dk, specielt den under: Massive datamængder – datastrukturer, processering og fortolkning. Det overordnede formål med Massive data – processering, visualisering og fortolkninger er at udvikle nye ydelser og software-værktøjer så 1. Dansk erhvervsliv bliver i stand til at udnytte de massive mængder data de ligger inde med – som i dag kun i neddroslede versioner kan processeres, visualiseres og fortolkes. 2. At skabe nye anvendelser af data til processering, visualisering og fortolkning ved til fulde at udnytte multi-core processorer tilgængelige i moderne pc'ere. Disse to formål kommer af tilsvarende to aktuelle og alvorlige barrierer for den videre udvikling af pervasive computing og dets anvendelser indenfor næsten alle områder af det danske samfund og industri såsom; energi, finans, grøn-it, industri, Internet, new media, oplevelser, produktion, sundhed, service, telekommunikation, transport, o.a. De to barrierer er: a) Mængden af information og data stiger så kraftigt i takt med udbredelsen af pervasive computing at den ikke kan udnyttes med traditionel software udvikling.		

b) Den enkelte processor-kerne som størstedelen af softwaren i den danske industri er udviklet til, er ikke blevet hurtigere siden årtusindeskiftet - se figur 1.

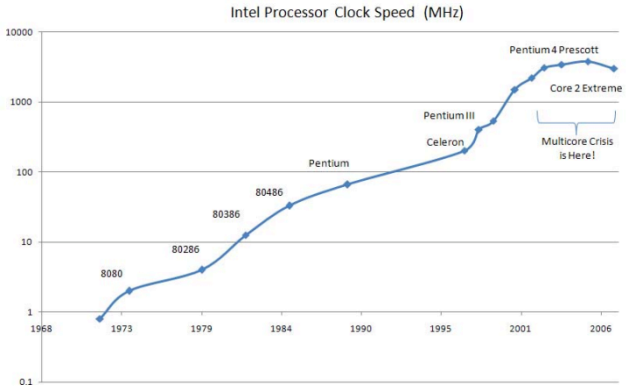


Figure 1. CPU clock speed growth, smoothspan.wordpress.com

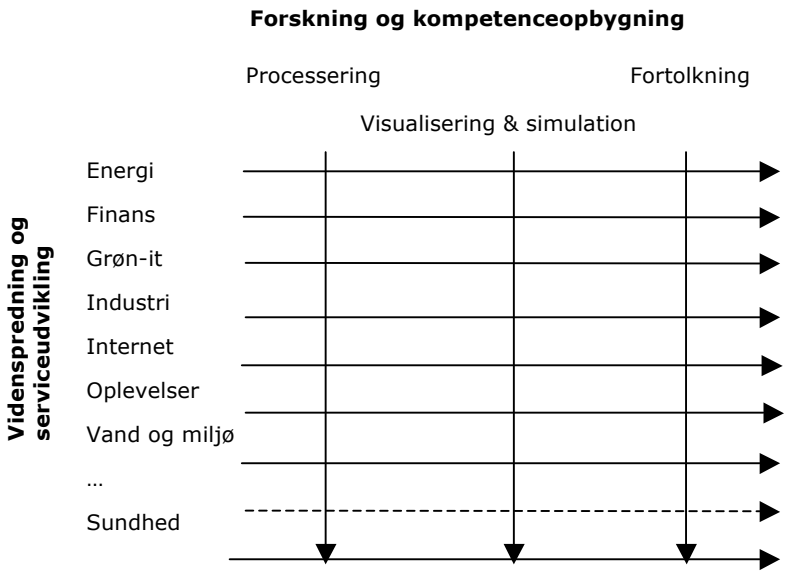
Mængden af data der i dag opsamles og lagres vokser eksponentiel – og meget hurtigere end muligheden for at bruge disse data med traditionelle midler. I 2000 viste et UC Berkeley studie at mængden af digitale data der blev lavet i løbet af to år var større end den samlede mængde digitale data, der eksisterede hidtidig. Mængden af data vokser hurtigere end både hastigheden på selve data-adgangen og beregningshastighed. Man kan sige at afstanden mellem data og den potentielle anvendelse af data bliver større og større. Den traditionelle måde at konstruere software på tager ikke højde for denne afstand – og i mange tilfælde kan eksisterende software således slet ikke bruge større mængder data. Der mangler konkrete kompetencer både nationalt og internationalt i at konstruere software der kan udnytte den voksende mængde af data.
Når data fx er for stor til ram og må ligge på harddisken bliver hastigheden på data-adgang ca. 1.000.000 gange langsommere. Dette er en absolut show-stopper medmindre software og data er designet således at det udnytter nogle af de andre karakteristika ved en harddisk. Fx at man ved hvert data-forespørgsel kan få de omkringliggende data ”gratis”. Søgning på Internettet er et godt eksempel på en problemstilling hvor den massive mængde data allerede nu er en udfordring. Fra 2000 til 2003 er mængden af overflade-data på Internettet tredoblet fra ca. 50 terabytes til 160 terabytes. For at kunne processere så mange data har Google udviklet et nyt paradigme til distribuerede beregninger; MapReduce, som muliggør effektive algoritmer på store datamængder på mange maskiner.
Beregninger på store mængder af data skal ses relativt i forhold til den disponible lagermængde og kravet til beregningstid for en given anvendelse – således kan interaktive applikationer med feedback til brugeren i realtid i mange situationer skabe helt nye anvendelser af de massive data.
Et eksempel på dette er MR scanning af patienter. Hidtidigt har man ”i blinde” scannet i stort område i patienten i håb om at indfange det interessante område for diagnosen. Rekonstruktions-algoritmen fra afmåling af det magnetiske felt til billeddata har simpelthen været for langsom, til at gøre dette mens patienten ligger i MR scanneren. På Datalogisk Institut, Aarhus Universitet har man udviklet en interaktiv rekonstruktions algoritme ved at udnytte de moderne grafikkort til andre beregninger – og derved givet radiologerne et helt nyt værktøj til diagnosticering.
Den store barriere i form af massive datamængder, og den eksponentielle vækst af data, skal ses i sammenhæng med ønsker om at lave effektive beregninger på den tilgængelige hardware. Dette er den anden store aktuelle barriere, nemlig at den traditionelle fordobling i hastighed på den enkelte processor hver 18. måned ikke længere gælder på grund af fysiske grænser – de bliver for varme og kræver

	for meget plads til lynhurtigt mellemlager. I stedet har samtlige chip producenter siden 2003 lavet "multi-core" processorer, hvor flere enkelte stående processorer (kerner) sidder tæt sammen. Den store barriere er at software skal være lavet specifikt til at udnytte flere kerner – og det er en væsentlig udfordring for både programmør, software værktøj og vedligehold af færdige programmer at gøre dette. Man skal forstå både, hvordan forskellige problemer kan paralleliseres og hvordan performance karakteristika er på de mange forskellige typer af multi-cores. Udnyttelsen af multi-core processorer kræver spidskompetencer der ikke i dag er tilstede i det danske erhvervsliv. Aktivitetsområdets grundlæggende hypotese, er at problemet med afstanden mellem mængden af data og maskinernes beregningskapacitet kan løses ved at formulere nye og bedre software metoder og algoritmer. Samt at den herigennem bedre udnyttelse af tilgængelige datamængder og beregningsenheder skaber muligheder for helt nye anvendelser indenfor en lang række brancher. Det er aktivitetsområdet formål, at undersøge disse hypoteser og i sidste ende at udvikle nye teknologiske services indenfor forskellige brancher til gavn for dansk erhvervsliv og samfundet som helhed. **Målgrupper** Idet ovennævnte problematikker med store datamængder skaber nye behov og muligheder indenfor alle dele af samfundet, dækker målgruppen for dette aktivitetsområde mange brancher bl.a.: • Energi – ex. forsyningsvirksomheder og vindmøllefabrikanter, hvor store datamængder anvendes dels til analyse og dels til realtids energioptimering og monitorering. • Finans – ex. forsikring og analyseinstitutter, som via store datamængder kan skabe nye services såsom benchmarking services. • Grøn-it – området er globalt i kraftig vækst, og omhandler dels at udvikle nye teknikker/metoder til at begrænse IT's energiforbrug samt store it-driftscentre samt nye virksomheder med fokus på udnyttelse af it til energioptimering baseret på store datamængder. • Design – Nye CAD værktøj til dansk design der skaber realistiske live-billeder af de store mængder af 3D data. • Industri – ex. Video analyse af produkter. • New media – Formidling af store datamængder på en let tilgængelig måde på nye online platforme. • Oplevelser – der er et stigende potentiale og behov i at kombinere fysiske og virtuelle verdener til nye oplevelser, f.eks. i form af udendørs computerspil eller udendørs, interaktive museumsbesøg. Den slags anvendelser vil dels ofte kræve opsamling og udnyttelse af store mængder positioneringsdata samt beregninger i realtid til skabelse af realistiske interaktioner; målgruppen dækker hele oplevelsesindustriens interessenter heriblandt spilfirmaer og museer. • Sundhed – ex. virksomheder indenfor medicinsk billedbehandling og nye anvendelser som 3d visualiseringer og kirurgiske simulationer, deriblandt også offentlige institutioner på afdelinger på hospitaler. Alexandra Instituttet har allerede i skrivende stund fået positive tilkendegivelser fra en række virksomheder og institutioner. Samtidig er der kommet en række positive tilkendegivelser via dialogen på bedreinnovation.dk. Derudover vil der løbende blive foretaget screenings-aktiviteter i forhold til at afdække behov indenfor de forskellige brancher og opbygge interessentgrupper.
Aktivitetsplanens indhold	Aktivitetsplanens indhold kan opdeles i to hovedområder: 1. Forskning og kompetenceopbygning, som primært vedrører den forskning, som instituttet selv udfører, og den kompetenceopbygning, som dels sker gennem videnhjemtagning fra nationale og internationale

forskningssamarbejdspartnere og andre former for studier.

2. Videnspredning og serviceudvikling, som omhandler dels screening af forskellige brancher med henblik på afdækning af bruger behov, udvikling af konkrete teknologiske services og afprøvning igennem konkrete pilot cases.

Der er naturligvis et tæt samspil mellem disse to områder, som illustreret i nedenstående matrix. De konkrete cases indenfor de forskellige domæner forventes at være i samarbejde med relevante virksomheder og brugere og at involvere flere forskningsmæssige deltemaer.



Som illustreret i figuren, kan forskning og kompetenceopbygningen omkring de massive datamængder og effektive beregninger opdeles i følgende deltemaer; processering, visualisering & simulation og fortolkning. De væsentligste aktiviteter omkring disse beskrives kort i det følgende.

Processering

De massive datamængder komme i tre relative størrelser;

1) Data der er så store at de aldrig lagres, men skal bearbejdes med det samme. Dette kræver udvikling af en bestemt type algoritmer der kun behøver at aflæse data én gang – men som alligevel kan udtrække fornuftig information. Dette kaldes også for **streaming software og algoritmer.**
2) Data der er gemt på harddiske eller netværksmedier. Harddisken er desværre ca. en million gange langsommere end ram, og er derfor en kontant show-stopper for algoritmer der ikke er designet efter et lavt antal forespørgsler til disk af større portioner data. Denne type data kræver **io-effektive software og algoritmer.**
3) Data der ligger i hukommelsen, men som skal processeres effektiv på de tilgængelige multi cores – både generelle (x86 baserede) og mere applikations-specifikke (GPU'er, Cell SPE'er) Dette kræver **parallel programmerings** teknikker og indgående kendskab til **hardware performance.**

Ovenstående tre kategorier fokuseres nærmere i følgende tre aktiviteter:

- **Streaming på mobile enheder:** Sensor data som fx video, billeder, net-data og lokation på moderne mobiltelefoner er et af de områder hvor mængden af indkommende data lang overstiger muligheden for at lagre denne og hvor beregningsenheden er relativt lille. Samtidigt er potentialet for anvendelsen stor indenfor mange grene af det danske erhvervsliv da den moderne

mobiltelefon er blevet til den computer vi altid har på os. Et scenarie i den nærmeste fremtid vil være at samtlige telefoner er stedsbestemte og transporterer video til nettet af de samme begivenheder. Anvendelsen af disse data kunne fx være til "live-streaming af koncerter med navigations muligheder igennem interpolation af samtlige video signaler" eller "auto generering af geoinfo" – både til koncertgæster, festivalpersonale og redningspersonale. Sådanne anvendelser kræver nye algoritmer og software udvikling der tager udgangspunkt i effektiv behandling af streaming-data.

- **Hadoop:** Der er store mængder af data i det danske erhvervsliv der i dag blot ligger gemt på eksterne lagrings medier, og ikke bliver processeret fordi der mangler kompetencer til at konstruere io-effektiv software og algoritmer. Vi vil samarbejde med Madalgo, den førende forsknings-gruppe indenfor massive data algoritmer, og bringe deres teoretiske resultater ud til praktisk anvendelse. Konkret vil vi opbygge kompetencer i anvendelsen af *Hadoop*, en open source implementation af Map-Reduce programmeringsmodellen - udviklet af Google til batch processering af meget data-intensive beregninger på distribuerede systemer.
- **Multi-cores CPU og General Purpose GPU:** Som allerede beskrevet kan man i dag ikke bare købe en ny computer og forventer at programmer kører hurtigere. De skal være skrevet specielt til at udnytte de parallelle processorer. Dette er en stor barriere for traditionel software udvikling, der siden ikke siden 1970'erne med "structured programming" har oplevet så stor en omvæltning. Dette kræver helt nye kompetencer i forhold til parallel-programmering og dyb indsigt i moderne parallel hardware.
  Vi vil oparbejde konkrete kompetencer i udvikling af software til multi-core processorer, herunder de mange forskellige hardware arkitekturer der effektivt kan løse specialiserede problemer. Specielt vil vi udforske den moderne stream processor (fx Tesla fra nVidia) der bygger på at bruge programmerbare grafikkort (GPU'er) til generelle beregninger – den mest kost-effektive processor i dag, takket være spilindustrien. Vi vil opbygge viden om anvendelsen af konkrete programmeringssprog som CUDA og OpenCL til at løse beregningsproblemer for den danske industri. Da GPU'en endvidere er ekstremt velegnet til interaktive applikationer med visualisering vil vi opstarte en række F&I projekter til afklaring af potentialet indenfor en række områder (se nedenfor) der i sig selv har potentiale som konsulent tjenester. DHI har indenfor deres anvendelsesområde, og i specifikke software applikationer, opbygget kompetencer med brugen af OpenMP til parallel programmering af CPU'er. I samarbejde vil vi kvalificere og generalisere disse kompetencer - deriblandt se på anvendelsen af GPU'er som parallel processerings-platform indenfor vand, miljø og sundhed.

Ovenstående punkter fokuserer på at opbygge kompetencer indenfor generelle værktøjer til processering af massive datamængder. Oven på disse værktøjer vil vi anvende visualisering, simulation og fortolkning til mere specifikke målgrupper og anvendelser.

Visualisering og simulation

Visualisering af store mængder af data handler både om at formidle en stor mængde information på en nem forståelig måde, men også at de store mængder data kombineret med muligheden for hurtige beregninger kan give anledning til helt nye anvendelser. Simulation kan enten være en overbygning på visualisering af eksisterende data eller kan i sig selv generere visualiseringer til et specifikt domæne. Igennem en simulering vil man "groft sagt" generere store mængder af data – som man løbende vil udføre effektive beregninger på. Et eksempel er visualiseringen af medicinske data i 3D – næste logiske skridt er at bygge fysiologisk og funktionelle egenskaber over på de morfologiske data for at kunne simulere fx kirurgiske indgreb. Der fokuseres på følgende aktiviteter:

	• **Interaktiv Infographics:** Specielt i mediebranchen mærker man den voksende mængde af information som "information-overload", kampen om læsernes opmærksomhed og den massive mængden af data, der ligger bag analyser og statistikker. Mediebranchen har et behov for at kunne skabe fortællinger med betydning for den enkelte læser, som samtidigt er let tilgængelige og med rum til fordybelse. Samtidigt vil man gerne kunne skabe fortællinger, der krydser medie-platforme som web, mobil, stor-skærme mm. Tekstuelle fortællinger er delvist succesfulde, hvad dette angår – men vi ser et stort potentiale i udviklingen af interaktive illustrationer og informations grafik til formidling af kompleks information – henover web, mobiler og storskærme. Det interaktive element gør at informationen kan tilpasses den aktuelle læser, skaber mulighed for navigation i data, samt at visualiseringen kan være baseret på løbende opsamling af data. Aktiviteten vil tage udgangspunkt i projekt-samarbejdet omkring "klimafortællinger" til COP15 i samarbejde med bl.a. Digital Urban Living, Jyllands-posten, og update. • **Medicinsk visualisering og simulering:** Medicinsk billeddannelse igennem fx MR og CT har igennem udviklingen af stadig bedre udstyr stået overfor en eksponentiel udvikling af tilgængelig data. Samtidigt bliver den kliniske brug af medicinsk billeddannelse til diagnosticering stadig mere udbredt. Arbejdsgange skal generelt effektiviseres og behandling af data til automatisk billedbehandling og diagnose relevant visualisering er derfor efterspurgt. Data fra medicinsk billeddannelse bliver også i højere grad tredimensionale hvilket skaber store muligheder for visualisering og kirurgisk træning igennem simulatorer. Kirurger bliver i dag uddannet i et mesterlæreprincip som har en række udfordringer: o Patienter kommer ikke i en naturlig kronologi i forhold til den pædagogiske træning. Et veltilrettelagt curriculum ville fx præsentere simplere tilfælde før de sværere. o Mange tilfælde eller kirurgiske strategier optræder meget sjældent, eller er endda unikke. Når den erfarne kirurg også skal opretholde sine egne færdigheder, vil den studerende i mesterlære oftest ikke få chancen for at forsøge sig med disse vanskelige tilfælde – indtil den dag hvor han står alene. Resultatet er en øget risiko for komplikationer når yngre kirurger opererer på patienter. Kirurgiske simulatorer er et interaktivt værktøj hvor den studerende opbygger vigtige erfaringer og bliver evalueret. På Alexandra Instituttet har vi opbygget spidskompetencer i forhold til kirurgisk simulation og gode samarbejdsrelationer til mange kirurger, deriblandt hjertekirurger, ørekirurger, kæbekirurger og gynækologer. Der findes endnu ikke danske firmaer der arbejder med kirurgiske simulatorer, og vi har en forventning om at det danske sundhedssystem vil skulle bruge simulatorer indenfor de næste 5 år til bl.a. akkreditering og curriculum baseret træning. Vi vil derfor støtte eksisterende firmaer indenfor medicinsk visualisering med overbygningen til simulator baseret træning, og kunne lancerer spin-off virksomheder til kirurgisk simulation. Vi vil opbygge generelle kirurgiske simulatorer som kan tilpasses en konkret type operation – heriblandt en simulator af sutur af bristninger i mellemkødet ved fødsler. • **Simulation af visuelle fænomener:** Formålet i denne aktivitet er at konstruere en software-komponent der kan beregne meget præcis gengivelse af visuelle fænomener (global illumination) meget hurtigt ved hjælp af GPU acceleration samt at bringe dette i anvendelse i forbindelse med realistisk billeddannelse til interaktiv design. Hastighed i rendering og visuel kvalitet er de to vigtigste parametre når en 3D grafiker arbejder med realistisk visuel gengivelse, til fx. film, reklamer, trykmateriale, produkt visualisering osv.

Flere store chipproducenter (Intel's larabee, Nvidia's Optix, CausticRT fra Cuastic Graphics) har proklameret at den nuværende teknik til grafisk rendering i hardware (rasterizing) vil blive erstattet eller suppleret fra teknikker i raytracing indenfor den nærmeste fremtid. Hurtig raytracing vil kræve udnyttelsen af de moderne processorer på store datasæt. Vi vil samarbejde med den førende forsker indenfor raytracing, Henrik Wann Jensen fra University of California samt DTU IMM. De generelle metoder fra aktiviteten vil konkret blive bragt i pilotanvendelse til realistisk billeddannelse til interaktiv design**:** Når designere af fx. møbler, lamper eller arkitektur formgiver ved hjælp af "computer aided design" i dag forgår dette med en skarp opsplitning imellem abstrakt formgivning og en virkelighedstro visuel gengivelse. Det betyder, at arbejdet med sammenspil mellem form, materiale, lys og skygger i de færdige produkter/bygninger ikke er understøttet af det primære værktøj designeren arbejder med – og derfor oftest ikke gennemtænkte. Vi vil konstruere et prototypeværktøj der giver mulighed for interaktiv design af geometri, lys og materiale - og som gradvist (progressivt) forbedrer detaljeniveauet efterhånden som designeren lader elementer hvilke. Dette vil give muligheder for at se hvordan et bygningsværk fremtræder på forskellige tidspunkter af dagen, hvordan glasmosaikker kan skabe spændende brændpunkter på andre overflader, hvordan naturlig belysning kan finde vej ind i arkitekturen, hvordan kunstlys kan benyttes som kilde til indirekte belysning uden skær (som fx. PH lamper) og hvordan reflekterende og refrakterende (glas) materialer kan bruges bevidst.

Fortolkning

I takt med at datamængder vokser, bliver der i stigende grad behov for at kunne fortolke data på en (semi) automatiseret måde og ofte i realtid. I mange konkrete anvendelser vil der derfor blive behov for nye værktøjer/metoder til effektiv søgning, filtrering, gruppering, genkendelse og beslutningsstøtte baseret på store datamængder. Konkrete cases, som vi vil fokusere på i den sammenhæng er a) effektive søgealgoritmer til søgning i Internet-baserede data kombineret med realtids positioneringsdata samt b) nye metoder og anvendelser af brain-computer interface med udgangspunkt i EEG baserede data.

Der fokuseres derfor på følgende aktiviteter:

- **Afdækning og vidensopbygning omkring de mest anvendelsesorienterede state-of-the-art algoritmer:** herunder datadriven machine learning, pattern recognition, feature extraction, clustering, data-mining og decision support. Aktiviteterne vil dels bestå af videnhjemtagning fra nationale og internationale partnere, litteraturstudier, udvikling af prototyper, samtidig med afdækning af behov hos relevante danske virksomheder og institutioner. DHI har indenfor deres anvendelsesområde, og i deres konkrete software produkter, arbejdet med emner indenfor beslutningsstøtte – og er derfor en værdifuld partner for opsamling af viden og generaliseringen af viden til gavn for SMV'er indenfor andre anvendelsesområder. Igennem den bredere vidensopbygning vil Alexandra Instituttet ligeledes kunne bibringe DHI ny viden indenfor "fortolkning" til gavn for DHI's projekter.
- **Integration og fortolkning af positioneringsdata:** Udviklingen i pervasive computing i retning af "the Internet og things", hvor alverdens apparater og sensorer er koblet på Internettet, gør at der vil ske en langt større integration mellem den fysiske verden og den digitale. Specielt vil positioneringsdata fra fysiske objekter stige markant og spille en større og større rolle i en lang række af fremtidens nye anvendelser.
  Vi vil i den forbindelse specielt fokusere på effektive søgealgoritmer til

søgning i Internet baserede data kombineret med realtids positioneringsdata fra fysiske objekter såsom personer, biler, o.l. Baseret på denne basale viden kombineret med afdækning af brugerbehov, vil vi fokusere på forskellige anvendelser og udvikle prototyper. Første to anvendelsesområder vil være omkring Internet/mobile baserede sociale netværk samt efterretning/overvågning. I begge tilfælde vil en effektiv anvendelse af realtids positioneringsdata kunne gøre en forskel.

- **Brain-computer interface:** Elektroencefalografi (EEG) er en teknik til at registrere hjernens elektriske aktivitet. Den består i, at man sætter elektroder på hovedets overflade og måler spændingen. Idet hjernens elektriske aktivitet er forskellig alt efter hvilken aktivitet/tankevirksomhed personen er i færd med, vil man i princippet kunne "læse personens tanker", hvis man ellers er i stand til at fortolke EEG mønstrene. EEG mønstre er dog dels meget støjfyldte, dels kender man i dag ikke den eksakte sammenhæng mellem mønstre og tankevirksomhed. Forskningsgrupper bl.a. i Danmark, Kina og USA, er dog indenfor de seneste år nået så langt, at de første konkrete anvendelser begynder at dukke op. I Danmark arbejder man bl.a. med EEG fortolkning på Ålborg Universitet og på DTU. Det vurderes, at der er unikke muligheder for nye anvendelser baseret på denne teknologi, som kan skabe nye forretningsmuligheder for danske virksomheder.
  Aktiviteten omfatter: a) at knytte samarbejde med relevante forskningsgrupper i Danmark og udland, b) at opbygge viden om state-of-the-art og c) at udvikle prototyper af nye anvendelser; i første omgang med fokus på handikappede og computerspil.

Resultater

Aktivitetsområdet Massive Datamængder vil overordnet resultere i følgende:

- Et idégenereringskatalog med en oversigt over idéer til nye anvendelser indenfor forskellige domæner.
- Software prototyper af udvalgte konkrete anvendelser.
- En software værktøjskasse med implementationer af udvalgte algoritmer.
- En håndbog, der dokumenterer "best practice" og operationelle retningslinier for anvendelsen af metoder og algoritmer indenfor de tre hovedområder af massive data: a) processering, b) visualisering og simulation, og c) fortolkning.
- En afdækning af forskellige anvendelsesdomæners fremtidige behov samt en oversigt over relevante fremtidige teknologiske services rettet mod disse domæner.

Kommercielle ydelser

De kommercielle ydelser som udvikles gennem aktivitetsområdet forventes at omhandle:

- Rådgivning og undervisning, bl.a. i form af en række generelle kursus tilbud i eksempelvis: CUDA parallel programmering, visualisering af store datamængder, Map-Reduce programmering med Hadoop.
- Konsulentydelser, eksempelvis i form af idégenerering og konceptudvikling rettet mod et bestemt anvendelsesområde eller i form af brugerinddragelse, feltstudier og prototypeudvikling
- Konkrete produktudviklingsprojekter for virksomheder.
- Virksomheders licensering af software komponenter.
- Eventuel etablering af nye spin-off virksomheder, baseret på kerneteknologier og konkrete anvendelser. En forudsætning for dette, vil være at spin-off virksomheden ikke er i konkurrence med andre eksisterende danske virksomheder.

Tidshorisonten for markedsmodningen af de nævnte ydelser vil være 3-5 år, dvs. fra slutningen af den ansøgte resultatkontrakt horisont.

	Forventede samarbejdspartnere Aktivitetsområdet involverer følgende samarbejdspartnere: • Center for massive data algorithmics (Madalgo), Aarhus Universitet. • eScience center. Københavns Universitet • Institut for Informatik og Matematisk Modellering (IMM), DTU • Computer Graphics Laboratory, Computer Science and Engineering, University of California, San Diego (Henrik Wann Jensen) • Department for Health Science and Technology, Aalborg University, Center for Sensory and Motor Interaction. • Klinisk Institut, MR-center og Onkologisk afdeling, Århus Universitets Hospital • Datalogisk Institut, Aarhus Universitet • Århus Tandlægehøjskole. • Øre-næse-halskirurgiske klinik, Rigshospitalet • Børnehjertekirurgisk afdeling, Skejby sygehus. • Surgical Simulation group, Inria, Frankrig. **International videnhjemtagning** Kompetenceopbygningen baseres i høj grad på international videnhjemtagning af forskellig karakter: • Samarbejde med konkrete udenlandske forskningsmiljøer, se ovenstående. • Studier af relevante forskningsartikler publiceret i anerkendte tidsskrifter • Deltagelse i open source udvikling omkring Hadoop. • Deltagelse i udvalgte relevante internationale konferencer. • International publicering af egne forskningsartikler med peer review.
Koordinering og samspil med andre F&I-aktiviteter	Projekter til medfinansiering. • Simulation af sutur af fødselsbristninger, FTP (søgt) • Digital Urban Living, nationalt center og EBST (projekt) • BodyExplorer, HTF eller region (planlagt) • ONE – FP7 ansøgning (søgt) • RTI, Train Your Brain (søgt) • <<slettet Koordinering med andre af Alexandra GTS aktivitetsforslag: • Der er en tæt sammenhæng med software infrastruktur til beregning på store mængder data og software platform til multi-core beregninger. • Der er en tæt sammenhæng mellem aktiviteten vedr. ”Integration og fortolkning af positioneringsdata” med RK forslaget vedr. ”Context-aware interaction and services” • I forbindelse med screening af brancher, behovsafklaring og konkrete cases, vil der være et tæt samarbejde med forslaget vedr. ”Helhedsorienterede innovationsprocesser baseret på integration af IKT, brugerinddragelse og forretning”. Generelt vil der være en tæt koordinering mellem Alexandra Instituttets fire aktivitetsforslag som illustreret i de indledende afsnit af dette dokument. Derudover er der aftalt samarbejde med DHIs resultatkontrakt vedr. Informations- og Kommunikationsteknologi. De to RK aktivitetsområder adresserer nogle af de samme emner specielt indenfor den del af processering, som omhandler multi-core processorer og general purpose GPU. Der er dog ingen overlap i de respektive aktiviteter, men i høj grad synergier, idet de to institutter adresserer emnerne fra forskellige vinkler. Alexandra Instituttet udvikler nye generiske metoder og algoritmer, som med fordel kan bringes i anvendelse indenfor DHI’s domæne, mens DHI fokuserer på udvikling af domænespecifikke matematiske modeller og numerisk analyse. Generelt vil de to institutter indlede et samarbejde, som kan udmønte sig i følgende former: • Fælles vidensopbygning og –udveksling

	Der nedsættes et fagligt forum på tværs af de to RK aktiviteter, hvor faglig viden udveksles ex. i form af fælles workshops. • Identifikation af og afprøvning på fælles cases Igennem RK projekterne identificeres og gennemføres konkrete aktiviteter på anvendelses-cases indenfor DHIs domæne. På nuværende tidspunkt er der identificeret én potentiel case: Massiv databehandling af LIDAR data (Light Detection And Ranging) • Fælles videnspredningsaktiviter I takt med at de to institutter opbygger en fælles vidensplatform og får konkrete erfaringer igennem test-cases, vil der blive mulighed for at arrangere fælles vindespredningsarrangementer. • Fælles nye F&I projekter De to institutter vil søge midler til nye supplerende F&I aktiviteter. Konkret vil man undersøge mulighederne for at et fælles innovationskonsortium muligvis med fokus på optimal anvendelse af LIDAR data. • Fælles kommercielle ydelser Begge institutter har en strategi med udvikling af software komponenter med åbne grænseflader. Institutterne vil undersøge mulighederne for samspil mellem software komponenter, som dermed kan spille sammen i større systemer, og på sigt føre til mulige fælles kommercielle ydelser.
Formidlings- og spredningseffekt:	Generelt vil aktivitetsområdet benytte de formidlings- og spredningsmekanismer, som Alexandra Instituttet råder over. • Landsdækkende kontakt til SMV'ere gennem instituttets vidensprednings-infrastruktur, som dels består af afdelinger i Århus, København, Herning og Sorø, dels består af en række formaliserede samarbejder med erhvervscentre rundt omkring i landet. Herigennem opbygges en interessentgruppe af virksomheder, hvor de primære aktiviteter dels vil være en række vidensprednings-seminarer , dels gennemførelse af en række idégenereringsworkshops med udvalgte virksomheder. • Formidling af viden gennem instituttets mange netværk; bl.a. igennem vores medlemsnetværk og vores højteknologiske netværk og innovationsnetværk. • Internetbaseret formidling via alexandra.dk samt dedikerede webportaler. • Publicering i dels populærvidenskabelige skrifter samt til forskningskonferencer og internationale forskningstidsskrifter. Et egentlig salg af teknologiske services baseret på den nye viden forventes først i den sidste del af resultatkontraktperioden. I løbet af den første del af perioden vil videnspredningen primært bestå i at opbygge netværk af interessenter samt at afdække interessenters behov. Aktivitetsområdet vil i den forbindelse afholde en række idégenereringsforløb med virksomheder, som har et behov for udvikling/afklaring af beregning, visualisering og fortolkning af store datamængder i forbindelse med deres produktudvikling. Herigennem vil behovet for egentlige udviklingsprojekter og anvendelse af den nye viden komme. Endelig deltager Alexandra Instituttet i nationale og internationale møder, seminarer og konferencer samt leverer artikler til målgruppens relevante fagblade og internationale tidsskrifter. Kvantitative mål for videnspredningen er: • Samlede antal virksomheder, som er blevet serviceret af aktivitetens vidensprednings, vil i år 3 være i størrelsesorden 100 virksomheder. Man vil samtidig løbende henvise virksomheder til instituttets relevante netværk såsom: Infinit, NFBI og sundhedsitnet. • Den forventede omsætning i året efter periodens afslutning forventes at være i alt: 1 mio. kr. Heraf ca. 50% til små og mellemstore virksomheder. Hefter forventes omsætningen at stige med ca. 20% om året over en årrække. • Aktivitetsområdet vil afholde 13 idégenereringsworkshops og 2-4

	videnspredningsseminarer over resultatkontraktperioden, herunder 1-2 fælles videnspredningsaktiviteter med DHI indenfor processering af massive datamængder med både teknisk bredde, og konkrete anvendelser i DHI's software produkter. • Det forventes at der laves i størrelsesorden 7 publikationer i løbet af resultatkontraktperioden.
Centrale kompetencer involveret i F&I-projektet	• Vicedirektør, Martin Møller • F&I chef, Softwareinfrastrukturer, Peter Andersen • Senior F&I specialist, Jesper Mosegaard • F&I specialist, Peter Trier • F&I specialist, Karsten Noe • F&I specialist, Jakob Fredslund • F&I ingeniør, Jerker Hammarberg
Milepæle år 1	**F&I-indikatorer** • State-of-the-art beskrivelser baseret på international videnhjemtagning og litteraturstudier af: o Streaming algoritmer til mobile enheder, Batch processering af data-intensitive beregninger via Hadoop, Multi-cores og general purpose GPU, Interaktiv infographics, Simulation af visuelle fænomener, Algoritmer til fortolkning af data • Deltagelse i 1-2 internationale forskningskonferencer • Opstart af udvikling af software toolkits herunder toolkits til medicinsk visualisering og simulation. • Kontakt etablering til centrale nationale og internationale forskningssamarbejdspartnere herunder etablering af fagligt forum med DHI. • Opstart til 1 nyt forskningsprojekt **Målgruppeservice** • Screening af brancher mht. behovsafklaring gennemført. • 1. version af behovsafklaringsdokument færdigt. • Mindst 2 præsentationer ved eksternt arrangerede videnspredningarrangementer såsom erhvervsklubber, konferencer, virksomhedsarrangementer, o.a. • Definition og opstart af to første case-anvendelser • 3 idégenereringsforløb gennemført med udvalgte virksomheder.
Milepæle år 2	**F&I-indikatorer** • Opdatering af state-of-the-art beskrivelser • 1. version af software toolkits klar dækkende centraler elementer af processering, visualisering og simulation samt fortolkning herunder: o Streaming algoritmer, Eksempelkode til Hadoop anvendelse o CUDA og OpenCL eksempelkode, Centrale software komponenter til interaktiv infographics anvendelser, Software komponenter til medicinsk visualisering og simulation herunder konkrete værktøjer til kirurgisk simulation, Prototype software komponent til avanceret simulation af visuelle fænomener, Centrale software komponenter og prototyper til anvendelse af EEG baseret interaktion. • 2-4 internationale publikationer • Fælles F&I aktiviteter igangsat med udvalgte nationale og internationale forskningssamarbejdspartnere • 1 forskningsseminar med nationale og internationale deltagere • 1 nyt forskningsprojekt ansøgt og godkendt **Målgruppeservice** • Opdatering af behovsafklaringsdokument. • 4 idégenereringsforløb gennemført med udvalgte virksomheder fra interessentgruppe.

	• Første to case-anvendelser gennemført og pilottestet. • 1. version af beskrivelse af relevante, fremtidige teknologiske services færdig (se afsnit om resultater under aktivitetsplanens indhold). • 1-2 videnspredningsseminarer med relevante faglige oplæg fra aktiviteten. Her præsenteres relevant generel faglig viden samt demonstreres konkrete eksempler på anvendelse. Heraf ét seminar i samarbejde med DHI. • Mindst 4 præsentationer ved eksternt arrangerede videnspredningarrangementer såsom erhvervsklubber, konferencer, virksomhedsarrangementer, o.a. • Projektwebsite etableret med relevant information og publikationer. • Mere end 10 virksomheder henvist til Infinit, NFBI eller sundhedsitnet.
Milepæle år 3	**F&I-indikatorer** • Opdatering af state-of-the-art beskrivelser • 2. version af software toolkits færdig dækkende centraler elementer af processering, visualisering og simulation samt fortolkning. • 6-8 internationale publikationer • Fælles F&I aktiviteter igangsat med udvalgte nationale og internationale forskningssamarbejdspartnere • 1 forskningsseminar med nationale og internationale deltagere • 1 nyt forskningsprojekt ansøgt og godkendt **Målgruppeservice** • Opdatering af behovsafklaringsdokument. • 5 idégenereringsforløb gennemført med udvalgte virksomheder fra målgruppe. • Fire case-anvendelser gennemført og pilottestet. • 2. version af beskrivelse af relevante, fremtidige teknologiske services færdig. • 1-2 teknologiske services pilottestet og evalueret. • 1-2 videnspredningsseminarer med relevante faglige oplæg fra aktiviteten. Her præsenteres relevant generel faglig viden samt demonstreres konkrete eksempler på anvendelse. Herunder ét seminar i samarbejde med DHI. • Mindst 4 præsentationer ved eksternt arrangerede videnspredningarrangementer såsom erhvervsklubber, konferencer, virksomhedsarrangementer, o.a. • Mere end 15 virksomheder henvist til Infinit, NFBI eller sundhedsitnet.