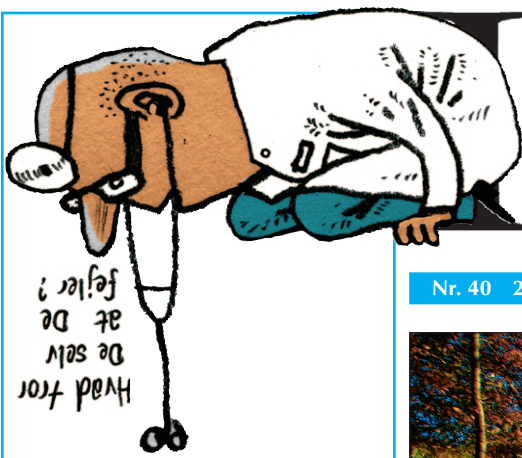




Hos lægen

Her har vi alle brug for
the Doctor Drug
Side 4



Gelotofobi

Når man er angst for
de andres latter
Side 7



Forfædre

Rod i stamtræet hos
skandinaverne
Side 5



God smag

Kan den lokaliseres
i hjernen?
Side 11

UDLÆR

weekendavisen
ANNO 1749

Nr. 40 2. - 8. oktober 2009

REDAKTØR: TINE EIBY

FOTO: KARSTEN SCHNACK



Forskere har målt lysindfaldet i skove og ganget det med vandtilførslen for at danne sig et detaljeret billede af plantevæksten i terrænet. Her et glimt fra Furesøen i skoven ved Næsseslottet.

Terabytes. Mellem år 2000 og 2002 producerede mennesker lige så mange data, som menneskeheden havde gjort siden Stenalderen. Store datamængder ændrer videnskabens vilkår.

Lysindfald ganget med vand

Af MARKUS BERNSÉN

DET var et slid at være oceanograf i 1990'erne. Oceanografer studerer havet, og mange af dem interesserer sig for undersøiske jordskælv, fordi de kan fortælle en masse om kontinentalpladernes dynamik og historie. Jordskælv er en god mulighed for at indsamle data. Og i Californien, der er hjemsted for nogle af verdens mest respekterede forskningscentre for oceanografi, plejede et undersøisk jordskælv at smælde som startskuddet på et udmattende kapløb.

For når de registrerede et stort jordskælv, ringede varslingsstjenesterne oceanograferne op – på alle tidspunkter af døgnet – og kort efter bakkede biler ud af indkørsler hele vejen ned ad den amerikanske vestkyst. Med et øje på armbandsuret og en fod på speederen satte forskerne kurs mod kysten. Derude skiftede de bilerne ud med både og fortsatte i retning af

skælvets epicenter, mens nogle vred sig ned i dykkerdragter eller miniubåde og bad stille bønner til, at der stadig var noget at hente dernede. At havtemperaturen ikke var normaliseret, eller at sprækkerne i havbunden ikke havde lukket sig igen. For hvert minut risikerede værdifulde data at gå tabt. Ofte var de få målinger, det faktisk lykkedes forskerne få med hjem, fulde af fejlkilder, fordi de var så længe undervejs.

Jo, det var et slid at være oceanograf dengang i 1990'erne. Gode data kostede sved, tårer og masser af benzin. Men for ti år siden skete der noget, der fra den ene dag til den anden ændrede vilkårene. I dag bliver forskere i Californien sjældent vækket midt om natten af startskuddet fra et jordskælv. Det skyldes ikke, at jagten er indstillet – tværtimod, oceanograferne nyder større bevågenhed end nogensinde – men den er blevet automatiseret.

På havbunden langs den amerikanske vest-

kyst ligger i dag et netværk af sensorer, som sender målinger tilbage til forskerne. Det mindste spjæt i undergrunden registreres, og målingerne strømmer ind i en digital bunke, der bare vokser og vokser. Udstyret er ganske vist dyrt, men de enkelte målinger er det ikke længere. Oceanograferne har langt flere data, end de ved, hvad de skal gøre af.

Omtrent samtidig begyndte deres kolleger i biologi, kemi, medicin, geografi – ja, selv sociologi og historie – at opleve noget lignende.

Deres måleudstyr begyndte at blive afløst af instrumenter, der registrerede 100 gange så meget som det gamle. Ved årtusindskiftet blev størrelsen på forskernes databaser samtidig mangedoblet, de gik fra gigabytes til terabytes af hukommelse. Ifølge en studie fra Berkeley University producerede mennesker

FORTSÆTTES SIDE 2



A HEREFORD BEEFSTOUW

Kbh.: Åbenrå 8 - 33 11 91 90 og ved Tivoli: 33 12 74 41. Herning: 97 12 35 44
Århus: 86 13 53 25. Kolding: 75 52 00 87. Odense: 66 12 02 22. Aalborg: 98 12 75 22
Skive: 97 52 53 25. V/Kruså: 74 67 30 00. Kastrup Lufthavn: 32 48 31 41
Göteborg: (46) 3177 50 441. Hong Kong: (852) 2136 0695. Nuuk: (299) 32 42 22

FORTSAT FRA FORSIDEN

Lysindfald ganget med vand

mellem år 2000 og 2002 lige så mange data, som menneskeheden havde gjort alle de foregående år.

Big Data, som fænomenet kaldes i udlandet, er fantastisk for forskningen – lige når det kommer til data, er mere altid bedre. Men størrelserne er også ved at ændre de metoder, som videnskabsfolk har forfinet gennem hundreder af år. Nogle taler om »teoriens endeligt«, og datalogerne bliver stadig mere uundværlige for naturvidenskabsfolk. For uden dem giver bunkerne simpelthen ikke mening. Mange forskere oplever for tiden, at store datamængder måske nok er bedre, men også... anderledes.

»GODT. Så vi er enige om, at adressen er hemmelig?« I telefonrøret høres lyden af fingre, der slår i computertaster. »Så sender jeg den nu.« Klik. En ny e-mail dukker op i Weekendavisens indbakke. Afsender: Lars Arge, professor og leder af Danmarks Grundforskningsfonds center MADALGO, *Center for Massive Data Algorithmics*.

I emnefeltet står adressen på en hjemmeside, som Lars Arge som sagt skal holde hemmelig lidt endnu. Ideen er, at alle danskere skal have adgang til det landkort, der ligger på siden, men der er stadig noget jura, der skal falde på plads. Kortet er en komplet model over det danske terræn, hvor samtlige bakker, kløfter og folder i landskabet er tegnet ind, og derfor kan det blandt andet bruges til at regne ud, hvordan regnvand fordeles sig i landskabet. Det kan også simulere vandstandsstigninger i havene og vise, hvilke dele af landet der først bliver oversvømmet. Øverst er en barre, som man kan trække til hver side med musen. »Vandstand«, står der over den.

Målingerne til terrænmodellen er lavet med to meters mellemrum, og den samlede mængde data er enorm. Størrelse betyder noget for den slags projekter. I højre side af skærmen kan man vælge en mindre detaljeret version af kortet, hvor det er stykket sammen af målinger med 80 meters mellemrum.

»Vælg den version og sæt så vandstanden til to meter,« siger Lars Arge. Skærmen fryser et par sekunder, mens computeren simulerer en vandstandsstigning på to meter. På det nye kort har kystlinjen flyttet sig flere steder. Omridset af Sydsjælland er uldent og lyseblåt.

»Okay, prøv nu at zoome ind på Mandø. Den ligger lige syd for Esbjerg.« Det tager lidt tid at finde øen, kun en lille stribe af den stikker op af Vesterhavet. Ved en vandstands-

stigning på bare to meter vil Mandø altså blive næsten helt oversvømmet.

Men nu beder Lars Arge mig vælge den datatunge indstilling med målinger med to meters mellemrum. Skærmen fryser, og så glider vandet væk fra øen igen. En god bid af Djursland er stadig oversvømmet, men Mandø holder stand. Hvad skete der?

»Der ser du fidusen ved store data,« siger Lars Arge. »Opløsningen på 80 meter registrerer ikke digerene på Mandø. Derfor ser det ud, som om den bliver oversvømmet, selv om den slet ikke gør det.«

I nogle tilfælde er resultatet kun pålideligt, hvis datamængden er meget, meget stor. Og i terrænmodellens tilfælde er der faktisk ingen øvre grænse for, hvor mange ekstra lag af data, der kan lægges ovenpå. Byplanlæggere og ingeniører kan måle på, hvor højt vindmøller stritter op i landskabet, eller hvor langt støjen fra nye motorveje eller signaler fra telefonmaster rækker. Det kan lægges oven på modellen. I USA har forskere målt lysindfaldet i skove og ganget det med vandtilførslen for at danne sig et detaljeret billede af plantevæksten i terrænet. Det er endnu et lag data. Kommunernes oplysninger om kloakkernes tilstand er et andet. Geologernes viden om jordlagene et tredje.

»I dag er det vigtigste ikke, om der er data nok på et bestemt område. Det vigtigste er, om de data, der er, overhovedet kan beregnes.«

Forsikringsbranchen er allerede i fuld gang med at lægge deres egne beregninger oven på terrænmodellen. Deres mål er at tilføje et detaljeret »risikokort« over Danmark, som kan bruges til at udregne individuelle forsikringspræmier. Snart vil der findes risikodata for hver enkelt dansk husstand, udregnet efter dens placering i terrænet, afstand til hav, søer, vandløb, politistation, brandstation, S-togsstation og mange, mange andre faktorer. Store datamængder ændrer forsikringsagenters arbejdsvilkår: I gamle dage blev præmierne med udregnet efter forsikrings-tagers postnummer. Snart bliver de udregnet automatisk efter deres husnummer.

DEN endnu hemmelige terrænmodel afløser generalstabskortet, som Forsvaret leverede den første udgave af i slutningen af 1800-tallet. Dengang blev militærkompagnier sendt ud i landskabet på lige rækker, hver soldat med sin sigtelinje og notesblok. I elleve år,

mellem 1891 og 1902, sled det danske forsvar med at indsamle data til kortet. Ligesom Californiens oceanografer var de nødt til at være i nærkontakt med materien for at skaffe data, og med så mange involverede var den fuld af småfejl.

Det nye kort er lavet med en laserscanner og et lavtgående propelfly. Nærhed er unødvendig: Så snart flyet er i luften, og apparatet er tændt, vælter målingerne ind. Udfordringen ved terrænkortet anno 2009 har ikke været at skaffe data, men at holde styr på de enorme mængder, der tikkede ind helt af sig selv.

De fleste målinger blev faktisk lavet for flere år siden, men forskerne har først fået skik på dem for nylig. Store datamængder er ofte ubrugelige, hvis forskerne ikke får hjælp af folk som Lars Arge til at sortere i dem. Dataloger laver algoritmer, der genkender mønstre i de uoverskuelige bunker og fisker de røde tråde op til forskerne. Det kaldes nogle gange *data mining*. Gode algoritmer er i høj kurs.

I de sidste ti år har der nemlig åbnet sig en kløft mellem den bunke data, som forskerne har til rådighed, og den regnekraft, som er nødvendig for at trække brugbar viden ud af den. Kløften udvider sig. Datamængden vokser hurtigere end regnekraften, og flere videnskabelige gennembrud afhænger udelukkende af datalogers evne til at sortere i de hastigt voksende bunker.

Thore Husfeldt, der er lektor på IT-Universitetet i København og docent ved Lunds Universitet, mener, at der er gledet »en linse af algoritmer« ind mellem videnskabsfolk og det område, de undersøger. Uden algoritmer er genstanden simpelthen svær at få øje på. Thore Husfeldt forsker i algoritmer for såkaldt »hårde problemer«. Den teoretiske gren af datalogien har fundet ganske mange konkrete datamængder, som simpelthen er for komplekse til at beregne.

»I dag er det vigtigste ikke, om der er data nok på et bestemt område. Det vigtigste er, om de data, der er, overhovedet kan beregnes.«

DET bliver dataloger, der gennem en ny algoritme engang puster liv i de første komplette modeller af menneskets dna og får fjerne galakser til at glide i fokus midt i det slørede mylder af observationer. Også selv om de intet ved om biologi eller astronomi. Den første store succes var *The Human Genome Project*, som i 2000 begyndte at kortlægge menneskets mere end 20.000 gener. Havde kemikere og biologer ikke allerede indset det, blev det her klart, at dataloger ikke kunne undværes på store projekter.

Cluster Exploratory er et andet internationalt projekt, hvor neurologer skal bruge teknologi fra IBM og Google til at lave computersimuleringer af hjernen og nervesy-



stemet. Tidsskriftet *Wired* hyldede sidste år *Cluster Exploratory* som et mønstereksempel på, hvordan datamylderet er ved at ændre forskernes fremgangsmåde. »Der er ingen grund til at klamre sig til gamle vaner,« skrev bladets chefredaktør Chris Anderson under overskriften *The End of Theory*, Teoriens endeligt. »Det er på tide at spørge: Hvad kan videnskaben lære af Google?«

Algoritmer som dem, der gør Google i stand til at sortere data på internettet, har allerede ændret flere forskeres gamle vaner. Før *Big Data* opstillede de først en hypotese for, hvordan tingene hang sammen. Så designede de måske et eksperiment og drog i felten for at indsamle data til at efterprøve hypotesen med. Hvis det ikke lykkedes, begyndte de forfra: ny hypotese, nyt eksperiment, ny tur i felten.

Men det klassiske kredsløb brænder sammen, når man tilføjer tilstrækkelig store mængder data. Gode algoritmer gør hypoteser og teorier overflødige. I stedet for at gøre sig antagelser har nogle forskere nu fået muligheden for bare at åbne for datasluserne og lade computerprogrammer finde orden i mylderet.

Clark Glymour, der er professor ved MIT, indså det allerede i 2004. I en artikel fra dengang sammenligner han videnskabens jagt på ny viden med at finde en nål i en høstak: I gamle dage fiskede man et eller andet ud af høstakken og undersøgte det grundigt for at se, om der var hul i den ene ende. Hvis der var det, kunne man formode, at der var tale om en nål. Hvis ikke, begyndte man forfra med at rode. »Fortsæt til du finder nålen, eller til civilisationen når sit endeligt.« Sådan lød devisen i gamle dage, ifølge Glymour. Den nye teknologi gav derimod forskerne mulighed for at »bevæge sig væk fra det gamle paradigme og omfavne automatisering, som svarer til at sætte ild til høstakken og puste asken væk for at blotlægge nålen.«

FALSIFICERET



Tyresæd

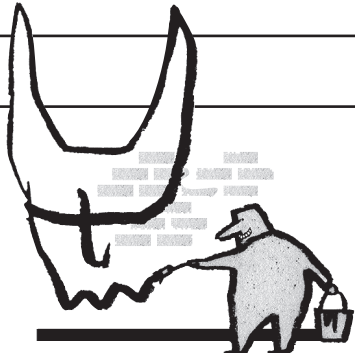
DET lød ellers som en rigtig god idé, da det for tre år siden blev muligt at kønssortere tyresæd og dermed sikre, at køerne producerer langt flere kviekalve end tyrekalve.

Fra naturens side vil en ko få omtrent lige mange af hver slags, men da det naturligvis kun er kvierne, som – efter selv at have fået deres første kalv – kan producere mælk, er der stor forskel på

værdien af afkommet, og mange tyrekalve bliver derfor aflivet umiddelbart efter fødslen eller solgt billigt til opfødning.

Ved kønssorteringen udnyttes det, at man kan binde et fluorescerende farvestof til sædcellernes dna, og da celler med X-kromosom (hun-celler) indeholder 3 procent mere dna end dem med Y-kromosom (han-celler), bliver hun-cellerne mere fluorescerende og kan derfor opsamles maskinelt. Teknikken er ikke helt perfekt, da omkring hver tiende celle fejlsorteres, men det sikrer alligevel, at køerne producerer omkring 90 procent kviekalve og dermed får større værdi som avlsdyr. Et stort antal landmænd, især i USA og Canada, greb derfor muligheden, og – *surprise, surprise* – her et par år senere er køer frembragt ved kønssortering nu begyndt at dukke op i besætningerne i titusindvis.

Desværre sker denne eksplosion af køer på et tidspunkt, hvor de amerikanske farmere i forvejen har



alt for mange, hvilket har fået mælkepriserne til at styrtedykke med 40 procent alene inden for det seneste år. Så alt imens sammenslutningen af amerikanske mælkeproducenter forsøger at reducere overproduktionen ved at betale farmerne for at slagte deres besætninger, myldrer de nye køer frem i staldene.

Selvom historien minder om en moderne udgave af troldmandens lærling, mener en talsmand for en storleverandør af kønssorteret sæd dog, at teknologien vil vise sig at blive en fordel, men indrømmer, at det indtil videre ikke er gået helt, som han havde håbet. *jopp*

The New York Times, 28. september.



Tuberkulose

SAGEN om den ægyptiske mumie Irtyersenu, som også går under tilnavnet dr. Granvilles mumie, er nu endeligt afsluttet. Mumien, der i sin tid blev fundet ved Thebes og menes at stamme fra omkring 600 f.Kr., blev ved Londons Kongelige Selskab i 1825 beskrevet som en kvinde, der var død i 50-års-alderen. Hun var den første mumie,

som undergik en videnskabelig obduktion, og dr. Augustus Granville konkluderede dengang, at hun var død af kræft i æggestokkene.

Men for 20 år siden blev sagen genåbnet. Dr. Granvilles mumie har altid undret ægyptologer og patologer. Dels fordi mumiens organer ikke som normalt var taget ud og balsameret særskilt, dels fordi hele mumien var klistret ind i en særlig voksagtig substans, som har været uhyre svær at arbejde med. Da resterne af mumien blev genundersøgt i 1990'erne, kunne man imidlertid konstatere, at den svulst, man i sin tid havde baseret diagnosen på, var godartet. Og det har siden været et stort mysterium, hvad Irtyersenu så døde af.

Nu viser nye undersøgelser, gennemført af forskere fra University College London og University of Birmingham, at hun ganske enkelt døde af tuberkulose.

eiby

New Scientist, 30. september