



Humaniora og de digitale Medier

Finnemann, Niels Ole

Published in:
Kampen om Disciplinerne

Publication date:
2015

Document version
Peer-review version

Citation for published version (APA):
Finnemann, N. O. (2015). Humaniora og de digitale Medier. I D. B. Pedersen, F. Stjernfeldt, & S. Køppe (red.), *Kampen om Disciplinerne: Viden og videnskabelighed i humanistisk forskning* (1 udg., s. 317-341). Hans Reitzels Forlag.

12

Humaniora og de digitale medier

Niels Ole Finnemann

Indledning

Computeren har i forskningssammenhænge især været set som et redskab, der har muliggjort en stadig voksende vifte af nye metoder. Det gælder inden for alle naturvidenskaber og en voksende del af samfundsvidenskaberne, mens der inden for humanvidenskaberne derimod har været en udbredt skepsis over for de metodologiske potentialer. I store dele af humaniora har digitale metoder frem til for nylig således kun spillet en marginal rolle, mens "Humanities Computing" (fra 1960'erne), den informationsvidenskabelige biblioteksforskning (fra 1970'erne) og humanistiske informations- og medievidenskab (fra 1980'erne) såvel som andre mere spredte miljøer, der har orienteret sig mod digitale medier, har befundet sig i relativ isolation fra mainstream humaniora. Gennem den sidste halve snes år er der imidlertid indledt et begyndende, men også ret vidtfor grenet opbrud, der både kommer til udtryk i en bølge af "Digital Humanities"-initiativer og forskellige former for "Computer Mediated Communication", "Information Studies" og "New Media Studies". Der er i mange tilfælde tale om nye forskningsfelter, men der spores også beslægtede tendenser inden for eksisterende fag (sprogvidenskab, historie, litteratur, kunsthistorie, etnografi, filosofi).

I Danmark blev der i 2011 iværksat et nationalt, tværuniversitært "Digital Humanities"-projekt, DIGHUMLAB, der især har rettet fokus mod de metodiske spørgsmål, der rejser sig i forbindelse med håndtering af de hastigt voksende mængder af digitale kildematerialer. Der fokuseres foreløbig på tre forskellige typer: digitale kopier af sproglige korpora, arkiverede webmaterialer og eksperimentelle forskerdata. Det fælles element i det danske projekt er især knyttet til de forskningsinfrastrukturelle behov i feltet

Kampen om disciplinerne.indd 317 28/03/15 11.37

318 Kampen om disciplinerne

mellem forskernes fra projekt til projekt skiftende søge-, analyse- og fremvisningsbehov

og de databevarende institutioners langsigtede service og bevaringsopgaver, der rækker ud over de aktuelle og specifikke forskningsønsker.

Da brugen af softwarestøttede metoder også producerer nye forskningsdata og software-tools, opstår der yderligere nye spørgsmål omkring håndtering og bevaring af disse.

I international sammenhæng omfatter Digital Humanities en endnu bredere vifte af strømninger, der ofte også diskuteres under termer som digital turn, e-research, og i-research, ligesom der er overlap og berøringspunkter til

“Big Data”-bølgen, der ikke kun handler om marketing.

Man kan spørge, om der er en fællesnævner i disse opbrud. Da alle de nævnte termer på den ene eller anden måde refererer til “det digitale”, er det nærliggende at tage udgangspunkt i spørgsmålet om, hvorvidt “det digitale” er forbundet med en identificerbar paradigmatisk bestemthed, og om det i givet fald er en bestemthed på epistemologisk, teoretisk eller metodologisk niveau, om det er en bestemthed på medieteoretisk niveau, eller om det konstitutive ved det digitale i sidste ende alene er kravet om manifestation i det binære alfabet, dvs. en bestemmelse på materialeniveau.

De følgende afsnit vil behandle disse spørgsmål med afsæt i den brede Big Tent-forståelse af digital humaniora (Svensson 2012), der inkluderer alle håndte studier af menneskelige anliggender ved hjælp af softwarestøttede metoder. Brede definitioner kan i sagens natur ikke være særlig præcise for de enkelte strømninger, der indbefattes, og teltforståelsen har mødt en del kritik. En stor del af kritikken har siget på at give en mere præcis paradigmatisk bestemmelse af det digitale, mens de fleste synes at hælde til at fastholde “the humanities” som en central term (Presner et al. 2009; Ramsay 2011; Hayles 2012; oversigt i Finnemann 2014a).

I Humanities Computing-traditionen, der opstod i 1960'erne, var målet at bruge computationelle metoder til at give et mere objektivt videnskabeligt grundlag for humaniora. Man knyttede her an til den klassiske computerforståelse med hovedfokus på det regelbestemte. Computeren betragtes som en logisk maskine, der kan udføre alle mulige former for beregning på alle mulige mængder af velordnede, homogene datasæt. Den utopiske drøm kom især til udtryk i de skiftende teorier om kunstig intelligens, men tanken genfindes også i en voksende strøm af praktisk bundne computerarkitektur-opfattelser (Haugeland 1985; Bannon & Pylyshyn 1989). I Humanities Computing-traditionen lægger man vægt på at fremstille digitale kopier af ikke-digitale origi-

Kampen om disciplinerne.indd 318 28/03/15 11.37

12 | Humaniora og de digitale medier 319

nalværker og korpora på en måde, der faciliterer anvendelsen (og udviklingen af) computationelle metoder.

Ideen om “computationelle metoder” har i dag mistet sin prægnans. Den klassiske forståelse blev udfordret af Human-Computer Interaction-traditionen og teorierne om neurale netværk i 1980'erne, ligesom forudsigelighedsparadigmet har måttet vige for netværksorienterede kompleksitetsteorier og kausalitetsforklaringer viger for korrelationsstudier inden for den nye Big Data-tradition. Computerens betydning som kommunikationsmedium og internettets betydning som medie for både privat og offentlig kommunikation har samtidig dannet grundlag for en bred vifte af kvantitative og kvalitative metoder udviklet til studiet af mere heterogene, digitalt fødte materialer. Mens Humanities Computing kunne udvikle en relativt veldefineret og fælles forståelse af computerne, er der betydelige divergenser i opfattelsen af, hvad der kunne karakterisere det digitale i Digital Humanities.

Spørgsmålet er imidlertid ikke blot, om der er en digital fællesnævner, der kan begrunde en afgrænsning af et sådant nyt digitalt humaniora. Spørgsmålet er tilmed, om der er en humanvidenskabelig fællesnævner. Forløber digitaliseringen på en sådan måde, at det giver mening at udsondre et særligt digitalt humaniora? Eller vil digitaliseringen snarere overskride grundlaget for at opretholde

humanioras separation fra den øvrige samfunds- og naturvidenskabelige forskning og fra den verden, som denne forskning forholder sig til?

Når det kommer til stykket, er humaniora en forholdsvis løs samling af fagligheder og metoder relateret til studiet af mennesket, dets virksomhed og produkter. Måske er humanioras fællesnævner i dag alene den at være tankebon, at afgrænse en gruppe af fagligheder fra de samfundsfaglige og naturvidenskabelige forståelser af menneskelig virksomhed.

Som indgang til denne diskussion er det nærliggende at kaste et blik på den mest generelle definition af computeren, nemlig den engelske matematiker Alan Turings nu klassiske beskrivelse af principperne for en universel computer fra 1936. Ved "universalitet" forstod Turing slet og ret, at den samme maskine skulle kunne udføre en hvilken som helst beregning, der kunne udføres ved et endeligt antal operationer. Den centrale pointe er, at computerens beregningsmæssige universalitet kun er mulig, hvis den selv arbejder helt uafhængigt

af enhver specifik regneregul.

For at indfri dette krav må maskinen være konstrueret på en sådan måde, at dens funktionelle arkitektur ikke er indbygget i den invariante hardware, men derimod leveres som løbende input og specificeres i samme (binære) format

Kampen om disciplinerne.indd 319 28/03/15 11.37

320 Kampen om disciplinerne

som alle de data, der skal behandles. Som følge heraf kan en computer ikke blot rekonfigureres via nye input, det er også altid muligt at implementere nye ideer, der ændrer computers funktionelle arkitektur. Man kan således også indskrive nye epistemologier i den funktionelle arkitektur.

For så vidt man tager afsæt i begrebet om den universelle computer, vil det med andre ord aldrig være muligt at tilskrive dette medie en bestemt epistemologisk,

teoretisk eller metodisk værdi. Det vil omvendt altid være nødvendigt at repræsentere såvel programmer som data i et fysisk format, der kan processeres og redigeres af skridtvis, mekanisk vej. Det "digitale" konstitueres således ved repræsentation af både data og programmer i det binære alfabet.¹ Nu er de computere, vi bruger, sjældent universelle i Turings forstand. Det er som regel mere hensigtsmæssigt at indbygge formålsspecifikke restriktioner i den invariante fysiske arkitektur eller at kodificere og låse mange bitsekvenser fast som automatiske procedurer, men det ændrer ikke ved det generelle kendetegn. Selv om en dedikeret maskine, f.eks. en automat, teoretisk set godt kan være bundet til bestemte epistemologiske teoretiske eller metodiske præmisser, kan disse præmisser ikke danne grundlag for en mere generel forståelse af digitale medier. Tværtimod er konsekvensen af den universelle computers egenskaber, at de digitale mediers historie bliver en historie om stadig flere forskellige funktionelle arkitekturer baseret på en voksende vifte af ideer og forestillinger om menneskelig viden, tænkning, behov, længsler, drømme og begær.

Digitaliseringens historie synes at føre i retning af en stadig udvidelse af de synlige potentialer, fordi der til stadighed kan bygges nye betydningslag ind i de digitale medier. Der tegner sig således både i forskningsssammenhænge generelt og også mere specifikt inden for Digital Humanities en voksende

emne mæssig og faglig spredning og i en voksende metodisk, teoretisk og epistemologisk differentiering (Finnemann 2014a).

Der gives i det følgende derfor blot nogle udvalgte pejlepunkter i det, der foreløbig tegner sig som en fortsat differentieringshistorie, der lejlighedsvis antager karakter af niveauspring. Der er flere kilder til denne differentiering. For det første præges udviklingen af softwarestøttede metoder naturligt af forskernes forskellige faglige udgangspunkter og framing af forskningsspørgsmål

1 Her og i det følgende bruges betegnelsen "binært alfabet" frem for den gængse beskrivelse af de to bits som numeriske værdier, fordi de to notationsenheder ligesom alfabetets bogstaver udelukkende er defineret ved en fysisk værdiforskel, hvorimod der er en betydningsdimension

knyttet til definitionen af en numerisk værdi.

Kampen om disciplinerne.indd 320 28/03/15 11.37

12 | Humaniora og de digitale medier 321

med de tilhørende epistemologiske, teoretiske og metodiske traditioner. Det kan illustreres med nogle få eksempler. I slutningen af 1980'erne og starten af 1990'erne udvidede Jay Bolter (1991), George P. Landow (1992), og Michal Joyce (1990) HCI-traditionens begreber om hypertext og interaktivitet med nye litterære betydningsdimensioner. Digital æstetikteori er siden blevet videreudviklet

med bidrag af bl.a. Oliver Grau, Jean Cubitt, Janet Murray og mange andre. Lev Manovich (2001) forsøger at kombinere et visuelt filmisk perspektiv med et idealiseret begreb om modulære databaser, der erstatter værkbegrebet. Kirschenbaum et al. (2010) introducerer juridiske efterforskningsmetoder rettet mod digitale kilder som litteraturanalytisk og historisk kildekritisk metoder. Herhjemme introducerer Bøgh Andersen (1991) og Finnemann (1994) semiotiske og medieteoretiske perspektiver i computerforståelsen.

Hayles (2012) tilføjer en anden differentieringsparameter, idet hun skelner mellem assimilationsstrategier, der fører traditionelle metoder over på digitale emner, versus distinktionsstrategier, der retter sig mod nye metoder, nye typer forskningsspørgsmål og måske nye felter. Alan Liu (2012) plæderer for en yderligere dimension, idet han kritiserer Digital Humanities for at have et for snævert medieteknologisk perspektiv og mangle en kritisk dimension, der i hans perspektiv vil være en forudsætning for, at Digital Humanities kan "redde humaniora ved at tage førertrøjen på", en opgave, der ikke blot bliver taget op af et voksende antal forskere, men som også synes at blive pålagt Digi tal Humanities udefra. Faglige differentieringer udgør som sagt kun en af kilderne til differentiering.

Verdens beskaffenhed er en anden kilde til voksende metodisk, teoretisk og epistemologisk differentiering. For det første fordi digitale medier giver mulighed for nye former for observation, der spænder fra scanning af kroppens indre til scanning af signaler fra det ydre univers og alt derimellem ved hjælp af et hastigt voksende antal digitale sensorer, der også tillader stadig højere opløsning

og stadig mere granuleret findeling. De nye observationsmuligheder åbner en ny fødekæde af information, der danner afsæt for nye videnskabelige analyser og erkendelser på snart sagt alle områder. For det andet fordi de digitale mediers udbredelse har medført, at digitale materialer og metoder i

dag som nævnt produceres af en langt mere heterogent sammensat aktørkreds, ligesom digitaliseringen breder sig ud i stadig mere forskelligartede sociale og naturmæssige sfærer. Den tid, hvor forskerne selv suverænt kunne definere de digitale formater og på forhånd sikre deres analytiske anvendelighed, er således løbet ud.

Kampen om disciplinerne.indd 321 28/03/15 11.37

322 Kampen om disciplinerne

Denne udvikling har især taget form gennem opbygningen og udbredelsen af internettet, der muliggjorde overgangen fra 1900-tallets stand alone-computere til 2000-tallets netværksforbundne digitale medier. I gennembrudsårene i 1990'erne var der tale om et relativt åbent og bredt samspil mellem civile, institutionelle og kommercielle aktører – de sidste fik først adgang i USA i 1993. I dag drives internettet hovedsagelig af kommercielle aktører, der baserer deres forretningsmodeller på en række nye former for interaktion mellem borgerne og de kommercielle online services.

Denne udvikling bidrager utvivlsomt til en vis homogenisering og standardisering af de dominerende formater og genrer, dvs. der er tale om et pres i retning af massemedie-relationer med få afsendere og mange modtagere, men nettet er dog endnu karakteriseret ved en principiel enhver-til-enhver-struktur, der blandt manifesteres i den såkaldte "lange hale" af netsteder, der hver især bruges af få, men som tilsammen overgår de store udbydere og muliggør, at nye ideer og software kan udbredes fra et hvilket som helst punkt, herunder ideer og software, der kan inddrages i den funktionelle arkitektur på den enkelte maskine eller i det samlede system.² Den internationale forskningsverden profiterer som mange andre miljøer og grupperinger stærkt af nye open source-kulturer, der fortsat er mulige, men en stadig mere omfattende del af de hastigt voksende mængder af digitale kildematerialer befinder sig i dag under en lille gruppe kommercielle vidensmonopolers fulde kontrol.

Man kan i dag med god grund tage afsæt i den antagelse, at digitaliseringen i 2000-tallet vil brede sig til alle samfundets sfærer, og at en hastigt voksende del af samfunds- og kulturlivet vil udspille sig i en tiltagende mangfoldighed af digitale genrer, der artikuleres på et voksende antal digitale medieplatforme. Samfundslivet vil i stigende grad blive artikuleret i digitalt fødte kildematerialer. De ældre medier kan overleve – ofte i digitaliserede former – hvis de kan tilpasse sig eller snarere forny sig i mere specialiserede nicher.

Kulturens artefakter vil i stadig højere grad indeholde digitale ingredienser, og kulturens tekster vil blive indskrevet i digitale online hypertekster, der tilbyder interaktiv, multimodal og trinløst variabel kommunikation på akserne offentlig-privat, lokal-global og fra hvem til hvem. Dermed ændres dagsordenen for hele humaniora, inklusive også den klassiske Humanities Computing, ² "Den lange hale" anvendes ofte som beskrivelse af det fænomen, at mængden af forskelligartede enkelttilfælde overgår mængden af ensartede tilfælde. Den tilhørende matematiske statistik kaldes med lidt forskellig matematisk udformning undertiden også the power law of distribution, Zipfs lov eller Pareto-distribution.

Kampen om disciplinerne.indd 322 28/03/15 11.37

12 | Humaniora og de digitale medier 323

der hovedsagelig har beskæftiget sig med digitalisering af analoge primærkilder for at studere de digitale kopier med "computational methods" (Unsworth

2002; McCarty 2002; Hockey 2004; Finneemann 2014a).

I de kommende år forskydes vægten fra digitale kopier i retning af studiet af digitalt fødte primærkilder. Dette indebærer også nye metodologiske udfordringer og impulser, ligesom der åbnes for nye fortolkninger af, hvad der kan forstås ved en computer, ved digitalisering og ved netværksforbundne digitale medier. I sidste ende udfordres også 1900-tallets begreber om information og medier. De følgende afsnit vil handle om udvalgte pejlemærker med hovedvægt på de spørgsmål, der knytter sig specifikt til studiet af digitalt fødte materialer og metoder udviklet i relation hertil.

Digitale kopier og digitale originaler

Da softwarestøttede analysemetoder kun kan anvendes i studiet af digitale materialer,

har forskerne som regel også været nødt til at definere og konstruere datamaterialets formater. Man har således i høj grad været henvist til at arbejde med digitale kopier af ikke-digitale originalkilder. Erfaringer fra Humanities Computing-traditionen har vist, at disse digitaliseringsprojekter rummer betydelige udfordringer og kræver, at man definerer nogle frames eller overordnede synsvinkler, der tjener de specifikke forskningsformål, man har for øje. Dermed begrænses muligheden for at anlægge andre analytiske vinkler. Noget lignende gælder også i de tilfælde, hvor man har brugt digitale medier til at indsamle originale digitale data via scanning af signaler fra det ydre rum eller kroppens indre, og det gælder endelig de hastigt voksende samlinger af de data, vi efterlader os i brugen af de netværksforbundne digitale medier. Til disse indbyrdes forskellige typer datamateriale kommer så de ligeledes voksende mængder af – især – webbaserede data, der produceres af alle mulige aktørtyper med hver deres mere eller mindre specialiserede brug og måske produktion af software. Datamaterialets formater er nært forbundet med dets herkomst og de epistemologiske, teoretiske og metodiske antagelser, der er lagt til grund, et tema, der har fået fornyet aktualitet i relation til big data-forskningen der tages op nedenfor. Begrebet om digitale kopier har særlig relevans, når det drejer sig om studier af historiske kilder, det være sig historiske, sproghistoriske, litterære eller kunst-historiske kilder, men det har også en mere implicit dimension, fordi mange digitaliseringsprojekter tager afsæt i forestillinger og modeller, der er udviklet i forhold til de ikke-digitale originalers former.

Kampen om disciplinerne.indd 323 28/03/15 11.37

324 Kampen om disciplinerne

Spørgsmålet om, hvad der er en digital kopi, og hvad der er en digital original, er således ofte vanskeligt at besvare. Mange applikationsprogrammer efterligner ikke-digitale formater og søger at fastholde originalernes kendetegn. Det gælder ikke blot scanningsprogrammer, men alle programmer, der anvendes til digitalisering. Digitale kopier kan samtidig være kopier på forskellig måde. Vi kan både scanne en bogside som billede og som tekst, eller den kan transformeres til lyd. Den digitale kopi er således aldrig blot en kopi, den er altid en transformeret version. Kopibegrebet giver dog stadig mening, fordi

den digitale kopi, modsat det født digitale udtryk, er en kopi af en ikke-digital original. For de fleste humanister er kildematerialets herkomst/proveniens af indlysende grunde vigtig for betydningen af forståelsen. Her vil distinktionen mellem digital kopi og original være central.

Den efterlignende relation, der er et af målene med digitalisering af historiske kilder som f.eks. af kulturarvsmaterialer, genfindes i øvrigt også i de grafiske brugergrænseflader og mange andre områder i andre dele af softwarehistorien. Brugen af kendte forestillinger i modelleringen af nye materialer kendes ligeledes fra anden teknologi-historie (path dependency), fra mediehistorien (remediering eller refunktionalisering) og fra de æstetisk faghistorier (filmens tidlige år er fyldt med filmet teater).

Men slørede grænser betyder ikke fravær af forskel. Path dependency kan lægge et konserverende præg over forståelsen og tilgangen til nye potentialer, men udgør ingen sikker bastion mod omstyrtende udviklinger. Filmen er ikke længere kun filmet teater. Musikbranchen er radikalt omdannet, og selv om 1900-tallets massemedier generelt har haft svært ved at indoptage de digitale mediers potentialer, drives de den vej, fordi de over hele verden er presset af de nye globale aktører som Google, Facebook, Amazon, Netflix og mange andre. Selv om forholdet mellem digitale kopier af ikke-digitale originaler og digitale originaler ofte er uklart, er der tre vigtige forskelle:

1. Når man arbejder med digitaliserede kopier, kan man benytte hypertext, interaktivitet og dynamiske scripts som hjælpemidler i analyse og præsentation, men disse features er aldrig en del af originalmaterialet. Det er de til gengæld, når man arbejder med digitalt fødte kilder (f.eks. webmaterialer), der som følge heraf kan være væsentlig mere komplekse end andre digitale materialer og indbyrdes også heterogene (Brügger & Finnemann 2013).

2. Digitale materialer kan for det andet udelukkende tilgås ved hjælp af

Kampen om disciplinerne.indd 324 28/03/15 11.37

12 | Humaniora og de digitale medier 325

søgemaskiner, der både er nødvendige for at fremfinde materialet og for at præsentere det på en eller anden måde, hvad enten denne måde er bestemt af softwareproducenten eller udvalgt og indstillet af aktøren.

Digi tale medier har således altid en indbygget metodologisk komponent.

De har altid en dobbelthed mellem det usynligt lagrede format

(binære sekvenser) og den sanselige tilgængeliggørelse. Denne dobbelthed kendes ikke fra andre eksternaliserede medier. Både film og tv medieres ganske vist også gennem ikke sanseligt tilgængelige formater, men her er relationen mellem den sanseligt og ikke sanseligt tilgænge lige form ikke en variabel. For digitale medier gælder det derimod, at den samme kilde altid kan repræsenteres på en flerhed af måder. Den kan omvendt aldrig repræsenteres uden filtre. Det gælder naturligvis også, hvis man arbejder med digitale kopier, men det gælder ikke, hvis man arbejder med den ikke-digitale original.

3. Født digitale materialer har for det tredje en underliggende, usynlig symbolsk arkitektur som en integreret del af materialet. Digitaliserede kopier har også en sådan underliggende arkitektur, men den er ikke er del af originalen. Den er defineret a posteriori af den instans, der er ansvarlig for digitaliseringen. For digitalt fødte materialer kan man ikke

i samme forstand ignorere den underliggende arkitektur, der både kan mobiliseres som betydningsvariabel via hypertext, interaktivitet, dynamiske scripts og i forhold til det variable fremvisningsregister.

Disse forskelle har direkte konsekvenser for forståelsen af datamaterialer. De er i dag langt mere heterogene og komplekse end de traditionelle forestillinger om databaser og andre former for velordnede, gennemklassificerede, homogene datasæt. De har også konsekvenser for kravene til softwarestøttet analyse og indirekte derigennem også for spørgsmålet om, hvordan vi kan organisere infrastrukturer omkring vores kilder og analyseredskaber. Mens der i f.eks. medievidenskaben kan være gode grunde til at udvikle interoperationelle standarder for digitaliserede mediearkiver på tværs af mediernes fysiske kendetegn (f.eks. digitaliserede aviser, tv, radio og webarkiver), vil det formentlig være vanskeligt at genbruge de mere avancerede softwareredskaber, der udvikles i forhold til de specifikke medier på tværs af mediernes forskellige fysiske egenskaber. For webarkivernes vedkommende vil interoperationelle løsninger mellem de forskellige nationale webarkiver måske være at foretrække. Selv om der blandt andet af sproglige og kulturelle grunde er – ofte oversete – stærke

Kampen om disciplinerne.indd 325 28/03/15 11.37

326 Kampen om disciplinerne

nationale særpræg i netbrugen, er der også tale om en tendens i retning af mere grænseoverskridende netbrug.

De netværksforbundne digitale mediers udbredelse medfører en hidtil uset eksponentiel vækst i mængden af stadig mere forskelligartede og stadig mere heterogene digitale materialer. Alene mængderne sprænger gængse forestillingsrammer.

Selv om der er store usikkerheder i opgørelsen, er det generelle billede, at dataproduktionen siden årtusindskiftet har nået et trindhøjere niveau (Kitchin & Lauriault 2014). Det antages af nogle, at den mængde informationer, der er produceret de seneste to-tre år, er større end den mængde, der er produceret i hele menneskehedens hidtidige historie.³ Uanset usikkerheden ved disse opgørelser er der tale om overvældende mængder dokumentationsdata både i henseende til, hvad der dokumenteres, hvordan dokumentationen frembringes – og hvad man kan gøre ved den foreliggende dokumentation, mens den produceres, og hvordan man kan bevare den, når den er produceret. Nogle af disse datamængder danner også afsæt for udvikling og afprøvning af en række nye forskningsmetoder, blandt andet de metoder, der ofte omtales under fællesbetegnelsen Big Data.

Big Data

Ønsket om at håndtere, beregne og analysere store datamængder har været en hoveddrivkraft i udviklingen af digitale medier fra den spæde begyndelse sidst i 1930'erne. Frem til slutningen af 1990'erne var den altdominerende bestræbelse baseret på produktionen af standardiserede datasæt, der kunne analyseres med et begrænset antal veldefinerede og hver især meningsfulde parametre. I 1980'erne udvikledes ideen om neurale, parallelprocesserende netværk (konnektionisme), der skulle gøre det muligt at lade computerne identificere

genkommende mønstre ud fra enkeltcases – ikke standardiserede datasæt – ved hjælp af læringsalgoritmer. Selv om der her arbejdes med mindre ensartede datasæt, er de fortsat udvalgt af forskerne ud fra antagelser om relevante relationer mellem de forskellige cases. De eksempler, der bruges til at træne computerne, er forskellige instanser af det samme generiske fænomen. I løbet af århundredet udvikles der endelig statistiske metoder, hvor man kompenserer den statistiske usikkerhed ved brug af repræsentativt udvalgte samples.

3 For et overview over opgørelsesmetoder, se Dienes (2012) og Hilbert (2012).

Kampen om disciplinerne.indd 326 28/03/15 11.37

12 | Humaniora og de digitale medier 327

De senere års voldsomt stigende interesse for Big Data bygger på et andet udgangspunkt. Selv om forskerne (eller andre analytikere) stadig vælger de data sæt, de vil analysere, sker dette valg hverken med sigte på at have et fuldstændigt systematisk datasæt, et eksemplarisk udvalgt sæt eller et repræsentativt

sæt. Man tager så at sige revl og krat af eksisterende data, for så vidt datamængderne skønnes tilstrækkelig store til rumme interessant ny viden. Google Flu Trends (GFT), der sigter på at forudsige influenza-epidemier ud fra hyppigheden af internetsøgninger på bestemte, udvalgte søgeord, er et ganske illustrativt og letforståeligt eksempel. Metoden har været under udvikling siden 2004, i kølvandet på SARS-influenzaepidemien i 2002-2003, og blev indtil 2013 anset for at være den mest effektive forudsigelsesmetode for influenza i henseende til tidlig varsling. GFT kan betragtes som en specialiseret videreudvikling af Googles oprindelige "Zeitgeist"-statistikker over mest benyttede søgeord i en given periode. Google Flu Trends demonstrerer de vigtigste kendetegn ved Big Data, som de f.eks. beskrives i Viktor Mayer-Schönberger og Kenneth Cukier (2013), der ser et konceptuelt nybrud på fire punkter:

1. Big Data er baseret på store data mængder eller mere præcist på alle de data, der findes, frem for repræsentativt udvalgte datasæt – samples. Googles GFT udvælger og indsamler alle forekomster af et givet sæt af søgeord i en given region i en given periode.
2. De benyttede data er irregulære, rodede data. Det kan f.eks. ofte være dataspor afgivet af mange forskellige kilders indbyrdes uafhængige adfærd. For GFT vides det f.eks. ikke, hvem der sidder bag de maskiner, hvorfra der søges, om det er patienter, pårørende, studerende, journalister eller måske robotter.
3. Data analyseres med sigte på at finde korrelationer frem for årsager. Relationen mellem søgeord på Google og influenzasygdommen er ikke kausal.
4. De datamængder, der indsamles, skal "dataficeres", dvs. behandles som kvantitative, numeriske data uanset deres oprindelige formål og meningsbestemte brug og uafhængigt af den kontekst, hvori de er blevet brugt. GFT baserer sig på frekvensen af forekomster af mekanisk identificerede forekomster af søgeord, dvs. der måles på ordenes binære repræsentationsform – som det er gængs i mekaniske søgemaskiner – og ikke på ordenes mening. Der er i GFT som i de fleste Big Data-analyser tale om genbrug

Kampen om disciplinerne.indd 327 28/03/15 11.37

328 Kampen om disciplinerne

af data, der er skabt til andre formål af mange indbyrdes uafhængige parter.

Begrebet "dataficering" refererer hos Mayer-Schönberger og Cukier – lidt uklart – både til den kvantitative betragtning og "det numeriske" og modstilles begrebet digitalisering, der forstås som en proces, hvor man konverterer analog information til nuller og ettaller, så den kan behandles i en computer, hvad der i deres termer også må være numerisk. Der er imidlertid (uanset om man kalder de to bits for numeriske eller for betydningsløse bogstaver) ingen forskel på, om man konverterer bogstaver eller tal til det binære alfabet.

Der er derimod forskel på, om man konverterer bogstaver og tal som billeder eller med – redigerbar – karaktergenkendelse, der gør enhederne operative, hvad enten de er bogstaver eller tal. Det er imidlertid væsentligt vanskeligere at konvertere komplekse sprogregler end matematiske regneregler. Mens regne arkene virker fortræffeligt inden for de finite størrelses domæne, halter sprogoversættelse langt bagefter. For Big Data er det afgørende imidlertid, at der kan søges på vilkårligt udvalgte bitmønstre i det materiale, der undersøges, men man vælger naturligvis ofte at skræddersy søgerutinerne i forhold til de forventninger, der ligger bag det enkelte projekt. Kitchin (2013, 2014) opregner flere andre kendetegn som karakteristiske, heriblandt:

- Høj processeringhastighed, ikke mindst i forhold til analyse af data, der skabes i realtid eller nær-realtid
- Diversitet i varieteter, omfattende både strukturerede og ikke-strukturerede data
- At man søger "alle data", der foreligger for hele populationen, frem for ordnede data
- Høj opløsning og entydig indeksering af alle enheder.

Der er her tale om kendetegn, der kan være mere eller mindre karakteristiske. Det er f.eks. næppe realistisk at forestille sig den samlede population af websider som objekt, blandt andet fordi webmaterialet udvikler sig hurtigere end den mulige samtidige analyse, og fordi mange essentielle interaktive og dynamiske

features vil unddrage sig undersøgelse. Webmaterialer og arkiverede webmaterialers særlige karakteristika behandles nærmere i Brügger og Finnemann (2013).

Det er ikke kun Googles søgemaskine, der samler interessant viden. Der

Kampen om disciplinerne.indd 328 28/03/15 11.37

12 | Humaniora og de digitale medier 329

ligger også interessant viden i Googles bogprojekt, hvor man er i gang med digitaliseringen af alle verdens bøger. Hver af disse bøger har jo haft sit eget formål, men den samlede mængde af digitaliserede bøger kan nu bruges til at studere helt andre ting.

Det er især blevet udnyttet af to yngre naturvidenskabelige forskere fra Harvard, Erez Aiden, uddannet specialist i genomstudier, og Jean Baptiste Michel, der dels er iværksætter, dels forsker med speciale i kvantitative metoder. De studerer kultur via et korpus af Google Books, der foreløbig omfatter digitale kopier af ca. 25 procent af verdens bøger (Aiden & Michel 2013). Google Books er en interessant kilde, fordi materialet strækker sig over århundreder.

Det er ikke bare big data, men long data. Blandt deres eksempler er en frekvensanalyse af forekomsten af de stærke verber i engelsk med kilder, der går tilbage til 800-900-tallet. (Aiden & Michel 2013, 23, 38-39, 42-43). De mener at kunne påvise, at de hyppigst brugte stærke verber også vedbliver at være stærke, mens sjældnere brugte verber skifter til at blive svage verber, der følger standarden. Hvis denne konklusion er på linje med gængse antagelser i sprogvidenskaben, bidrager analysen primært med en mere solid empiri, da man har undersøgt alle de skriftlige spor, der findes. Det mere interessante spørgsmål er: Hvad nu hvis det stred mod tidligere forskning?

Big Data-forskning beskrives ofte som teorifri og datadreven (modsat hypotesedreven) forskning og med en undertone af, at man her nærmer sig en dogmefri, objektiv sandhed. Termen blev boostet i slagfærdig Silicon Valley-retorik med Chris Andersens artikel "The end of Theory: The data deluge makes the scientific method obsolete", publiceret i Wired Magazine 16. juli 2008. Termen har siden bredt sig til mange områder – datadreven socialvidenskab, datadreven journalistik og ikke mindst datadreven marketing, men den er også blevet genstand for mere dybtgående formuleringer (f.eks. hos Hey et al. 2009).

En kritisk diskussion af forholdet mellem hypotese og datadreven forskning i biologien findes bl.a. i Leonelli (2012). Mens man her på den ene side ser kortlægningen af det humane genom som afsæt for nutidens Big Dataforskning, trækkes der på den anden side tråde tilbage til tidligere former for induktiv-empiristisk forskning baseret på opbygning og klassifikation af datakorpora. Der peges i det perspektiv på tre særlige bio-big-data-karakteristika. For det første de mere komplekse biologiske data, der skabes med henblik på at forstå relationer mellem de forskellige biologiske organisationsniveauer (genom,

Kampen om disciplinerne.indd 329 28/03/15 11.37

330 Kampen om disciplinerne

celle, organisme, population, økosystem). For det andet de uhyre omfattende socioøkonomiske interesser knyttet til de mulige biomedicinske opdagelser og endelig for det tredje den hastige udvikling af de computerbaserede lagringssøgnings-, analyse- og visningsmetoder (Leonelli 2012, 2-4). Det specifikke – hypotesefri datadrevne – Big Data-element knyttes til den induktive tilgang til eksisterende data og datamaskinernes betydning i processen med at udtrække meningsfulde mønstre, men det påpeges, at hverken forskernes induktive slutninger eller den mekaniske analyse kan betragtes som hypotesefri, og der peges også på, at der er en tendens til at integrere andre metoder og tilgange. Trods den megen hype om den datadrevne, hypotesefri Big Data-forskning synes de fleste forskere i dag mere forsigtige eller mere kritiske. Hvis Big Data er "rodede data", der tilmed kun kan studeres for korrelationer, hvordan bedømmer man så disse korrelationers gyldighed?

Aiden og Michel afviser således den teori- og hypotesedrevne Big Data-forestilling om datadrevet forskning. De anvender selv evolutionær teori i deres tolkning af long data, men de fastholder samtidig, at der er tale om et konceptuelt brud. Big Data repræsenterer et uhyre reservoir af hidtil uudnyttet viden, der kan ekstraheres med de metodiske optikker, der udvikles i forhold til "rodede" datasæt, der ikke er tilrettelagt af forskerne, selv om forskerne naturligvis foretager identifikation og udvælgelse. Aiden og Michel forestiller

sig en historie om menneskets fortid set gennem robotter. De udvikler således en evolutionært funderet, kvantitativt orienteret optik, kaldet culturomics, der retter sig mod studiet af historiske forandringer på basis af store digitaliserede datamængder. Et humaniora, der ikke afprøver disse perspektiver på de hastigt voksende mængder digitaliserede kulturarvsmaterialer, kender ikke sin besøgstid.

Aiden og Michels projekt er interessant, fordi de sammen med Google har udviklet og offentliggjort en "N-gram viewer", der kan levere automatiserede korrelationsindeks for vilkårligt valgte søgeord inden for Googles bogkorpus, der sigter på at indbefatte digitale kopier af alle verdens trykte bøger. N-gram vieweren er frit tilgængelig, og man kan søge i forskellige – til dels selvvalgte - dele af dette korpus (Figur 12.1).

Kampen om disciplinerne.indd 330 28/03/15 11.37

12 | Humaniora og de digitale medier 331

Figur 12.1. Søgning på Frankenstein, Albert Einstein og Sherlock Holmes (books.google.com/ngrams).

N-gram viewerens forside vises med to fiktive og en real person som default. N-gram vieweren er ligeglad, den søger ikke på ordenes betydning eller status, men på de binære repræsentationsmønstre. Selv om der tegner sig tydelige mønstre i store datamængder, er der også tydelige usikkerheder. Big data er ikke "the end of theory" og heller ikke "the end of interpretation". Det kan illustreres med en N-gram søgning på Big Data i USA hhv. Storbritannien, hvor man burde forvente en høj korrelation (Figur 12.2 og 12.3).

Figur 12.2. Søgning på Big Data: American English, 1950-2008.

Kampen om disciplinerne.indd 331 28/03/15 11.37

332 Kampen om disciplinerne

Figur 12.3. Søgning på Big Data: British English, 1950-2008.

Som graferne viser, dukker søgeordene op i forskellige tiår, og for det andet falder et højdepunkt i USA samtidig med et lavpunkt i Storbritannien og omvendt. Det er ikke særlig sandsynligt, at Storbritannien skulle halte et tiår efter USA i forhold til den faglige interesse for Big Data. Det er også svært at forstå den amerikanske kurves fald op til 2008, hvor termen fik et veritabelt gennembrud.

Der er en række mulige fejlkilder at tage i betragtning. Søgetermerne har måske været brugt med forskellig betydning over tid i hhv. USA og England. Der kan også have været brugt andre termer om samme fænomen i det ene korpus (f.eks. "massive data", der har en højere forekomst end "big data" i begge korpora gennem hele perioden). De to korpora kan også være uens sammensat i henseende til repræsenterede fagområder eller disproporcionale mængder af inkluderede kilder. Selv om det totale kildemateriale er omfattende, er litteraturen om Big Data det ikke. Disse fejlkilder peger endelig tilsammen på en bagvedliggende fejlkilde, nemlig at forskerne kan mangle den fornødne baggrundsviden om termernes historie, om deres betydning og den kulturelle kontekst. Når det gælder de trykte mediers historie, omfatter det også viden om de trykte mediers rolle i forskellige samfund og til forskellige tider i det samme samfund, ligesom de trykte mediers betydning varierer i forhold til andre mulige medier.

Big Data fører altså direkte frem til de fortolknings spørgsmål, der hidtil

har udgjort en humanvidenskabelig hjemmebane. Men det foregår i forhold til nye typer kildemateriale og er forbundet med nye softwarestøttede analysemetoder.

Fortolkningen får til gengæld en ny og mere omfattende betydning

Kampen om disciplinerne.indd 332 28/03/15 11.37

12 | Humaniora og de digitale medier 333

i takt med udbredelsen af statistiske metoder, netværksanalyser, korrelationsstudier og studier af messy korpora, der ikke kan tilskrives validitet og reliabilitet via de gængse repræsentative samplingmetoder.

Ovenfor blev Googles Flu Trend brugt til at introducere Big Data. I marts 2014 publicerede Lazer et al. (2014) en lille artikel i Science, der opridses de – stadig mere tydelige – usikkerheder, der knytter sig til denne metode.

Lazer et al. (2014) identificerer en række systematiske fejlkilder:

1. Søgeordene, der bruges, kan have forskellig relevans i forskellige regioner, og afgrænsningen bør ske relativt til andre kontekstuelle forhold (årstider osv.).
2. Google-søgemaskine justeres konstant, så “samme søgning” giver forskellige resultater fra gang til gang.
3. Google og andre aktører tilføjer fordrejende forstærkere, f.eks. Amazons service, der multiplicerer forekomsten af søgetermer, eller Googles egen forretningsmodel, der kobler annoncer på søgeordene.
4. Google offentliggør ikke viften af søgetermer og vanskeliggør dermed efterprøvningen af resultaterne.

De konkluderer, at Big Data-analyser aldrig kan stå alene og bør kombineres med både traditionelt, repræsentativt forankret statistik og andre analysemetoder.

Der er behov ikke bare for analyse af big data, men også af “small data” og “all data”.

Datamaterialer har med andre ord ingen særlig “objektiv” eller teori- og hypotesefri status, men skal betragtes på linje med alle andre kildematerialer, de skal forstås ud fra deres herkomst, den materielle overlevering, proveniens og formålet med deres produktion og transformation, dvs. deres historie. Der findes ingen “rå data”, data er altid allerede kogt. Lignende konklusioner nås bl.a. i Gitelman (2013), Wouters et al. (2013) og Lesage & Hackett (2014). På den anden side adskiller digitale materialer sig fra andre materiale typer ikke blot ved deres eksplosivt voksende mængde, der omhandler stadig flere forskelligartede fænomener og processer af enhver art, men også ved det særlige manipulationsrepertoire. Digitale medier er ikke blot bestemt på det elementære materialeniveau af kravet om binær repræsentation, der er også en række bestemmelser på medieniveauet. De omfatter dels en række invariante betingelser, der gælder alle digitale medier: Alle data og regler manifesteres i det binære alfabet; der kan implementeres nye ideer i den funktionelle arkitek-

Kampen om disciplinerne.indd 333 28/03/15 11.37

334 Kampen om disciplinerne

ture; digitale materialer og redskaber kan indeholde hypertextuelle, interaktive og andre dynamiske features; alle transaktioner efterlader digitale spor, der kan lagres og gen-repræsenteres eller slettes. Digitale materialer er usynlige, de skal opsøges og repræsenteres for at blive tilgængelige for menneskelig sansning og fortolkning.

For det andet omfatter dette repertoire et register af endnu ikke udviklede,

mulige genrer, metoder og anvendelsesmuligheder. De grundlæggende variable registre er hypertext, interaktivitet og multimodalitet. Med internettet kommer også den trinløst variable kommunikation på skalaerne privat-offentlig, lokal-global og fra hvem til hvem.

Digitale medier er plastiske på en måde, der mere ligner sproget end andre mekaniske maskiner, inklusive radio og tv, men de adskiller sig fra sproget, fordi vi kan eksternalisere alle mulige dynamiske operationer på den ene betingelse, at de kan mekaniseres og udføres i et endeligt antal skridt.

Man kan endelig hævde, at det fundamentale kendetegn ved "computational methods" i dag ikke er beregning, men derimod søgning. Den konstitutionelle betydning af søgning er i sidste ende forankret i, at enhver operation i en computer vil starte med en søgeinstruks, der udpeger det næste skridt eller den næste sekvens af skridt. Forskydningen fra computation til search som det fundamentale metodologiske udgangspunkt indebærer også en dramatisk ændring i den grundlæggende forståelse af og tilgang til digitale materialer, fordi man nu kan benytte enhver konstellation af data som søgemønster, ofte også på tværs af de benyttede data- og databaseformater, som det f.eks. kendes i den i dag udbredte webbaserede fritekstsøgning.

Overgangen fra enkeltstående til netværksforbundne digitale medier åbner også for nye metodologiske potentialer som f.eks. websurveys og big datastudier på Google-søgeord og andre webmaterialer og en række andre former for crowdsourcing, men også for nye former for forskningssamarbejder og udveksling af forskningsresultater. Der åbnes også for studiet af de nye digitalt fødte genrer, der dannes omkring de mere komplekse og indbyrdes heterogene digitale materialer – herunder de i dag hyppigt studerede materialer på de sociale medie-platforme, de omfattende mængder af digitale spor, der afgives ved enhver brug af nettet.

Både mængden, forskelligheden indbyrdes og heterogeniteten i de digitale materialer fremmer behovet for metodologiske nybrud, hvor der endnu kun er taget embryonale tilløb. Selv om "social data" og "big data" i dag er de hyppigst omtalte datatyper, vil vi se nye metoder til analyse af alle former for data: "big

Kampen om disciplinerne.indd 334 28/03/15 11.37

12 | Humaniora og de digitale medier 335

data", "small data", "broad data" og "long data" – kort sagt "all data". Kvantitative og kvalitative analyser kan kombineres på stadig flere måder, f.eks. i form af websurveys, der tillader integration af traditionelle surveysspørgsmål med hypertextuelle

(f.eks. betingede og demonstrative) og interaktive (uddybende)

undersøgelser, ligesom det er muligt at følge op med traditionelle kvalitative interviews. Websurveys rejser samtidig nye spørgsmål omkring validitet og reliabilitet.

Information er altid medieret.

Om behovet for et nyt mediebegreb

Digitale medier er som alle andre medier og teknologier menneskeskabte. De er frembragt af menneskelige anstrengelser for at tilfredsstille menneskelige behov og længsler. Mens studiet af digitale medier således er et genuint human- og

samfundsvidenskabeligt anliggende, har naturvidenskaberne spillet en væsentlig historisk rolle både i formuleringen af de behov, der har drevet udviklingen af digitale medier, og i selve udformningen af disse medier. Det gør sig ikke blot gældende i den forstand, at der siden 1930'erne har været lagt stor vægt på beregningskapacitet og hastighed, men også i den grundlæggende forståelse af computeren som en "autonom", objektiv entitet, ligesom faget computerscience/datalogi, der beskæftiger sig med artefaktiske materialer og metoder, næsten altid er placeret på naturvidenskabelige fakulteter. Datalogien som fag beskæftiger sig næppe mere med teknologiens fysiske egenskaber, end for eksempel lingvistikken beskæftiger sig med talens og skriftens fysiske egenskaber. Hvis man ved humaniora forstår de videnskaber, der beskæftiger sig med det menneskelige og det menneskeskabte, burde det ikke blot omfatte samfundsvidenskaberne, men også fag som datalogi og den generelle lære om videnskabelige metoder, for slet ikke at tale om studiet af de menneskelige aktiviteter virkninger på naturgrundlaget. Det ville så kræve, at Beethovens musik og menneskets introspektive, selvrefleksive og ekspressive kapaciteter også blev betragtet som dele af den naturvidenskabelige forsknings forståelse af naturgrundlaget. Den humanistiske forskning skulle samtidig beskæftige sig med, hvordan denne del af naturen er forbundet med naturgrundlaget i øvrigt. Den største udfordring er måske, at en mediering mellem fakulteterne kun vil foregå, hvis man opgiver de enkelte fakulteters respektive, autonome domskriterier. Det er ikke enkelt at ophæve den konceptuelle distinktion mellem fakulteterne.

Kampen om disciplinerne.indd 335 28/03/15 11.37

336 Kampen om disciplinerne

Spørgsmålet er, om de eksisterende fakultetsopdelinger ligefrem er skadelige for forskningen. Der er tegn på, at det er tilfældet i en række tværfaglige forskningsgrene. Forskningen retter i disse år stærkere fokus mod grænselandet mellem humaniora og de andre fakulteter. Vi ser det f.eks. i den stærkt voksende interesse for forholdet mellem det biologiske og det kognitive, hjernen og bevidstheden, mellem biologisk og social arv, mellem driftsmotiver og fornuftmotiver

i det politiske og kulturelle, mellem det private og det offentlige, ligesom vi bruger de digitale medier til at flytte grænserne mellem det, vi har i hovedet, og det, vi lagrer i eksterne medier uden for os selv. Måske vi – eller vores børn og børnebørn - er ved at flytte fra en mere introvert, litterær kultur til en mere ekstrovert og også mere kropsligt orienteret, digital kultur? Hvis man i 1900-tallet tænkte det fysiske, det biologiske, det psykiske, det sproglige, det litterære, det historiske osv. for sig som lagdelte, parallelle universer i en form for psykofysisk parallelisme (Neumann 1932), har forskningen i dag forsigtigt bevæget sig i retning af psykofysisk interaktionisme. Denne udvikling befordres især af, at man nu er i stand til at scanne hjerneaktiviteter og påvise mentale inputs effekter på neurofysiologiske processer, men den befordres også af vores brug af digitale medier i en bredere forstand, fordi digitaliseringen gør sig gældende på tværs af de tidligere domænegrænser. De digitale medier har egenskaber, der overskrider det mediebegreb, der blev formuleret i relation til 1900-tallets dominerende massemedier (Finnemann 2014b), men de overskrider også den parallelle, immaterielle forståelse

af informationsbegrebet, der blev formuleret i relation til fortolkningen af termodynamik

og kvantefysik og i relation til den digitale informationsteknologi.

Da man i 1900-tallets informationstænkning flyttede *res cogitans* ind i *res extensa*s i tid og rum udstrakte verden, erstattede man den frie cartesianske tanke med et fysisk defineret informationsbegreb, idet man betragtede informationen som en mekanisk refleks af den fysiske verden – med den kvantemekaniske ubestemthed som undtagelse.

Det termodynamiske og kvantefysiske informationsbegreb var smalt i den forstand, at det kun drejede sig om den information, et fysisk system afgav til en iagttager af systemet i en slags en-til-en-relation. Shannon generaliserede denne tilgang og betragtede al information som fysisk medieret, idet han samtidig så helt bort fra informationens mulige indhold. Mediet er i Shannons teori alene til stede som neutral forudsætning og som kilde til støj. Noget mere omfattende begreb om mediering af information, viden og kognition kom ikke på tale.

Kampen om disciplinerne.indd 336 28/03/15 11.37

12 | Humaniora og de digitale medier 337

Den fundamentale distinktion mellem informations- og mediebegrebet kan ses som en forsinket og forskudt cartesiansk arv grundlagt i distinktionen mellem *res extensa* og *res cogitans*. Cartesianismen udelukker, som det bl.a. er påpeget af Guillory (2010), et mediebegreb, der kunne mediere udvekslingen mellem de to sfærer. Mens Turing selv var mest optaget af den universelle computer,

der beskriver, hvordan alle mulige beregningsopgaver kan udføres på samme maskine, er det snarere hans "choice"-maskine, der er det paradigmatisk mest træffende billede af udveksling mellem fysisk mekaniske og symbolske processer. Mens den universelle computer tænkes som en automat, tillader choice-maskinen udefrakommende nye input, der forandrer dens fremtidige processer. Den inkluderer på paradigmatisk vis det omstridte, men stadig mere påtrængende begreb "downward causation" (Andersen et al. 2000).

Grænsen mellem det fysiske og det mentale nedbrydes gradvis indefra fra begge sider. Mens der i humaniora sker en forsigtig nyorientering i retning af at spørge til naturgrundlaget i menneskets virksomhed, forlader naturvidenskaben i et vist omfang den deterministiske naturlovstænkning – dels til fordel for mere specifikke, nytte- og evidensbaserede forståelser, dels i erkendelsen af, at ny viden undertiden fører til forkastelse af tidligere teorier, og som en konsekvens af, at både den biologiske og den kulturelle udvikling er karakteriseret ved udvikling af trindhøjere organisationsformer, der ikke lader sig forklare med de anerkendte fysiske parametre. Omvendt tager de nyere teorier om netværk og kompleksitet næsten udelukkende afsæt i eksisterende kompleksitetsniveauer

og udelader spørgsmålet om, hvordan man kan forklare stigende kompleksitet og interrelationer mellem højere og lavere organisationstrin. At hjernen på en gang er medie for bevidstheden, men ikke operativt påvirkelig af bevidsthedsindhold er i alle tilfælde et – ikke længere plausibelt – dogme (se Kapitel 11).

I dag kan man således med solidt belæg hævde, at al menneskeligt tilgængelig information er fysisk medieret. Medieringen har altid ikke blot en fysisk dimension, men også en kodningsdimension og en indholdsdimension

(Finnemann 2014b). Det menneskelige bevidsthedsliv, tankeprocesser, følelsesregistre og hvad der ellers kan betragtes som mentale aktiviteter og – indre – psykiske tilstande har en neurofysiologisk realisering. Vi kan udtrykke – eksternalisere – de mentale processer på forskellig måde ved hjælp af et – op gennem historien – voksende repertoire af medier, fra det talte sprog over skrift, bogtryk, analogt elektroniske medier til de digitale medier for nu at nævne de fem vigtigste, epokedannende kultur- og mediehistoriske fornyel-

Kampen om disciplinerne.indd 337 28/03/15 11.37

338 Kampen om disciplinerne

ser. Hvert nyt medie bidrager på den ene side til at udvide den menneskelige kommunikations- og handlingsrækkevidde i henseende til tid og sted, men de bidrager på den anden side også til at udvide det, vi kan opleve i og vide om os selv og verden, og de muliggør nye vidensformer ved at udvide vores observations- og iagttagelsesmuligheder.

De digitale mediers egenskaber er kun en af flere anledninger til at reformulere mediebegrebet. En anden anledning er behovet for et mediebegreb, der kan gøre det muligt at forankre det "rene" informationsbegreb i den samme "beskidte", tid- og stedbundne fysiske verden, som bevidstheden også befinder sig i.

Den socialkonstruktivistiske bølge, der har domineret den human- og samfundsvidenskabelige tænkning gennem de seneste 30 år, har leveret centrale bidrag til denne proces. Det gælder f.eks. kritikken af determinisme og fokuseringen på naturvidenskabernes sociale karakter. Men konstruktivismen kommer samtidig til kort i sin – mere eller mindre radikalt formulerede subjektivistiske – vægring ved at anerkende menneskets erkendelsesadgang til naturgrundlaget. Denne begrænsning er nært forbundet med en problematisk forståelse af det situerede i den menneskelige eksistens. At information, viden og handling altid er situeret og kontekstuel betyder ikke, at mennesket er indespærret i et situeret "her og nu". Menneskets erindring og sprog vil altid overskride denne moderne individualistiske horisont. Denne overskridelsesmulighed udvides i en eller flere henseender ved hvert nyt hukommelses- og kommunikationsmedie, vi selv skaber, inklusive de digitale, der ikke blot er 2000-tallets nye mediehistoriske omdrejningspunkt, men også det videnshistoriske og forskningsmetodiske omdrejningspunkt.

Mens forskningen i stigende grad henter sin magt og legitimitet fra det, den påstår at vide, er forskningen selv i sidste ende funderet i forholdet til det, den ikke ved. De helt store udfordringer ligger her måske ikke så meget i de enkelte fagvidenskaber som i det forhold, at de alle handler om et og samme univers. Et skridt kunne være at indrette teorierne derefter. Man kan måske også formode, at jo flere data vi får, jo mere går det i den retning.

Kampen om disciplinerne.indd 338 28/03/15 11.37

Litteratur

- Aiden, E. & Michel, J.-B. (2013). *Uncharted. Big Data as a lens on Human Culture*. New York: Riverhead Books.
- Andersen, P.B. (1991). *A Theory of Computer Semiotics*. Cambridge: Cambridge University Press.
- Andersen, P.B. et al. (2000). *Downward Causation. Minds, Bodies and Matter*. Aarhus: Aarhus Universitetsforlag.
- Bannon, L. & Pylyshyn, Z. (red.) (1989). *Perspectives on the Computer Revolution*. NorwoodNJ: Ablex.
- Berners-Lee, T. (1999). *Weaving the Web: The Original Design and Ultimate Destiny of the World Wide Web by its Inventor*. San Francisco: Harper.
- Bolter, J.D. (1990). *Writing Space: The Computer, Hypertext, and the History of Writing*. Mahwah: Lawrence Erlbaum Associates.
- Brügger & Finnemann (2013) XXXXXXXXXXXXXXX
- Burgess, J. & Puschmann, C. (2014). "Metaphors of Big Data". *International Journal of Communication*, 8:1690–1709 1932–8036/20140005
- Dienes, I.A. (2012). "Meta Study of 26 'How Much Information' Studies: Sine Qua Nons and Solutions". *International Journal of Communication*, 6: 874–906 1932–8036/20120874
- Digital Humanities Lab, Denmark, DIGHUMLAB, [dighumlab.com/EU COST action IS1004](http://dighumlab.com/EU_COST_action_IS1004). 2011-2015. Web-based data-collection - methodological challenges, solutions and implementations.
- Finnemann, N.O. (1994). *Tanke, sprog og maskine. En teoretisk analyse af computerens symbolskeegenskaber*. København: Akademisk Forlag.
- Finnemann, N.O. (2005). *Internettet i mediehistorisk perspektiv*. København: Samfundslitteratur.
- Finnemann, N.O. (2011). "Mediatization theory and Digital Media". *Communications*, 36:67-89.
- Finnemann, N.O. (2014a). "Digital Humanities and networked digital media". *MedieKultur*, 57: 94-114.
- Finnemann, N.O. (2014b). "Digitization: New Trajectories of Mediatization". I: Lundby, K.(red.). *Handbooks of Communications*, 21: 297-322. Berlin: De Gruyter Mouton.
- Gitelman, L. (red.) (2014). *"Raw Data" is an Oxymoron*. Cambridge, MA: The MIT Press.
- Google Flu Trends (GFT), www.google.org/flutrends/. Verificeret 29. juli 2014
- Google Zeitgeist, arkiv for årene 2001-2008, www.google.com/intl/en/press/zeitgeist/archive.html
- Guillory, J. (2010). "Genesis of the media concept". *Critical Inquiry*, 36: 321–362.
- Kampen om disciplinerne.indd 339 28/03/15 11.37

340 Kampen om disciplinerne

- Haugeland, J. (1985). *Artificial Intelligence. The very Idea*. Cambridge MA: The MIT Press.
- Hayles, N.K. (2012). *How We Think. Digital Media and contemporary Technogenesis*. Chicago: The University of Chicago Press.
- Hey, T. Tansley S. & Tolle, K. (2009). *THE Fourth Paradigm. Data Intensive Scientific Discovery*. Redmond, Washington: Microsoft Research.
- Hilbert, M. (2012). "How to Measure 'How Much Information'? Theoretical, Methodological, and Statistical Challenges for the Social Sciences". *International Journal of Communication*, 6: 1042–1055 1932–8036/20121042 (verificeret 31.7 2014).
- Hockey, S. (2004). "The History of Humanities Computing". I: Schreibman, S., Siemens, R. & Unsworth, J. (red.). *Companion to Digital Humanities. Blackwell Companions to Literature and Culture*. Oxford: Blackwell. Consulted on: 15 September 2013, www.digitalhumanities.org/companion/
- Joyce, M. (1989). *Afternoon, a Story*. CD-ROM Mac/Win edition (June 1, 1999): 300 pages. Watertown, MA: Eastgate Systems Inc.
- Kirschenbaum, M.G., Ovenden, R. & Redwine, G. (2010). *Digital Forensics and Born-Digital Content in Cultural Heritage*. Washington DC: Council on Library and Information Resources.
- Kitchin, R. (2013). "Big data and human geography: opportunities, challenges and risks". *Dialogues in Human Geography*, 3(3): 262–267.
- Kitchin, R. & Lauriault, T.P. (2014). *Towards critical data studies: Charting and unpacking data assemblages and their work*. Pre-print version of chapter to be published in Eckert, J., Shears, A. & Thatcher, J. (red.). *Geoweb and Big Data*. University of Nebraska Press.
- Landow, G.P. (1992). *Hypertext. The Convergence of contemporary critical theory and technology*. Baltimore: The Johns Hopkins University Press.

- Lazer, D., Kennedy, R. King, G. & Vespignani, A. (2014). "The Parable of Google Flu: Traps in Big Data Analysis". *Science*, 343, 14 March 2014. www.sciencemag.org. Verificeret 29.juli 2014.
- Leonelli, S. (red.) (2012). Special Issue of Studies in the History and the Philosophy of the Biological and Biomedical Sciences: Part C 43:1. www.sciencedirect.com/science/journal/13698486/43/1 verificeret 29.7 2014
- Lesage, F. & Hackett, R.A. (2014). "Between Objectivity and Openness—The Mediality of Data for Journalism". *Media and Communication*, 2(2): 42-54.
- Liu, A. (2012). "The state of the Digital Humanities: A Report and a Critique". *Arts and Humanities in Higher Education*, 11(1-2): 8-41. Consulted on: 23 September 2013, ahh.sagepub.com/content/11/1-2/8
- Mahrt, M. & Scharnow, M. (2013). "The Value of Big Data in Digital Media Research". *Journal of Broadcasting & Electronic Media*, 57(1): 20–33.

Kampen om disciplinene.indd 340 28/03/15 11.37

12

- Manovich, L. (2001). *The Language of New Media*. Cambridge, MA: The MIT Press.
- Mayer-Schönberger, V. & Cukier, K. (2013). *Big Data. A Revolution That Will Transform How We Live, Work and Think*. London: John Murray.
- Mazlish, B. (1995). *The Fourth Discontinuity. The Co-Evolution of Humans and Machines*. New Haven: Yale University Press.
- McCarty, W. (2002). "Humanities Computing: Essential Problems, Experimental Practice". *Literary Linguistic Computing*, 17(1): 103-125.
- Neumann, J. von (1932). *Mathematische Grundlagen der Quantenmechanik*. Berlin: Springer.
- Presner, T, Johanson, C. et al. (2009). *The Promise of Digital Humanities - A Whitepaper*. White Paper from the UCLA Center for Digital Humanities.
- Ramsay, S. (2011). "Who's in and who's out". Blog entry, paper presented at The History and Future of the Digital Humanities panel at the MLA Convention. Consulted on: 31 October 2013, lenz.unl.edu/papers/2011/01/08/whos-in-and-whos-out.html
- Schnapp, J., Drucker, J., Burdick, A., Lunenfeld, P. & Presner, T. (2012). *Digital Humanities*. Cambridge, MA: The MIT Press.
- Svensson, P. (2012). "The Digital Humanities as a Humanities Project". *Arts and Humanities in Higher Education*, 11(1-2): 42-60. Consulted on: 23 September 2013, ahh.sagepub.com/content/11/1-2/42
- Turing, Alan, M. (1936). "On Computable Numbers – with an Application to the Entscheidungsproblem". *Proceedings of the London Mathematical Society*, series 2, 42: 230-265 with corrections in 43 (1937): 544-546. Reprint in Davis (red.) (1965), pp. 115-154.
- Unsworth, J. (2002). "What is Humanities Computing and What is not?". *Jahrbuch für Computerphilologie*, 4, computerphilologie.uni-muenchen.de/jg02/unsworth.html. Verificeret 16. oktober 2013.
- Wouters, P., Beaulieu, A., Scharnhorst A. & Wyatt, S. (2013). *Virtual Knowledge. Experimenting in the Humanities and the Social Sciences*. Cambridge, MA: The MIT Press.