

Edb og historie:

Frå tilfeldig verktøybruk til eigen disiplin?

Arne Solli, vitskapsteoretisk innlegg til dr. art -grad.

Innleiing

Datamaskinar og databehandling har blitt daglegdags innafor historiefaget, både for profesjonshistorikarane og amatørhistorikarane. Bruken femnar vidt; frå tekstbehandling til statistisk databehandling, biletbehandling, tekstmerking, utvikling av pedagogisk programvare og programutvikling for kjelderegistrering og analyse. Men særleg har databaseprogram for å skrive av kjelder fått stor utbreiing mellom historikarar. Brukargruppa er særst heterogen m.o.t. kvalifikasjonar både i historie og informasjonsteknologi. I løpet av 1990-talet har den personlege datamaskina endra kvardagen til dei fleste innafor historiefaget, sjølv professorane må ”føre inn” kladden sin sjølv no. Gleda av å kunne beherske avansert teknisk verktøy har vore stor, naturlegvis speidd i sinne og frustrasjon når ting går galt. Men det synest vere ei stor semje om at datamaskiner er eit nyttig verktøy.

På bakgrunn av denne fellesstemte gleda fell det naturleg å setje søkjelys på grenselandet mellom informasjonsteknologi og historie. Kva karakter har bruken av informasjonsteknologi i historiefaget? Og korleis kan grenselandet mellom informatikk og historie karakteriserast. Det kan skiljast mellom to hovudsyn som klarast har krystallisert seg i synet på bruk av standard programvare, såkalla ”office-program” eller kontorstøtteprogramvare. Lærebøker i historie og edb framhevar bruk av standard programvare som databaseprogram og rekneark som gode og eigna verktøy i historisk forskning. Men det har vore stilt kritiske spørsmål om i kva grad standard programvareverktøy har dei eigenskapane og funksjonane som trengst i samband med historisk forskning. Dette er verktøy utvikla for å løyse informasjonsbehov i næringslivet, på kontoret i små og store verksemder – i dag. Det er primært ikkje forskingsverktøy.

Meiningsskilnadane omkring programvareverktøy har skapt to frontar: Temaorientert databehandling mot kjeldeorientert databehandling i historie. Eg har valt å sjå på noko uklare frontlinene i eit nytt perspektiv – verktøyorientert mot fagdisiplinorientert. Og usemja stikk djupare enn berre synet på standard programvare. For det er òg usemje om dei underliggjande modellane i standard programvare set for strenge restriksjonar på historikaren i forskingsprosessen. Dessutan har det vore etterlyst større vilje til teoretisk refleksjon og fokus på utvikling av eigne metodar og verktøy spesielt for historiefaget. Eg har derfor ein følelse av at usemja om bruk av programvare for kontorbruk kan dermed ha flytt fokus frå problem av meir grunnleggjande metodisk og teoretisk karakter. Det ligg eit anna sett av store utfordringar i å studere, utvikle, dokumentere og bruke eigne metodar, framfor kritikklaut å bruke standard programverktøy og dermed implisitt, overføre metodar, teknikkar og modellar frå andre fag- og verksemdsområde. Enkelte hevdar at desse utfordringane er både omfattande og særleine slik at ein eigen fagdisiplin er naudsynleg. Har historie i høve til informasjonsteknologi særleine, viktige og store nok utfordringar til at det kan forsvarast å etablere ein eigen disiplin? Eller er dette grenselandet først og fremst prega av lite

problematisk verktøybruk, slik trongen for ein eigen fagdisiplin mellom historie og informatikk er langt mindre?

Bruken av informasjonsteknologi i historiefaget er gjennomgripande og derfor er avgrensingar naudsynt. Informasjonsteknologi vert nytta både i forskning, formidling, og undervisning. Eg vil utelate historie og informasjonsteknologi i samband med formidling, undervisning og historie som fortid, som blir teknologihistorie. Kva står då att? I all hovudsak vil eg drøfte forholdet mellom historie som forskning og informasjonsteknologi. Og med forskning vil eg i stor grad vere oppteken av prosessane i skjeringspunkta mellom kjelder, analyse og problem. Vi er på andre ord på grunnplanet for historie som forskning.

Bruk av edb i historieforskning hadde sin spede start på slutten av 1960-talet. Men databrukande historikarar organiserte seg først i 1986 og danna *Association of History and Computing* (AHC) i 1987.¹ Organiseringa av databrukande historikarar avfødde først årlege konferansar og med etterfølgjande publisering av utvalde konferanseinnlegg. Fem-seks år seinare kom dei første lærebøkene spesifikt for historie og databehandling. Konferanseinnlegg og lærebøker vil vere mine hovudkjelder for å drøfte syna på høvet mellom informasjonsteknologi og historisk forskning.

Bruk av standard programvare

Svært mange - historikarar og andre nyttar dagleg Internett til kommunikasjon og informasjonssøking på Internett. Då nyttar ein program som epost-klientar og Web-lesarar. Dette er heilt uproblematisk bruk av standard programvare. Det same gjeld tekstbehandlingsprogram for skrivning. Men mange historikarar brukar også rekneark som MS-Excel, databasar som MS-Access og statistikkprogram som SPSS i arbeidet med kjelder og analyse av kjeldematerialet. Det er ved denne typen kontorprogramvare, database- og reknearkprogram at usemja tek til. I verktøyorienteringa, i motsetnad til disiplinorienteringa, har ein vore oppteken av å vise at denne type standard programverktøy let seg godt sameine med historisk forskning.

Ei av dei første av fleire lærebøker spesifikt for historikarar om bruk av datamaskinar var *Computing for historians. An introductory Guide* frå 1993 av Evan Mawdsley og Thomas Munck.² Forfattarane legg stor vekt på å syne korleis standard programvare kan vere særst nyttige og allsidige verktøy for historikarar. Databaseprogram er historikaren sitt hovudverktøy. I databasesystemet kan fleire operasjonar gjennomførast; registrering av kjelder, koding og kategorisering. Etter at fleire ulike kjelder registert og koda, kan databaseprogrammet nyttast for å kople opplysningar om t.d. same biologiske person. Mawdsley og Munck er tydeleg på at standard verktøy, standard programvare, kan løyse dei fleste oppgåvene.³ Men poengterer at det er mange problem på vegen frå kjelde til ferdig database, og at det sjeldan er **eitt** svar på problema som oppstår. Men dei legg vekt på at

¹ Heimeside: <http://grid.let.rug.nl/ahc/>

² Mawdsley og Munck: 1993. Mawdsley og Munck tilhøyrer historikarmiljøet ved University of Glasgow (Department of Modern History). Mawdsley er primært oppteken med sovjetisk historie, Munck med sosialhistoriske tema frå Europa i tidleg moderne tid. Også Greenstein tilhøyrde dette miljøet tidleg på 1990-talet.

³ Mawdsley og Munck: 1993, s. 187. "This book has outlined some of the basic types of software that are most likely to be of relevance to historians."

mange operasjonar som før 1990 kravde særskild programmering, no kan løysast heilt utan programmering. Dette heng saman både med auka funksjonalitet i standard programvare og at standard programvare er særskild brukarvennleg.

Dei er positive til databaseprogram bygt på relasjonsmodellen, og meiner at bruk av denne databaseteknologien kan gje resultat som ikkje hadde vore mogeleg å nå utan bruk av databasar.⁴ Korleis underbyggjer dei denne påstanden? For det første er det eit resultat av sjølv registreringsprosessen. Datamaskinen i seg sjølv krev nøyaktigheit, struktur og systematikk ("accuracy and rigorous logic"). Historikarar som brukar datamaskin må derfor i større grad ta stilling til eigenskapar ved kjeldematerialet; om opphavsforhold, om varierende presisjonsnivå mellom kjeldedata og om kjelde sitt egentlege føremål. Forskaren blir dermed meir på medviten om styrkjer og veikskap i kjeldematerialet.⁵

Mawdsley og Munck er positive til at bruk av datamaskiner strukturerer historikaren i forskingsprosessen. Eller sagt på ein annan måte: Det at dataverktøy er "firkanta" er positivt på ein "uryddig" forskar. Karakteristisk for verktøyorienteringa er at ein stort sett nyttar den programvaren som kommersielt sett er i vinden. Tidleg på 1990-talet nytta mange derfor Wordperfect, Lotus-1-2-3 og Paradox. På slutten av 1990-talet har Microsoft-produkta Word, Excel og Access dominert marknaden, historikarane og lærebøkene. Dei kommersielle salssuksessane har såleis styrt historikarane sine val av verktøy. Med dette meiner eg at verktøybruken har vore tilfeldig.

Eg har late Mawdsley og Munck representere verktøyorienteringa, men liknande synspunkt finn ein t.d. hos Greenstein i *A historians' guide to computing* (1995) og Gunnar Thorvaldsen i *Databehandling for historikere* (1999).⁶ Harvey og Press i *Databases in Historical Research* (1996) prøver å bygge bru mellom kjeldeorientert og temaorientert databruk, eller mellom verktøyorienteringa og disiplinorienteringa som eg har kalla det.

Fokuset på standard, kommersielle kontorverktøy og på sjølv bruken av verktøya kan ha samanheng med fleire forhold. For det første er historikarar ofte orienterte mot tema, problemstillingar og historiske resultat av forskinga. Metoden og teknikken er underordna og det gjeld òg bruken av informasjonsteknologi. Dermed er datahistorikaren oppteken av å bruke verktøy, ikkje finne fleire metodeproblem. Konsekvensen av dette har dels ført til mangel på dokumentasjon. George Welling påviser fleire arbeid der det i forordet blir klart uttrykt at forskingsarbeidet føresette bruk av datamaskiner, men samstundes som at metodane ikkje er drøfta og heller er ikkje databruken dokumentert.⁷ Mangel på drøfting av metode og dokumentasjon fører dermed til at hjulet blir funne opp mange gonger både av ulike historikarar og av same historikar. Her må eg innrømme at kjenner eg meg godt att.

Men bruk av standard programvare har ein klar fordel i opplæringsamanheng, der ein skal lære nye historiestudentar bruk av dataverktøy. Denne er rimeleg i innkjøp, og studentane kjenner brukargrensesnittet frå før. Dette reduserer kostnadane både ved innkjøp og opplæring. Dessutan stødde det skulesstyresmaktene i fleire land på 1990-talet vektlegging av

⁴ Mawdsley og Munck: 1993, s. 165.

⁵ Mawdsley og Munck: 1993, s. 168 og s. 186-188.

⁶ Mawdsley og Munck: 1993, Greenstein: 1994, Harvey og Press: 1996 og Thorvaldsen: 1999.

⁷ Welling: 1998, s. 11

overførbare ferdigheiter og generell dataopplæring innafor fleire fag.⁸ Grunnen til dette var at på 1990-talet var det naudsynt å leggje basis opplæring i bruk av dataverktøy i undervisning på høgskule og universitetsnivå fordi studentane ikkje hadde fått slik opplæring i grunnskulen på 1980-talet. Koenrad de Smedt meiner at slike kurs i databruk under merkelappen humanistisk eller historisk informatikk kan øydeleggje truverde til humanistisk informatikk som fagdisiplin.⁹

Men er verktøybruken så uproblematisk som lærebøker og verktøyorienteringa kan gje inntrykk av?

Frå verktøybruk til fagdisiplin

Ovanfor såg vi korleis databrukande historikarar framheva bruken av standard dataverktøy utan særskild programmering. Utvikling av grafiske brukargrensesnitt frå kring 1990 gjorde dei generelle dataverktøy langt meir brukarvennlege enn tidlegare. Funksjonar som tidlegare kravde spesialkunnskap i programmering blei i løpet av 1990-talet tilgjengeleg på maskinar som fekk plass på eit lite skrivebord.

Men kan dei fleste behova historikarane har i si forskning løysast med standard programvare? I kva grad har historisk forskning eigenskapar som krev særskilde og eigne modeller, eigne teknikkar, eigne omgrep og eigne verktøy? Under den stort sett problemfri overflata har vi vore innoom ei par ikkje heilt uproblematisk teknikk: Koding og kategorisering av kjelldata og såkalla "record linkage" (kjeldelenking). Standard programvare er spesielt utvikla for næringslivet i notida og dei behova offentleg sektor og større og små verksemdar har. I kva grad skil historie seg både frå samfunnsvitskapar og private verksemdar ("notida") med omsyn til bruk av informasjonsteknologi? Dette er spørsmål som utfordrar verktøyorienteringa.

Manfred Thaller er den som med størst styrke har påpeika trongen for ein eigen disiplin, eit eige fagfelt.¹⁰ Thaller har kalla faget for "Historische fachinformatikk", historisk informatikk på norsk. Innafor dette fagfeltet mellom historie og informasjonsteknologi skal ein arbeide med utfordringane som overstig verktøyorienteringa. Men er det verkeleg trong for ein eigen disiplin?

Kvifor eigen disiplin?

Manfred Thaller har sidan 1970-talet hatt sitt virke hovudsakleg i Tyskland, men har vore heilt sentral i oppbygginga av fagmiljø i "Historisk informatikk" i store delar av Europa, og særleg i Aust-Europa. Thaller var òg leiar på HIT-senteret her i Bergen på slutten av 1990-talet.

I følgje Thaller treng vi ein eigen fagdisiplin av fleire grunnar. For det første er eit eige teoretisk grunnlag for historie og informasjonsteknologi naudsynt for å føre ein akademisk diskusjon. For det andre har historie eigenskapar som skiljar det frå andre fagdisiplinar intellektuelt. For det tredje er ein klart definert fagdisiplin viktig i konkurransen om

⁸ Sjå t.d. *IT i norsk utdanning 1996-99*, <http://odin.dep.no/ufd/norsk/satsingsomraade/ikt/014005-991110/index-dok000-b-n-a.html>

⁹ Smedt: 2002

¹⁰ Også Tito Orlandi, Ingo Kropac, Lou Burnard og Jan Oldervoll har gitt uttrykk for synspunkt som fell saman med Thallers.

budsjettmidlar. Det er naudsynt for å markere seg og få status i høve til løyvande styresmakter og byråkratiet.¹¹ Dette er Thaller's tre påstandar. Eg vil drøfte dei to første. Behovet for teori som kan strukturere ein akademisk diskusjon og b) Historie skiljar seg frå andre disiplinær som har konsekvensar for bruk av datateknologi.

Kvifor treng vi teori for å strukturere den akademiske debatten? Dei første internasjonale konferansane for databrukande historikarar på slutten av 1980-talet var i følgje Thaller prega av stor velvilje og forståing og få kritiske røyster. Deltakarane og prosjekta deira fekk ikkje den naudsynte faglege kritikken fagleg utvikling er ein føresetnad for. Gleda over endeleg å møte nokon med liknande interesser og utan dårleg skjult skepsis var så stor at konferansedeltakarane la att den kritiske evna heime. Thaller nemner tre eksempel på emne på slutten av 1980-talet som burde ha vore debattert, men som i liten grad førte til diskusjon:¹²

- Den underliggjande føresetnaden om at berre kvantitative metodar og studiar er den einaste måten å nytte datateknologi på i historiefaget.
- Påstanden om at system for kunstig intelligens kan nyttast uendra på historiske data.
- Den unisone semja om at den relasjonelle datamodellen og relasjonsdatabaser er best eigna for utvikling av databasar for historiske data.

For det første etterlyser Thaller eit rammeverk for å føre diskusjonar omkring slike og tilsvarende tema. Påstandane er vanskeleg å diskutere utan eit rammeverk i form av metodar og omgrep. Thaller avviser ikkje påstandane som usanne, men etterlyser eit felles, kvalifisert grunnlag for å diskutere dei. Eg vil gå inn i diskusjonen av desse problema, men berre kort konkretisere dei.

Thaller meiner at historikarar som nyttar kvantitative metodar i si forskning har adoptert samfunnsvitskaplege metodar utan å drøfte skilnadane mellom historie og samfunnsvitskap. Han meiner ei drøfting av konsekvensane av ei slik metodeoverføring kan og bør skje innafør ramma av historie og informatikk, altså historisk informatikk.

Når Thaller syner til Kunstig Intelligens (AI), så var dette 'hype' på 1980-talet. KI var då særleg i vinden innafør diagnostikk og styring av prosessar (t.d. gass- og straumforsyning). Verksemdar med aktivitetar og fagområde som er milevidt frå historie som fag. Dataverda er stadig full av tilsvarende nye og 'hype' trendar. 'Hype' trendar i dataverda og grenselandet mellom informasjonsteknologi og humaniora (undervisning) har i følgje Koenraad de Smedt ein lei tendens til å fokusere på tekniske og administrative aspekt og i mindre grad på nye innovative undervisningsmetodar.¹³ Ankepunktet til både Thaller og de Smedt gjeld med andre ord at historikarar og humanistar lett blir blenda av teknologi. Men adaptasjon av teknologi og metodar treng kritisk drøfting, særleg når desse er utvikla innafør andre fagfelt.

Relasjonelle databasesystem (RDBMS) var også 'hype' i dataverden på 1980-talet og er framleis dominerande som databasesystem. Relasjonsdatabaser byggjer på relasjonsmodellen. I relasjonsmodellen vert data representert med tabellar og der samband mellom tabellar blir representert ved eksplisitte tabelldata. Ein tabell tilsvarende ei mengde i

¹¹ Thaller: 1989. Sjå også Thaller: 1991 og Thaller: 1994.

¹² Thaller: 1989, s. 2.

¹³ Smedt: 2002. Ein liten sidemerknad: "Bransjen" er klar over denne kritikken, og det kan vere ein av grunnane til at studiesystemet ClassFrontier har vore så ivrig etter å bli integrert mot "vårt" KARK. KARK vil gje ClassFrontier eit etterlengta pedagogisk innhald.

matematikkens mengdelære.¹⁴ I ein relasjonsdatabase for ei folketeljing vil ein dermed kunne ha ein tabell for personar, ein annan tabell for hushald, ein tredje tabell for skip, ein for institusjonar, hus eller gard. Relasjonsbasen krev dermed at ei kjelde blir broten opp i tabellar av ulik storleik og at sambandet mellom desse blir representert ved såkalla primærnøklar. Men kvifor er det viktig å kjenne den underliggjande modellen? Alle dataverktøy – både programmeringsspråk og sluttbrukarverktøy som databaseprogram – har ein underliggjande modell. Modellar som er basis for dei datastrukturane vi får tilgang til i verktøyet. Thallers poeng er at den underliggjande, implisitte datamodellen ikkje må avgrense historikaren. Han meiner at den underliggjande datamodellen strukturerer implisitt måten forskaren ”ser” kjeldematerialet og fortida på. Og særleg har relasjonsmodellen eit sett av strenge restriksjonar som ikkje høver for historiske kjelder og data. Historiske data har ikkje alltid ein unik id, primærnøkkel, som personnummer eller kontonummer. Sjølv nokre typar historiske kjelder ser ut som tabellar i utgangspunktet er langt frå alle det. Ved val av dataverktøy må historikarane kjenne den eller dei underliggjande datamodellen og vurdere denne i høve til kjelda og det verktøyet skal brukast til. Utan ei kritisk vurdering av modellane er det ein fare for at den underliggjande modellen i verktøyet formar fortida og historikaren og ikkje omvendt.

Thallers første hovudpoeng er altså at ein manglar ein infrastruktur for å drøfte denne typen problemstillingar. Drøfting av problemstillingane krev eit felles metodisk og teoretisk fundament. Dette fundamentet manglar, og etablering av eit eige fagfelt kan gje rom for ei kritisk drøfting av utfordringar i grenselandet mellom informatikk og historie. Diskusjonen som eg i utgangspunktet presenterte som ein diskusjon kring standard programvare ser vi gjeld også overføring og bruk av metodar og modellar frå andre fagområde.

Kva er særskild med historie?

Thallers andre ankepunkt dreier seg at teknikkar, metodar og verktøy frå andre fag og samfunnsområde ikkje direkte kan overførast til historie kritikklaust. Historie har ein rekkje særeigne kvalitetar på ulike nivå. Særeigne eigenskapar som gjer at historie skil seg frå både samfunnsvitskap og private verksemdar bruken av informasjonsteknologi. Kva for eigenskapar er særskild for historie og i kombinasjonen av historie og informatikk?

Tid er viktig for historikaren, og eg vil syne med eit lite enkelt eksempel kor problematisk noko så basalt som mål for tid kan vere i samband med datamaskiner. I dr. avhandlinga ville eg finne ut om konfirmasjonsalder vart endra i løpet av 1800-talet. Eit enkelt spørsmål var: Vart konfirmantane eldre eller yngre? All litteratur tyda på at konfirmasjonsalder ikkje endra seg i perioden. Var dette rett? På basis av lister over konfirmantar der konfirmasjonsdag og fødselsdag/dåpsdato var kjent ville eg rekne ut konfirmanten sin alder. Dette er enkelt, differansen mellom dei to datoane gjev alderen (dvs. litt i underkant dersom berre dåpsdato er gitt, men det er uvesentleg). Reknearkprogrammet MS-Excel har ein funksjon som heiter `datodiff(frådato;tildato;resultatformat)`, slik at eg valde å nytte dette dataverktøyet. Reknearkfunksjonen `datodiff` returnerer `datodifferansen` i t.d. dagar. Umiddelbart fekk eg eit problem! MS-Excel ville ikkje godta datoar eldre enn 1-1-1900. Dinest vart datoar som ikkje var registrert med fire siffer for år automatisk endra til ein

¹⁴ Date: 1986, s. 99.

dato på 1900-talet eller på 2000-talet, dvs. datoen i kyrkjeboka frå 1880-talet var 18-02-86 (t.d. 18-02-1886) tolka reknearket som 18-02-1986, jf. figur 1.

Figur 1 Standard verktøy (MS-Excel) og tolking av datoar

Standard verktøy og datoar i historiske kjelder

- 10-Nov-1901 - OK
- 31-10-1899 – ”feil”, tolkar ikkje datoar før 1-1-1900
- 18-feb-86 – automatisk tolka som 18-02-1986
- 13-05-12 – automatisk tolka som 13-05-2012
- 31-juni-1834 – feil, berre 30 dagar i juni
- 29-02-1833 – feil, 1833 ikkje skotår
- __-04-1913 – ”feil”, dag manglar
- 18-__-1845 – ”feil”, månad manglar
- ”Botolphi 1561” – blir ikkje tolka, fast dato
- ”7de søndag eft. trin. 1757” – blir ikkje tolka, rørleg
- Tidsrom: ”7. søndag etter trinitatis 1698” til ”2. søndag i advent 1713”
(Ekstra problem: Frå juliansk til gregoriansk kalender)

Dette er ein standard tolkefunksjon i MS-Excel som gjer at tosifra år mindre enn 29 blir tolka som på 2000-talet, tosifra år større enn 29 blir tolka som eit år på 1900-talet. Denne funksjonen måtte eg slå av, elles fekk eg øydelagt mine data. Men alvorlegast var at datoar eldre enn 1-1-1900 ikkje blei tolka som ein dato, men som ein vanleg tekststreng, og programmet ville derfor ikkje rekne datodifferansen fordi datoar eldre enn 1-1-1900 som pr. definisjon var ugyldige. Men programmeringsspråket Visual Basic har ein funksjon som reknar ut datodifferansar. Dermed kunne eg skrive eit lite Visual Basic program som eg kalla frå reknearket. Dette vesle programmet inneheld kall til Visual Basic sin datofunksjon. Denne datofunksjonen returnerte datodifferansar for datoar både på 1700-talet og 1800-talet. Flott! Men så oppdaga eg at nokre datoar ville ikkje funksjonen akseptere, t.d. 31-juni-1834 eller 29 februar 1833. Dette var datoar oppgitt i kjelda, men det er ikkje gyldige datoar. Juni har berre 30 dagar, 1833 var ikkje skotår. Men desse stod i kjelda, eg kunne ikkje endre dei. Dinest var ein del datoar ufullstendige, t.d. i kjelda mangla dag i månaden, eller månad, t.d. ”__-04-1833” eller ”18-__-1845”. Desse problema løyste eg ved å setje inn medianen som verdi, t.d. 15 for dag i månaden og månad 06 for månad. Eit lite spesialprogram gjorde nytten.

Eksempla syner at heilt elementære, grunnleggjande problem ein historikar støyter på i kjelder krev spesialhandsaming i vanleg programvare. Implisitt i standard programvare ligg ei rekkje automatiske tolkingar av tid som skapar nye problem framfor å løyse dei. Datoen

'24.12' vil automatisk bli til 24.12.2003, om ikkje denne funksjonen blir slått av. Følgjande åtvaring og hjartesukk fann eg nettopp i debattforumet til Digitalarkivet:

Ellers påpeker Kari Thingvold noe viktig: Microsoft har desverre til de grader overlesset Excel, og forsåvidt også Word, med autofunksjoner at de nesten kan være farlige å bruke i dataregistreringssammenheng. En mengde autofunksjonsbrytere må stå i OFF for å være sikker, og til og med da kan du få problemer.¹⁵

Men om vi går attende til kyrkjebøkene så er det sjeldan at prestane brukte datoar før 1815. Ei kyrkjeleg handling, t.d. dåpen eller vigsel, vart då tidfest i høve til kyrkjeåret, t.d. ”7de søndag efter trinitatis. Døbte Anders Olsens Grevstads barn Marthe.” Vanlegvis vil historikaren (demografen) vere interessert i datoen slik at hendinga kan relaterast til kalenderdato, særleg sidan dei fleste tidsangivelsar i kyrkjeåret er rørlige. Dato for 7de søndag trinitatis er avhengig av plassering av påske i det aktuelle året og den er igjen avhengig av astronomiske forhold. For historikarar som arbeider med denne typen kjeldematerialet ville det derfor vore heilt sjølvstakt at datamaskinen kunne tolke tidsangivelsar som ”7de søndag eft. trin. 1757” og konvertere den automatisk til kalenderdato i det aktuelle kalendersystemet (Gregorianske kalender), og samstundes også rekne tidsdifferansar mellom slike tidsangivelsar, t.d. skilnaden i tid mellom ”7. søndag etter trinitatis 1698” og ”2. søndag i advent 1713”. Dette finst det naturleg nok særskilde dataprogram for, som konverterer datoar frå eit kalendersystem til eit anna, men ønskjemålet ville naturlegvis vere at dette var ein automatisk funksjon slik at konverteringa skjedde utan at ein manuelt måtte skrive inn datoen på nytt i eit separat dataprogram. Men slike generelle rutinar er ikkje ein del av standard programvare.

Problemet over relaterer seg til opplysningar om tid i historiske kjelder, om historiske kalendersystem og korleis representere desse elektronisk. Som eksempel på heilt grunnleggjande problem i arbeidet med å leggje til rette ei kjelde for analyse eller utgjeving vil eg peike på seks ulike problemområde:

1. Særskilde teikn i historiske tekstar som ikkje lenger er i bruk i ”levande” skriftspråk. Fagmiljøet på nordisk i Bergen arbeider med å få t.d. O med kvist inn i standardar for teiknsett (Unicode).
2. Måleeiningar og verdssystem, t.d. volum, avstand og pengeverdiar (1 Rd 4 ort 9 sk.). Transformasjonsreglar må spesifiserast slik at enkle logiske og matematiske operasjonar kan utførast på verdieiningane.
3. I ein historisk kontekst må identifisering av biologiske og fysiske objekt må skje ved hjelp av namn, personnamn og geografiske namn. Same fysiske objekt kan vere benemnt forskjellig, Er ”Marthe” lik ”Marta”? Krev utvikling av metodar for namnstandardisering.
4. Geografiske namn og fysisk utstrekning. Er Stranda prestegjeld = Stranda tinglag = Stranda kommune. Problemet er at det geografiske namnet representerer ulike fysiske realitetar. Til dømes har NSD (Norsk samfunnsvitenskapleg Datatjeneste

¹⁵ Debatt i Digitalarkivet om ”Kilderegistering, standarder, metoder, problemer”, <http://digitalarkivet.no/cgi-win/WebDebatt.exe?slag=listinnlegg&debatt=brukar&temanr=17998&sok=standard&startnr=&antall=&spraak=>. Kari Thingvold hadde fått '1/2' omgjort til 1.februar 2003.

definert eit tabellverk med geografiske koordinat for dei norske kommunane etter 1838.

5. Yrke og yrkestitlar. Gamle yrke som forsvinn og nye kjem til. Mange problem særleg knytt til plassering i sosiale kategoriar på grunnlag av yrkesnemningar.
6. Og som eg har gitt eksempel på: Opplysning om tid og kalendersystem.

Problemet for historikaren i høve til samfunnsvitaren og næringslivet er at våre målesystem ikkje følgjer desimalsystemet eller andre notidige standardar for t.d. tidsmåling. Beskrivelsen av historiske objekt inneheld historiske, utdaterte målesystem, lokale verdissystem og varierende presisjonsnivå. Ofte er det vanskeleg å avgjere om ulike måtar å beskrive eigenskapar ved eit historisk objekt er ein skrivefeil eller om variasjonen er meiningsberande. Identifisering av fysiske og biologiske objekt er basert på varierende og omtrentleg grunnlag, ikkje unike personnr., kundenr., kommunenr. o.l. Dette fører til at heilt elementære logiske operasjonar som er A lik, mindre enn eller større enn B ikkje nødvendigvis har eit svar. Heilt elementære matematiske operasjonar som addisjon krev transformasjonsreglar som er avhengig av den historiske konteksten, tid og rom. Dette kompliserer heilt enkle basale matematiske og logiske operasjonar på eigenskapane ved objekta. Standard programvare baserer seg på at alle målesystem følgjer desimalsystemet og at kalendertid er uttrykt korrekt etter den gregorianske kalender (I bruk i Norge frå 1700). Få samfunnsvitskaplege analysar treng å ta omsyn til at pengeeiningane var annleis før 1876. Og eg trur neppe at Microsoft eller eit av dei andre store programvarehusa finn det kommersielt forsvarleg å inkludere støtte for historiske målesystem i sine dataverktøy. Vi kan dermed trygt slå fast at historie har ei rekkje særeigne problem knytt til representasjon og formalisering av relativt elementære objekt og heilt enkle eigenskapar ved disse objekta som storleik, verdi, utstrekning og tidsbestemming.

I denne gjennomgangen har eg synt nokre problem med å representere historiske data i standard verktøy. Vi kan kalle dette representasjonsproblemet. Metodisk skapar dette nye problem: Søking, gjenfinning, samanlikning av data krev særskilde metodar, t.d at likskap blir til 'nesten lik' og andre grader av likskap, jf. punkt 3 over. Og i høve til kvantitative metodar og i høve til matematiske operasjonar krev våre data ei rekkje transformasjonar slik at dei kan behandlast med desimalsystemet, jf. punkt 2 over. Eit anna sett av problem gjeld modellnivået, det å kunne modellere historiske objekt. Eg kan ikkje gå inn på desse her men berre nemne fire problemområde.

1. Strukturproblemet. Oldervoll har peika på at mange historiske kjelder kan ha ein kompleks struktur som er vanskeleg å representere med tabellar i relasjonsdatabasar.
2. Ei anna problem har eg kalla sekvensens semantikk. Relasjonsmodellen byggjer på mengdelæra, og element i ei matematisk mengde har ikkje ei bestemt rekkjefølgje. Tolking av mange liste-type kjelder byggjer på semantiske referansar der rekkjefølgja er avgjerande for tolkinga.
3. Problemet med doble, kryssande hierarki i hierarkiske datamodellar. I strenge hierarkiske datamodellar er kryssande hierarki ikkje lov. Problemet oppstår når vi i ein hierarkiske datamodell vil modellere både kjelda og dei saksforholda kjelda

fortel om. Ein entitet kan ikkje vere medlem av to hierarki samtidig. Dette er sær problematisk dersom ein vil lage ein modell som både tek omsyn til kjeldeteksten sin struktur, samtidig som ein modellerer det semantiske innhaldet i kjeldeteksten. Eller sagt på ein annan måte: Dersom ein vil la ein person både vere underelement av ei kjeldeside og samstundes vere medlem i eit hushald eller eit gardsbruk.

4. Modellering av tid, tidsperiodar og tidsseriar i datamodellar generelt. I generelle datamodellar er tid implisitt lik "notid" og tid derfor ikkje med i definisjonen av det ein kallar eit atomisk datum. Eit utviding av t.d. relasjonsmodellen der tid er med inkludert som eit elementært datum er ei skikkeleg intellektuell utfordring.

Vi kan dermed trygt slå fast at historie har særskilde og eigne problem i høve til informatikk og informasjonsteknologi på representasjonsnivået, metode eller algoritmenivået og på modellnivået.

Pragmatikarar og puristar

Eg vil no vende attende til disiplinorienteringa og verktøyorienteringa – eller puristar og pragmatikarar som også er nytta som merkelappar. Det eg har kalla verktøyorienteringa og disiplinorienteringa står ikkje nødvendigvis i skarp motsetnad til kvarandre. Eg trur det er best å karakterisere skilnadane som ulik vektlegging og dels skuldast usemja ein skilnad mellom ein sentral-europeisk og ein anglo-amerikansk forskingstradisjon. I tabell 1 har eg prøvd å setje opp hovudskilnadane i vektlegging og fokus mellom "pragmatikarar" og puristar.

Tabell 1 Verktøyorienteringa og disiplintilnærminga. Vektlegging og fokus.

Verktøyorienteringa – "Pragmatikarar" Temaorientert	Disiplintilnærminga "Puristar", Kjeldeorientert
Standardisering	Standardisering
Resultatorientert	Prosess- og problemorientert
Bruk av metodar (kvantitative)	Studie av metodar og teknikkar
Struktur eit "gode"	Struktur (modellar) eit "onde"
Orientert mot tema, fortida="verda"	Kjeldeorientert
Kjelder er til for (eigen) analyse	Også: publikasjon av kjelder
Bruk av standard programvare	Programvareutvikling
	Eigen fagdisiplin ein føresetnad

Det synest vere ei felles oppfatning av at arbeid med standardisering er nyttig og viktig.¹⁶ Eit eksempel på standardisering er HISCO. The North Atlantic Population Project arbeider med å lage ein kodenstandard for yrke i folketeljingar i Europa og Nord-Amerika. Ein slik sams standard vil gjere komparasjon av sosiale struktur og næringsstruktur over tid og over landegrensar mogeleg.¹⁷ Men nasjonale tiltak er meir vanleg. I Noreg er HISTFORM utarbeidd som standard for registrering av folketeljingar. Standard 4G er utforma for å registrere kyrkjebøker.¹⁸

¹⁶ Greenstein: 1991.

¹⁷ Atlantic Population Project, sjå <http://www.nappdata.org/>

¹⁸ Nygaard og Registreringssentral for historiske data: 1995, Thorvaldsen: 1996 og Samarbeidsrådet for historielagene i Oppland, *Standard 4G*, <http://home.online.no/~ddarkiv/>

Verktøyorienteringa synest eg er klarare orientert mot å få fram nye og originale forskingsresultat.¹⁹ Måten dette er gjort på spelar mindre rolle. Disiplintilnærminga er klarare på at prosessen og problema ein støyte på undervegs må dokumenterast. Resultat av prosessen bør innehalde beskrivelse av metodar ein nytta. Resultatet av forskinga ligg dermed klarare på generell metodeutvikling spesifikk for historie. Verktøyorientering brukar metodar for å løyse eit historisk problem, med særleg vekt på kvantitative metodar. I disiplintilnærminga er ein meir oppteken av å studere, utvikle og dokumentere metodar. Det er og eit høgare ambisjonsnivå i høve til kva datamaskinen kan nyttast til – ikkje berre statistikk.

Verktøyorientering synest primært å vere meir orientert mot historiske tema og den ”verda” kjeldene beskriv. Dette er blitt kalla ”issue-oriented”, ”object-oriented” og ”topic-oriented”. Det eg har kalla disiplintilnærminga blir og kalla ”source oriented computing”. Disiplinorienteringa er derfor kritisert for å primært vere oppteken av kjeldene. Kritikarar av den kjeldeorienterte retninga meiner at historieforsking sjeldan tek til med ei kjelde, men med eit historisk problem. I verktøyorientering har kjeldene ei klar sekundær rolle i høve til sak, tema, hypotesar, debattar og problemstilling. I disiplintilnærminga er ein klarare på at kjeldene skal publiserast som t.d. digitale diplomatariske utgåver. Same kjelda skal kunne nyttast til ulike problemstillingar og ulike forskningstema.

Dels kan ein og sjå på skilnadane som nivåskilnader i utdanning av ein historikar. Verktøyorientering høver på grunnivå der ein skal lære enkel bruk av datamaskinar i historiefaget, medan disiplintilnærming høver betre på mastergradsnivå og i eiga forskning.

Den tydelegaste skilnaden mellom frontane har vore synspunkta på bruk av standard programvare. Desse skilnadane er dermed grunnlaget for debatten om bruk av standardprogramvare eller utvikling av eigna programvare. Mangelen på eigna programvare har og ført til utvikling av eige programsystem for historikarar. Thaller har med prosjektet ”Historical Work Station” og programsystemet Kleio hatt som målsetjing å utvikle eigna programsystem for historikarar. Jan Oldervoll har utvikla programsystem som CensSys for registrering og analyse av kjelder, og WebCens for søking, analyse og presentasjon av historiske kjelder på Internett. WebCens er databasemotoren i Digitalarkivet. Men som eg prøvd å vise med denne figuren stikk skilnadane djupare, samstundes som forholdet mellom verktøyorientering og disiplinorienteringa er klart symbiotisk.

Med utgangspunktet i desse to orienteringane vil eg no prøve å komme nærare kjernen i grenselandet mellom historie og informatikk. Kva er egentleg historisk informatikk?

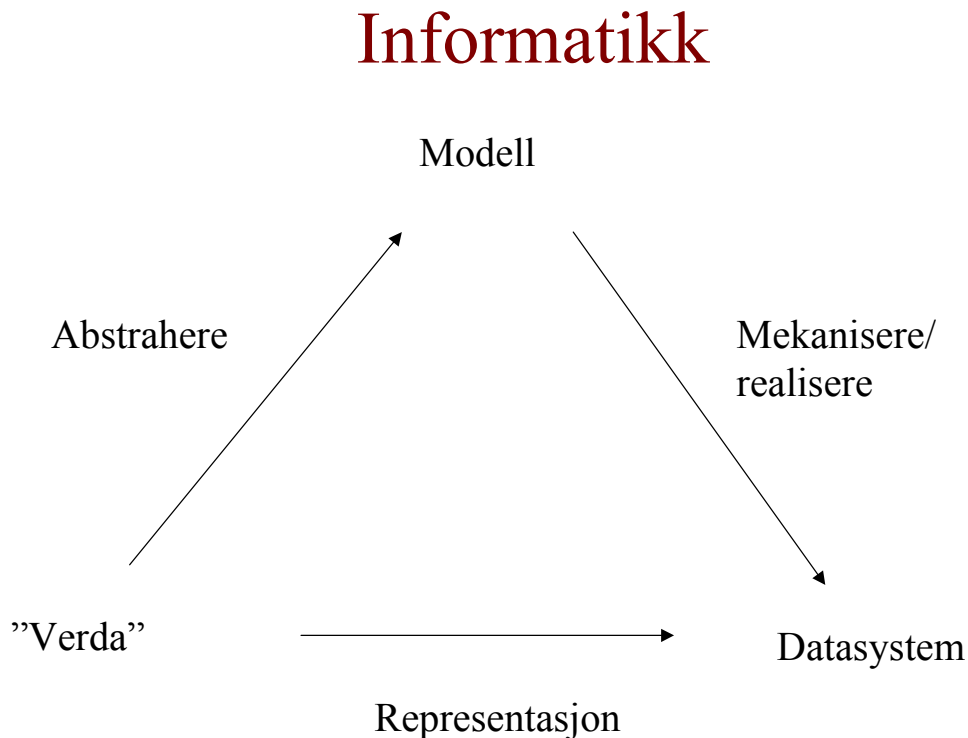
Historie og informatikk

Historie har eg tidlegare avgrensa til historisk forskning. Men kva er informatikk? Informatikk som fag har ikkje **ein** fagleg eller teoretisk basis. Informatikk og utvikling av informasjonsteknologi baserer seg på mange fagområde; matematikk, logikk, lingvistikk og psykologi for å nemnde nokre. Slik sett er informatikk eit sær samansett fagfelt, og derfor finst det ulike definisjonar av kva informatikk (’computer science’) er. Ein utbreidd definisjon av faget informatikk seier at informatikk er ein vitenskap om abstraksjonar og mekanisering av abstraksjonar, jf. figur 2. Informatikk er dermed å utvikle ein modell for eit problem

¹⁹ Welling: 1998, s. 11, Harvey og Press: 1996, s. 6-7.

(abstrahere) og deretter finne og nytte eigna mekaniske teknikkar for å løyse problemet (realisere).²⁰

Figur 2 Informatikk: Abstrahering og realisering



Abstrahering innebær ein forenkling, der vi erstattar den komplekse og detaljerte røyndomen, med ein modell som gjer oss i stand til å forstå den og løyse det aktuelle problemet. Modellomgrepet i informatikk har klart mest likskap med det Kjellstadli kallar formelle modellar.²¹ Skjematisk kan vi tenkje oss abstrahering og realisering som i figur 2

”Verda” eller problemet for ei privat eller offentleg verksemd som skal løysast vere høgst ulikt. Problemet kan vere eit institutt som vil ha eit informasjonssystem for halde orden på studentane. Det tenkte studentsystem skal svare på enkle spørsmål som: Kor mange studentar tek modul HIS102? Eller løyse enkle oppgåver som: Send e-post til alle studentane på modul HIS101 som også tek modul HIS114. Studentsystemet må derfor innehalde namn, adresseinformasjon, kva fag/emne studenten tek og spørjerutinar for å løyse desse spørsmåla.

Første steg frå problem til datasestem er abstrahering, dvs. å trekkje ut frå ”verda” det som vi treng for å løyse problemet. Det viktigaste i denne prosessen er avgrensing, utveljing og forenkling. Vi definerer dei objekta og dei eigenskapane som datasestemet må innehalde for å løyse oppgåvene vi vil ha det til å gjere. I studentsystemet vil student og modul vere typiske objekt. Namn og adresse er typiske eigenskapar vi vil ta med. Vi tek ikkje med studenten sin hårfarge om det ikkje er ein eigenskap vi vil ha nytte av seinare.

²⁰ Aho og Ullman: 1995, s. 1-

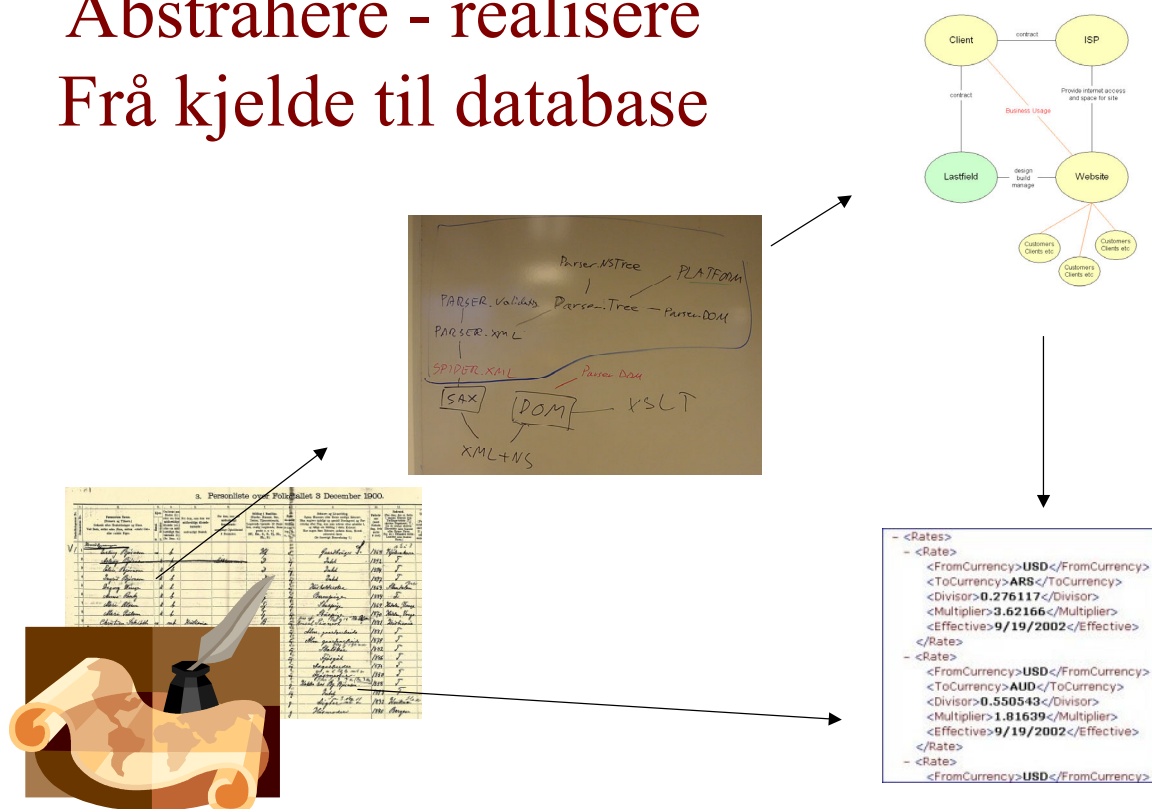
²¹ Kjellstadli: 1992, s. 136.

Deretter må vi beskrive modellen. I første omgang kan vi bruke eit grafisk språk, bokstar og piler, som berre andre innvidde menneske forstår. Bokstar og piler representerer objekt, samband mellom objekt og type samband mellom objekt, dvs. mening. Men til slutt må vi nytte eit språk som også datamaskina forstår. Når vi beskriv modellen og løysinga av problemet må vi derfor bruke eit språk som både menneske og maskin ”forstår”, eit programmeringsspråk. Programmeringsspråk er ein måte å beskrive objekt eller entitetar som studentar og modular vert kalla på modellplan. Og med programmeringsspråk kan ein også beskrive ein algoritme for å få datamaskinen til å telje opp talet på studentar i ein modul. Datamaskinen inneheld dermed to representasjonar, for det første ein representasjon av modellen vi har laga av ”verda”. Dette kallar vi dataprogram. Dinest ein representasjon av eigenskapar ved objekta i ”verda” som er sjølve data.

Sidan historie handlar om menneske i fortida og deira etterlatenskapar kan ”Verda” i figur 2 både vere historisk kjelder i seg sjølv og personar og ting i den fortidige verda som kjeldene beskriv eller er levningar av. Figur 3 syner prosessen med å lage ein datamodell for ei kjelde, dinest realisere denne og til slutt skrive av kjelda sitt tekstlege innhald.

Figur 3 Abstrahere – realisere, frå kjelde til database.

Abstrahere - realisere Frå kjelde til database

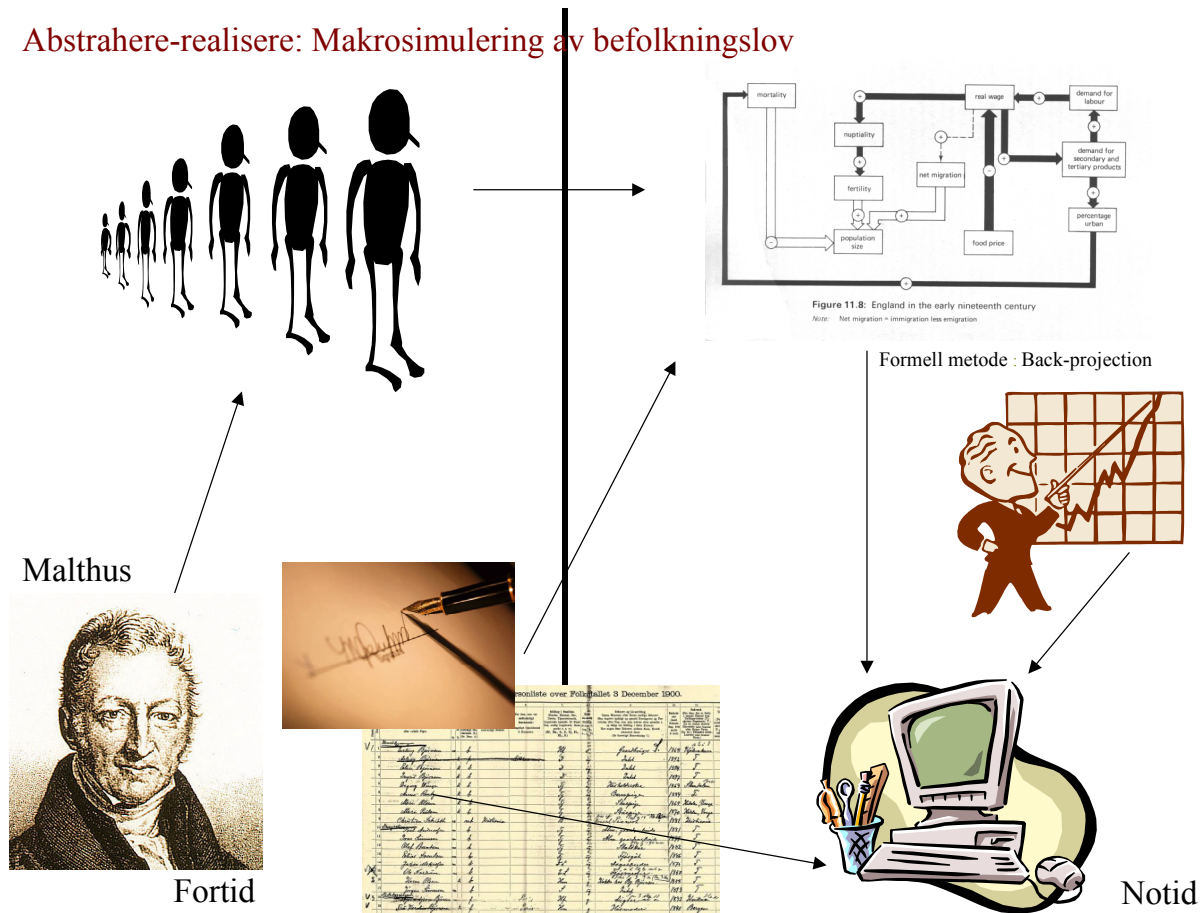


Men ”verda” kan òg vere tankar og idear dvs. abstraksjonar i utgangspunktet, som t.d. Malthus sin befolkningsteori. Men ”Verda” kan og bestå av notid. Vi kan og tenkje oss at problemet i figur 1 er ei historisk problemstilling vi ynskjer å få svar på. For å få svar på denne nyttar vi m.a. historiske kjelder. ”Verda” for ein historikar kan dermed bestå av ein kombinasjon av fortid og notid og konkrete kjelder og abstraksjonar. Denne breidda av ulike

element ein historikar si "verd" er viktig å understreke, jfr. figur 4. Både den kjeldeorienterte skulen og verktøyorienteringa synest å ha oversett nettopp dette poenget. Den kjeldeorienterte skulen har hatt fokus på å lage representasjon av kjelda, medan verktøyorienteringa har klart fokus mot den verda kjelda fortel om (personar, skip, institusjonar).

Figur 4 Abstrahere-realisere: Teoriar, fortidas røyndom, demografiske metodar og kjelder.

Abstrahere-realisere: Makrosimulering av befolkningslov

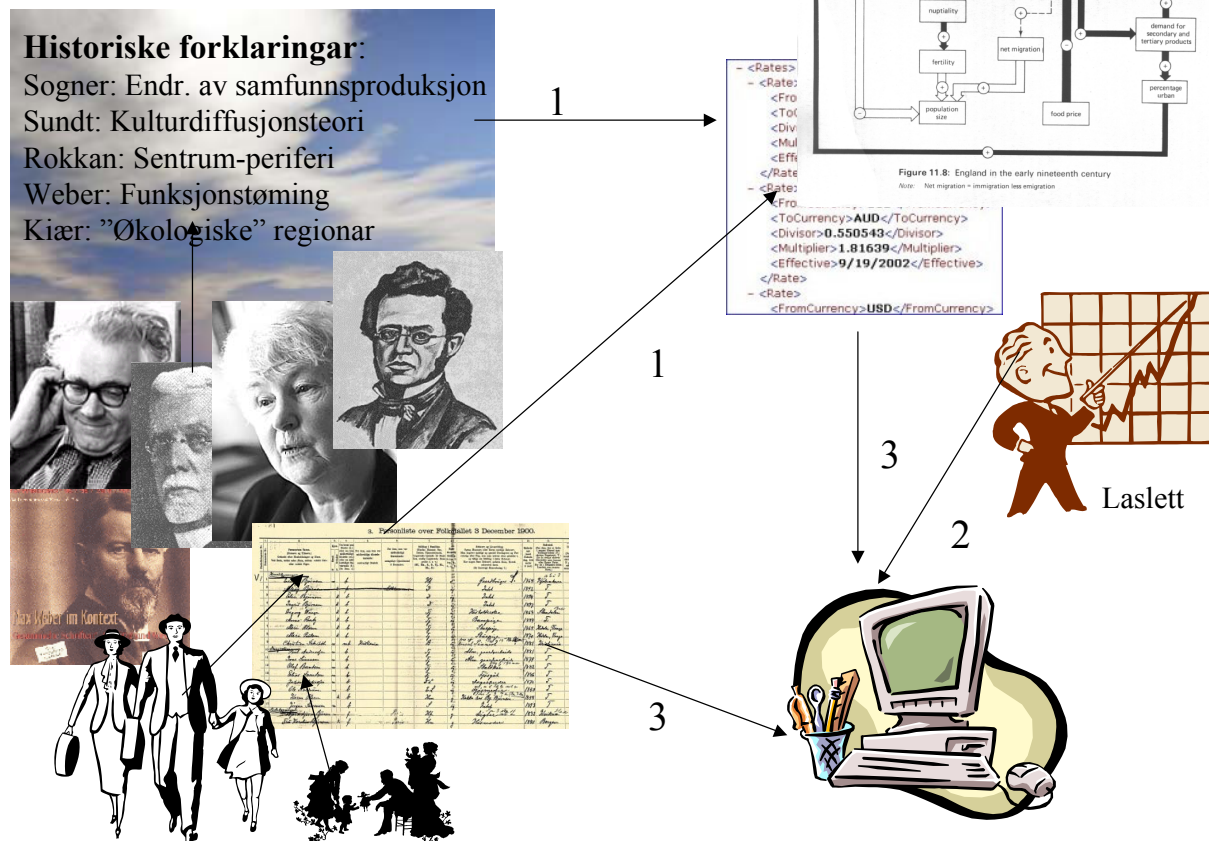


Kva er kjernen i disiplinen historisk informatikk?

- 1) For det første dreier faget seg om studie, utvikling og bruk av formelle metodar for å beskrive historiske objekt. Historiske objekt kan vere ting som framleis eksisterer – levningar, skriftlege og andre, eller ting som kjelder berettar har eksistert, t.d. personar, gardar eller skip. Men historiske objekt kan vere langt meir enn sjølve kjelda og det konkrete tekstlege innhaldet. Det kan og vere teoriar, modellar eller idear, som fortida, historikarar før oss har hatt. Dersom den formelle beskrivelsen gjeld konkrete levningar kan metodar frå informatikk nyttast til å publisere ein digital representasjon av levningen.
- 2) For det andre inkluderer fagområdet studie av, utvikling av og dokumentasjon av metodar og teknikkar som vi appliserer på historiske objekt, t.d. transformasjonsreglar, namnstandardisering, kjeldelenking og statistikk.
- 3) Og for det tredje omfattar historisk informatikk realisering/mekanisering av modellar, metodar og data i form av dataprogram. Desse tre elementa i fagområdet historisk informatikk er framstilt i figur 5.

Figur 5 Historisk informatikk: Modellering, formalisering, realisering av forklaringar, objekt og metodar.

Historisk informatikk



Historie og informatikk er i dette perspektivet langt meir integrert og samanvove med heile forskingsprosessen. Langt enn berre det å lage enkle registreringsskjema for ei kjelde. Denne modellen for den historiske forskingsprosessen skil seg klart frå Harvey's modell, der databasane først og fremst er representasjon av kjeldata. Det er ingenting i vegen for å nytte modelleringsteknikkar frå systemanalyse på historiske problemstillingar uavhengig om ein har tenkt å bruke datamaskin eller ikkje. Modelleringsteknikkane dreier seg om abstrahering som forenkling og bruksområdet er derfor vidt. Kjellstadli argumenterer også for bruk av enkle formelle modellar som verktøy for hypotesegenerering.

Konklusjon

Utgangspunktet for dette innlegget var å bruke usemja kring bruken av standard programvare analytisk for å komme til kjernen i fagfeltet mellom historie og informatikk.

Vi har sett at historiske kontekstar er kompliserte på ein annan måte enn notidas kontekstar. Tid skapar utfordringar på mange måtar. Det er usedvanleg viktig å forstå at bruk eller adaptasjon av ny teknologi i form av dataverkty også inkluderer metodar og modellar frå andre fag og samfunnsområde. Nokre gonger er overføring av metodar, teknikkar og modellar uproblematisk, andre gonger ikkje. Ein vil sjeldan vite i starten av eit forskingsprosjekt kva problema eller utfordringa ligg i. Ei kontinuerlig kritisk drøfting av verktøybruk, modellar og metodar er derfor naudsynt. Det er og viktig at datamaskiner blir tilpassa historikaren sine

behov, og ikkje omvendt. Særleg i desse to forholda ligg basis for eigen fagdisiplin: Historisk informatikk – eit fag i grenselandet mellom historie og informatikk.

Historisk informatikk: Abstrahering av historiske problem og historiske objekt, formalisering av historiske metodar og mekanisering av problem, objekt og metodar ved hjelp av datamaskin.

Aho, Alfred V. og Jeffrey D. Ullman (1995). *Foundations of computer science*. New York.

Date, C. J. (1986). *An introduction to database systems*. Reading, Mass.

Greenstein, Daniel I., red. (1991). *Modelling Historical Data: Towards a Standard for Encoding and Exchanging Machine-Readable Texts*. Halbgraue Reihe zur Historischen Fachinformatik. St. Katharinen.

Greenstein, Daniel I. (1994). *A historians' guide to computing*. Oxford.

Harvey, Charles og Jon Press (1996). *Databases in historical research theory, methods and applications*. Basingstoke.

Kjeldstadli, Knut (1992). *Fortida er ikke hva den en gang var : en innføring i historiefaget*. Oslo.

Mawdsley, Evan og Thomas Munck (1993). *Computing for historians : an introductory guide*. Manchester.

Nygaard, Lars og Registreringsentral for historiske data (1995). *Histform : norsk standard for registrering og utveksling av nominative folketellingsdata for årene 1865-1910*. Tromsø.

Smedt, Koenraad de (2002). "Some reflections on studies in humanities computing". *Literary & linguistic computing : journal of the Association for Literary and Linguistic Computing*. 17 (1).

Thaller, Manfred (1989). "The Need for a Theory of Historical Computing". *History and computing II*. P. Denley, C. Harvey og S. Fogelvik. Manchester, Manchester University Press: s. 2-11.

Thaller, Manfred (1991). "The Need for Standards: Data modelling and Exchange". *Modelling Historical Data*. D. I. Greenstein. St. Katharinen. A11.

Thaller, Manfred (1994). *What is source oriented data processing; what is a historical computer science?* Social Science History Association, Atlanta, Georgia.

Thorvaldsen, Gunnar (1996). *Håndbok i registrering og bruk av historiske persondata*. [Oslo].

Thorvaldsen, Gunnar (1999). *Databehandling for historikere*. Oslo.

Welling, George M. (1998). *The Prize of Neutrality*. Groningen.