

Historia i en digital värld

Jessica Parland-von Essen och Kenneth Nyberg

med fördjupningsartiklar av

*Jenny Bergenmar, Julia von Boguslawski, Lars Borin,
Helena Holm-Cüzdan, Richard Johansson, Markus Kainu,
Sakari Katajamäki, Juha Lehtonen, Mats Malm,
Yrsa Neuman och Helena Strömquist*

digihist.se

Version 1.0, februari 2014

Licens: [CC BY-NC 3.0](https://creativecommons.org/licenses/by-nc/3.0/)

Innehåll

Förord	5
Inledning	7
Kapitel 1. Den digitala revolutionen	10
<i>Från stordatorer till sociala medier</i>	10
<i>Konsekvenser för samhälle och forskning</i>	12
<i>Hur ska vi förhålla oss?</i>	14
Kapitel 2. Den digitala världen	16
<i>Den gutenbergska parentes</i>	16
<i>Bit bang</i>	19
<i>Kultur, kommunikation och medier</i>	22
<i>Webben som medium</i>	24
<i>Den digitala texten</i>	26
<i>Publicerandets ekonomi</i>	29
<i>Den ökande mängden digital information</i>	35
<i>Det digitala samhället</i>	39
Kapitel 3. Forskarvärldens respons	42
<i>Att möta förändring</i>	42
<i>Digitala humaniora: definitioner och debatter</i>	44
<i>Digital historia: en inringning</i>	48
<i>Betydelse(r) för forskningen</i>	50
Kapitel 4. Historieforskning i den digitaliserade världen	54
<i>Betydelser av digitalisering</i>	54
<i>Att hitta bland material</i>	56
<i>Digitalisering av källor</i>	59
<i>Fördjupning: Digitala textkritiska utgåvor</i>	71
<i>Fördjupning: Textkorpusar för historikerbruk – ett inifrånperspektiv</i>	76
<i>Fördjupning: ProBok – en proveniens- och bokbandsdatabas</i>	80
<i>Källkritik och hänvisningar på webben</i>	89
<i>Källkritik och hänvisningar i material i digitalarkiv</i>	99

Kapitel 5. Metoder inom digital historia	105
<i>Miljöer, verktyg och arbetssätt</i>	105
<i>Att strukturera information</i>	107
<i>Databaser</i>	109
<i>Big data</i>	116
<i>Fördjupning: Digitala textarkiv och forskningsfrågor</i>	121
<i>Fördjupning: Kulturomik: Att spana efter språkliga och kulturella förändringar i digitala textarkiv</i>	126
<i>Fördjupning: Open research methods in computational social sciences and humanities: introducing R</i>	139
<i>Visualiseringar</i>	148
<i>Kollektivt arbete (crowdsourcing)</i>	155
<i>Fördjupning: Transkribering av manuskript och förstaupplagor med talkokrafter</i>	158
<i>Om kritisk granskning</i>	167
Kapitel 6. Digitalbaserade material och långsiktigt bevarande	170
<i>Den förändrade offentligheten</i>	170
<i>Vad är forskningsdata?</i>	173
<i>Arkivering av digitalbaserade material</i>	177
<i>Urval – vad ska sparas?</i>	179
<i>Långsiktigt digitalt bevarande</i>	183
<i>Fördjupning: Bevaringssynpunkter på några vanliga filformat</i>	188
Kapitel 7. Att förmedla historia i en digital värld	191
<i>Att arkivera och publicera forskningsdata</i>	191
<i>Det nya publiceringslandskapet</i>	195
<i>Open access i teori och praktik</i>	199
<i>Fördjupning: Open Access-böcker – vad säger forskningen?</i>	205
<i>Nya publiceringsformer</i>	210
<i>Nya undervisningsformer</i>	220
Kapitel 8. Historikerrollens förändringar	225
<i>Vad har egentligen hänt?</i>	225
<i>Forskarens roll och expertens auktoritet</i>	228
<i>Framtidens historiker</i>	231

Förord

Jessica Parland-von Essen och Kenneth Nyberg

En av de grundläggande förändringsprocesser som påverkat alla vetenskapliga discipliner under de senaste decennierna är digitaliseringen av både själva forskningsarbetet och den vetenskapliga kommunikationen. Vid universiteten har medvetenheten om de utmaningar detta medför ökat på senare år, men förhållningssättet till informationsteknologiska metoder blir ofta okritiskt eller aningslöst på grund av att tekniken är så tillgänglig att vi inte ser hur avancerade program formar allt det vi gör med den. Därmed är det viktigt att historiker, liksom de som ansvarar för insamling och bevarande av kulturarvet, har en grundläggande förståelse för de tekniska aspekterna av sitt arbete och de frågor detta aktualiserar för forskning och samverkan både inom och utom kulturarvssektorn.

Arbete med digitala material och verktyg kräver mycket reflektion och omsorgsfull planering. Varje projekt är unikt, men det finns ändå anledning att känna till en del grundläggande saker. Eftersom vi alla arbetar med datorer numera gäller detta i praktiken var och en som arbetar med forskning. Med tanke på att vi i dag har mera potentiella digitala källmaterial än någon annan typ av källor, och att framtidens historiker kommer att ha främst digitala material för att rekonstruera eller förklara vår tid, är detta ett brännande problem även för kulturarvsinstitutioner som arkiv, bibliotek och museer.

Internationellt är digital humaniora (Digital Humanities, ofta förkortat DH) redan ett stort och växande fält och i såväl Finland som Sverige pågår arbete baserat på avancerade tekniska lösningar, till exempel Zacharias Topelius Skrifter (<http://topelius.fi/>) och den svenska Litteraturbanken (www.litteraturbanken.se/). På engelska finns vid det här laget en omfattande, delvis ganska specialiserad litteratur, men på svenska saknas till och med mer översiktliga arbeten som ger en bred ingång till ämnet.

Syftet med denna bok är att fylla en sådan funktion och att samtidigt ta upp de mer kunskapsteoretiska och principiella metodologiska frågor som den digitala utvecklingen väcker. Fokus ligger huvudsakligen på historisk forskning och bevarandefrågor knutna till material som används där, men stora delar av boken bör ha relevans även för det humanistiska fältet i vidare mening.

Projektets målsättning är en publikation, gärna både digital och tryckt, som i första hand ger en bred översikt men på ett antal punkter också innehåller fördjupningar av varierande omfattning. Rikligt med länkar och hänvisningar leder läsaren vidare till mer specialiserad litteratur och externa digitala resurser för ytterligare fördjupning och/eller exemplifiering av de företeelser som behandlas.

Den text som presenteras här publicerades först som en serie blogginlägg på webbplatsen <http://digihist.se> från maj 2013 till februari 2014, där det gavs möjlighet för intresserade att lämna respons på materialet i form av kommentarer, förslag till ändringar/tillägg eller tips på länkar och litteratur. Dessa synpunkter, tillsammans med de som lämnas på föreliggande samlingsversion av materialet, kommer vi att ta med oss när vi våren och sommaren bearbetar alla texterna inför publicering i digital och, förhoppningsvis, tryckt form i slutet av 2014 eller början av 2015.

Den som vill lämna respons kan göra det i webbplatsens kommentarfält eller genom att höra av sig direkt till någon av oss. Kontaktuppgifter (e-post, Twitter m.m.) finns på <http://digihist.se/om-oss>. Stort tack för ditt bidrag!

Anmärkning

För varje kapitel och fördjupning anges en eller flera ansvariga författare. Avsnitt inom kapitel har, när inget annat anges, skrivits av kapitelansvarig(a) författare.

Inledning

Jessica Parland-von Essen

En ovanligt stor publik hade samlats för att åhöra en disputation i Åbo den 28 mars 1663. Utöver studenter och professorer hade en mängd andra ståndspersoner infunnit sig för att höra Enevald Svenonius presidera vid försvaret av sin senaste text. Svenonius var smålänning och hade studerat i såväl Åbo och Uppsala som i Wittenberg. Han var en ung och lovande professor i teologi, men nu väntade man sig något särskilt – och det fick man sannerligen.

I publiken fanns som väntat Svenonius äldre före detta kollega, vice kanslern och biskopen Johannes Elai Terserus. Han närmade sig de sextio och hade också studerat ute i Europa, förutom i Tyskland också i Nederländerna och England. Terserus hade nyligen, som biskopar vid denna tid brukade, på egen bekostnad låtit trycka upp en liten förklaring till katekesen, vilken han hoppades kunna sälja vid sina biskopsinspektioner. Han hade gett exemplar, eller delar därav, till bland andra Svenonius för kommentarer, men utan att särskilt notera den kritik som då rests; ”det skulle falla mig alltför svårt och nesligt nu vid denne åldern och Embetet kasta mitt arbete under deras Censur”, som han själv förklarade.

Svenonius blev uppbragd; för inte så länge sedan hade Terserus själv beskyllt andra för häxkonster och nu kom han med egna tolkningar av heliga texter, tolkningar som kunde anses irrläriga. Svenonius hade därför författat en motskrift på svenska kallad *Warning för affall ifrån Troone*, för att ”ej allenast uti sina Academiska övningar at sätta sig emot dessa meningar och wara ungdomen för de samma”. Arbetet fick den effekten att Terserus blev arg och klagade hos konsistoriet, och som väntat infann sig också biskopen på Svenonius disputation. Det hela utmynnade i ett veritabelt gräl som gjorde att rektor tvingades avbryta hela tillställningen efter att båda parter förfallit till olämpligt beteende och börjat gå till personangrepp.

Senare behandlade konsistoriet saken och på dess uppmaning beslöt man dra in de knappt 400 böcker som fanns. Hela affären slutade med att kanslern Per

Brahe såg till att Terserus fick lämna sina ämbeten och flytta till Stockholm, vilket var något närmast oerhört på den tiden. Religionen var ett särskilt känsligt ämne, och det ansågs att en stor skada hade skett när saken på detta sätt disputerades offentligt. Speciellt illa var det förstås att det nu fanns kontroversiella böcker på svenska som kunde bli tillgängliga för en betydligt större krets, än de som normalt tog del av dylika diskussioner förda på latin i lärda kamrar. Terserus *Förklaring* vann ändå en viss spridning bland prästerskap och ståndspersoner, vilka kunde väntas läsa den med omdöme. Den är också en av de äldre böcker som tryckts på svenska inom nuvarande Finlands gränser.

Intressant är dessutom det faktum, att Henrik Gabriel Porthan ägnade stora delar av flera nummer av *Tidningar utgifna af et Sällskap i Åbo* hösten 1772 åt att presentera fallet. Det var nämligen strax efter det att Gustav III hade begränsat tryckfriheten. Hela historien och dess betydelse för Porthan, likaväl som för oss i dag, är att den handlar om kontroll av information och kunskap, om hur forskarna kan och måste testa tankar och idéer, och om hur samma vetenskapsmän sedan går ut med sin kunskap och sina tolkningar för att försöka förändra världen med sina bidrag.¹

Under 1600- och 1700-talen, ja ända fram till slutet av 1900-talet, har vi levt i en tid där böcker och andra trycksaker har varit den otvivelaktigt viktigaste och effektivaste kanalen för kunskapsspridningen. Inte ens den bästa forskare eller professor har genom sina föreläsningar och brev kunnat få ett sådant genomslag som en bästsäljande bok som spridits till otaliga bibliotek och kanske slitits i läsarnas händer under decennier. Samtidigt har makthavarna försökt styra bokmarknaderna, men den allra största faktorn har varit de kostnader som är förknippade med produktionen och distributionen av trycksaker. Dessa har gjort

¹ *Tidningar utgifna af et Sällskap i Åbo*, no 39–43, hösten 1772 (Digitala tidningsbiblioteket <http://digi.lib.helsinki.fi>), Matti Klinge: *Kungliga Akademien i Åbo 1640–1808*. Helsingfors 1988, s. 407–413, Matti Klinge: *Professorer*, Helsingfors 1989, s. 26–30, Tuija Laine: *Kolportöörejä ja kirjakauppiaita, Kirjojen hankinta ja levitys Suomessa vuoteen 1800*. Helsinki, 2006, s. 103 och Helsingin yliopiston opettaja- ja virkamiesmatrikkeli 1640–1917 (http://www.helsinki.fi/keskusarkisto/virkamiehet_2/index.htm) samt databasera Henrik (<http://dbgw.finlit.fi/henrik/index.php>), Fennica (<https://fennica.linneanet.fi/>) och Libris (<http://libris.kb.se/>).

förhandsgallring till en viktig faktor inom vetenskaperna, också om man frånsäger andra motiv att begränsa tillgång till distributionskanaler, vilket onekligen hör till den akademiska världens tråkigare traditioner.

Det är därför bra att komma ihåg att webben och många av dess egenskaper är resultat av många forskares och idealisters arbete för en bättre värld. Många drag i den nya digitala värld som denna bok behandlar liknar, återskapar eller förbättrar just vetenskapliga processer, medan andra egenskaper undergräver traditionella akademiska tankesätt. Inte minst hotas den fullständiga expertauktoritet den akademiska världen tidigare kunnat ge sina medlemmar. Att förneka denna utmaning av en tidigare självskriven status är inte fruktbart. Tvärtom är det fruktbart och värt att se den fria och mer demokratiska tillgången på information och kunskap som en god sak. Men man bör kunna spelets regler så gott det går och ta sitt ansvar genom att bidra på ett positivt sätt till utvecklingen. Det kräver att man är öppen för nya arbetssätt och diskurser.

Också om man inte aktivt deltar, bör man som forskare i dag vara medveten om vad den digitala världen för med sig, vad där pågår och hur man kan delta och påverka. Dagens informationsteknologi erbjuder dessutom fantastiska möjligheter för var och en att söka och hitta både information och kontakter runt om i världen. Det är en sak som man inte har råd att förbise, och som man inte ens kan komma undan genom att ägna sig åt böcker eller gamla papper och föremål.

Många av de frågor som gäller historikerna och deras forskning i övergången till det digitala är mycket komplexa och tiden kommer att utvisa hur forskningen och forskarna finner sin plats, och hur till exempel publicerandet kommer att se ut i det långa loppet. Då det gäller den flyktiga digitala världen är det ändå bättre att man försöker tänka efter före. En forskare behöver kunna verifiera och säkra sina källor, metoder och andra material och sprida sina resultat. I en värld där saker förändras med ljusets hastighet är det en verklig utmaning.

Kapitel 1. Den digitala revolutionen

Kenneth Nyberg

Från stordatorer till sociala medier

En av de mer genomgripande förändringsprocesser som präglat tiden efter andra världskriget är *den digitala revolutionen*. Den började, åtminstone i det allmänna medvetandet, med datorer och programvara för affärstillämpningar och spel under 1960-tal och tidigt 1970-tal. Efterhand blev datorerna och annan elektronisk utrustning alltmer av en konsumentprodukt, en förskjutning som accelererade först med mobiltelefonernas intåg och därefter den snabba etableringen och spridningen av internet i breda kretsar under sent 1990-tal.

Sedan millennieskiftet har både hård- och mjukvara i ökande utsträckning inriktats på mobil användning genom en konvergens mellan datorer, telefoner och nya produktkategorier som spelkonsoler och surfplattor. Detta har blivit än mer uttalat under de senaste fem-sex årens explosionsartade tillväxt av *sociala medier*, ett samlingsbegrepp för olika nätbaserade miljöer för social interaktion och informationsspridning i realtid. Några av de mest kända exemplen på sociala medier eller nätverk är i skrivande stund Facebook, Twitter, Wikipedia och YouTube, men det är ett landskap präglat av snabb förändring. Bloggar (av *weblogs*) är en publiceringsform som också ofta, men inte alltid, räknas till denna kategori.

Den nya tekniken i allmänhet, och sociala medier i synnerhet, har på ett oerhört kraftfullt sätt förändrat hur främst yngre människor interagerar med omvärlden. I vidare mening har detta bidragit till att medier blivit en allt viktigare del av människors liv, och det till en sådan grad att det moderna samhället knappast ens är begripligt om man inte beaktar just mediernas roll. Framför allt mobiltelefoner, men också andra tekniska lösningar används dels för att upprätthålla en intensiv kommunikation med vänner och familj, dels för mediekonsumtion eller andra tjänster som i varierande utsträckning förutsätter

tillgång till internet. Dessa båda användningsområden är ofta inte separata aktiviteter, utan intimt sammanvävda med varandra genom sociala medier som något av de ovan nämnda.

Två enkla statistiska uppgifter får här räcka som exempel på det genomslag som den nya tekniken och de sociala medierna har fått:

- År 1997 använde hälften av alla svenska 15-åringar internet. Åldern vid vilken denna frekvens uppnåddes hade 2004 sjunkit till nio år. År 2011 var det hälften av alla treåringar som använde internet.²
- Tredje kvartalet 2011 skickade amerikanska ungdomar i åldern 13–17 år i genomsnitt 3 417 SMS per månad, dvs. drygt 100 om dagen.³ Färskasvenska siffror tycks inte finnas tillgängliga och det finns anledning tro att de är något lägre än de amerikanska, men inte mycket.

Sammantaget visar dessa siffror, och många andra som skulle kunna anföras, att ungdomars (och i hög utsträckning även vuxnas) kommunikations- och medievanor genomgått mycket stora förändringar under de senaste decennierna. Det innebär också att det sammanhang i vilket historier bedriver forskning och, kanske framför allt, undervisning, är ett helt annat än för bara 10–15 år sedan. Allt tyder också på att den snabba omvandlingen kommer att fortsätta i oförminskad eller ökande takt framöver.

Det bör poängteras att det inte bara handlar om att ungdomar idag använder en annan teknisk pryl för underhållning och kommunikation än de gjorde för en generation sedan, utan om en mycket mer djupgående samhällelig förändring. Genom den ständiga tillgången till information och möjligheten att omedelbart

² Erik Forsberg, "Hälften av treåringarna använder internet", *Internetstatistik* 2011-11-16, <http://www.internetstatistik.se/artiklar/halften-av-trearingarna-anvander-internet/> (hämtad 2012-02-23). *Internetstatistik* är en webbplats som drivs av .SE, Stiftelsen för Internetinfrastruktur, den organisation som (lite förenklat) ansvarar för driften av de svenska delarna av internets grundläggande infrastruktur.

³ "New Mobile Obsession: U.S. Teens Triple Data Usage", *Nielsen Wire* 2011-12-15, http://blog.nielsen.com/nielsenwire/online_mobile/new-mobile-obsession-u-s-teens-triple-data-usage/ (hämtad 2012-02-23). The Nielsen Company är ett stort amerikanskt företag specialiserat på mätningar av mediekonsumtion från TV-tittande till internetanvändning.

sprida nyheter eller andra data i socialt formade nätverk, skapas helt nya mönster för hur människor bygger upp sin förståelse av omvärlden och interagerar med den. Traditionella auktoriteter undermineras när informationsflödet inte kan kontrolleras på samma sätt som tidigare vilket får politiska, ekonomiska och kulturella följder: "den arabiska våren" 2011, den snabba försvagningen av konventionella massmedia – dagspress, bokutgivning, musik- och filmdistribution – och så vidare. Allt tyder på att vi sammantaget står inför en radikal omstöpning av grundläggande samhällsstrukturer som, även på relativt kort sikt, är mycket svår att överblicka.

Denna utveckling har vid olika tidpunkter omväxlande kallats dator-, IT- eller den elektroniska revolutionen. Idag används alltmer ett begrepp som möjligen blir det mer långsiktigt gångbara: den digitala revolutionen. För historiker, arkivarier, lärare och andra aktualiserar den många intressanta frågor om samhällets förändring, vart vi är på väg och hur kunskaper om det förflutna kan hjälpa oss att bättre förstå det som händer. Men hur påverkar den digitala tekniken och dess samhälleliga följder historikerna själva, deras arbetssätt och forskningens villkor? Vilka nya möjligheter uppstår och vilka nya (eller gamla) problem behöver vi förhålla oss till? Det är några av de frågor som vi vill ta upp i denna kortfattade översikt.

Konsekvenser för samhälle och forskning

Den samhällsutveckling som beskrivits här har självfallet påverkat svenska historiker, men digital historia existerar inte som ett fält eller inriktning i Sverige idag.⁴ Många arbetar på olika sätt med digitala och andra tekniska verktyg, men sällan tas ett helhetsgrepp om vad de betyder för verksamheten idag och på längre sikt. Allmänt sett tycks traditionen av teknikanvändning vara svagare i historia än i angränsande ämnen som antikvetenskap, arkeologi och litteraturvetenskap. På grund av språkvetenskapliga frågor knutna till latin och

⁴ Det bör dock påpekas att inom det större fältet digital humaniora finns det vid Umeå universitet en internationellt framstående miljö, HUMlab, för utforskning av frågor som rör skärningspunkten mellan "humaniora, teknologi och konst" (<http://humlab.umu.se>, 2013-01-16). För mer om begreppen digital historia och digital humaniora, se kapitel 3.

grekiska har till exempel antikvetare i decennier använt den mest avancerade tillgängliga tekniken för databehandling av källtexter.⁵ Inom arkeologin har på liknande sätt geografiska informationssystem, GIS, länge varit ett viktigt verktyg, och där har mer generellt olika slag av teknisk utrustning spelat en större roll i forskningsarbetet än de gör i historieämnet.

Inte desto mindre är det uppenbart att även svenska historikers arbete, liksom forskningens villkor i större perspektiv, i allt högre grad påverkas av den digitala utvecklingen. Ur ett samhällsperspektiv kan exempelvis konsekvenserna för den akademiska kunskapens betydelse bli mycket långtgående. Den mest grundläggande och omvälvande följden av den digitala utvecklingen är ju att information av alla slag sprids allt snabbare, längre, enklare och billigare. Det är med andra ord ett kommunikativt genombrott av samma dignitet i mänsklighetens historia som talet, skriften eller tryckkonsten. Medan den sistnämnda innebar en stark standardisering och därmed möjlighet till centralisering av informationsflödet, betyder den digitala tekniken att flödet decentraliseras och blir mindre kontrollerbart på både gott och ont. Det finns å ena sidan en oerhörd demokratisk kraft i en sådan utveckling, men den kan också bidra till en fragmentering i subkulturer, att osanningar och odemokratiska åsikter sprids och att fenomen som näthat får större spelrum.

En annan konsekvens är att den några hundra år gamla idén om individuell upphovsrätt utmanas i en kultur där kopiering och modifiering av andras skapelser (oavsett medium) är oändligt mycket lättare än tidigare. Det är en utveckling som ibland sammanfattas under slagordet *everything is a remix*, och som på en mer fundamental nivå bidragit till att själva idén om ett sökande efter gemensamma "sanningar" om verkligheten undergrävs. Istället tillskrivs mångfalden av uttryck för individuella tolkningar av världen ett egenvärde i högre grad än tidigare.

Dessa tendenser skapar en delvis helt ny situation för historiker och andra forskare att verka i. När den moderna vetenskapen växte fram var det den som

⁵ Greg Crane, "Classics and the Computer: An End of the History", i Susan Schreibman, Ray Siemens och John Unsworth (red.), *A Companion to Digital Humanities* (Malden, MA: Blackwell 2004), s. 46–55.

stod för en kunskapsdemokratisering och utmaning mot befintliga auktoriteter, men idag är det ofta forskarna som utgör auktoriteter genom sin ställning som experter. Denna position undergrävs nu, hävdar vissa, genom att information så snabbt och enkelt går att få fram via nätet, till exempel hos Wikipedia. (Om information är detsamma som kunskap kan dock diskuteras, och vi återkommer därför till den frågan längre fram.)⁶ Å andra sidan erbjuder webben och andra digitala kanaler helt nya möjligheter för forskare att synliggöra sin verksamhet och nå ut med sina kunskaper. Det kan i bästa fall leda till att historiker får en tydligare och mer given roll i människors ögon, genom att deras arbete avdramatiseras och blir mindre svårbegripligt för utomstående. Genom den typen av kanaler kan det rentav bli möjligt för forskare att mycket enklare än hittills samarbeta med lekmän i så kallade *crowdsourcing*-projekt (även det återkommer vi till i ett senare kapitel).

Hur ska vi förhålla oss?

Genom åren har många profetior uttalats om de omstörtande följder den nya teknologin skulle få, även inom högre utbildning och forskning, men hittills har de i stor utsträckning kommit på skam. Även om datorer blivit vanliga redskap i olika delar av verksamheten har de inte varit mer än just hjälpmedel för att göra det man redan gjorde förut på ett bättre, snabbare och mer effektivt sätt. Ett kvalitativt nytt drag i det som hänt sedan millennieskiftet, och i synnerhet de senaste fyra–fem åren, är emellertid de sociala och ekonomiska konsekvenserna av den teknologiska utvecklingen i takt med att tekniken blivit alltmer mobil och tillgänglig (jfr ovan). Det gör, menar vi, att vi som lärare och forskare mer systematiskt måste börja diskutera hur vi ska förhålla oss till den digitala utmaningen och dess följder för hur vi arbetar med historisk kunskapsbildning.

Vi skulle kunna svara på dessa förändringar och förväntningar genom att säga att systemet har fungerat bra i generationer och att det inte finns någon anledning att ändra på det. Det är en inte helt ovanlig hållning, särskilt då

⁶ Sven-Eric Liedman, *Ett oändligt äventyr. Om människans kunskaper* (Stockholm: Bonniers 2001).

personer som arbetar med historia och historiska material gärna slår vakt om traditioner och beprövade modeller. Det är också en förståelig reaktion, särskilt inför det slag av oreflekterad teknikentusiasm där "nytt" definitionsmässigt innebär "bättre" än "gammalt" oavsett de faktiska meriterna. Ställda inför den samhällsomvandling vi nu är mitt uppe i, vore dock en sådan hållning inte förenlig med den insikt vi som historiker har om att samhällen alltid förändras. Den modell vi haft för vad forskning och undervisning är må vara utmärkt på många sätt, men den har utvecklats i ett specifikt historiskt och socialt sammanhang som nu håller på att omvandlas i grunden.

Vi behöver därför på nytt ställa de fundamentala frågor som aldrig bör glömmas bort: Vad är syftet med vår verksamhet? Hur ska vi uppnå detta syfte utifrån de förutsättningar som råder idag? Vilka metoder och förhållningssätt är de mest ändamålsenliga och varför? Gör vi inte detta kommer vi sannolikt att få det allt svårare att för det omgivande samhället begripliggöra värdet av vårt arbete. På sikt, och kanske snabbare än vi tror, kan vi då uppfattas som irrelevanta av beslutsfattare och andra, vilket skulle kunna få mycket olyckliga konsekvenser.

Det ovan sagda ska inte uppfattas som att vi är odelat positiva till den tekniska utvecklingen och dess samhälleliga följder, utan som ett uttryck för övertygelsen att vi aktivt måste förhålla oss till denna utveckling för att inte överflyglas av den. Vi måste fråga oss vad som händer med själva kunskapen och hur framtidens historia ska skrivas. Bara om vi förstår de nya förutsättningarna, miljöerna och verktygen har vi några som helst möjligheter att i viss mån påverka dem och få till stånd en kritisk diskussion om för- och nackdelar med olika redskap och metoder. Vi måste också kunna föra en diskussion om på vilket sätt vårt tidigare kulturarv bäst digitaliseras och hur vi kan förhindra att vi förlorar för framtiden viktiga digitala material. Vi hoppas att denna bok både ska underlätta förståelsen och bidra till den kritiska diskussionen om historia i en digital värld.

Kapitel 2. Den digitala världen

Jessica Parland-von Essen

Den gutenbergska parentesen

Tryckkonsten som spred sig i Europa från slutet av 1400-talet till 1600-talet bidrog med nya aspekter till den skrivna kommunikationen och genom detta till hela kulturen. Under upplysningens sekel var både böcker, press och småtryck redan en relevant del av kultur och samhälle i Europa. Boksamlingar och bibliotek växte och i synnerhet ansedd och viktig information kunde ordnas och sparas. Idéer om allt större frihet och folkupplysning spred sig. Läsesällskap och folkupplysande tidskrifter blev möjliga i takt med att de ekonomiska förutsättningarna blev bättre. I bibliotek och encyklopedier ville man samla all världens kunskap. Man klassificerade, strukturerade och indexerade växter, djur, människor, religioner, vetenskaper, böcker. Man samlade och ordnade förutom böcker föremål av alla slag: antikviteter, insekter, kartor, stenar, mynt och man kartlade länder, stjärnor och planeter. Kunskapen blev strukturerad i olika discipliner och fakulteter, kontrollerade först av vetenskapliga akademier och sedan under följande sekel av universiteten. Och i böckerna fanns sanningen. Den romantiska tilltron till det individuella geniet gjorde att de gamla antika och religiösa auktoriteterna fick ge vika för nya intellektuella storheter, som gav ut tjocka böcker baserade på vetenskapliga metoder.

Eftersom det är dyrt att göra böcker, att trycka upp och distribuera dem, var man tvungen också inom det vetenskapliga publicerandet att gallra *innan* man publicerade i tryck. Det som en gång var tryckt kunde dessutom sprida sig och bli läst och tolkat utom kontroll, utan tillräcklig kännedom om kontext och kritisk förståelse (att beakta är att censur i praktiken aldrig gällt ägande och läsning av texter inom samhällseliten förrän under 1900-talet, utan endast "allmänheten" eller "folket" samt bokbranschen och pressen). Universitet och lärda sällskap fick således en allt viktigare roll av garanter för kvalitet, för att det

som trycktes var rätt och sant. Uppgiften blev allt viktigare i och med att tryckning av texter på folkspråk blev vanligare. Man lät göra mycket viktiga och nyttiga böcker om ekonomi och medicin. Samhällsfrågor kunde tidvis ventileras flitigt i pressen, den nya offentlighet den tyske sociologen Jürgen Habermas skrivit om och som fanns både i talad, handskriven och tryckt form och kommunikationen flöt emellan dessa sfärer.

Under den tryckta textens era var det enkelt att jämföra texter och man lyckades inom den framväxande bokbranschen lansera upphovsrätten för författare, först endast för några år, så att upphovsmannen kunde få en skälig ersättning för sitt arbete. Mecenaternas betydelse minskade; det borgerliga samhället trodde på individens arbetsinsats och på det skapande geniet, som närmast ur intet med sin egen tankes kraft kunde upptäcka stora sanningar om världen och människan. Till detta hörde också den ökade betydelsen av "originalet" eller "originaltexten", något som förut inte varit särskilt relevant i normala fall. Handskrivna texter och gamla manuskript var aldrig fullständigt identiska, men tryckkonsten skapade en illusion om en originaltext, en text som skapats och auktoriserats av en enda individ. (Bokhistoriker vet att det verkligen rör sig om en illusion, även tryckta exemplar till och med inom samma upplaga kan skilja sig ansevärt från varandra.)

Det tog ett tag innan översättningarna nåddes av samma syn, men så småningom blev också översättningar "auktoriserade" och mer trogna författarens version. Längre hade det varit norm att översättningar tvärtom lokaliserades och förkortades enligt översättarens eller förläggarens eget omdöme. Upphovsrätten var till en början en nationell angelägenhet, men detta var ohållbart: stora mängder piratupplagor trycktes upp i till exempel Belgien och USA och spreds över stora områden. Slutligen gjordes ett internationellt avtal om saken, den s.k. Bernkonventionen 1886. Författarna hade nu uppnått absolut äganderätt till sin text. I och med nöjesindustrins starka lobbying har upphovsrätterna sedermera förlängts till flera decennier (sju för närvarande, nio har diskuterats) *efter upphovsmannens död*.

Medierna påverkar alltså kulturen. Boktryckarkonsten har betytt mycket för vårt kultur. Då böckerna blev vanligare kunde folkspråken normaliseras och stavningen blev enhetligare inom olika nationalspråk. Hela idén om nationalspråk och nationer blev mer aktuell. Folkbildningen blev möjlig på ett nytt sätt. Läsesällskap och allmänna bibliotek uppstod. Tidningspressen skapade en offentlighet som i teorin överskred tid och rum. Världen och kunskapen blev enhetligare och möjlig att organisera och skenbart kontrollera. Samtidigt började informationen drastiskt ackumuleras. Småningom kunde man inte läsa alla böcker, man kunde inte längre känna till all forskning eller ens alla discipliner i stort, man kunde inte ha läst all skönlitteratur. Mängden information ökade närmast exponentiellt. Så bibliotekariernas makt minskade inte i takt med att böckerna blev fler, utan deras metoder och förmågor blev allt viktigare. Klassificering, katalogisering och indexering var det enda sättet man kunde orientera sig bland alla trycksaker.

Livet i den gutenbergska galaxen, som Marshall McLuhan kallat den⁷ eller innanför *den gutenbergska parentes* (ett begrepp tillskrivet den danske professorn L.O. Sauerberg) var en tid då författare och verk var entydiga begrepp, då original var något som kunde användas om en text, trots att de postmoderna filosoferna och konstnärer som Andy Warhol problematiserade det. Auktoritet och expertis erhöles genom en kontrollerad (ut)bildning, som i ett gammalt korporativt system. Referee-rutiner och dyra tidskrifter gjorde att de mest ansedda forskarna, dvs de som en gång uppnått en god status, hade mycket stor kontroll över hur forskningen utvecklades.

All humanistisk forskning krävde *a priori* stora mängder tid, läsning av enorma mängder böcker och artiklar och källor, skrivning, avskrivning, kopiering, omskrivning. Allt detta är arbete som ger en mycket tid att tänka på enskilda detaljer och stora helheter. Samtidigt var forskaren tvungen (åtminstone i sina första undersökningar) att begränsa sin forskning *utgående* från källorna, helt enkelt av arbetsekonomiska skäl. Vi har lärt oss att i början av

⁷ Marshall McLuhan, *Gutenberggalaxen. Den typografiska människans uppkomst* (Stockholm: PAN/Norstedt 1969).

en undersökning presentera alla källor (antagande att de är enhetliga helheter, typ vissa böcker, en brevsamling, ett arkiv eller t ex vissa tidningar under vissa år) och motivera avgränsningar och diskutera källkritiska frågor *på ett allmänt plan* för varje käll- eller materialgrupp. Vi måste då ta ställning till källornas och samlingarnas uppkomst och proveniens. Alltför sällan har man ändå i allmänhet djupare reflekterat över de åtgärder som vidtagits av arkiv och den gallring och rekonstruktion av historien som gjorts i alla år av professionella arkivarier.

Bit bang⁸

Den moderna datorn som uppfinning brukar tillskrivas Alan Turing, som under andra världskriget var ett kodknäckargeni i brittisk tjänst, även om olika typer av mekaniska räknemaskiner utvecklats både i teori och praktik åtminstone sedan Pascals dagar. Turing skapade datavetenskapen under 1930-talet och under samma period kom också den första digitala elektriska räknemaskinen. Under åren har flera lager av kod och programmering kommit ovanpå den ursprungliga koden av ettor och nollor. I takt med att datorerna har blivit snabbare och minnet radikalt mycket effektivare har många system och program vuxit sig nästan oöverskådligt stora och komplicerade.

Ett viktigt steg i utvecklingen var hopkopplandet av datorer till nätverk. Internet var från början ett projekt finansierat av ARPA, Advanced Research Projects Agency, och som var en del av USA:s försvarsmakt. Målsättningen var att skapa ett system som kunde trygga kommunikation och dataöverföring också under undantagsförhållanden, då en del av systemet kanske var satt ut spel. Detta kallas robusta system och en grundidé är förstås att de är distribuerade, alltså geografiskt utspridda på olika datorer som kan utbyta information med varandra. De första datorerna som kopplades samman fanns vid forskningsinstitut på den amerikanska västkusten. ARPANET föddes kring 1970, över tio år efter att projektet startats. Systemet som växte fram för att överföra informationen kallas Transmission Control Protocol/Internet Protocol TCP/IP.

⁸ Uttrycket "Bit bang" i den betydelse det används här har tillskrivits Yrjö Neuvo i Martti Häikiö & Essi Isotalo, *Bit bang. Yrjö Neuvo ja digitaalinen kumous* (Helsinki: SKS 2013) s. 11.

Det togs i bruk 1983. Idén är att varje dator har en egen adress som består av en sifferserie (IP) och informationen packas i olika paket som skickas genom nätverket, varje paket den snabbaste lediga vägen, vilket ger både snabbhet och gör systemet robust. Vid denna tid var utvecklingsarbetet redan starkt förankrat i den akademiska världen, eftersom många forskare såg nyttan med ett dylikt kommunikationssystem, som sedan mot slutet av 1980-talet också öppnades för kommersiella aktörer och ARPANET lades ner. Ännu i dag består internet av många nätverk som är länkade till varandra. Forskare och myndigheter behöver ofta mycket stor kapacitet för dataöverföring varför man i Finland upprätthåller det snabba FUNET-nätet, som också är kopplat till det europeiska GEANT.⁹

Internet omfattar alltså mer än det vi kallar webben, och kan alltså också användas till annan kommunikation än överföring av information i hypertextformat. Det som avses med "webben" är nämligen i princip det material vi kommer åt via våra webbläsare i dag. World Wide Web utvecklades vid det europeiska forskningsinstitutet CERN under slutet av 1980-talet av Tim Berners-Lee. Det bygger på ett särskilt protokoll, Hyper Text Transfer Protocol, HTTP, som förutom den enkla texten också består av en hel del information om hur texten ska visas. Detta förmedlas med en kod som kallas html (Hyper Text Markup Language) och det trevliga med den är att den kan göras rätt enkel, så att den till stora delar består av engelska. Dagens webbsidor är ändå ofta väldigt proffsigt och komplicerat byggda, så html-koden är inte särskilt lättbegriplig för en lekman. Men den är ingalunda hemlig eller dold, utan kan studeras via webbläsaren, som brukar ha en möjlighet där det står "visa källkod" eller motsvarande.

Den allra viktigaste egenskapen hos HTTP och html är ändå möjligheten att skapa länkar till andra filer. Koden gör också att man kan bädda in bilder, video, ljud eller länkar till andra html-filer i den bild som visas i webbläsarens fönster på skärmen. Varje sida på webben är en egen html-fil med en egen adress eller stig. Den kallas Uniform Resource Locator eller URL-adress, men som sagt består

⁹ Melih Bilgils korta video *History of the Internet* är en bra introduktion till ämnet: <http://www.youtube.com/watch?v=9hIQjrMHTv4> (2014-03-04).

den ofta i praktiken av en hel del olika filer som webbservern plockar ihop och webbläsaren sedan visar enligt html-direktiven. Varje webbserver där webbsidor finns har en egen särskild databas och ett eget namn och en egen adress. Eftersom de så kallade IP-adresserna är rätt otympliga sifferserier länkas de med domännamn enligt internationella avtal och system, vilket gör att vi oftast slipper fundera på IP-adresserna och kan operera med vad vi kallar webbadresser i vardagligt tal.

Mängden information på webben ökar snabbt för att det i detta distribuerade system finns allt fler aktörer och innehållsproducenter. År 2001 beräknades det finnas 550 miljarder webbsidor, de flesta av dem fanns då i den så kallade djupa webben som inte indexerades av sökmaskiner eller kanske ens är fritt tillgängliga. I mars 2009 indexerades över 25 miljarder webbsidor.¹⁰ Man räknar med att webben i juni år 2012 hade 2,4 miljarder användare.¹¹

En av de absolut största konkreta skillnaderna mellan den gutenbergska tiden och den digitala världen finns i kostnadsstrukturen och de försvinnande små, ibland i stort sett obefintliga, resurser som i dag krävs för att publicera sig. I teorin kan man göra det helt gratis åtminstone i stora delar av Europa och Nordamerika. Det enda man behöver är kanske ett bibliotekskort, eller att man köper en kopp kaffe på ett internetkafé, det vill säga tillgång till internet. Dessutom behöver man elementära kunskaper i webbanvändning och information eller lust att skaffa information om hur man exempelvis startar en blogg. I och med att tjänster och teknik blir allt mer användarvänliga, blir webben allt mer tillgänglig som medium. Och då också som kommunikations- och potentiell publikationskanal.

¹⁰ World Wide Web, Wikipedia på engelska http://en.wikipedia.org/wiki/World_Wide_Web. 16.6.2013.

¹¹ Internet World Stats <http://www.internetworldstats.com/stats.htm> Källorna till statistiken finns något vidlyftigt beskrivna här <http://www.internetworldstats.com/surfing.htm>. Mer specifik data finns på Wikipedia i artikeln Global Internet usage http://en.wikipedia.org/wiki/Global_Internet_usage. 16.6.2013.

Kultur, kommunikation och medier

För att kunna greppa vad vi egentligen står inför i dag och i framtiden, är det nödvändigt att titta lite närmare på vad det är för en förvandling och utveckling vi har för händer. Vad innebär förändringen, om vi granskar den ur ett analytiskt perspektiv? Det står klart att teknisk utveckling ofta har kulturella och ekonomiska dimensioner. Då det handlar om kommunikationsteknik kan effekterna för kultur och kunskap vara mycket djupgående.

Kommunikation är en förutsättning för kultur i ordets vidaste bemärkelse. En kultur måste ha flera deltagare och kulturen inbegriper språk, men också andra sätt att kommunicera individer emellan. Kommunikation förutsätter att information rör sig, att någon medvetet eller omedvetet sänder ett budskap bestående av ett eller flera tecken till någon annan. Den som tar emot budskapet tolkar meddelandet utgående från kontexten och sin egen kunskap. Detta är den klassiska informationsteoretiska modellen. Den är enkel, väldigt enkel. Kraftigt förenklad innehåller den endast några element och en enkelriktad process: Avsändare – meddelande – mottagare. Och någon sorts kodning och kodavläsning. Ju mer identisk informationen i avsändarens och mottagarens huvud, dess bättre, mer lyckad kommunikation.

Problemet med modellen är att mening inte uppstår så rätlinjigt. Betydelser är komplexa i sig, inte minst är själva skapandet av betydelse allt annat än enkelt. Enligt kultursemiotikern Juri Lotman är redan det utgående meddelandet en sorts kompromiss: då vi formulerar ett ord eller en mening fastslår vi en (ibland flera) specifika betydelser, vad vi avser med vårt uttalande, och samtidigt avgränsar vi ordets och begreppets alltid mer eller mindre diffusa betydelser och tvingar in dem i ett visst sammanhang, som *varje gång* blir unikt. En kontext kan aldrig upprepas helt identiskt. Om inte annat förändrats är tidpunkten en annan, liksom erfarenheterna som föregått yttrandet.

Mottagaren, den som tolkar meddelandet, gör sedan dessutom sin egen tolkning utgående från sig själv och den information och förmåga han eller hon har. Den tolkningen blir också mer eller mindre olik för varje gång, fast meddelandet är det samma. Man kan tydligt se att det blir så då meddelandet är

komplext: varje gång man läser om en bok eller ser om en film gör man det på ett nytt sätt, nya tolkningar och associationer uppstår.

Tolkande är liksom kommunicerande en grundläggande mänsklig egenskap. Vi försöker tolka och förstå allt vi förnimmar: stjärnorna på himlen blir stjärnbilder, liksom vi försöker förstå ord och bilder; fågelsången vi lyssnar till för tankarna till en fågel, till sommaren eller något annat för oss betydelsefullt. Vi söker mening i livet. Vi söker struktur och ordning för att greppa vår värld. Information som vi inte kan tolka blir brus och oväsen, den irriterar oss och vi försöker ignorera den om den inte öppnar sig för oss och delar med sig av budskapet. I ett främmande språk söker vi desperat efter hållhakar, bekanta ord eller mönster som kan ge oss någon information. För att förstå behöver vi en kontext, ett sammanhang: Vad talar du om? Vilket språk talar du? Vad vill du?

Ett medium är egentligen det som är "mellan" dem som kommunicerar. Man avser då oftast något som gör att man kan kommunicera med någon annan utan att båda är närvarande samtidigt. Ur ett semiotiskt perspektiv är definitionen lite problematisk: för semiotikern kan mediet vara talet i sig självt, lika väl som ett trafikmärke, en tv-reklam eller en tavla. Det handlar i grund och botten om samma sak: om att förmedla ett budskap, att ge något en betydelse utanför objektet självt. Betydelsen varierar beroende på kontext och kultur: att lämna kvar mat på tallriken betyder en sak i en kultur och en annan i någon annan kultur. Genom en handling vill vi kommunicera något, men om man inte är medveten om de andras kultur kan det gå fel ...

Ändå har själva mediet också relevans för hur betydelser och tolkningar kan överföras, eftersom det skapar begränsningar lika väl som möjligheter. Under de senaste seklerna har till exempel den tryckta texten förstärkt vissa drag i den västerländska kulturen. Det skrivna språket har gjort att meddelanden har kunnat bevaras över tid och rum. Komplex information har kunnat spridas, sparas och organiseras i stor omfattning. Den tryckta texten och bilden har gjort att detta kunnat göras än mer effektivt.

Inom filosofin har John Searle närmast sig förhållandet mellan språket och kommunikationen genom begreppet yttrande (utterance), en handling där en

tanke får en specifik betydelse i sin specifika kontext då den uttalas med språkliga medel, som tal. I paritet med detta har Peter L. Shillingsburg lanserat en skriftteori (script theory), som innefattar skrivandet som en motsvarande handling.¹² Processen är givetvis besläktad, det handlar om att förvandla tankar till ord och meningar, text. Men samtidigt är skrivandet annorlunda: den skrivna texten saknar tonfall, intonation. Det moderna normaliserade skrivspråket saknar otaliga nyanser, samtidigt som texterna materialiserats och kan flyttas och kopieras utom skribentens kontroll. Detta bidrar till det nödvändigtvis längre formatet. För att vara exakt och minska risken för fatala feltolkningar, måste den normaliserade texten vara omständlig och utförlig.

Webben som medium

Mediets betydelser för kulturen har diskuterats också av historiker, så som Lucien Febvre, Henri-Jean Martin, Harold Innis Adams och Elisabeth Eisenstein. På senare tid har viktiga namn på området varit till exempel Robert Darnton och Ted Striphas. Gemensamt för alla dessa är att de ser vilka tekniska och ekonomiska effekter själva mediet haft för samhällsutvecklingen.

Sedan tryckprodukterna fick en viktig roll i vår kultur har kommunikationen under en lång tid blivit allt mer enkelriktad. Det vi i vanligt tal tidigare avsåg med media, det vill säga massmedierna, är påfallande mycket av en-till-många karaktär. Med undantag av enskilda ring in-program har det varit journalister eller myndigheter som valt varje ord som sagts eller tryckts och nått ut till publiken. Få har haft tillgång till att publicera sig och uttrycka sin åsikt inför en större publik utan mellanhänder.

I dag har vi ett nytt snabbt och demokratiskt medium: webben. Webben är i regel mycket snabbare och framför allt mer demokratiskt än de tidigare massmedierna. Vem som helst kan publicera en "nyhet". Det finns stora skillnader i hur långt enskilda individers röst bär på webben beroende på de nätverk var och en har, men en stor nyhet sprids oberoende av detta ofta väldigt

¹² Peter L. Shillingsburg: *From Gutenberg to Google. Electronic representation of literary text.* (Cambridge : Cambridge University Press 2006).

snabbt med hjälp av mikroblogger eller andra sociala medier tack vare andra människors nätverk.

Om man i dag är uppkopplad på internet är det sannolikt att man nås av nyheter den vägen först. Dagarna då alla visste samma saker efter kvällens huvudnyhetssändning på huvudkanalen är förbi. Den enhetliga offentlighetens tid börjar vara slut. Nyhetsflödet liksom den allmänna debatten är kaotiska och innehåller i dag också sannolikt mera desinformation. Det korta formatet och snabba tempot leder till missförstånd och feltolkningar. Lyckligtvis följs ändå mikrobloggernyheter rätt snabbt av länkar till utförligare uppgifter, ofta producerade av kommersiellt eller offentligt finansierade journalister och medier. I längden brukar fel rättas till.¹³

Vi har plötsligt ett lättillgängligt medium som förutom att vara potentiellt snabbt, vilket telefonen eller redan telegrafan var, nu även kan nå en mycket stor publik. Mediehusens och journalisternas tidigare kontroll över innehållet, som tidigare kraftigt styrdes av ekonomiska faktorer är nu radikalt förminskad i samhällen där internetpenetrationen är hög. Vi har i dag en fjärde statsmakt som granskar den tredje. Samspelet mellan sociala medier och traditionella medier är komplext och de nya affärsmodellerna söker ännu sin form.

Det nya mediet har visat sig ha en del positiva effekter. En av dem manifesterar sig allra tydligast i Wikipedia. Det har visat sig att folk hellre är konstruktiva och delar med sig av sitt kunnande än är destruktiva. Det "goda" har hittills vunnit över det "onda" på webben eftersom de allra flesta människorna hellre ger en positiv bild av sig själva också för främlingar till och med som anonyma eller praktiskt taget anonyma. Dessutom har det visat sig att sociala belöningar är mycket mer effektiva incitament och viktigare för de allra flesta än ekonomisk vinning. Att öppenhet och samarbete är effektivare och oftast leder till bästa resultat har bevisats många gånger i praktiken inom programmeringsbranschen. Vi har i långa tider låtit oss luras av teorier om

¹³ Se Farida Vis forskning <http://www.sheffield.ac.uk/is/staff/vis>.

ekonomisk rationalism. Webben fungerar inte på samma sätt som tidigare medier.¹⁴

Webbens många-till-många-karaktär har lett till att informationsflöden idag är kaotiska och slumpartade. En del nyheter sprider sig som löpeld över webben via de nätverk som privatpersoner utgör. Vissa spärrar finns ändå, både tekniska som är uppsatta av politiska skäl och språkliga. Men många internetanvändare är i dag aktiva innehållsproducenter. Folk bloggar, laddar upp bilder, delar bokmärken och kompletterar information på olika plattformar: genom att skriva kommentarer på bloggar och nyheter, skriva texter till Wikipedia eller fyller i och korrigerar uppgifter i olika databaser och register. I synnerhet det sista är en möjlighet för forskare. Men också vad gäller forskningsmetoder och publicering innebär det förstås djupgående förändringar i historikernas arbete.

Den digitala texten

Man kunde alltså lätt tänka att den digitala texten (i vid bemärkelse) har inneburit en textens befrielse, att den dynamiska virtuella texten är fri från begränsningar. Vi kan välja mellan otaliga bokstavstyper, oändliga layouter och vi kan skapa och manipulera bilder på ett sätt som gör att man lätt känner sig omnipotent. Med hjälp av datorn och webben kan vi spränga många gränser som tidigare medier haft. Nu när den tryckta textens hegemoni är bruten, blir dess betydelse för kulturen tydlig: vi ser nu hur ideologier och ekonomi direkt påverkat och påverkats av kommunikationens form, av det tryckta mediet. Den digitala världen ter sig vid första anblicken som den ultimata frihetens kultur.

Men som Jaron Lanier påpekat är denna frihet en chimär. När vi kastat tryckpressens bojar och simmar i den digitala textens hav, ser vi inte att också detta hav styrs av strömmar, av molekylernas bestämda kemiska relationer och havets ekologiska balans. Dels beror detta på att allt är så nytt, i en brytningstid är det svårt eller omöjligt att urskilja detaljer och de stora sammanhangen och se deras relationer. Men det beror också på att teknologin och de logiska

¹⁴ Clay Shirky, *Here comes everybody* (London: Penguin 2009), Malcolm Gladwell, *The tipping point. How little things can make a big difference* (New York: Little Brown 2001).

strukturerna som producerar och presenterar texterna är väl dolda för oss. I sin användarvänlighet har tekniken blivit oväsentlig för alla dem som inte direkt arbetar med utveckling och underhåll av informationsteknologi. Eller snarare: man upplever den som oväsentlig eller onödig att förstå sig på. I dag ingår ännu inte tillräckligt med it-kunskaper i den grundläggande utbildningen på högskolenivå, inte teknologins historia eller grundprinciper. Det är kanske inte nödvändigt att alla har den kunskapen, men den borde ingå i all högre utbildning, också inom humaniora. Det beror på att de tekniska strukturerna påverkar vår kunskap och vår kultur, hur den struktureras och de mekanismer som driver fram betydelser och kommunikation i vårt samhälle. Således påverkar den också oss själva, hur vi fungerar och uppfattar oss själva.¹⁵

Eftersom vi vanligen nås av färdiga produkter inser vi oftast inte att det bakom varje lösning finns många val, som påverkar inte bara hur texten ser ut, utan också dess strukturer och till och med vårt eget beteende. Det kunde ha funnits många fler alternativ till hur man programmerar datorer och hur man strukturerar informationen i dem. De strukturer som i dag dominerar, såsom "filer" och "webbsidor", är bara ett av många alternativ. Dessa strukturer kommer att påverka vår kultur lika mycket som tryckpressen.

Det finns emellertid en annan aspekt av den digitala texten som är ännu mer problematisk för humanisten. Den sammanhänger med föregående resonemang, om att vi behandlar datorn och dess program som en svart låda, där det pågår något för oss irrelevant. Vi är nöjda med att vi på datorskärmen ser det som vi önskar, men som de facto är mycket mer predestinerat och påverkat av våra tidigare erfarenheter av datorer än vi vill medge. Det är viktigt att komma ihåg, att datorskärmen inte är ett papper utan den visar en representation, en vald bit av det som pågår inne i ett mycket komplicerat system. Och detta system opererar inte med betydelser, för betydelserna finns bara inne i våra hjärnor. Samtidigt är de betydelser vi tillskriver det vi ser på skärmen åtminstone delvis producerade av datorn.

¹⁵ Jaron Lanier, *You are not a gadget. A manifesto* (London: Vintage 2011).

Låt oss för en stund återgå till semiotiken och Juri Lotman för att undersöka hur betydelser egentligen uppstår. Lotman (och andra lingvister och semiotiker) talar om språkets pragmatiska och syntagmatiska dimensioner, och hur de två möts i yttrandet eller skrivandet och tolkandet. Tecknet har två (eller tre, beroende på om man håller sig till den europeiska eller amerikanska semiotiska skolan) dimensioner – varav datorn kan operera endast med en: den representerande symbolen. Alla andra aspekter raderas ut i datorns svarta låda, som omöjligen kan ta i betraktande innehållet, betydelsen eller det representerade. Inne i datorns system frikopplas tecknet från sin betydelse och genomgår olika processer varefter en ny representation spottas ut på skärmen. Trots att man kan utveckla systemen så att datorer kan hantera osäkerhet och olika slag av ambiguitet i högre grad, kan de ändå aldrig tänka, eftersom tänkandet kräver medvetenhet om tecknens betydelser. Datorns själlösa räkneoperationer kan aldrig likställas med en människas skapande handling då vi "spottar ur oss text". Då vi låter datorn hantera tecknen som om de vore tomma, utan innebörd, försvinner den kreativa dimensionen. Om vi låter det ske i en för oss svart låda, utan att vara medvetna om hur den fungerar eller påverkar tecknen är vi inte vetenskapliga. Vi måste minnas att datorn i grunden är en räknemaskin, "a computer". Den kan inte resonera eller tänka.

Men kod är också språk. Den datorkod programmerare oftast jobbar med är, trots att den inte syns för användaren, också för människor läsbar och begriplig text. Trots att datorn i grunden endast opererar med ettor och nollor är datorsystemen i dag uppbyggda så, att kodare arbetar på en nivå med text som är en kombination av logiska utsagor, siffror och engelska som sedan maskinöversätts till binär kod. För att programmet ska fungera måste man följa syntaxen till punkt och pricka. Programmeringsspråk finns det väldigt många av och de fungerar på olika nivåer. De innehåller alla var sin uppsättning med möjliga funktioner och operatorer. Ofta finns boolesk logik inbakad: om detta och detta kriterium uppfylls gör detta, om inte gå vidare så och så. De olika programmeringsspråken har sina för användaren dolda strukturer, som också

påverkar vad man kan göra på sin dator och vad som inte är möjligt. De påverkar vad vi ser på vår skärm och hur vi kan interagera med innehållet.

Datorernas språk har var sin egen vokabulär, olika funktioner och sina egna begränsningar. Eftersom det i grunden handlar om att formulera entydiga kommandon och processer för en maskin, är de i sig mer logiska och transparenta än de vaga och undflyende betydelserna och strukturerna i de naturliga språken. Men transparenta är de och deras fulla betydelse för vår kultur endast i teorin. Det skulle krävas väldigt mycket forskning i mjukvara för att egentligen kunna få syn på hur datorkoden formar vår kultur.

Publicerandets ekonomi

Det finns olika former av publicering och grader av offentlighet i dag, liksom det funnits förr. Muntlig kommunikation har alltid varit den flyktigaste men också den billigaste av de språkliga kommunikationskanalerna. Tidigare, då de andra kanalerna var få och långsamma, fungerade brev ofta som halvvoffentlig kommunikation. Brev kunde cirkulera bland familjemedlemmar, vänner och bekanta eller läsas högt, till och med i kyrkan, om de var av särskild vikt. Även mer hemliga och privata brev och skriftliga meddelanden har förstås också behövts. Tillgången till skriven text kunde man väl ha, fast man inte själv var läs- eller skrivkunnig. Högläsning var ett viktigt medium, det var en form av publicerande och offentliggörande på ställen där man hade åhörare till stades.

Vad menar vi med publicering? Före pressens tidevarv, under en epok då all text var antingen handskriven eller huggen i sten (alternativt broderad, målad eller dylikt) var begreppets innebörd annorlunda än i dag. Det kunde ha betytt att en text blivit uppläst på en offentlig plats inför en folksamling till exempel på ett torg eller i en kyrka. Om man med publicerad avser att vem som helst åtminstone i teorin haft tillgång till viss skriven text, innebär det vid denna tid ändå inte att särskilt många människor faktiskt tagit del av just exakt den texten. Stora delar av litteraturen faller då egentligen utanför kategorin "publicerad", inte ens Bibeln var väl i så fall egentligen "publicerad" under en väldigt lång tid.

Det är kanske fruktbarare att i stället fokusera på spridningen av en text, alltså på hur utbredd och använd den faktiskt varit. Detta är å andra sidan praktiskt taget omöjligt att veta med säkerhet. Ibland är det möjligt för en insatt historiker att bedöma huruvida en text varit allmänt känd eller inte. Sist och slutligen hör hela begreppet publicering i sin moderna betydelse ändå tydligt ihop med den tryckta texten: en publicerad text har under lång tid i praktiken betytt att den är tryckt antingen i en tidning eller på andra ark, kanske bunden till en eller i en bok. En publicerad, skriven text har varit tillgänglig för "vem som helst", den har varit "offentlig". Uppfattningen har varit så etablerad att till och med upphovsrätten delvis baserar sig på tidpunkten för publicering, i praktiken året för tryckning.

Den gamla betydelsen av publicera, offentliggöra, betecknas av Svenska Akademiens ordbok som föråldrad, markerad med ett kors. De nya betydelserna av ordet är mer kopplade till tryck, även om alternativet "i skrift" erbjuds. I praktiken betyder det ändå, sedan 1700-talet, att en text blivit tryckt och distribuerad. Denna definition av publicering betyder att det faktiskt har funnits svåra hinder att övervinna och höga trösklar att komma över innan man kunnat få en text publicerad. De har behövts pengar för tryckning och någon som finansierar det hela om man inte själv varit stadd vid kassa. Det har behövts professionella sättare och tryckare, senare också förläggare och redaktörer. Ofta har det behövts tillstånd för tryckning av censor eller universitet, av kyrka eller krona. Distributionskanaler har inte heller varit alla förunnade. Man har helt klart varit beroende av att ha både pengar och kontakter för att ha en chans att kommunicera genom detta medium.

En stor och tidig gallring var alltså en nödvändighet i den gamla textekonomin. Förlagen som så småningom etablerade sig som en viktig mellanhand och finansiär var i Sverige och Finland rätt nära kopplade till tryckerierna ännu på 1800-talet och i vissa fall ända in på 1900-talet.

Förläggarna var på ständig jakt efter storsäljare som gav utsikter att få intäkter och minimerade oftast risktagning.¹⁶

Lagerhållningen har också utgjort en viktig begränsning, som inneburit en mycket stor risk och kostnad för bokhandlare. Därför utvecklades också olika former av förhandsprenumerationer, något som i Norden särskilt kom att bli avgörande för dagstidningarna, som i motsats till pressen i många andra länder inte baserar sig på försäljning av lösnummer. I den digitala världen förändras situationen dramatiskt. Eftersom kostnaderna för publicering är nästan obefintliga och tillgången till olika nivåer av offentlighet i praktiken obegränsad kan vem som helst publicera sig. Kostnaderna för lagring av en enskild text är också så marginell att den gärna finansieras av företag som ett sätt att få synlighet och tillgång till information om användare.

Ämnet har behandlats rätt ingående av Chris Anderson i boken *The long tail*.¹⁷ Han visar hur webben skapar en situation där en nästan oändlig mångfald skapar en lång "svans" av marginella texter som ändå fyller en funktion. Utöver de bästsäljande titlarna finns i dag utrymme för en i praktiken oändlig mängd mindre säljande texter eller andra verk, sådana som tidigare i den fysiska världen helt fallit utanför och därför aldrig nått ens publicering. Det hela bildar i stort sett en exponentiell kurva, vilket gör att "den långa svansen" sammanlagt kan inbringa ett företag som iTunes lika mycket inkomster som toppen, det vill säga de enda skivor som tidigare platsade i en TopTen-skivaffär. Chris Anderson sammanfattar svansen och dess betydelse ungefär såhär (något förenklat):

¹⁶ Om bokbranschen i Finland finns två mycket bra undersökningar: Tuija Laine, *Kolportöörejä ja kirjakauppiaita. Kirjojen hankinta Suomessa vuoteen 1800*. Suomalaisen Kirjallisuuden Seuran toimituksia 1098. (Helsinki 2006) och Jyrki Hakapää, *Kirjan tie lukijalle. Kirjakaupan vakiintuminen Suomessa 1740-1860*. Suomalaisen Kirjallisuuden Seuran toimituksia 1160 (Helsinki 2008). Om förhållandet mellan författare och förläggare på 1800-talet kan man läsa i Pia Forssell, *Författaren, förläggarna och forskarna. J.L. Runeberg och utgivningshistorien i Sverige och Finland*. Skrifter utgivna av Svenska litteratursällskapet i Finland 726. (Helsingfors 2009).

¹⁷ Chris Anderson, *The long tail. Why the future of business is selling less of more*. Hyperion books (New York 2006). Grundidén presenterade han i en artikel med samma namn i tidskriften *Wired* i oktober 2004 <http://www.wired.com/wired/archive/12.10/tail.html> (26.6.2013).

- På varje marknad finns det mycket mer utrymme för nischade varor än för storsäljare. Detta förverkligas exponentiellt i takt med att produktionskostnaderna blir billigare och allmänt tillgängliga.
- Kostnaden för att nå dessa nischer har sjunkit drastiskt. Det beror på att systemen för digital distribution förbättrats: sänkta lagringskostnader, billigare teknologi, tillräckligt utbredda bredbandsförbindelser etc. Det är nu möjligt att erbjuda ett enormt mycket bredare sortiment än någonsin förut.
- Att bara ändra utbudet räcker inte till för att ändra efterfrågan. Konsumenterna måste också kunna hitta sina nischer. Det behövs därför system för rekommendationer, rankning och dylikt för att kunderna skall hitta fram till svansen. Ju bättre och mer utbredda sådana verktyg blir, desto fler och snävare nischer kan uppstå. Anderson kallar verktygen för "filter".
- Försäljnings- eller konsumtionskurvan kommer allt närmare en traditionell exponentialkurva ju bättre dessa filter fungerar och alla "konstgjorda" hinder som utgjorts av knapphet försvinner.

Särskilt svårt har det varit för musik- och underhållningsbranschen med de nya medierna. Musikindustrins försäljning har rasat i USA från 800 miljoner år 2000 till 500 miljoner dollar 2007. Vinsterna beräknades sjunka kring en tredjedel mellan åren 2000 och 2012. Inom gutenbergsfären är det tidningarna som råkat i det värsta blåsvädet, både upplagor och annonsintäkter har sjunkit. Annonspriserna är mycket lägre på webben än på papper. Fast tryckning och distribution åtminstone vad gäller *Hufvudstadsbladet* i Finland utgör nästan 80 % av kostnaderna; ändå är det den redaktionella sidan man många gånger försöker skära ner på eller effektivisera. I USA beräknas 80 % av tidningarnas intäkter komma från annonser, som därmed nästan liksom kommersiell radio och tv är annonsfinansierad. Tidningspressen har hittills haft svårt att riktigt hitta sin nya form i den nya ekonomin, även om den knappast kommer att försvinna helt. Det stora problemet för tidningarna är ändå det korta formatet

som fungerar minst lika bra på webben, där folk nödvändigtvis inte är benägna att betala för tillgång till nyhetsnotiser. Spärrar för avgifter minskar lätt synligheten och får läsarsiffrorna att bokstavligen krascha (siffror som 90 % har förekommit) även om man vid *Helsingin Sanomat* haft positiva erfarenheter och så kallade betalmurar blir allt vanligare. Att utveckla det digitala innehållet blir viktigt för att kunna införa avgifter på innehållet. Dessutom finns sannolikt en växande efterfrågan på både databasjournalistik och längre texter av hög kvalitet.¹⁸

För forskare som publicerar sig i text är det av en viss relevans att studera hur den kommersiella journalistiken anpassar sig eftersom utvecklingen kan förebåda hur det går för bokförlag och spridningen av text. En enorm debatt väcktes av Clay Shirky genom blogginlägget om en förestående tidningsdöd "Newspapers and thinking the unthinkable" där hans budskap gick ut på att vi behöver journalism men inte tidningar. Shirky har försökt nyansera problematiken genom att peka på den samverkan som faktiskt finns mellan så kallad gammel- och nymedia. Hans målsättning har egentligen varit att frigöra journalismbegreppet från de traditionella medierna, vilket han bland annat gjort genom att diskutera källskydd. I Finland har hovrätten ansett att källskydd kan

¹⁸ Ken Auletta: *Googled. The end of the world as we know it*. London 2009, s. 230–231; Robert MacMillan, U.S. "Newspaper circulation plunge accelerates", Reuters 26.10.2009, <http://www.reuters.com/article/idUSN2633378520091026>; Malte Thustrup och Victor Bendz, "Livet efter döden: Svenska tidningar anpassar sig till en digital framtid", Mittuniversitetet, Fakulteten för naturvetenskap, teknik och medier, Institutionen för informationsteknologi och medier (2012), <http://urn.kb.se/resolve?urn=urn:nbn:se:miun:diva-18001>; Panu Moilanen, *Seitsenpäiväiset sanomalehdet verkossa. Lehtien näkemyksiä toiminnastaa verkossa. Tuotteistaan ja asemastaan*. Tietojenkäsittelytieteiden julkaisuja. Tutkimuksia TU-23. (Jyväskylä, 2004). "Päätoimittaja Pentikäinen: Hesarin maksumuuri toimii juuri niin kuin toivottiin" FNB 3.4.2013, *Kauppalehti*, <http://www.aamulehti.fi/Talous/1194805368699/artikkeli/paatoimittaja+pentikainen+hesari+n+maksumuuri+toimii+juuri+niin+kuin+toivottiin.html>; Dan Sabbagh, "Times paywall: numbers are out (should we charge for this?)" Beehivecity 2010, <http://web.archive.org/web/20100721121922/http://www.beehivecity.com/newspapers/times-paywall-the-numbers-on-the-street-should-we-charge-for-this180712/> (samtliga hämtade 26.6.2013).

utsträckas till bloggare. Det innebär att journalistiskt publicerande inte längre måste definieras genom publicering i traditionella medier.¹⁹

Hittills har det varit något lättare för bokbranschen, trots att försäljningen av e-böcker för många av bästsäljarna på engelska redan överstiger försäljningen av tryckta böcker i USA.²⁰ Förlagen har ändå lyckats fylla sin funktion som förmedlare och kvalitetsgranskare. De etablerade förlagen kommer alltså sannolikt allt mer att satsa på bästsäljarna, medan det å andra sidan skapas möjligheter för nischade småförlag att leta fram och föra fram speciallitteratur av olika slag. Dessutom kommer print-on-demand och e-böcker att göra det mycket förmånligare för alla aktörer inom bokbranschen, eftersom man slipper risktagning gällande tryckning och distributionskostnader. Men i dag erbjuds alternativa kanaler att publicera och distribuera också längre texter. Det är rätt enkelt och förmånligt att ge ut böcker på eget förlag både som traditionella böcker och som e-publikationer. Det betyder förstås att det också kommer ut avsevärt mycket mer böcker av sämre eller riktigt usel kvalitet. Bredden blir allt större och det blir också allt viktigare att hitta i textmängderna.

Den i genomsnitt "försämrade kvaliteten", som ger alla blommor möjlighet att blomma, skapar förutsättningar för allt fler subkulturer, som dessutom kan vara beroende av språkkunskaper snarare än geografi, och gör att en begåvad författare eller en bra text som av olika skäl inte blivit accepterad av förlag, med god tur kan bli berömd och få stor spridning så att säga förbi de traditionella kanalerna.²¹ Samtidigt har vi hamnat i en situation där man betalar för den tekniska tillgången till information, men mer sällan betalar för själva innehållet. Läget är något absurt, men hur framtiden kommer att te sig är för tidigt att säga.

¹⁹ Clay Shirky: Newspapers and thinking the unthinkable, Clay Shirky 13.3.2009 <http://www.shirky.com/weblog/2009/03/newspapers-and-thinking-the-unthinkable/> . Lähdesuoja suojelee Kasvia – Hovioikeus kumosi uhkasakon. Yle 2.7.2010, http://yle.fi/uutiset/kotimaa/2010/07/lahdesuoja_suojelee_kasvia_hovioikeus_kumosi_uhkaskon_1803521.html.

²⁰ Frederic Lardinois: Amazon Kindlebooks now outsell printbestsellers 2 to 1. *Read Write Web* 25.10.2010, http://www.readwriteweb.com/archives/amazon_kindle_books_now_outsell_print_best sellers_t wo_to_one.php.

²¹ Nassim Nicholas Taleb skriver om ett sådant fiktivt fall i *The black swan. The impact of the highly improbable* [2007] (London, 2 uppl. 2010).

Inom det vetenskapliga publicerandet finns det dessutom egna mekanismer och behov. De ofta mycket lönsamma stora vetenskapliga förlagen möts av en ökande konkurrens genom alternativa publiceringssystem. Det digitala publicerandet möjliggör dessutom dokumentation av hela forskningsprocessen, länkning, flexibilitet och transparens på sätt som kommer att - och på vissa områden redan har - revolutionerat forskningen. Vi kommer att behandla detta i ett senare kapitel, men i detta skede är det viktigt att understryka att de ekonomiska konsekvenserna av digitaliseringen är av genomgripande karaktär för vår kultur. Historien och historikerna står på inget sätt utanför detta.

Den ökande mängden digital information

Det väsentliga är förstås att det vid sidan av den tryckta texten uppkommit en digital textmassa som dessutom växer närmast exponentiellt. Webben i sig själv är oöverskådlig, ingen vet exakt hur den ser ut eller är strukturerad. Eller som Luke Tredinnick uttryckt saken:

The telephone, television and radio all resulted in a decline in the importance of writing in the transmission of knowledge and information. Participation in culture became less a matter of mastery over a new oral discourse. If this is the case, then the digital age can perhaps be understood as inaugurating a secondary literacy. As digital technologies become more central to culture, writing has reasserted itself as the dominant mode of knowledge and information transmission. Through e-mail, wikis, blogs, social networking and instant messaging, text is becoming integral to culture, work and social relationships. We put text to more and different uses than ever before, distorting it towards communicative ends in more inventive ways. But this re-emergence of writing has brought with it a challenge to the stability of textual artefacts. Text has become a more mutable and malleable medium, torn from the control of the printing houses and publishers. As we move to a secondary literacy, so our assumptions about what text represents become less secure.²²

I dag är det fler människor än någonsin förut som producerar text och som också lägger ut text på nätet. I och med att textmängderna ökat lavinartat och publicering är möjligt för så många människor har den kulturella mekanismen

²² Luke Tredinnick: *Digital Information Culture: The individual and society in the digital age*. Chadnos Publishing. (Oxford 2008) s. 59.

kring publicerande och konsumtion av text förändrats. Själva det normaliserade skriftspråket har fallit sönder i oändliga genrer och stilar, inte minst tack vare kreativa och fördomsfria ungdomar, så kallade digitala infödingar, som hämningslöst testar de nya mediernas gränser. Också kommunikation med bilder integreras smidigt i kommunikationen.

Som Shirky uttryckt det: förr gällde "gallra först, publicera sedan", som nu har ersatts av "publicera först, gallra sedan".²³ Det är lätt att himla sig över alla för en själv meningslösa bloggar, men som Shirky påpekar är de knappast meningslösa för dem som skriver dem och många har ändå sina nischade läsare, som kanske bara är bloggarens närmaste vänner. Tanken på att folk lägger ut väldigt personliga saker på webben är för många som är ovana otrevlig, men man kunde nog på många sätt egentligen jämföra webben med gatan eller ett kafé där man befinner sig i en sorts halvoffentlighet. Skillnaden är förstås att text och bilder på webben finns kvar och om någon är intresserad kan den ofta gräva fram mycket information om en enskild person. Hur illa det är, är en mångfasetterad fråga. De flesta har väl oftast nytta både av att bli hittade och att hitta information om människor på webben.

Men webben förändras också hela tiden och man räknar sällan med att något finns där för evigt. Den digitala texten är föränderlig och den blir per definition aldrig färdig, det finns på sätt och vis varken original, kopior eller slutgiltiga versioner. Den digitala texten på skärmen är alltid endast en representation, ett fönster in i datorns system. Både medier och programvara föråldras och information blir ofta otillgänglig av sådana orsaker. Vi återkommer till denna problematik längre fram. Här är poängen den digitala textens och informationens efemära karaktär; den är instabil och undflyende. I detta läge försvinner också den tydliga roll upphovsmannen haft i vår kultur sedan romantiken. Det finns mycket text i dag vars upphovsman man inte känner till, eller som har flera upphovsmän. Eftersom en text kan leva och ständigt förändras, vilket i många fall är en bra sak – jämför bara ett tryckt uppslagsverk

²³ Clay Shirky, *Here comes everybody. How change happens when people come together*. Penguin. (London 2009) s. 96–108.

med ett på webben – fördunklas betydelse av upphov och texten börjar leva ett eget liv. Roland Barthes utrop om att upphovsmannen är död har blivit ännu tydligare i den digitala världen. Trots allt behövs ändå människor för att skapa nytt och stå för kreativitet och innovationer.

Clay Shirky och andra, såsom Charles Leadbeater, ser väldigt positivt på utvecklingen. De anser att kollektiv kunskap kan ackumuleras på webben, att öppenheten och den generositet som webben möjliggör gynnar mänskligheten och världen. Då vi kan handla utan för stora omkostnader, handlar vi gärna rätt och moraliskt. Det finns också mer cyniska och kritiska röster såsom Jevgenij Mozorov, Jaron Lanier och Nicholas Carr. Mozorov har närmast diskuterat politiska dimensioner, hur webben inte alls är så fri som det lätt ter sig. Lanier och Carr har på olika sätt kritiserat den digitala kulturen. Carr utgår från det faktum att digitala material och webben är för snabb och splittrad för traditionell bildning. Han menar att webben på allvar hotar fördumma mänskligheten. Lanier, som är mer tekniskt bevandrad, framför en mer kunskapsteoretisk kritik och pläderar för att det är bara människor som kan äga kunskap eller vara kreativa. Han påminner oss om att datorerna bara är sofistikerade räknemaskiner och att det i bakgrunden dessutom ofta finns dolda kommersiella intressen.

År 2008 pågick en stor diskussion om den s.k. Googlegenerationen. Det fanns åsikter och forskning om hur i synnerhet ungdomar använder webben. Att använda moderna elektroniska medier skiljer sig givetvis från att skriva på maskin eller för hand, att söka information på ett bibliotek är annorlunda än att söka den på webben, att läsa på datorskärm är annorlunda än att läsa i en bok (liksom det är annorlunda att skriva för webben än för tryck, med det ska vi återkomma till). Liksom tekniska nyheter och nya modetrender brukar, åstadkom förstås också denna förändring förtvivlade svador om dekadens och regeneration, om hur dagens ungdomar inte kan söka information eller orkar läsa böcker. Gör Google oss dumma? frågades det.²⁴

²⁴ Nicholas Carr: Is Google making us stupid? *The Atlantic*. July/August 2008 <http://www.theatlantic.com/magazine/archive/2008/07/is-google-making-us-stupid/6868/> (26.6.2013).

Det typiska webbläsandet är mycket fragmentariskt och sporadiskt jämfört med det monografiläsande vi tidigare sysslat med. Största delen av materialet endast ögnar man igenom och sällan kollar man fler än de tio första sökträffarna i en webbsökning. I stället gör man kanske en ny sökning. Folk är dessutom usla på att använda de finesser som erbjuds vid sökningar, såsom avancerad sökning eller andra avgränsningar. Webbsökningar ska gå i rasande takt, tålmodet är väldigt kort hos de flesta. Nicholas Carr har gått ut som en av de mest kända kritiska rösterna. Han har skrivit om hur förståelsen förflackas, allt blir ytligt och man helt slutar tänka ordentligt då man använder webben för mycket. Vi lär oss sämre då vi läser på datorskärm och vi riskerar enligt honom att tappa förmågan till djupläsning. Carr oroar sig särskilt för hur snabbt man lär sig använda digitala medier och att detta syns i hjärnans strukturer. Frågan man kanske kunde ställa sig är om inte detta är ett helt adekvat sätt att bete sig i en situation med extremt informationsöverflöd.²⁵

Bakom denna oro finns kanske en idé om människan som den dator vars hårdskiva kan bli full eller som kan fungera endast enligt ett givet mönster. En parallell är hur man under Freuds tid gärna kunde likna människans psyke vid den allmänna ångmaskinen. Antagandet om att det att man lär sig ett beteende är det samma som att glömma bort ett annat baserar sig mig veterligen inte på någon psykologisk forskning, utan på ett filosofiskt antagande som låter vackert och logiskt. Det är ändå knappast omöjligt att upprätthålla båda färdigheterna: webbeteende och djupläsning. Det är i själva verket omöjligt att tänka sig en forskare som inte kan tänka i flera led. Men det skulle säkert vara viktigt att man vid planering och praktik inom utbildningen, från dagis till högskola, medvetet skulle odla båda färdigheterna hos unga.

Och visst visar forskning att man läser på ett annat sätt på en vanlig datorskärm – i normala fall, om man inte bestämmer sig för att fokusera och läsa

²⁵ En presentation av Google Generation-forskningsprojektet finns som podcast från den 18.1.2008: <http://www.jisc.ac.uk/podcasts/the-google-generation-myth-or-reality-18-jan-2008> Se även projektets webbplats <http://www.jisc.ac.uk/whatwedo/programmes/resourcediscovery/googlegen.aspx> och Nicholas Carr, *The Shallows: what the Internet is doing to our brains*. (New York 2010). Carrs blogg finns på <http://www.roughlytype.com/> (26.6.2013).

en lång text, vilket nog går, men det kräver lite övning. Internet och Google har blivit ett hjälpmedel bland andra för informationssökning. För många människor är det en ersättning för inget alls (eftersom man mer sällan tog sig tid att gå till biblioteket för att kolla små saker man var osäker på) och innebär alltså en ökad tillgång till information. Men för forskarna blev det ett komplement till litteratur, kataloger och bibliotek. Det blir också ett allt viktigare komplement för forskarna i och med att digitaliseringen av böcker framskrider. Det är onödigt att bestrida det. Indexerade kataloger i all ära, men att helt låta bli att göra fulltextsökningar på webben inklusive digitaliserade böcker vore ju bara dumt. Också för den humanistiska forskningen innebär de stora datamängderna utmaningar som har konsekvenser även för metodfrågor.

Det digitala samhället

Diskussionen om hur livet och världen förändras i och med webben har blivit mindre affekterad och mer substansorienterad och nyanserad med tiden. Överoptimistiska åsikter, som Clay Shirkys mest radikala idéer,²⁶ har tonats ner, samtidigt som de som gått ut med påståenden om att ingenting förändras egentligen har fått se sig överbevisade. Det har visat sig att digital delaktighet inte är självklart, utan att det finns massor av människor som frivilligt eller ofrivilligt ställs utanför delar av eller till och med hela den digitala kulturen. Samhället och kulturen ter sig mer fragmenterade. Webben är oöverskådlig och förutom "objektiv" information finns där mängder av vinklad och falsk information. I dag kan man hitta "bevis" för vad som helst på webben. Uppdelningen mellan den fysiska (den riktiga, analoga, pålitliga) och den virtuella (digitala, osäkra) verkligheten lever fortfarande kvar.²⁷

Mycket av det vi uppfattar som förändringar i kulturen handlar trots allt om att vi tack vare webben blivit medvetna om saker vi tidigare kunnat förbise. Ett exempel kunde vara näthatet, som framställs som om det till sin essens skulle

²⁶ Clay Shirky's *Here comes everybody* och i synnerhet *Cognitive Surplus* genomsyras av en tanke om att *hela* världen kommer att förändras och bli mer demokratisk i rätt snabb takt.

²⁷ Se till exempel Eric Schmidt & Jared Cohen, *The New Digital Age. Reshaping the future of people, nations and business*. (London: John Murray 2013) s. 256-257.

skilja sig från annat hat och mobbning, som varit mer osynliga och lättare att ignorera. Nätverk, som alltid funnits,²⁸ blir synliga, saker som tidigare endast existerat muntligt finns plötsligt i skrift. Utvecklingen löper också parallellt med enhetsspråkens sönderfall, vars tolkningar som antingen en kreativ och positiv process versus dekadens, brukar avslöja rätt mycket om tolkarens personlighet och ideologi. De största konkreta förändringarna är hastigheten med vilka en del sociala processer sker och oberoendet av geografiska avstånd. Falsk information har alltid funnits, men den kan nu spridas snabbare och man är kanske mer medveten om den eftersom den finns öppet åtkomlig i skriven form. Men man bör komma ihåg att också korrekt eller "objektiv" information i dag finns tillgänglig för allt fler.

Socialt och kulturellt engagemang ter sig ofta flyktigare på webben. Man talar pejorativt om gillakultur som om samhällsengagemanget skulle ha minskat.²⁹ Också här kan man ändå fråga sig hur mycket det handlar om verkliga förändringar. I vilken grad har dessa eventuella förändringar sin grundorsak i webben och inte i annan samhällsutveckling? Ensaksrörelserna som fenomen är äldre än tillgången till Internet. Men de är kanske mer ändamålsenliga i ett samhälle fullt av mer information än någon kan ta till sig?

Att folk påstås vara allt större individualister och inte lika ofta vill förbinda sig i till exempel medlemskap eller förtroendeuppdrag i föreningar eller politiska partier, tolkas ibland som en ökande egoism. Massorganisationsprincipen, ett samhällsfenomen som varit otroligt viktigt i det nordiska samhället och för medborgarskapet så som det fungerat under 1900-talet,³⁰ håller på att mattas ut och delvis ersättas av andra former av engagemang. Det moderna engagemanget ser möjligen annorlunda ut, men sociala medier är en nyhet som underlättat engagemang i sig. Att man på webben lätt kan byta fokus från en samhällsfråga till en annan gör också att kunskap och perspektiv kan sprida sig

²⁸ På denna punkt missar Luke Tredinnick, som tolkar nätverken som ett nytt fenomen.

²⁹ Ellinor Skagegård: "Äkta gillande och rent geschäft." UNT 23.9.2012.
<http://www.unt.se/kultur/akta-gillande-och-rent-geschäft-1865737.aspx> (6.7.2013).

³⁰ Henrik Stenius, *Frivilligt-jämligt-samfällt. Föreningsväsendets utveckling i Finland fram till 1900-talets början med speciell hänsyn till massorganisationsprincipens genombrott*. (Helsingfors: SLS 1987)

och utbytas mellan olika ensaksrörelser. Också inom politiken kan vi alltså ana samma sönderfall som på andra håll: de stora koherenta berättelserna får ge vika för en allt mera fragmenterad verklighet. Både poststrukturalismen och postmodernismen är ändå som fenomen äldre än webben. Men webben erbjuder en miljö där dessa teorier kan realiseras på ett synligt sätt.

Clay Shirky har hävdat att den nya informationsteknologin, som är genuint interaktiv, har frigjort enorma resurser bara genom att tränga undan tv-tittandet, som tidigare passiviserade människor för en stor del av deras fritid.³¹ Detta är en potentiell resurs också för forskare. Samtidigt måste man vara medveten om hur mycket formatet och redskapen påverkar informationen redan då den föds. Människor är i allmänhet inte medvetna om hur mycket vi påverkas av exempelvis både verbala och grafiska element då vi svarar på en enkät eller gör en uppgift på dator.

En mycket viktig skillnad för forskaren är att man i och med användningen av komplicerad teknologi inte automatiskt har insyn i de begränsningar eller andra problem den kan medföra. All sorts kritik är därför svårare. Forskaren använder sig av vad som kunde kallas fönster, för att ta del av information, utan att se de underliggande strukturerna, ofta utan att ens vara medveten om dem.³² Trots att vi talar om "digital" information är det vi ser på skärmen analogt. Som forskare måste man vara medveten om att de osynliga, genomskinliga, strukturerna som finns inbäddade i de informationsteknologiska medierna finns där och också påverkar vad vi ser och hur vi ser det. Samtidigt har förändringarna i medierna först i efterhand på sätt och vis synliggjort de gamla mediernas egentliga effekter på kulturen. Det är brytningen och förändringen som gör mediernas kulturella och samhällseliga mekanismer synliga.

³¹ Jfr <http://www.flickr.com/photos/93878419@N00/96808416/> (hämtad 2014-03-04).

³² Colleen A. Reilly, "Teaching *Wikipedia* as a mirrored technology". *First Monday* Vol. 16:2011, <http://firstmonday.org/htbin/cgiwrap/bin/ojs/index.php/fm/article/view/2824/2746> (6.7.2013).

Kapitel 3. Forskarvärldens respons

Kenneth Nyberg

Att möta förändring

Ett ständigt återkommande tema i framställningen hittills har varit frågan om hur djupgående de förändringar som nu pågår egentligen är och i vilken mån de representerar något nytt eller inte. Den frågan, och olika svar på den, har präglat även historikers och andra humanisters respons på den digitala utvecklingen. De som på olika sätt arbetar med kunskapsbildning om det förflutna är mer medvetna än de flesta om hur föränderliga mänskliga samhällen är, och för oss historiker är det en huvuduppgift att undersöka, analysera och förklara just samhällslig förändring. I anslutning till den utveckling vi beskrivit i tidigare avsnitt kan man dock fråga sig om historiker är bättre än många andra på att upptäcka och förstå de förändringsprocesser vi själva lever mitt i, särskilt sådana som kan komma att omskapa hela grunden för det fält vi verkar inom.

Det är ganska uppenbart att den digitala revolutionen är en historisk process av den digniteten, men här liksom i andra sammanhang finns det en tendens att diskussionen domineras av de med starka åsikter i ena eller andra riktningen. Lite tillspetsat finns här å ena sidan de stockkonservativa vilka ser internet som bara ett praktiskt sökverktyg och sociala medier som trams, och å andra sidan de teknikfrälsta vilka tror att digital hård- och mjukvara är lösningen på alla problem och att allt är nytt under solen. Sanningen lär snarare ligga någonstans däremellan: helt uppenbart finns stora möjligheter och fördelar med den nya tekniken som vi bör ta vara på, men den har också risker och nackdelar som det gäller att vara medveten om och försöka undvika.

Vi bör alltså, som alltid, inta ett kritiskt förhållningssätt till den digitala utvecklingen och dess konsekvenser. Att låtsas som att ingenting har hänt, att denna utveckling inte på många sätt innebär en stor utmaning för historisk forskning och för kulturarvssektorn, är emellertid inget alternativ. Dels bör det

mer principiellt vara självklart att vi alltid är öppna för att ta till oss nya metoder, verktyg och perspektiv som kan främja den historiska kunskapsbildningen. Dels är det en fråga om trovärdighet; om vi inte aktivt förhåller oss till utvecklingen i vår samtid, kommer historisk och annan humanistisk forskning snart – och kanske snabbare än vi tror – att uppfattas som irrelevant och det samhälleliga stödet för den därmed att undergrävas. Mer krasst uttryckt kan man också se en tydlig digital profil som en konkurrensfördel i en situation där de ekonomiska förutsättningarna för humanistisk forskning blir allt kärvare.³³

Alla dessa skäl, och många andra som mer specifikt rör de nya möjligheter den digitala tekniken erbjuder inom enskilda forskningsfält, ligger bakom den snabba expansion av forskning med digitala inslag som vi sett på senare tid. Delar av den har rötter som går decennier tillbaka, andra befinner sig fortfarande under formering och är mer direkt kopplade till de allra senaste årens utveckling på webben och i sociala medier. Det finns en mängd olika inriktningar, angreppssätt och strömningar, med mer eller mindre komplicerade inbördes relationer, som skulle kunna vara intressanta att ta upp här. Mest relevant för våra syften är dock det framväxande fält som benämns digital historia (eng. *Digital History*), vilket utgör ramen för många av de avsnitt och kapitel som följer.

Digital historia är i sin tur en del av, eller nära förbunden med, en bredare och i sig mycket mångfasetterad rörelse kallad digital humaniora (*Digital Humanities*). Bådadera, men särskilt den senare, förkortas ofta (i olika sammanhang) "DH", och det är i sig betecknande då gränsen mellan de båda är tämligen porös. Ett av DH:s kännetecken är just tvärvetenskapligheten och mer allmänt uppluckringen av de traditionella disciplinerna och akademiska hierarkierna. För att förstå fältet digital historia och hur det förhåller sig till befintlig historisk forskning är det alltså viktigt att först förstå vad digital humaniora är (definitioner) och vad det handlar om (debatter).

³³ Det bör kanske påpekas att detta är mest påtagligt i de länder (till exempel USA, Storbritannien och Spanien) där mycket stora stora nedskärningar i anslagen till universitet och högskolor och till grundforskning gjorts under senare år.

Digital humaniora: definitioner och debatter

Digital humaniora är på flera sätt ett mångtydigt och flerskiktat begrepp, där allra minst tre olika betydelser kan urskiljas. För det första kan det användas som en rent generell beteckning på humanistisk forskning som på något vis bedrivs med digitala metoder; det är en definition som, beroende på tolkning, kan bli så generös att den omfattar nästan *all* humanistisk vetenskap. Det säger i sig något om den betydelse den digitala utvecklingen redan har fått för oss som humanistiska forskare, utan att de flesta av oss kanske funderat så mycket över konsekvenserna. Samtidigt finns det många aspekter av de konsekvenserna en så vid definition inte fångar; den täcker bredden snarare än djupet – vissa skulle säga radikaliteten – i utvecklingen och dess följder.

För det andra kan alltså digital humaniora också stå som samlingsnamn för ett antal mer eller mindre löst sammanhängande forskningsinriktningar vilka kännetecknas av att de bygger på mer omfattande eller systematisk användning av digitala verktyg och metoder, men där sådan användning i sig inte uppfattas innebära något kvalitativt annorlunda jämfört med annan forskning. För det tredje, slutligen, finns det en mindre men snabbt växande grupp av aktiva utövare vilka själva gärna definierar sig som just digitala humanister snarare än exempelvis historiker eller litteraturvetare. För dem handlar digital humaniora om mer än en samling nya forskningsverktyg; snarare innebär det ett visst (och delvis nytt) förhållningssätt präglad av starkt samhällsengagemang, tvärvetenskaplighet och kritik mot traditionella akademiska strukturer och praktiker i en situation där digitaliseringen på nytt har aktualiserat de grundläggande frågorna om vad kunskap och bildning är.

När vi talar om digital humaniora i det följande syftar vi i första hand på de två senare betydelserna, med viss tyngdpunkt på den sistnämnda, då det är i spänningsfältet mellan dessa två – eller snarare längs det spektrum som de representerar – som en ny "digital historia" håller på att växa fram. Det är också en utgångspunkt som hamnar rätt nära en av de oftast citerade definitionerna av begreppet, den i *Digital Humanities Manifesto* från 2008: "Digital Humanities is

not a unified field but an array of convergent practices". Som helhet är det dock ett oerhört mångfasetterat och föränderligt landskap vi rör oss i, och det som följer kan därför inte bli något annat än en snabbskiss över några drag som för oss tycks särskilt relevanta här.³⁴

Terminologin i sig är, som alltid, belysande. Det som nu alltmer allmänt kallas *Digital Humanities* har, med varierande grad av överlappningar, under det senaste decenniet också gått under namn som *humanities computing*, *the computational turn*, *e-humanities* och *computational humanities*. I sin nuvarande betydelse tycks termen *digital humanities* ha etablerats på allvar i samband med utgivningen av den inflytelserika antologin *A Companion to Digital Humanities* 2004. Ursprungligen skulle den ha hetat *A Companion to Humanities Computing*, men förlaget ansåg att titeln gav ett för snävt tekniskt intryck och en av redaktörerna, John Unsworth, föreslog då *Digital Humanities* istället.³⁵ Som Kathleen Fitzpatrick påpekar: "Digital humanities thus grows specifically out of an attempt to make 'humanities computing,' which sounded as though the emphasis lay on the technology, more palatable to humanists in general."³⁶

³⁴ *A Digital Humanities Manifesto*, <http://manifesto.humanities.ucla.edu/2008/12/15/digital-humanities-manifesto/> (2013-07-21); jfr *The Digital Humanities Manifesto 2.0*, <http://manifesto.humanities.ucla.edu/2009/05/29/the-digital-humanities-manifesto-20/> (2013-07-21). Patrik Svensson, föreståndare för HumLab vid Umeå universitet, publicerade 2009–2010 en artikelserie i fyra delar i *Digital Humanities Quarterly* där han tar ett helhetsgrepp på både fältet och termen *Digital Humanities*. Alla är läsvärda och har citerats flitigt, men se särskilt Patrik Svensson, "The Landscape of Digital Humanities", *Digital Humanities Quarterly* 4:1 (2010), <http://digitalhumanities.org/dhq/vol/4/1/000080/000080.html> (2013-07-10).

³⁵ Susan Schreibman, Ray Siemens och John Unsworth (red.), *A Companion to Digital Humanities* (Oxford: Blackwell, 2004), <http://www.digitalhumanities.org/companion/> (2013-07-10). Historien om hur titeln kom till publicerades först i Matthew G. Kirschenbaum, "What Is Digital Humanities and What's It Doing in English Departments?", *ADE Bulletin* 150 (2010), <http://mkirschenbaum.files.wordpress.com/2011/03/ade-final.pdf> (PDF, hämtad 2013-07-10). Angående *the computational turn* se David M. Berry (red.), *Understanding Digital Humanities* (Houndmills: Palgrave Macmillan, 2012).

³⁶ Kathleen Fitzpatrick, "The Humanities, Done Digitally", *The Chronicle of Higher Education* 2011-05-08, <http://chronicle.com/article/The-Humanities-Done-Digitally/127382/> (2013-07-10). I en av Patrik Svenssons tidigare nämnda artiklar diskuteras relationen mellan dessa två begrepp; se hans "Humanities computing as digital humanities", *Digital Humanities Quarterly* 3:3 (2009), <http://digitalhumanities.org/dhq/vol/3/3/000065/000065.html> (2013-07-10).

Både som kompromiss och som paraplybegrepp står alltså *digital humaniora* för olika betydelser och bär på inneboende konflikter.³⁷ En sådan är den som Fitzpatrick antyder, den i själva uttrycket uppenbara spänningen mellan *digital* respektive *humaniora*, de "tekniska" och de "mänskliga" sidorna. Även om vi begränsar oss till det som ovan definierades som den aktiva DH-rörelsen, finns det en besläktad friktion som ständigt återkommer i försöken att ringa in och definiera vad DH är och inte är. Fitzpatrick igen:

It's clear that there's an overlap between [...] scholars who use digital technologies in studying traditional humanities objects and those who use the methods of the contemporary humanities in studying digital objects—but clear differences lie between them. Those differences often produce significant tension, particularly between those who suggest that digital humanities should always be about making (whether making archives, tools, or new digital methods) and those who argue that it must expand to include interpreting.

The terms of this tension should begin to sound a bit familiar: It's an updated version of the theory-practice divide that has long existed in other quarters of the humanities.³⁸

Dessa spänningar kan ibland ta sig mycket konkreta uttryck i den, trots allt, relativt begränsade miljö som digital humaniora (som inriktning eller rörelse betraktad) ännu utgör. Till exempel uppstod en bitvis mycket frän och infekterad diskussion på bloggar och Twitter i början av 2013 om vem som egentligen "räknas" som digital humanist eller inte. Debatten föranleddes av boken *Digital Humanities*, där ett avsnitt kallat "A Short Guide to the Digital Humanities" bland annat innehöll en auktoritativt formulerad definition av DH vilken många uppfattade som snäv och exkluderande.³⁹ Som Wikipedia

³⁷ Utöver hänvisningarna i övriga noter kan nämnas Jason Hepplers webbsajt *What is Digital Humanities?* (<http://whatisdigitalhumanities.com>, hämtad 2013-07-10). Den innehåller hundratals citat om vad digital humaniora är från aktiva utövare och vid varje besök på sajten väljs en av dessa formuleringar slumpvis ut och presenteras för läsaren.

³⁸ Fitzpatrick, "The Humanities".

³⁹ Anne Burdick et al, *Digital Humanities* (Cambridge och London: MIT Press, 2012), http://mitpress.mit.edu/sites/default/files/titles/content/9780262018470_Open_Access_Edition.pdf (2013-07-10), s. 121–135. Det aktuella avsnittet finns också tillgängligt som separat PDF på http://jeffreyschnapp.com/wp-content/uploads/2013/01/D_H_ShortGuide.pdf (2013-07-10).

konstaterar med hänvisning till antologin *Debates in the Digital Humanities*: "The definition of digital humanities is volatile and is highly contested."⁴⁰

En bidragande faktor till att definitionsfrågorna ibland blir så laddade är utan tvivel, som redan antytts, att det finns en urskiljbar gruppering av mycket synliga digitala humanister med en tämligen uttalat "ideologisk" profil. Om också i varierande mån präglas denna inriktning eller rörelse av tydligt artikulerade ideal om akademien som en lärande gemenskap med egalitära drag, där information ska flöda fritt och där man arbetar tillsammans i en anda av öppenhet, generositet och tolerans. Här finns kopplingar till såväl hackerkulturen som 1970-talets progressiva politiska rörelser, och för dem som har ett sådant synsätt på DH har fältet en demokratisk och emancipatorisk potential. För andra utövare, särskilt de som inte själva identifierar sig som digitala humanister men i stora drag arbetar med liknande verktyg och metoder, är sådana aspekter inte lika självklara och det kan bidra till friktion kring prioriteringar och annat.

Den mer synliga strömning av digital humaniora som här ringats in har i hög grad dominerats av amerikanska forskare, åtminstone fram tills nyligen, medan det som tidigare kallades *humanities computing* har djupa rötter även i flera europeiska länder (inte minst i Norden). Mer generellt har också kritik, både inifrån och utifrån DH-nätverken, riktats mot deras bristande representativitet ifråga om såväl genus som etnicitet; de som syns och hörs mest är, liksom i många andra akademiska sammanhang, framför allt anglosaxiska vita män. Ofta uppfattas det som särskilt problematiskt mot bakgrund av rörelsens annars så "progressiva" och egalitära profil. Kanske är det inte helt och hållet en slump att detsamma kan sägas om den amerikanska teknikindustrin, symboliserad av Silicon Valley, eftersom dess kultur (av ganska uppenbara skäl) åtminstone i vissa avseenden påverkat DH:s framväxt och utveckling.

För ett uppmärksammat blogginlägg med kritisk udd se David Golumbia, "[D]igital humanities: two definitions", *Uncomputing*, <http://www.uncomputing.org/?p=203> (2013-07-10).

⁴⁰ *Wikipedia*, u.o. "Digital Humanities", http://en.wikipedia.org/wiki/Digital_humanities (2013-07-10), och Matthew K. Gold (red.), *Debates in the Digital Humanities* (Minneapolis: University of Minnesota Press, 2012), <http://dhdebates.gc.cuny.edu> (2013-07-10).

Digital historia: en inringning

Liksom digitala humaniora är digital historia (i det följande DH) ett begrepp som ännu saknar en fast definition. I detta sammanhang syftar det emellertid på det fält som ligger i skärningspunkten mellan digital teknologi, sociala medier och vetenskaplig kunskapsbildning om människans förflutna. DH är ett försök till svar på den digitala utmaningen, där nya arbets- och kommunikationsformer för historiker utvecklas utan att de mest grundläggande målen för verksamheten egentligen har ändrats. Vid det amerikanska historikerförbundets årskongress 2012 lanserades DH på bred front, en parallell till hur digitala frågor anses ha fått sitt genombrott vid Modern Language Associations kongress tre år tidigare.⁴¹

Bland dem som driver på utvecklingen inom området kan särskilt nämnas Roy Rosenzweig Center for History and New Media (CHNM) vid George Mason University i Virginia. De stod bakom många av DH-sessionerna vid AHA-kongressen 2012, driver ett antal webbplatser och har utvecklat resurser för både forskning och undervisning med koppling till digital historia (i vid mening).⁴² Dess tidigare föreståndare Dan Cohen, som i april 2013 utsågs till chef för det nystartade Digital Public Library of America, är en förgrundsgestalt inom digitala humaniora mer generellt, och han spelar en nyckelroll inom nätverken på området. Andra lärosäten med tydlig DH-profil är University of Nebraska-Lincoln, University of Virginia och Stanford University.

Det som ibland kallats den digitala vändningen (eng. *the digital turn*) inom historia och humaniora lär bli av betydligt mer genomgripande natur än både

⁴¹ American Historical Association, "126th Annual Meeting", <http://www.historians.org/annual/2012/index.cfm> (2013-01-14); om MLA 2009 se Matthew G. Kirschenbaum, "What Is Digital Humanities and What's It Doing in English Departments?", *ADE Bulletin* 150 (2010), <http://mkirschenbaum.files.wordpress.com/2011/03/ade-final.pdf> (PDF, hämtad 2013-07-10).

⁴² CHNM:s webbplats nås på adressen <http://chnm.gmu.edu>. Exempel på resurser som tagits fram vid CHNM är referensverktyget Zotero (<http://www.zotero.org>), webbutställningsplattformen Omeka (<http://www.omeka.net>), forskarportalen *Digital Humanities Now* (<http://digitalhumanitiesnow.org>) och lärarportalen *Teachinghistory.org* (<http://teachinghistory.org>).

språkliga, narrativa och andra "vändningar" under de senaste årtiondena. Den parallell som bör dras är snarare till det paradigmskifte som ägde rum på 1960- och 70-talen när de kvantitativa metoderna gjorde sitt intåg i historieämnet. Det var en utveckling som knöt an till större samhällsförändringar och som radikalt ändrade fältet men inte helt ersatte de kvalitativa metoder som funnits tidigare. De digitala metoderna innebär på liknande sätt ett nytt inslag eller dimension i forskarnas arbete, men likväl inte en förändring som kullkastar tidigare verksamhet utan en (kraftfull) vidareutveckling av något redan befintligt.

William G. Thomas har kortfattat skisserat några drag i den historia som lett fram till dagens situation. De stora socialhistoriska projekten runt 1970, med deras användning av kvantitativa metoder och datorer, spelar där en viktig roll, och 1986 bildades The Association for History and Computing (AHC) med bas vid University of London. Under 1990-talets lopp försköts tyngdpunkten från (stor)datorer och deras betydelse som *computers* i bokstavlig mening, alltså räknemaskiner, till den framväxande webben som arbetsverktyg och kommunikationskanal. Ett pionjärprojekt i det avseendet var Edward L. Ayers *Valley of the Shadow*, ett digitalt arkiv rörande det amerikanska inbördeskriget som lades ut på nätet med början 1993. Det var en av de allra första webbplatserna över huvud taget och erbjöd med teknikens hjälp komparativa och interaktiva drag som inte hade varit möjliga att åstadkomma i text eller andra traditionella media. Sedan dess har internet i allmänhet och webben i synnerhet blivit allt viktigare på allt fler sätt för forskarna.⁴³

I Sverige har historiker ännu knappt börjat diskutera dessa förändringar och deras konsekvenser för forskning, undervisning och populärvetenskapligt arbete, åtminstone inte på något mer systematiskt vis. Det finns enstaka projekt som tydligt kan knytas till "digital historia", till exempel forskningsprogrammet [Early Monasticism and Classical Paideia](#) (MOPAI) i Lund som bland annat syftar till att med hjälp av databaser möjliggöra jämförelser och analyser av tusentals äldre texter och textfragment på språk som arabiska, grekiska och latin. I vidare

⁴³ William G. Thomas, III, "Computing and the Historical Imagination", i Schreibman et al, s. 59–62. Webbplatsen för *Valley of the Shadow* finns på <http://valley.lib.virginia.edu>.

mening, alltså digitala humaniora snarare än historia, finns också ett internationellt framstående tvärdisciplinärt centrum i Umeå, HUMlab, som är "en mötesplats mellan humaniora, kultur och informationsteknik". I sammanhanget kan också nämnas miljöer som Humanities Lab i Lund, ett resurscentrum för tvärvetenskaplig forskning med IT-inslag, eller projekt som Språkbanken och Litteraturbanken i Göteborg.⁴⁴

I Finland tycks situationen vara likartad, med den skillnaden att någon motsvarighet till HUMlab egentligen inte finns. Hittills har det som här kallats digitala humaniora resp. historia mest kommit till uttryck i form av enskilda projekt. Vid Humanistiska fakulteten vid Åbo Akademi pågår dock arbete med att ta ett större helhetsgrepp på digitala humaniora, bland annat i form av planer på *digital humanities* som ett nytt biämne från hösten 2014 i samverkan mellan humanister och datavetare. Såväl för Sveriges och Finlands del som övriga Nordens (och här får de som vet mer om läget i Danmark, Norge och Island gärna komplettera eller korrigera) tycks det alltså rimligt att anta, att det kommer hända mycket de närmaste åren samtidigt som digital historia ännu i högsta grad är i sin linda.

Betydelse(r) för forskningen

På ett övergripande plan sammanfattar man ibland den digitala utvecklingens effekter på forskningen med begreppet *big data*. Med det menas inte i första hand stora datamaskiner utan stora datamängder. Begreppet används med skiftande innebörd i olika sammanhang, men det syftar på att det som präglar vår digitala era är – ännu en gång – det gigantiska och ständigt växande flödet av data. Det kan ses som ett problem: Hur ska vi välja? Hur ska vi orientera oss? Men datamängderna erbjuder också möjligheter att få fram mönster, och därmed ställa frågor, som inte varit möjliga förut. Och vi kan göra det genom att vi har tillgång till alltmer kraftfulla datorer som kan bearbeta materialet. Det är

⁴⁴ HUMlab, <http://humlab.umu.se/sv/om-oss/vaar-verksamhet/>; Lund Humanities Lab, <http://www.humlab.lu.se>; Språkbanken, <http://spraakbanken.gu.se>; och Litteraturbanken, <http://litteraturbanken.se> (samtliga hämtade 2013-07-11).

dessa förhållanden som, på olika sätt, ligger till grund för alla de digitala metoder som behandlas i kommande kapitel.

Vi kan se en föregångare till dagens digitala historia i den dataintensiva och kvantitativt inriktade socialhistoria som fick ett uppsving under 1960-talets lopp och åren därefter. Det är kanske inte så förvånande att den typen av metoder först användes i befolkningsstudier och annan forskning där grundmaterialet består av siffror. Skillnaden är att detta, i någon mening, kvantitativa förhållningssätt nu tillämpas även på frågor som ofta uppfattas som kvalitativa till sin karaktär, det vill säga vi använder data till att visa på mönster i kvaliteter, egenskaper, snarare än i antal och populationer.

Därmed berör vi också en annan tendens i tiden, som inte bara fått genomslag inom digital historia utan mer generellt på en rad olika historiska forskningsfält, nämligen intresset för *spatiala* (rumsliga) perspektiv. När vi ska återge de kvalitativa mönstren gör vi det nämligen gärna, med stöd av de stora datamängderna och kraftfulla datorprogram, genom *visualiseringar* som visar hur det vi undersökt är distribuerat i rummet. Visualiseringar är, i sin tur, ytterligare en av de stora trenderna vi kan se inom digital historia och digital humaniora i övrigt. Det är, potentiellt sett, en mer radikal förändring av vad forskning handlar om än man först kan tro och vi återkommer därför till det längre fram.

Mer konkret tar sig dessa stora strömningar uttryck i två metodologiska huvudspår inom digital humaniora: *text mining* (TM) och *datavisualisering*. Det första handlar om hur man med hjälp av såväl kvalitativ som kvantitativ databehandling kan analysera stora mängder text i historiska källor (sedan den digitaliserats). Det kan vara en fruktbar metod, men det är viktigt att fundera på vad sådana analyser egentligen säger om djupare betydelser eller större historiska sammanhang och inte bara om orden eller tecknen i sig.⁴⁵

Det andra temat, som med fördel kan kombineras med det första, handlar om hur olika data och deras samband kan framställas och analyseras i visuell form

⁴⁵ Se t.ex. Jessica Parland-von Essen, "Ord och betydelser", *Essetter* 2012-01-08, <http://essetter.blogspot.com/2012/01/ord-och-betydelser.html> (2012-02-23).

och därmed möjliggöra resultat och slutsatser som inte fullt ut kan gestaltas i textform. Ett uppenbart exempel på datavisualisering är historiska kartor, men möjligheterna går mycket längre än så eftersom det egentligen rör sig om alla slags spatialiserade databaser där olika parametrar kan knytas till rumsliga mönster. Den typen av visualisering blir särskilt värdefull när man ska försöka åskådliggöra och undersöka mycket komplexa företeelser med relationer och/eller strukturer i tid eller rum (även tidslinjer är ju en form av visualisering).

Den digitala utvecklingen har alltså gett upphov till en mängd nya möjligheter i fråga om material och metoder för historiker. I vissa fall har dessa nya möjligheter främst betytt att "gamla" frågor kunnat besvaras snabbare eller enklare med nya metoder, men tekniken har också öppnat vägen för utforskning av nya frågeställningar eller aspekter av historien. Det har skett på flera sätt som det kan vara intressant att titta närmare på, inte minst med tanke på den tidigare refererade diskussionen om vad och hur mycket som egentligen är nytt med den digitala tekniken.

För det första finns det ett gränsland mellan vad som bara är mer effektiva sätt att besvara en gammal fråga och helt nya studier som inte kunnat genomföras utan den digitala tekniken. Det gäller främst vissa beräkningsintensiva, kvantitativa studier inom befolknings-, social- eller agrarhistoria vilka hade varit teoretiskt möjliga men i praktiken orimliga att genomföra utan kraftfulla datorer. Detsamma kan sägas om de resultat som *text mining* leder till inom lingvistik och litteraturvetenskap, där själva skalan på dagens textdatabaser och de verktyg som används för att analysera dem innebär helt nya forskningsmöjligheter. En del av den kunskap vi idag har om dessa aspekter är därför en direkt följd av den tekniska utvecklingen.

För det andra innebär vissa av de nya metoderna ett kvalitativt annorlunda sätt att både studera, rekonstruera och gestalta historiska skeenden och strukturer. Visualiseringar är kanske det tydligaste exemplet på det, särskilt de som erbjuder ett mått av interaktivitet som traditionella vetenskapliga texter inte förmår. Rumsliga relationer, förändring över tid, snabba växlingar mellan överblick och detaljgranskning blir på ett helt annat sätt tillgängliga för såväl

forskare som lekmän genom användningen av multimediala verktyg vilka i sin tur bygger på digital teknik – och ofta enkel åtkomst via webben.

För det tredje ger den digitala utvecklingen i sig upphov till nya studieobjekt för historikerna. Dels kan det handla om helt nya fenomen, som de aspekter av mänskligt liv som nu utspelas på webben eller andra delar av Internet; de har redan gett upphov till en ny disciplin som kallas *web science*, webbvetenskap, men kommer efterhand att bli allt viktigare även för historiker.⁴⁶ Dels kan det också handla om nya sammanhang där befintliga fenomen kan studeras, till exempel hur historiebruk och förhållningssätt till historisk kunskap kommer till uttryck i dataspel eller sociala medier. Sådana miljöer är minst lika intressanta för den typen av studier som dagstidningar och läroböcker traditionellt har varit.

Det kan med andra ord inte råda någon tvekan om att digital historia erbjuder mycket av värde för dagens och morgondagens historiker. Samtidigt är det viktigt att alltid vara medveten om varför man väljer en viss form eller en viss metod. Varken traditionella eller nya arbetssätt har ett egenvärde, utan de lämpar sig mer eller mindre väl för olika syften. Hur spektakulära olika visualiseringar än är, eller hur lockande andra digitala metoder än må vara, gäller fortfarande kraven på stringens och kritiskt förhållningssätt. I fråga om transparens och kontrollerbarhet hos källor har dessa krav snarare ökat, i och med att åberopat material numera ofta går att länka till direkt, i digital form och omedelbart tillgängligt. De nya verktygen bygger alltså vidare på de befintliga och kompletterar dem, de ersätter dem inte.⁴⁷

⁴⁶ Se Web Science Trust, <http://webscience.org>.

⁴⁷ Jfr Kenneth Nyberg, "Om digital historia efter Chicago", *Tidens skiften* 2012-01-13, <http://tidensskiften.se/2012/01/13/om-digital-historia-efter-chicago/> (2013-07-09).

Kapitel 4. Historieforskning i den digitaliserade världen

Jessica Parland-von Essen och Kenneth Nyberg

Betydelser av digitalisering

Digitaliseringen av själva forskningsarbetet har flera olika aspekter och dess genomslag inom historieämnet består därför av flera nivåer. Den första utgörs av historikerns *vardagsverktyg* och handlar om användning av datorer för ordbehandling, e-postkommunikation och informationssökning via webben. Den andra nivån kretsar kring *material*, dels omvandling av befintliga källor i digital form och dels användning av material som från början är digitalt. Ett tredje steg rör digitala *metoder*, där det hänt oerhört mycket på senare år. Vissa av dem innebär att datorer används för krävande beräkningar som inte tillför något kvalitativt nytt i metodväg, medan andra kan ge resultat som tidigare inte varit möjliga att få fram. De sistnämnda gör att forskarna kan svara på helt nya *frågor*, vilket kan betraktas som en fjärde dimension av digitaliseringen. Till samma kategori, nya frågor och ny kunskap, räknas också det akademiska studiet – med digitala metoder eller andra – av det digitala samhället i sig. Möjligen bör man också som en helt egen, och i så fall femte och sista, aspekt se de nya former för *kunskapsspridning* och samverkan med det omgivande samhället som digital teknik öppnar upp.

För historiker har den digitala utvecklingen också betytt en grundläggande förskjutning i fråga om tillgången på material: från brist till överflöd. Det gäller särskilt för dokumentationen av vår egen tid, då det produceras oöverskådliga mängder data för varje dag som går. Även för tidigare perioder – där det kanske finns relativt få källor bevarade – har dock digitaliseringen betytt att det material som finns kvar blir tillgängligt för forskningen på ett helt annat sätt. På grund av att sökmotorderna i dag är annorlunda påverkas också avgränsningar och urval av material, och därigenom forskningsmetoderna.

Både forskningsobjekt och arbetsmetoder redan har alltså förändrats på många sätt för de allra flesta historiker. I detta kapitel behandlas främst digitala material som källmaterial för historiker och de övriga aspekterna i de följande kapitlen.

Det finns två typer av digitala källmaterial vi som forskare ställs inför: de som är digitaliserade versioner av material som har eller har haft ett annat original och de som är födda i digitalt format och som således saknar original i egentlig betydelse, eftersom de i praktiken skapas på nytt varje gång man tar del av dem.⁴⁸ Den senare typen av material är ny och behandlas mer ingående i följande kapitel. Vi kallar denna typ av material digitalbaserade och den förstnämnda typen digitaliserade.

Både privata och offentliga aktörer har föresatt sig att digitalisera maximala mängder litteratur, till och med "allt som någonsin tryckts". Digitalisering kan ändå i praktiken betyda väldigt olika saker och att något finns digitaliserat behöver naturligtvis inte innebära att det finns fritt tillgängligt på nätet. Det beror inte minst på upphovsrättslagstiftningen, men det finns också andra ekonomiska orsaker till att hålla digitaliserade material bakom licens. Digitalisering är fortfarande dyrt och en stor investering, trots att kostnaderna sjunkit drastiskt i takt med att kvaliteten har blivit bättre. Förutom texter digitaliseras också andra typer av material.

I gemen kan man säga att fördelen den enkla digitaliseringen tillför är stor. Den första och för många forskare största nyttan är den dramatiskt förbättrade sökbarheten, som dock också får metodiska konsekvenser. Vi kommer därför att ägna en del utrymme åt frågor kring sökning och de informationsstrukturer som påverkar sökning och resultaten vid sökning i källmaterial.

⁴⁸ Matthew Kirschenbaum, "The .txtual Condition: Digital Humanities, Born-Digital Archives, and the Future Library" *Digital Humanities Quarterly* 7:1 (2013) <http://www.digitalhumanities.org/dhq/vol/7/1/000151/000151.html> (13.08.2013).

Att hitta bland material

Jessica Parland-von Essen

Ett grundbegrepp som är viktigt att känna till är metadata, som brukar definieras som information om data, vilket är en ganska klumpig definition. Enklast kan man exemplifiera vad det handlar om med en katalogpost i en bibliotekskatalog: den innehåller metadata. Metadata kan antingen beskriva *innehållet* i ett dokument eller en fil, så som upphovsman, upphovsrätt, tillkomstår, titlar, ämnesord eller dylikt, eller *andra aspekter som gäller själva objektet*. Till de senare hör på ett bibliotek till exempel uppgifter om var man kan hitta en bok. I den digitala världen är kraven på den senare typen av metadata ännu större: uppgifter om filtyp, filformat, administrativa uppgifter om olika rättigheter och tekniska uppgifter om proveniens eller vilka program man behöver för att kunna komma åt innehållet är mycket relevanta för att man ska kunna hantera digitala material. För att allt detta ska fungera i datorvärlden måste metadata vara mycket strikt och enhetligt formulerad, så att den kan behandlas maskinellt. För ändamålet finns därför en hel del olika riktlinjer och standarder.⁴⁹

Olika typer av kataloger, register och index har länge funnits för att ta fram information. Inom de olika minnesorganisationerna, alltså museer, bibliotek och arkiv, har man traditionellt haft olika sätt att ordna sina material. Det handlar dels om att ordna själva materialet, dels om att skapa verktyg och hjälpmedel att hitta fram på andra sätt än enligt den systematik samlingarna är ordnade. Som exempel duger igen biblioteket: böckerna ordnas enligt biblioteksklass och enligt författare – *en* bok kan stå endast på *ett* ställe. Med hjälp av kataloger har man sedan kunnat leta sig fram också andra vägar, till exempel direkt enligt författare.

I arkiven har man i stället ordnat materialen enligt proveniens, det vill säga enligt den kontext där ett dokument uppkommit. Men uppdelningen mellan

⁴⁹ Se till exempel Jenn Rileys och Devin Beckers visualisering "Seeing Standards" (2009–2010) <http://www.dlib.indiana.edu/~jenlrile/metadatamap/> (13.8.2013).

samlingstyper har inte heller varit helt konsekvent. I många museer finns det bibliotek och arkiv, i arkiv finns det föremål och i bibliotek finns arkivalier. Bibliotekarier och arkivarier har specialiserat sig på att ordna skrifter, så att man kan hitta vad som behövs. Visserligen handlade det länge om kunskap som till stora delar funnits i huvudet på folk, men strukturerna i materialet har ändå behövts för att hitta fram.⁵⁰

Den tryckta bokens parentes och det växande antalet böcker har gjort biblioteksmänniskorna till experter på att skapa enhetlig, normaliserad metadata. I arkiv har metadata tidigare inte värderats särskilt högt, eftersom arkivteorin har proveniensens som grundprincip. Detta har också gjort arkivpersonalen till verkliga grindväktare till informationen. Det viktiga har varit att betjäna myndigheter och forskare.

Pappersepokens fysiska, oföränderliga exemplar och sättet att strukturera informationen, som ändå i grunden är baserat på att hitta fram till ett fysiskt dokument är i dag föråldrad. Framför allt har dessa tidigare varit verktyg för personalens interna och professionella bruk, men i och med webben vill och kan allt fler användare själv söka i både material och olika typer av index. I dag pågår därför ett omfattande omstruktureringsarbete inom samtliga sektorer, men förändringen är omfattande och långsam, för den rubbar i grunden uppfattningen om vad det är vi förvaltar och hur vi gör det.⁵¹

Bibliotekssystemen är stora informationsresurser med förhållandevis god (enhetlig och väl strukturerad) kvalitet, men i grunden för allt finns fortfarande kortkatalogens sätt att hantera informationsförvaltningen. Ett arkiv består i sin tur av handlingar som uppstått inom en verksamhet. En myndighet, en person, en släkt eller en organisation är arkivbildare, vilket utgör grunden för

⁵⁰ Insamling, ordning och klassificeringssystem i sig har givetvis mycket viktiga och intressanta ideologiska dimensioner, en sak inte minst Michel Foucault pekat på. Som ett nordiskt exempel se Rainer Knapas, *Kunskapens rike. Helsingfors universitetsbibliotek – Nationalbiblioteket 1640–2010* (SLS, 2012) s. 116–119, 141–143 och 415.

⁵¹ Inom museivärlden har man kommit längst med att ta i bruk de nya systemen som baserar sig på grundligt gjorda semantiska begreppsmodeller där man skiljer på begrepp och namn (CIDOC CRM), inom biblioteken har man länge arbetat med modellen (FRBROO) men att byta system har varit lättare sagt än gjort. Inom arkiven har processen varit betydligt mycket svårare men verkar ha kommit igång på olika håll.

arkivsystemet. Sättet att ordna materialet är därför helt annorlunda, det baserar sig på ursprung, dokumenttyp och kronologisk ordning. Det är sammanhanget som ger varje dokument sin betydelse, och det reflekteras i dess placering, såsom breven i ett personarkiv, som kan vara ordnade enligt motpart i korrespondensen och sedan enligt datum. I arkiv behövs inga ämnesord som i en bibliotekskatalog, kontexten säger det som behöver sägas. Indexen är därför mer begränsade.

En museisamling liknar mer ett bibliotek än ett arkiv till sin idé och även i hur den administreras. Också en del av de samlingar som går under namnet "arkiv" i dag är de facto snarast samlingar, till exempel inom folkkultursarkiv. Därför finns det också inom arkivvärlden olika system med olika logik. I dag förväntar man sig ändå att sökande skall vara enkelt, det skall vara snabbt och helst ske via ett enda gränssnitt, alltså en och samma tjänst oberoende av var material finns eller hur det är ordnat. Kravet är fullständigt rimligt i sig. Det är därför projekt som Europeana, K-samsök, Söndera och det Nationella digitala biblioteket kommit till. Tekniken skall nog kunna klara det. Uppdelningen mellan de olika kulturarvsinstitutionerna är på sätt och vis konstgjord och ibland mer eller mindre slumpmässig.

Men det är fortfarande ett problem att man i minst ett par sekler ordnat sin information på så olika sätt. I dag borde man lära sig att tänka digitalt. Numera ordnas till exempel alla böcker i det franska nationalbiblioteket endast enligt akvissionsordning i magasinerna. Det räcker och är det allra mest effektiva.

Det är möjligt att bryta de gamla informationsstrukturerna inom alla sektorer och gå in för en rikare struktur. Men det kräver ett fördomsfritt förhållningssätt och noggrann semantisk analys av resurserna: vad betyder egentligen de olika elementen i alla kataloger och databaser? Vilken typ av information innehåller de? Det krävs också en del mod att bryta sönder de gamla strukturerna och arbetssätten. För gemene informationssökare är det helt egalt om materialet förvarats i ett arkiv eller katalogiserats i ett bibliotek. Man måste få fram det som är relevant.

Att tvingas till nya datamodeller inom olika branscher ger också nya möjligheter. Gemensamt för de nya modellerna är att de, till skillnad från många biblioteks- och arkivsystem som närmast motsvarar lite utvecklade elektroniska traditionella kortregister, faktiskt är genuina begreppsmodeller av verkligheten. Då man byggt upp dem har man ordentligt analyserat informationen och dess art. Resultatet är system som ser mer invecklade ut, men som i själva verket är mycket enklare och mer rationella. Framför allt innehåller det otroligt mycket mera information om relationer *mellan* olika entiteter, information som tidigare måste finnas i huvudet på användaren.

En viktig aspekt är till exempel att man skilt på benämning/titel/namn från representationen av själva begreppet/verket/personen, vilket gör att systemen kan hålla reda på olika varianter av samma verk eller olika namn på samma person mycket effektivare. Men avancerade, så kallade semantiska system, kan också beskriva roller och funktioner, alltså relationer mellan olika objekt. Det är ju rätt stor skillnad om en tavla är målad på eller föreställer en viss plats, något som visserligen ofta funnits dokumenterat, men inte kunnat utnyttjas fullt ut vid sökning.

Om man analyserar informationen i de olika gamla systemen, finner man förutom att den ofta är extremt inkonsekvent och bristfälligt ifylld, att den lätt passar in i dylika semantiska modeller. Och då försvinner också problemet med de i utgångsläget olika informationsstrukturerna. Fortfarande är det ändå ett problem, i synnerhet då det gäller arkivmaterial, att man borde producera mycket mera beskrivande metadata och bättre fungerande sökningar och andra tjänster för att man ska kunna använda det digitaliserade materialet fullt ut.

Digitalisering av källor

Jessica Parland-von Essen

Digitalisering av text

Inom kulturarvssektorn har man ofta förespråkat massdigitalisering och digitalisering av hela material och serier snarare än urval baserade på forskares

eller kunders behov och önskemål. Denna linje har motiverats med bästa input-output-nytta, alltså flest filer och gigabytes, frånvaro av en extra urvalsprocess och den största nytta man hävdar komma för forskare med obrutna helheter som är mycket använda och som genom digitalisering kan skyddas från slitage. Linjen har dessutom förstärkts genom resultatstyrningen inom den offentliga sektorn som belönar mängd fram om kvalitet. En stor del av materialen är därför digitaliserade från gamla mikrofilmer, vilket är både billigt och snabbt. Kort sagt skannar man gärna in bilder av dokument så automatiserat som möjligt.

Problemet var, och är det delvis fortfarande, att sökbarheten i sådana arkivmaterial är mycket dålig, inte minst för att metadata eller annan kontextuell information är bristfälliga. I praktiken har man gjort livet lite enklare för (släkt)forskare som kanske nu kan titta på dokument på skärm hemma istället för att behöva ta sig till en forskarsal. Å andra sidan vet varenda forskare att kvaliteten ofta är dålig på dessa material och det är mycket tröttsamt och arbetsdrygt att läsa oskarp text från en datorskärm. När man väl hittat det man söker, vill den seriösa forskaren dessutom ofta ändå granska originalet för att kunna göra källkritisk bedömning, eftersom det i metadata saknas information om papperskvalitet och dylikt. Äldre digitalisering gjordes svartvit och i värsta fall som tvåbit (alltså utan gråtoner) så all färginformation har försvunnit. Därför är det mycket problematiskt att man rutinmässigt belägger originalen till digitaliserade material med totalt användningsförbud vid vissa arkiv.

Denna form av digitalisering är i regel ändå bara det första steget vid digitalisering av text. Är texten maskinskriven eller tryckt kan den också läsas in maskinellt så att den tekniskt sett faktiskt blir en text istället för en bild. Processen kallas ofta *optical character recognition*, OCR. Sådan text är genast sökbar och betydligt mer hanterlig för en forskare, man kan till exempel kopiera den in i ett dokument. Problemet är att om bilden är av dålig kvalitet kommer den digitala texten sannolikt att innehålla många fel, även om den maskinella avläsningen hela tiden blir bättre. I det finska [Historiska tidningsbiblioteket](#) har man löst detta genom att använda sig av oskarp logik (fuzzy logic, sumea haku). Denna metod ger programmet möjlighet att förbise felen och söka textsträngar

som liknar på ett ungefär, i motsats till normal datorlogik, där allt antingen är ja eller nej, 1 eller 0.

I Australien – eller inom [Project Runeberg](#) – har man i stället använt sig av de möjligheter webben erbjuder, genom att låta användarna rätta korrekturfel direkt i databasen. Detta är förstås mycket klokt genom att kvaliteten hela tiden förbättras på ett för webben typiskt sätt och varje korrigerings behöver göras endast en gång. Erfarenheterna har varit goda.

Vad gäller handskrifter är situationen tillsvidare en annan. Själva texten måste uttydas av en människa och skrivas in manuellt i en annan fil. I detta fall är det den som gör arbetet som gör tolkningen av texten, som vanligen är betydligt mer utmanande då det gäller handskriven text. När texten skrivs in måste man välja hur man förhåller sig till stavfel, otydliga bokstäver eller saker som överstrykningar. För detta ändamål kan man använda sig av koder, förslagsvis TEI-kod (Text Encoding Initiative) som erbjuder en standardiserad form för att märka ut sådant i texten. Detta ger sedan möjlighet att presentera texten på olika sätt på en datorskärm enligt redaktörers eller användares preferenser.

Genom att presentera bilden och den inskrivna texten samtidigt för användaren erbjuder man möjlighet för forskaren att göra egna tolkningar, men man har ändå möjliggjort bra sökning och kopieringsmöjligheter. Att skriva in text på detta sätt är ungefär hundra gånger dyrare än att bara producera bilder av arkivdokument. Ett exempel på hur man med annotering med TEI-kod kan beskriva hur en handskrift ser ut:

```
Du har ju vågat en fräckhet som jag godtagit, <add place="below"
medium="pencil">ehuru</add> inte för att <subst><del>vad du
sagt</del><add place="below" medium="pencil">de infall du
haft</add></subst> varit sista ordet.
```

Den annoterade texten är svår att gestalta som ren kod, varför olika visningsprogram är nödvändiga för att man ska kunna använda den kodade texten på ett bra sätt, i synnerhet om man inte arbetat med annotering själv.

Till svårigheterna på handskrifts- och arkivsidan hör ytterligare katalogiseringstraditionen och den arkivvetenskapliga synen på dokument som diskuterades ovan. Ett arkiv har bildats under och genom en specifik verksamhet och det enskilda dokumentets plats i strukturen anger dess kontext. Till skillnad från bibliotek, då man beskriver varje bok på en egen katalogpost, beskriver man inom arkiv strukturen med en hierarkisk ordning, som i sig utgör beskrivning av det enskilda dokumentet. Några ytterligare särskilda klasser, ämnesord eller dylikt behövs inte. Det är förstås arbetsekonomiskt vettigt, men i digitala sammanhang – då man lätt har att göra med ett enskilt dokument helt lösryckt ur sitt sammanhang – blir det lätt svåröverskådligt.

Om man har texten ordentligt utskriven och annoterad blir den sökbar och kan användas eller ytterligare vidareförädlas på många sätt. Förutom de TEI-koder som ovan nämnades i samband med hur ett manuskript ser ut kan man markera till exempel namn, språk, orter, titlar eller vad man önskar i texten. Själva texten blir då en xml-fil som består av läsbar text, men som också kan visas snyggt och på flera olika sätt enligt hur man önskar på en skärm. Den som någon gång arbetat med enkel html vet idén i stort, principen är den samma. Man markerar delar av texten med taggar före och efter det annoterade avsnittet, det kallas att taggar öppnas och stängs. Dessutom finns det en hel del regler om hur man skall göra det för att följa standarden. Vissa taggar behöver inte stängas.

```
<name type="person">Thomas Hoccleve</name> <name  
type="place">Villingaholt</name> <name type="org">Vetus Latina  
Institut</name> <name type="person" ref="#HOC001">Occeleve</name>
```

Då man annoterat texten kan man också föra in en kopia i en databas om man vill, eller länka personnamn till auktoritetsregister eller orter till koordinater. Arbetet låter drygt, men det är förvånansvärt snabbt, förstås beroende på hur komplicerad kod man använder och hur många korrekturläsningar man gör. Då man läser korrektur på kodens grammatik kallas det *validering*. Det kan man göra automatiskt med ett valideringsverktyg som hittar fel i koden åt en. Då man

läser korrektur på själva texten kallas det för *kollationering*. Det ska helst göras av en annan person än den som skrivit in texten.

Det är i alla fall helt klart att för många forskningsändamål är dylikt berikande av text en stor hjälp och gör det möjligt att få fram strukturer och samband man annars kanske aldrig kunde ha fått syn på. Då texter är i digital form kan man göra omfattande sökningar i litteratur eller till exempel tidningsmaterial som av arbetskonomiska skäl annars blivit helt outnyttjade som källor. Poängen här är fulltextsökning, det vill säga att användaren kan söka inne i texterna, i dokumenten och inte bara i metadata, alltså kataloger och register, som ibland kan dölja information mer än att hjälpa en rätt.

De stora textmassorna är förstås en guldgruva för lingvister, vare sig de är annoterade eller inte. Stora textmassor ger möjligheter att undersöka språk och utveckla språkteknologi och de kallas korpusar då man upprätthåller dem i detta syfte. Automatisk översättning är till exempel en utmaning som många arbetar med och som skulle avhjälpa den språkförbistring som många gånger finns på nätet och som i Europa mest syns som en dramatisk ökning av det anglosaxiska inflytandet på snart sagt alla områden. Forskningen går också mycket snabbt framåt och också Google har hållit sig väl framme.

Historiker är ofta intresserade av enskilda uppgifter i de stora textmassorna, varför de gärna uppskattar tjänster som bygger på databaslösningar, där texten ordnats enligt innehållets struktur, vilket gör sökandet mycket effektivare. Databaser med uppgifter ur kyrkböcker eller dylikt passar bra att göra till databaser eftersom de är väl strukturerade, men det skulle vara önskvärt att man parallellt med texten kunde studera bilden av originalet.

Litteraturvetare är ofta mer intresserade av intertextuella relationer eller litterära processer, som på ett mycket bra sätt kan återges genom annotering. Den digitala textens många nivåer fungerar på ett liknande sätt som texten "på riktigt" gör då den blir till. Med hjälp av TEI är det till exempel lätt att jämföra olika varianter av en text, och dessa kan också behandlas som jämbördiga, då man slipper den tryckta bokens endimensionella eller betydligt klumpigare format.

Det finns alltså lika många behov och ansatser som det finns forskare och forskningsdiscipliner då det gäller annotering av text och det är knappast värt att annotera allt med tanke på alla eventuella framtida forskare. Det är ändå så, att man aldrig kan veta med säkerhet vad som anses relevant och viktigt för vetenskapen i framtiden. Som ett klassiskt exempel inom bokhistorien kan man nämna alla de skyddspapper och papperspärmar som tagits bort då man bundit in böcker, som i dag är mycket eftersökta och intressanta för forskare, eftersom de berättar väldigt mycket om bokmarknaden, ett område som tidigare inte uppfattades som särskilt kulturhistoriskt relevant. Och ändå är det ju så, att forskare läser, skriver och behandlar stora mängder text under sin forskning och att allt detta material också kunde göras till råmaterial för ny forskning. Det behövs mycket samarbete mellan arkiv och bibliotek och forskare för att digitalisering ska kunna göras på ett bra sätt. Med nära och integrerade samarbeten med forskare får också kulturarvsorganisationerna omedelbart mervärde för sina digitaliserade material. Material som bearbetats, strukturerats och berikats och som det finns forskning om kan också ge den mindre insatta användaren mycket mer än en massa bilder av dokument man inte kan gestalta betydelsen av på något sätt.

Digitalisering av bilder och bilder av föremål

Det finns två olika huvudtyper av digitala bilder: vektorgrafik och bitmaps (kallas också punktuppbyggda bilder eller rasterbilder). Vektorgrafiken baserar sig på geometriska figurer som datorn ritat upp enligt instruktioner i filen, vilket gör att man kan förstora bilder hur mycket som helst utan att de förlorar i skärpa. Vektorgrafik används därför i regel för ritningar som gjorts med datorprogram och filnamnet slutar ofta på .svg. Fotografier består däremot i normala fall av punkter (pixlar), där exakt färg för varje punkt finns angiven i filen (vars namn kan sluta på t ex .jpg eller .tiff). En del filformat kan innehålla både vektorgrafik och raster (t ex .psd-filer skapade med bildbehandlingsprogram).

Med resolution avser man hur många punkter det finns på en inch (25,4 mm) (på skärm ppi, i tryck dpi alltså dots per inch). Det handlar ofta om rätt stora mängder information varför en del filformat kan komprimeras. Detta behandlas närmare i avsnittet om bevaring av digitala material.

En bild kan sparas som enbitsbild, vilket innebär att varje pixel kan ha bara ett av två värden, antingen svart eller vit. De allra äldsta digitaliseringarna gjordes ofta som enbit eftersom förvaringsutrymmet var mycket dyrt tidigare. Fortfarande räcker enbitsbilder för till exempel digitalisering av stora mängder text, där man bedömer att ingen annan information är bokstävernas och de andra tecknens form, placering och storlek är av intresse. Gråskala ger dock redan betydligt mycket mera information. Då det gäller manuskript strävar man i vanliga fall i dag till färgbilder, som kan återge mycket mera nyanser och detaljer om till exempel pappret eller bläckets kvalitet. Det är viktigt att komma ihåg att information som förloras vid digitaliseringsprocessen inte kan återskapas annat än genom en ny digitalisering.

Det sägs ofta att datorskärmen kan visa fler färger än ögat kan uppfatta, till exempel över 16 miljoner. Färgerna på en skärm består i normala fall av rött, grönt och blått som grundfärger (RGB). Varje pixel har alltså ett värde för varje ljuslängd (färg) och färgen kan anges som tre siffror på en skala som varierar enligt hur många bitar man anvisar för varje värde (bitdjup). Det är bra att veta att till exempel de 16 miljoner utlovade färgerna alltid ytterst begränsas av datorskärmens och själva mjukvarans tre extremvärden: den mest röda, gröna och blåa färgen som skärmen kan visa.⁵² Dessa punkter kan variera i olika programversioner, filer och skärmar, varför datorn kan visa färger felaktigt om de tre inte utgår från samma inställningar. Vid digitalisering är det därför viktigt att man har kalibrerat både skanner eller kamera och skärm på bildbehandlingsprogrammet. Även hemma eller i forskarsalar borde man kalibrera skärmen om man vill göra noggrannare analys av materialet. För detta finns utrustning att köpa. Dessutom bör man veta att det alltså fortfarande finns

⁵² Man kan läsa mer om färgrymder till exempel i Wikipedias artikel om Gamut, <http://en.wikipedia.org/wiki/Gamut> (hämtad 2013-09-02).

en hel del färgtoner som inte kan återges på en datorskärm, eftersom de består av ljuslängder som befinner sig utanför den triangel som bildas av de angivna (dock varierande) extremvärdena av datorskärmens tre grundfärger.⁵³

På arkiv och museer har man ofta en något olika inställning till bilder såsom gamla fotografier. Inom museivärlden uppfattas bilder som föremål, vilka digitaliseras som sådana som *en del av dokumentationen*, medan arkiv ofta bara strävar till att *reproducera* dem som digitala versioner. Att digitalisera bilder kan verka enkelt, skanners är ju enkla att skaffa, men det är nog en betydligt mer svår process än så.

Som ovan framkom finns det en hel del saker att ta ställning till, man väljer (medvetet eller omedvetet) mjuk- och hårdvara för själva processen, liksom man måste bestämma hur mycket bilderna ska behandlas eller beskäras, man måste välja filformat, resolution och så vidare. I regel korrigerar man så lite som möjligt i hela bildens färgåtergivning och att ändra enskilda detaljer är förbjudet. Det innebär förstås i klartext att en digital bild är en tolkning i sig. Dessutom måste man fråga sig: en tolkning av vad? Av det som bilden föreställer? Av det hur bilden såg ut när fotografen nyss framkallat den (och var nöjd)? Eller kanske av bilden i det skick den råkade befinna sig den dagen man skannade den?

Det sista är det svar man ofta spontant skulle få också av en del människor i branschen, men efter någon eftertanke brukar de flesta vara överens om att man måste spara så mycket information som möjligt. Det är informationsinnehållet som är det som ska räddas. "Allt" ska sparas.

Men även skannern eller kameran som ibland används har ju gjort en tolkning. En tolkning sker varje gång man överför information från ett språk till ett annat, vare sig det gäller naturliga språk eller datorkod. Varje gång förlorar man sannolikt information. Den tolkning som gjorts vid överföringen i digital

⁵³ För intresserade finns det mycket läsning om ämnet på webben. Ett utförligt exempel är Pei-Ying Li, Ya-Ping Wang, Lang-Hsuan Kao, *Digitization Procedures Guideline: Color Management*, Taiwan e-Learning and Digital Archives Program, Taiwan Digital Archives Expansion Project (2011), <http://collab.teldap.tw/digitalLibrary/Color%20Management.pdf> (PDF, 18 Mb). En mer konkret inledning av Gareth Hawker finns i Artwatch UK-bloggen från den 10 januari 2011, <http://artwatchuk.wordpress.com/2011/01/10/10th-january-2011/> (hämtad 2013-09-02).

form är inte alltid optimal, utan mycket information kan finnas dold och kan tas fram genom att justera hur bilden visas. Denna information är ju också bra att spara, så att en framtida användare hittar den. Vissa saker anses alltså korrekta att göra med en bild då man digitaliserar den för att ta fram information, medan annat, så som att beskära den eller retuschera den anses absolut förbjudet. Slutresultatet är att man ofta har minst två versioner av en bild. Dessutom vill man gärna ha flera bilder av mindre format som man kan bläddra och hantera enklare på datorn. För att bilden skall vara till någon större glädje för en forskare borde det dessutom finnas så mycket uppgifter som möjligt om bilden, både om originalet och om vad bilden föreställer (vem, var, när, vad?). Sådana uppgifter är också viktiga för att man skall kunna söka och hitta bland bilderna. Uppgifterna kan antingen lagras i själva filen eller i en databas, helst båda.

Fotografering används i synnerhet då det gäller föremål och stora bilder såsom tavlor och kartor. Med skanner kan man uppnå stora resolutioner, upp till 24 000 ppi, men fotografering är många gånger snabbare och enklare med stora objekt. Vid fotografering kan i vissa fall uppstå upphovsrätt om fotografiet har verkstatus. Gränsen är något flytande, men det är alltid hyggligt att nämna fotografens namn om man har det då man använder en bild. I princip anses ändå att ny upphovsrätt inte uppstår vid digitalisering, som likställs vid en ny utgåva av originalet.

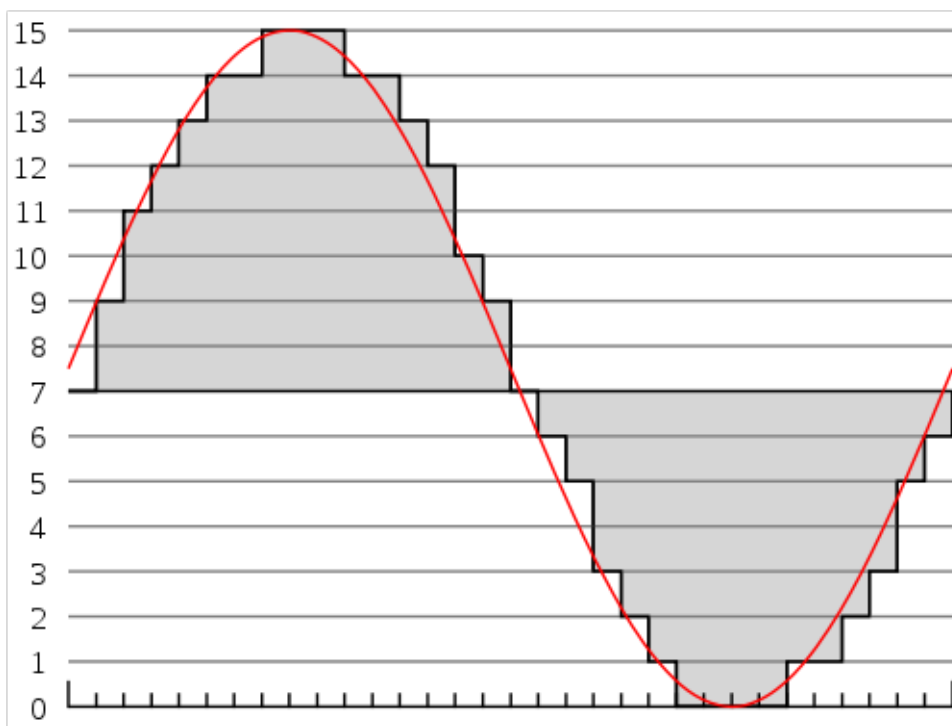
Då man beställer en bild för att använda den är det viktigt att kontrollera att man får tillräckligt med uppgifter om bilden, inte minst om upphovsmannen och innehållet. Det är också stor skillnad om bilden ska tryckas eller användas i något annat syfte, t ex på webben, vilket man bör beakta då man beställer en bild. Andra siffror, format eller färgprofiler etc behöver lekmän inte bry sig så mycket om i detta skede, eftersom de är så pass standardiserade tekniska saker.



Ett exempel på digitalisering i form av bilder av föremål: "Tyskt utmärkelsetecken för Luftwaffe. Instiftat 1936. Amémuseum, AM.086968." Från Digitalt museum, <http://www.digitaltmuseum.se/things/mrke/S-AM/AM.086968> (hämtad 2013-09-02, public domain).

Digitalisering av ljud och rörlig bild

Då det gäller digitalisering av ljud och rörlig bild gäller i princip alla de ovan diskuterade principerna och problemen. Ljud- och videoinspelningar på magnetband förstörs av sig själva rätt snabbt, redan inom några decennier, men det kan vara stor skillnad i kvaliteten så man kan sällan vara säker på den exakta livslängden. För CD-skivor räknar man inte heller med någon längre livstid. I praktiken är det så att information som går förlorad vid digitaliseringen ofta är förlorad för evigt.



Digitaliseringen tvingar till val eftersom den digitala informationen är diskret till sin karaktär. Vid digitalisering av ljud måste man bestämma sig för till exempel samplingsfrekvensen. Från Wikimedia Commons (användare Ktims), <http://en.wikipedia.org/wiki/File:Pcm.svg> (hämtad 2013-09-02, licens CC-BY-SA 3.0).

Då det gäller skapandet av ljudfiler finns det många parametrar man måste ta ställning till, som påverkar slutresultatet. Man måste definiera skala och format för både ljudstyrka (decibel), frekvenser (Hertz), samplingsfrekvens och bitdjup och förstås filformatet. I mån av möjlighet borde man använda samma apparatur vid digitaliseringen, som använts vid den ursprungliga inspelningen. Också då det gäller ljud är det mycket viktigt att man inte gör ingrepp för att "förbättra" ljudkvaliteten i samband med själva digitaliseringen. Till exempel eventuellt reducerande av brus ska göras som separata åtgärder, på kopior av den så kallade masterfilen eller arkivfilen, eftersom information alltid försvinner under processen och tekniken hela tiden blir bättre. Det händer ofta att man behöver gå tillbaka till en tidigare version av en inspelning för att förbättra

resultatet då tekniken förbättras. Då ett analogt original förstörts, är den ursprungliga digitaliseringen den ursprungligaste versionen. Den bör därför vara så rik på information som bara någonsin är möjligt. Även om det innebär att en hel del brus ingår och det inte låter särskilt bra. Användarkopior av materialet kan sedan behandlas på ett sådant sätt att ljudet låter bra för människoörat.

Filer med rörlig bild består av både en ljudfil och en serie bilder. Bilderna kan vara separata eller innehålla radsprång (*interlacing*), vilket innebär att de enskilda bilderna är strimlade i mycket smala rader som sedan är sammanflätade lite i otakt, vilket gör att bilderna kan vara färre och ändå ge rörelsen på filmen ett smidigt intryck. Vidare kan filer packas så att endast de ställen som förändrats från föregående bild ersätts i följande bild, så att resten av bildens kod i praktiken säger "ta detta område från den förra bilden". Man tar gärna till dylika tekniker för att få mindre filstorlekar, eftersom videofiler är mycket stora, hundratals gånger större än ljudfiler och tusentals gånger större än textfiler. Slutresultatet är att filer med rörlig bild är enorma och väldigt komplexa till sin struktur. Om filerna är väldigt komprimerade är de mindre pålitliga eftersom den enskilda bilden innehåller artefakter, alltså sådan information som datorn konstruerat. Å andra sidan kan man ju komma ihåg att också människans egen hjärna fyller i avsevärda delar av det vi tror oss se på ett liknande sätt (*unconscious inference*), men varje tolkning eller överföring av information förvanskar eller förändrar den ytterligare.

Beskrivande metadata är viktig för sökning i rörlig bild och ljud. Man har dessutom stor nytta av tidskodning. Om man har en ljudfil utskriven som text kan man göra kodningen i xml. Automatisk transkribering av ljud är ett viktigt område under kraftig utveckling, som kommer att hjälpa mycket vid sökning i framtiden. Många språkvetare är bekanta med talspråkskorpusar, där man sparar både ljud och text med ibland flera varianter av transkribering med inflikade tidskoder, så att man med rätt program kan plocka fram exakt rätt ställe i ljudfilen. Samma teknik kan givetvis användas på videofiler.

Fördjupning: Digitala textkritiska utgåvor

Jenny Bergenmar

Argumenten för att ge ut textkritiska utgåvor bygger ofta på författarskapets litterära och kulturella betydelse. "Topelius litterära verk är ett monument i ord över Finlands 1800-talshistoria och utgivningen är därför en nationell uppgift av nordisk bärvidd", skriver redaktionen för [Zacharias Topelius Skrifter](#). Det är författare som tillmätts denna betydelse som blir föremål för textkritiska utgåvor. Några av de mest påkostade nordiska utgivningsprojekten är följaktligen Topelius, [Henrik Ibsens skrifter](#) och Nationalupplagan av [Strindbergs samlade verk](#). Dessa tre projekt, liksom exempelvis [The Rossetti Archive](#) illustrerar väl den kanoniserande och kulturkonserverande funktion en textkritisk utgåva ofta fyller.

De nämnda utgåvorna är alla i någon mån digitala, några "born digital" (digitalbaserade), andra med en digital utgåva vid sidan av den tryckta utgåvan. Jerome McGann, pionjär inom digitala textkritiska utgåvor med The Rossetti Archive blickar självkritiskt tillbaka i en artikel om [Scholarly Editing in the Twenty-First Century](#), där han menar att projektet – liksom andra efterföljande som [The Walt Whitman Archive](#) – misslyckades med att skapa interaktiva utgåvor. Modellen var fortfarande lika statisk som i de tryckta textkritiska utgåvorna: editionsfilologen gör det textkritiska arbetet och presenterar resultatet. Man kan som användare titta på representationer av text eller bild, men inte göra något med dem.

Spetsar man till det kan man som Peter Shillingsburg säga att som regel är de författare som publiceras i digitala utgåvor döda, vita män, och de som ansvarar för utgåvorna är medelålders, vita män.⁵⁴ De principer för hur historia produceras som historikern och antropologen Michel Rolph Trouillot skrev om i *Silencing the Past: Power and the Production of History* (Boston: Beacon Press,

⁵⁴ Peter Shillingsburg, "Hagiolatry, Cultural Engineering, Monument Building and Other Functions of Scholarly Editing", *Voice, Text, Hypertext. Emerging Practices in Textual Studies*, ed. Mondiano, Searle & Shillingsburg, Seattle & London 2004.

1995) gäller även för historia i digitala medier. "Silences enter the process of historical production at four crucial moments: the moment of creation (the making of sources); the moment of fact assembly (the making of archives); the moment of fact retrieval (the making of narratives); and the moment of retrospective significance (the making of history in the final instance)" (s. 27). Det finns alltså goda skäl att fråga sig vilka tystnader som uppstår inom textkritiken, och Peter Robinson summerar i artikeln "[Editing Without Walls](#)" dess hierarkiska organisation såhär:

- *Utgivaren bestämmer vad som ska utges*
- *Utgivaren bestämmer vem han/hon ska arbeta med*
- *Utgivaren bestämmer hur han/hon ska edera*
- *Utgivaren samlar materialet*
- *Utgivaren skapar utgåvan med sina medarbetare*
- *Utgivaren behöver stort stöd från institutioner*

Efter detta klagörande av premisserna för många digitala textkritiska utgåvor skisserar Robinson en annan utgivningsform som gradvis växt fram, där bibliotek och andra aktörer tillgängliggör digitaliserat material på webben, forskare och studenter gör transkriptioner och beskrivningar baserade på och länkade till bilderna, andra forskare tar vid och kollationerar transkriptionerna och publicerar dem online, vilket kan ligga till grund för fortsatt textkritiskt arbete med kommentarer och textkritisk apparat.

Utgåva/arkiv/databas

Den förändrade organisationen inom textkritiska projekt som Robinson beskriver har väl inte fullt ut slagit igenom. Jag ska strax återkomma till det kollaborativa inslaget, men först några ord om en annan tendens i digitala textkritiska projekt generellt: avgränsningen mellan arkiv och utgåva är inte längre så tydlig. För editionsfilologer representerar arkiv en samling av texter

som inte är ederade och kommenterade. Paula Henrikson beskriver skillnaden så här i [Textkritisk utgivning. Råd och riktlinjer](#):

Det verkar också klokt att upprätthålla en rågång mellan "arkivet" och "utgåvan", där den senare ställer större krav på utgivarval och utgivaransvar, till exempel genom att erbjuda en etablerad text. Lika litet som den tryckta faksimilutgåvan kan för övrigt det digitala arkivet göra anspråk på objektivitet eller neutralitet. Bägge vilar (i likhet, förstås, med vilket som helst fysiskt arkiv) på val och urval och utgör i den bemärkelsen en tolkning av det föreliggande materialet.

Om man med arkiv menar en virtuell eller fysisk plats där materiella eller digitala artefakter *lagras* är förstås skillnaden mellan arkivet och utgåvan avgörande och uppenbar. Men i en digital kontext förekommer ofta utgåvans noggranna annotationer tillsammans med arkivets inklusiva ambitioner. När initiativet inte heller alltid ligger hos forskare och editionsfilologer utan hos bibliotek och institutioner kan det vara självklart att utgå från en specifik samling, som antingen kan representeras "dokumentaristiskt" eller kombineras med en textkritisk utgåva. Det digitala arkivet används ofta också till att skapa helt nya samlingar av exempelvis manuskript som varit spridda i en mängd olika fysiska arkiv, såsom [The Emily Dickinson Archives](#), som också beskriver sig som "A repository for the study of resources related to Emily Dickinson". Som Kenneth M. Price skrev redan 2009 har en mängd olika begrepp prövats för att beskriva digitala projekt med textkritiska inslag; se artikeln [Edition, Database, Archive, Thematic Research Collection: What's in a Name?](#)

Crowdsourcing

I många framväxande arkiv/digitala utgåvor är det kollaborativa inslaget starkt betonat och gränserna mellan editionsfilologen som expert och frivilliga bidragsgivare mindre skarpa. Det finns förstås både starka ekonomiska och demokratiska argument som talar för detta. Experter behöver frivilliga krafter både för att hitta dokument och för att transkribera dem. Finska litteratursällskapet erbjuder till sina medlemmar om att transkribera Aleksis Kivis texter, som Sakari Katajamäki skriver om i sin [fördjupningsartikel](#), visar också att crowdsourcing inte nödvändigtvis behöver innebära att man tummar

på kvaliteten. Det irländska projektet [Letters of 1916. Creating History](#) syftar till att representera vardagsliv under tiden kring Påskupproret 1916 genom en kombination av digitalisering av samlingar av brev, och crowdsourcing. Allmänheten uppmanas att bidra med brev och bilder från en period av sex månader före och efter upproret, men kan också bidra genom att transkribera redan uppladdade brev. Fler och fler projekt öppnar för denna möjlighet för allmänheten att bidra som "citizen humanists", exempelvis [The Shelley Godwin-Archive](#). Det blir också allt vanligare att digitala arkiv/utgåvor erbjuder möjligheter för användaren att delta i det textkritiska arbetet genom verktyg för textjämförelse som exempelvis [Juxta](#).

Det finns också initiativ som tar ännu ett steg bort från top-down modellen, såsom den beskrevs av Peter Robinson ovan. Litterära wikis tillåter alla registrerade användare att själva göra det textkritiska arbetet i form av annotationer och kommentarer. Ett framgångsrikt exempel är [pynchonwiki.com](#) där frivilliga har bidragit till att annotera åtta verk av Pynchon sida för sida. Denna öppna, ohierarkiska modell är starkt betonad i [The Digital Humanities Manifesto 2.0](#), där sociala medier beskrivs som kultur- och kunskapsproducerande laboratorier.

Kvantitativ och kvalitativ digitalisering

Digitala utgivningsprojekt kan vid första anblicken ha lite att göra med massdigitaliseringsprojekt som [Google books](#). Men var går gränsen mellan digitalisering, exempelvis genom att man skapar e-texter av scannade och OCR:ade böcker, och det textkritiska arbete vars viktigaste beståndsdelar är kollationering av text, förteckning av varianter och versioner, samt kommentarer till texten? I och med att fler digitala verktyg för textanalys utvecklas, måste också texterna vi använder för dessa analyser vara av någorlunda god kvalitet. [Hathi trust](#) ställer jämfört med Google books högre krav på textens tillförlitlighet och tillgänglighet. [Litteraturbanken](#) som har som uppgift att samla in och digitalisera svensk skönlitteratur, står ännu ett steg närmare den kvalitativa textkritiska utgivningen. Det betyder att OCR-fel

korrigeras och att emendationer (rättade textfel) redovisas, medan e-texten så långt som möjligt följer utgåvan i layouten. De textkritiska insatserna synliggörs i en redovisning för varje text av vilka ändringar som gjorts. Här framgår också vilket exemplar som legat till grund för digitaliseringen.

MARIE SOPHIE SCHWARTZ
*Amanda. Några strödda anteckningar ur
Stagnelius' lif*

Albert Bonniers förlag, Stockholm 1893.

Läs verket som ETEXT
Även tillgängligt som EPUB



Det exemplar som ligger till grund för
Litteraturbankens utgåva tillhör Göteborgs
universitetsbibliotek (Litt. Sv.).



I etexten har följande ändringar gjorts mot originalet:

- | | |
|----|---------------------------------------|
| 2 | Stående bild har vänts till liggande. |
| 25 | Stående bild har vänts till liggande. |
| 27 | tömde det – tömde det. |
| 28 | Behöfs inte!" – "Behöfs inte!" |

Litteraturbanken ger också ut andra ederade utgåvor i digital form, exempelvis Nationalupplagan av August Strindbergs Samlade Verk, och textkritiska utgåvor av Svenska vitterhetssamfundet. Som utgivare av svenska klassiker kan Litteraturbanken tyckas representera samma fokus på litterära monument i nationell kontext som många författarskapsutgåvor. Men vid sidan av de särskilda författarskapen ingår också enligt [utgivningsprinciperna](#) "tematiska eller genremässiga grupperingar", vilket öppnar för genrelitteratur (exempelvis skräck) som hamnat utanför kanon, samt "betydelsefulla översättningar från andra språk", vilket bidrar till att bryta det nationella fokus som många författarcentrerade utgåvor har.

Fördjupning: Textkorpusar för historikerbruk – ett inifrånperspektiv

Helena Holm-Cüzdan och Julia von Boguslawski

Under några år har det nu vid [Svenska litteratursällskapet i Finland](#) och [Institutet för de inhemska språken](#) (Helsingfors, Finland) pågått ett projekt där syftet har varit att skapa en finlandssvensk språkbank som en del av den svenska [Språkbanken](#) vid Göteborgs universitet. Tanken har varit att skapa en finlandssvensk textkorpus med både en modern och en äldre del, och på så vis göra ett större material tillgängligt för språkforskare och andra intresserade. I den finlandssvenska textkorpusens äldre del på adressen <http://spraakbanken.gu.se/korp/?mode=fisk1800> finns en samling texter som spänner över en tidsperiod på ca 200 år, från mitten av 1700-talet till mitten av 1900-talet. Materialet är indelat i följande genrer: Brev och dagbokstexter (som motsvarighet till moderna bloggar), skönlitteratur, sakprosa, tidskrifter, tidningar och myndighetstexter.

Urval

En fördel med en textkorpus är att det blir möjligt att behandla stora mängder data och göra sökningar i ett stort material som tidigare enbart funnits tillgängligt i arkiv i form av till exempel handskrifter eller mikrofilmer. Att detta material nu finns digitaliserat är ett framsteg, men som användare av textkorpusen måste man minnas att digitaliseringen föregåtts av en urvalsprocess och endast en liten del av allt det som finns i arkiv och bibliotek finns digitaliserat. Urvalet har påverkats av bland annat tekniska och upphovsrättsliga begränsningar.

Genrevis har vi följt indelningen i den moderna finlandssvenska korpusen för att det skall vara möjligt att så enkelt som möjligt göra jämförelser mellan äldre och nyare material. Vår strävan har varit att försöka hitta texter från alla decennier mellan 1750 och 1960 och att få en så stor geografisk spridning som möjligt. Då det gäller den geografiska spridningen har vi koncentrerat oss på

Finlands kustregioner där det bor flest svensktalande. Den geografiska täckningen är relativt bra vilket gör att både språkforskare och andra förhoppningsvis kan ha nytta av materialet. Könsfördelningen däremot är skev trots att vi strävat efter så många skribenter som möjligt av bägge könen. Dels deltog kvinnor inte i det offentliga livet i samma mån som män under 1700- och 1800-talen och dels finns färre texter av kvinnor bevarade över lag. Bland brevskrivarna och 1900-talsförfattarna är könsfördelningen jämnare. Också klassfördelningen är skev och de flesta skribenterna hör till de högre samhällsklasserna.

Materialet utgörs av många olika sorters texter och urvalet har gjorts efter lite olika principer beroende på genre. Digitalisering i allmänhet är en arbetsdryg process. Tryckt material behöver först skannas in innan man kan köra det igenom ett OCR-program (textigenkänning) och sedan måste dokumentet korrekturläsas. Korrekturläsningen kan vara mer eller mindre arbetsdryg beroende på hur bra OCR-programmet lyckats läsa texten. I synnerhet äldre tryck, dvs. allt fram till tidigt 1900-tal, är än så länge en utmaning för tekniken beroende på att papperskvaliteten och sättningen varierat mycket. På grund av detta har vi med en del av de äldre verken valt att göra stickprov och digitaliserat valda delar istället för hela verk. Eftersom arbetet är tidskrävande har tillgängligheten varit ett huvudkriterium när vi valt material.

För skönlitteraturens del har vi dels valt sådana verk som kan anses ingå i den finlandssvenska litteraturkanon, men också verk av mindre kända författare. Kategorin är speciell på så sätt att det tidigaste materialet är från år 1840, då den första svenskspråkiga romanen publicerades på östra sidan om Finska viken. Materialet från 1900-talet har valts ut delvis på upphovsrättsliga grunder, eftersom upphovsrätten i Finland är i kraft i 70 år efter författarens frånfalle. I fall då det varit lätt att nå innehavaren av upphovsrätten har vi gjort avtal med dem, men ofta har vi valt verk där upphovsrätten inte längre gäller.

För sakprosans del har vi använt oss av de texter som varit mest lättillgängliga. Största delen av texterna som ingår har givits ut i nytryck eller

funnits färdigt digitaliserade.⁵⁵ I de fall då avtal varit nödvändiga har vi haft gott samarbete med utgivarna. Materialet från 1700- och 1800-talen består till största delen av resebeskrivningar och texter som gäller ekonomiska frågor. Texterna från 1900-talet behandlar mera varierande ämnen och består till största delen av tidigare digitaliserade artiklar.

Brev- och dagbokstexterna härstammar nästan uteslutande från Svenska litteratursällskapet i Finlands arkiv och förlag, men en liten del av materialet kommer från privata brevsamlingar. Vissa begränsningar finns också eftersom denna typ av texter är av en mer privat natur, och en del arkivsamlingar inte alls är öppna för allmänt bruk. Dessutom påverkas urvalet i allra högsta grad av handstil och läslighet eftersom det tar tid att digitalisera handskrivna dokument. I övrigt har vi i mån av möjlighet använt oss av brev och dagböcker som redan tidigare givits ut på Svenska litteratursällskapetets förlag. Urvalet påverkas således även av vad man under tidigare år har ansett värt att publicera i bokform och av de olika redaktörernas språkliga linjedragningar, även om vi valt texter som ligger så nära originalet som möjligt.

Språket och ändringar som gjorts i texterna

En språkbank ger forskaren tillgång till ett betydligt större material än vad en enskild forskare skulle hinna samla ihop på egen hand. Lars Borin och Richard Johansson behandlar i sin artikel "[Kulturomik: Att spana efter språkliga och kulturella förändringar i digitala textarkiv](#)" möjligheter som ett stort digitalt material ger och även problem som kan uppstå gällande sökbarheten. Vissa problem uppstår redan i digitaliseringsprocessen. Gamla stavningsformer och handstilar går inte alltid ihop med den teknik som används för digitalisering och kodning och själva överföringen av de gamla texterna till digital form kräver att texterna anpassas för det nya formatet.

Variationen i språk och stavning är större ju längre bakåt i tiden man går, men så länge man arbetar med tryckt text finns inga egentliga tolkningssvårigheter. I

⁵⁵ De tidigare digitaliserade texterna härstammar bland annat från Projekt Runeberg (<http://runeberg.org/>), Litteraturbanken (<http://litteraturbanken.se/>), Zacharias Topelius skrifter (<http://www.topelius.fi/>) och från Svenska litteratursällskapet i Finlands förlag.

handskrifter däremot kan det bland annat vara svårt att avgöra ifall ett ord är skrivet med versaler eller gemener, ifall ord är särskrivna eller har skrivits ihop. Dels är det fråga om tidstypiska variationer men också skribentens utbildningsnivå. Då man transkriberar äldre text ingår alltid en viss grad av tolkning. De publicerade brev och dagböcker som vi använt är formade enligt olika redaktionella beslut och skiljer sig därför också ofta i någon mån från originalet.

För att man skall kunna koda texten så att den blir sökbar krävs att den som digitaliserar materialet gör vissa förändringar, som att inleda meningar med stor bokstav och avsluta dem med punkt. I större delen av de äldre texterna, i synnerhet brev och dagböcker från 1700- och 1800-talen, används både versaler och punkter varierande då en klar språkstandard saknades. Detsamma kan förekomma sporadiskt en bra bit in på 1900-talet, särskilt i privata texter, och i synnerhet om skribenten i fråga inte har gått särskilt mycket i skola.

Sammanfattning

Den finlandssvenska språkbanken, liksom språkbanker över lag är ett bra verktyg för snabba överblickar, men eftersom materialet genomgår vissa förändringar och en urvalsprocess kan det i många fall vara nödvändigt att också ta sig en titt på originalen. Åtminstone bör man vara medveten om att den digitaliserade versionen kanske inte stämmer överens med originalet till 100 procent och att materialet inte är fullständigt representativt för någon viss tidsperiod eller genre. Det här har beaktats i Språkbanken och den är uppbyggd så att det är lätt att se vilket originalet är och var det finns bevarat.

I och med det här projektet finns det nu en del historiskt material digitaliserat, men mycket återstår ännu att göra. Under projektets gång har det framkommit att det redan finns en hel del äldre material digitaliserat inom olika projekt och på olika forskares privata datorer. Tack vare gott samarbete har vi fått tillgång till en del och med bättre koordinering kunde allt större mängder material som digitaliserats på olika håll bli tillgängligt för fler användare i framtiden.

Fördjupning: ProBok – en proveniens- och bokbandsdatabas⁵⁶

Helena Strömquist

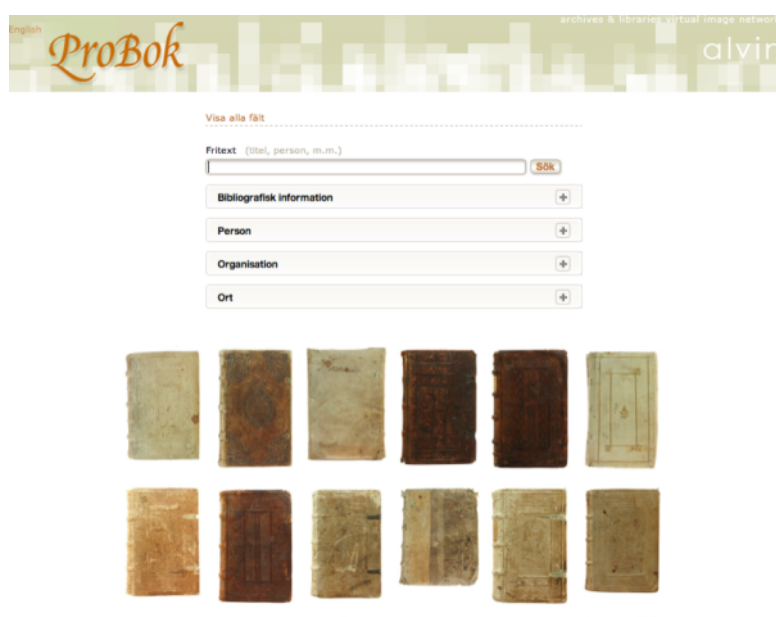
Åren 2010 till 2011 fick jag möjlighet att arbeta med utvecklingen av en databas som skulle dokumentera provenienser och bokband i forskningsbibliotekens samlingar av äldre tryck. Databasen heter ProBok (<http://probok.alvin-portal.org/alvin/>) och medel för projektet hade lämnats av [Riksbankens Jubileumsfond](#) efter ansökan av kulturarvsansvariga vid [Uppsala](#) och [Lunds](#) universitetsbibliotek, Per Cullhed och Björn Dal. Arbetet med databasen leddes i Uppsala av Maria Berggren och i Lund av mig Helena Strömquist. Stefan Andersson vid UUB ansvarade för de it-tekniska delarna av projektet.

ProBok innehåller böcker från handpress- och tidig maskinpressperiod (fram till 1870) och den dokumenterar böckernas inbindning och deras provenienser. I databasen kan man söka fram hur en bok ser ut, vilken bandtyp och material den har bundits i, hur den är dekorerad, vem som bundit in den, när och var det har skett. I proveniensdelen hittar man uppgifter om alla sorts spår av användning som lämnats i boken, från understrykningar och klotter i den tryckta texten över ägaranteckningar, exlibris till bibliotekssignaturer. ProBok är ett exempel på hur man genom digitalisering kan tillgängliggöra samlingar på ett nytt sätt och skapa metadata ur ett källmaterial som tidigare inte funnits att tillgå för vare sig forskning, för undervisning eller allmänheten.

ProBok-databasen är en del av arbetet med utvecklingen av en nationell plattform för forskningsbibliotekens samlade digitaliserade kulturarvsmaterial. Alvin-plattformen (archives & digital libraries virtual image network) byggs vid Enheten för digital utveckling vid Uppsala universitetsbibliotek. Fördelarna med en större plattform för digitaliserat kulturarvsmaterial är flera. Man kan dra nytta av tekniska lösningar, auktoritetsregister och andra funktioner som kan

⁵⁶ Innehållet i denna fördjupningsartikel publiceras, när inget annat anges, enligt licensen [CC BY-NC-ND 3.0](#).

vara gemensamma för alla databaser på plattformen. Alla databaser inom Alvin ingår också i en större kontext av digitaliserat material som användarna kan söka i, det kan vara bilder, bokband, provenienser och olika typer av handskriftsmaterial. Med Alvin bygger man alltså en digital infrastruktur med ett långsiktigt perspektiv på förvaltning, ekonomi och utveckling av digitaliserat kulturarvsmaterial som biblioteken gemensamt ansvarar för.



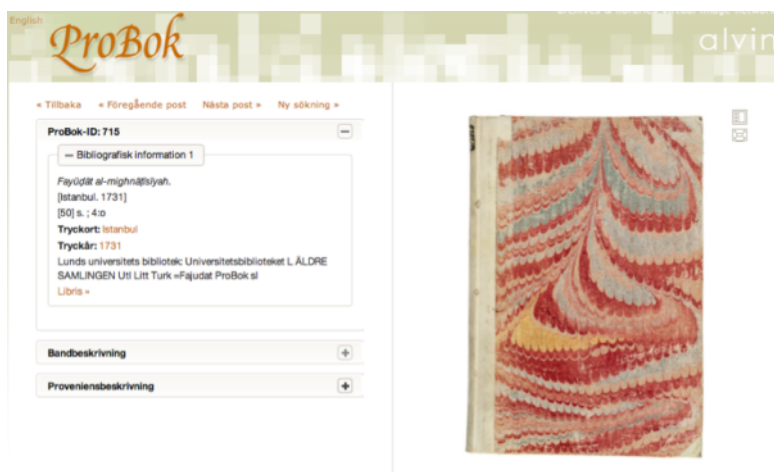
Skärmbild från databasen ProBoks söksida: <http://probok.alvin-portal.org/alvin/> (2013-10-10).

I ProBok-databasen digitaliseras boken i sin egenskap av fysiskt föremål och källdokument. Den traditionella verkrelaterade informationen finns med som en del av ProBok – i form av en bibliografisk post – men tyngdpunkten ligger på analysen och dokumentationen av det fysiska dokumentet, på boken som artefakt som ingått och ingår i ett bokhistoriskt kretslopp. Materiella egenskaper, band- och dekortechniker och form hos artefakten redovisas. I proveniensdelen redogörs för enskilda och institutionella ägare och de spår de lämnat i boken. Varje enskilt objekt i ProBok kopplas till aktörer (förläggare, tryckare, bokbindare, handlare och bokägare) som varit involverade i dess

publicering, produktion och spridning. På detta sätt ger ProBok möjlighet att följa bokens materiella produktion, spridningsvägar och reception både på mikro- och makronivå.

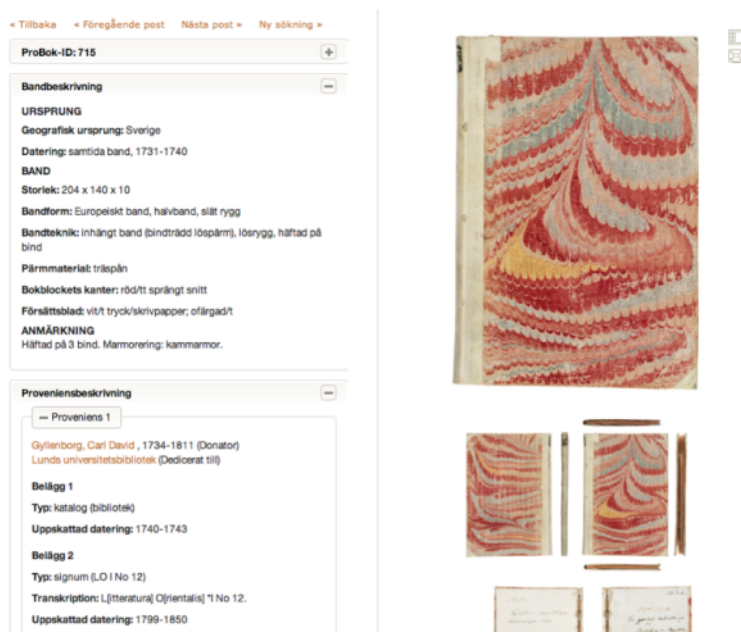
Två bokhistoriska modeller som beskriver processer inom grafisk kommunikation och det grafiska dokumentets livscykel har inspirerat ProBokformatet och kan appliceras på en ProBokpost. Den ena modellen är Robert Darntons kommunikativa kretslopp, den pekar ut aktörer i bokhistoriska processer. Den andra är Thomas Adams och Nicholas Barkers livscykelmodell vilken beskriver själva dokumentets livslopp.⁵⁷

ProBokposten består av tre delar: tryckbibliografisk information, banddokumentation och provenienshistorik. Den bibliografiska posten är utförd i Libris enligt katalogiseringsregler för äldre tryck och hämtas därifrån till ProBok.



Bibliografisk information i en ProBokpost, hämtad från Libris. ProBok 715, skärmbild: <http://probok.alvin-portal.org/alvin/record.jsf?pid=alvin-record%3A15715&cid=3> (2013-10-10).

⁵⁷ Robert Darnton, "What Is the History of Books?", *Daedalus*, Summer 1982, s. 65–83. Thomas Adams & Nicolas Barker, "A new model for the history of the book", Nicolas Barker (red.), *A potencie of life: books in society: the Clark Lectures 1986–1987* (London: British Library, 1993).



Bandbeskrivning och provenienshistorik i en ProBokpost. ProBok 715, skärmavbild: <http://probok.alvin-portal.org/alvin/record.jsf?pid=alvin-record%3A15715&cid=3> (2013-10-10).

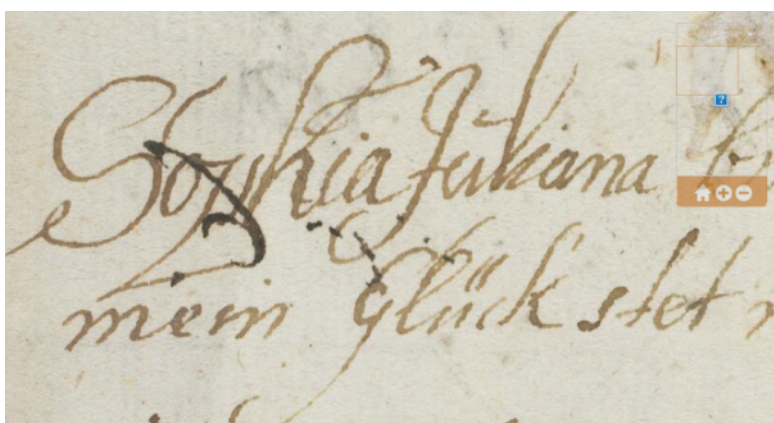
Bandbeskrivningen i posten har tre underavdelningar; den första dokumenterar bandets ursprung, den andra beskriver bandtyp, material, teknik och form, den tredje underavdelningen beskriver bandets dekor. I ProBokpostens tredje del redovisas bokens provenienser. Den bygger på dokumentation av proveniensbelägg, en materiell/teknisk beskrivning av beläggen samt attribuering av varje enskilt belägg till en person eller institution.

Varje ProBokpost innehåller också en utförlig bildredovisning med åtminstone 11 bilder per bok. Bilderna har mycket hög upplösning och utgör ett kraftfullt verktyg för visuella analyser av dokumentets materiella egenskaper och dess innehåll. Material, färger och tekniker liksom handskriftens detaljer kan studeras i kraftig uppförstoring i bilderna. Bilder av det här slaget är en av digitaliseringens stora fördelar. De ersätter givetvis inte dokumentet eller forskares behov av en fysisk bedömning av ett material men de bidrar till arbetet med att analysera och tolka dess egenskaper och innehåll. Bilder av hög

kvalitet är ett exempel på hur digitaliserat källmaterial utgör nya verktyg för humanistisk forskning.



En förgylld och ciselerad kant på boken i kraftig förstoring. ProBok 656, skärmavbild: <http://probok.alvin-portal.org/alvin/record.jsf?pid=alvin-record%3A15656&cid=17> (2013-10-10).



Sophia Juliana Forbus anteckningar på insidan av pärmen till sin bok. Samma som ovan, ProBok 656, skärmavbild: <http://probok.alvin-portal.org/alvin/record.jsf?pid=alvin-record%3A15656&cid=17> (2013-10-10).

För att skapa en sammanhållen uppsättning metadata i kravspecifikationen till ProBok fick vi göra ett omfattande materialbibliografiskt utvecklingsarbete. Termuppsättningen finns nu i en svensk och en engelsk version.

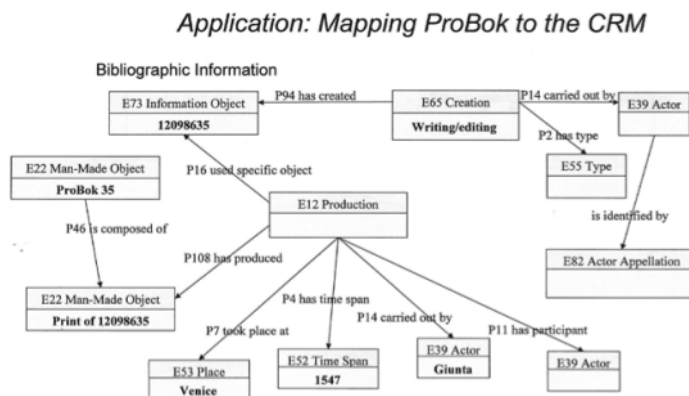
I ett framskridet skede av arbetet med ProBok inleddes ett europeiskt samarbetsprojekt för att ta fram en engelskbaserad tesaurs, *Language of Bindings* (LoB), för beskrivning och dokumentation av bokband. Arbetet leds av professor Nicholas Pickwoad vid [Ligatus Research Centre, University of the Arts London](#). Inom ProBok har vi aktivt bidragit till LoB:s innehåll och vi har också justerat en del termer i ProBok med anledning av det gemensamma arbetet.

Både ProBoks termer och LoB är ordnade enligt Simple Knowledge Organisation System (SKOS), ett formellt språk och en standard för beskrivning av taxonomier, tesaurer och kontrollerade vokabulärer inom kunskapsdomäner.⁵⁸ ProBoktermernas SKOS heter BokbandsSpråk. Att ordna termer i en SKOS innebär att data inom olika kunskapsfält och datasystem kan delas, användas, förmedlas och relateras till andra terminologier och databaser som data. Information i databaser kan mappas genom en SKOS, d.v.s. information i olika databaser med likartat innehåll görs sökbart parallellt med hjälp av frågespråk. Data kan också importeras och exporteras mellan datasystem.

Termer i *BokbandsSpråk* och *Language of Bindings* är ordnade under ett antal huvudkoncept. Dessa är hämtade från ICOMs CIDOC Conceptual Reference Model som är en formell ontologi för dokumentation av kulturarv.⁵⁹ När ett kunskapsområdes ontologi kompletteras med en formaliserad terminologi för detta område och databaser och andra informationssystem "mappas" mot dessa, kommer systemen att kunna kommunicera med varandra på ett meningsfullt sätt och i stor skala. Då har vi realiserat visionen om den semantiska webben.

⁵⁸ SKOS, Simple Knowledge Organization System
Primer: <http://www.w3.org/TR/2009/NOTE-skos-primer-20090818/> (2013-10-10).

⁵⁹ Definition of the CIDOC Conceptual Reference Model: http://www.cidoc-crm.org/docs/cidoc_crm_version_5.1-draft-2013May.pdf och http://www.cidoc-crm.org/official_release_cidoc.html (2013-10-10).



2

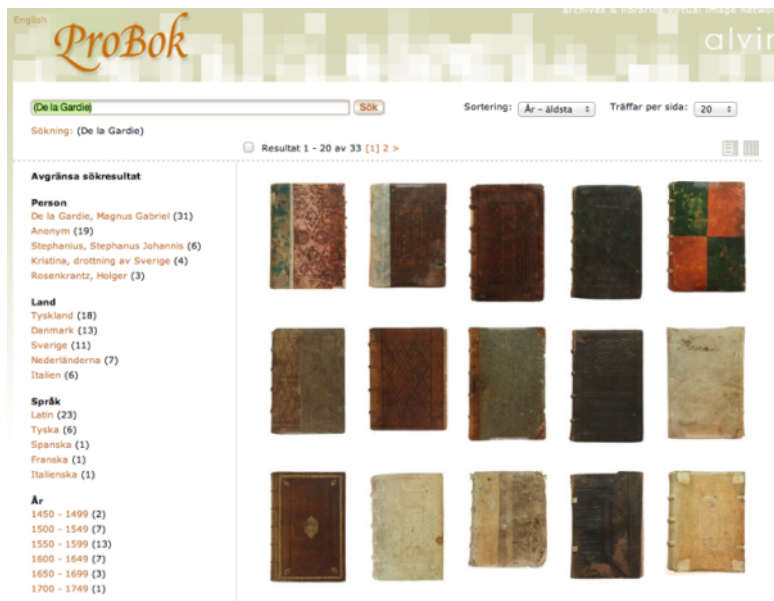
© Martin Doerr. Denna mappning av ProBok mot CRM är utförd av professor Martin Doerr i samband med en workshop som projektmedlemmarna deltog i vid Institute of Computer Science, Foundation for Research and Technology, Heraklion, Grekland, september 2011.

Att ta fram en terminologi i SKOS-form är ett omfattande arbete, men på lång sikt är det väl investerad tid. Metadata får ett större värde för forskningen och är hållbar i ett längre tekniskt/digitalt perspektiv. BokbandsSpråk innehåller omkring 300 termer, LoB ännu fler.⁶⁰

I ProBok kommer man i framtiden att kunna skapa digitala bibliotek av böcker som finns spridda över olika institutioner. Böckerna kan grupperas och studeras på en rad olika sätt beroende av vilka metadata som man väljer. Materiella egenskaper och framställningstekniker i bokbanden kan analyseras. Dekortekniker och ornament inom grafisk produktion kan knytas till en större kulturell kontext i tid och rum. Distribution och handel med böcker kan studeras genom priser och bokens varierande materiella former. Enskilda personers och institutioners böcker kan samlas ihop, grupper av bokägare kan sökas fram och deras bokinnehav studeras; kvinnor, präster och läkare för att bara nämna

⁶⁰ BokbandsSpråk är under publicering och är för närvarande inte tillgänglig. Vill man ta del av den kan man vända sig till författaren.

några. Läspraktiker kan studeras genom anteckningar och bokinnehav. Spridning och reception av texter kan bekräftas genom attribuering av enskilda böcker till personer. Nätverk, kontakter och kulturella praktiker kopplade till boken som materiellt objekt och förmedlare av idéer och kunskap kommer också att kunna utforskas.



En sökning i ProBok på "De la Gardie" ger 31 träffar, sökresultatet avgränsas i tid och rum, geografiskt och avseende språk.

ProBokdatabasen är resultat av ett samarbete mellan forskare och en kulturarvsinstitution där vi tillsammans lyckats förmedla och tillgängliggöra ett traditionellt biblioteksmaterial på ett nytt sätt. Länkningen till Libris är en mycket betydelsefull del av ProBokprojektet. På lång sikt, när databasen fyllts på, kommer den att synliggöra bibliotekens äldre samlingar på ett mycket effektivt sätt.

Flera av världens främsta och största universitetsbibliotek, bland andra Harvard, Princeton, Oxford, Gent och München, deltar idag i Googles biblioteksprojekt i vilket miljontals copyrightbefriade tryck digitaliseras i

fulltext och görs tillgängliga via internet.⁶¹ Det finns andra liknande projekt. Inom överskådlig framtid kommer väsentliga delar av de svenska forskningsbibliotekens äldre utländska samlingars tryckta texter att finnas tillgängliga i fulltext genom Google och andra kommersiella aktörer. Kvaliteten på den digitalt förmedlade texten varierar men den finns där för alla att ta del av i läsbar form.

För forskningsbibliotekets tryckta kulturarv handlar det framtida digitaliseringsscenariot därför mindre om att digitalisera "text on demand" i svartvita bläddringsbara filer som tillgängliggörs över internet på bibliotekens hemsidor och i digitala kataloger. Undantaget här är givetvis det nationella trycket som borde vara nationalbibliotekens ansvar att tillgängliggöra digitalt.

Det handlar mer sannolikt om att upprätthålla kunskap och kompetens om bibliotekens samlingar i sig, deras historiska kontext och den källinformation som boken som artefakt är bärare av. Det är i det här avseendet som bibliotekens samlingar är unika och inte kan ersättas av Googles eller andra kommersiella aktörers digitala filer. Avsevärda delar av bibliotekens äldre samlingar är också döda i ett verkperspektiv och har förlorat sitt informationsvärde som texter. Som historiska artefakter och som ett materiellt uttryck för intellektuell och konstnärlig utveckling och spridning av idéer i tryckt form under tidigmodern och modern tid utgör de emellertid ett betydelsefullt forskningsmaterial och kulturarv.

Det är också denna typ av information som databaser som ProBok kan förmedla och som därför är värdefulla för institutionerna själva, forskare, studerande och allmänhet.

⁶¹ Googles bokprojekt finns beskrivet på: <http://books.google.com/googlebooks/library.html> och <http://books.google.com/googlebooks/partners.html> (2013-10-10).

Källkritik och hänvisningar på webben

Kenneth Nyberg

Den grundläggande förändring, med en mängd uttryck på olika områden och nivåer, som internets framväxt representerar handlar i grund och botten om en sak: tillgänglighet. Vi har, som nämnts tidigare, gått från en situation av brist till överflöd på information eller, i historikernas fall, på källor. En del av materialet är digitalt från början ("born digital"), en del är digitaliserat från befintliga, fysiska källor – men de har alla det gemensamt att vi idag i ökande grad når dem via internet. I detta avsnitt ska vi titta närmare på vad detta betyder för källkritik och hänvisningar på webben, och i nästa mer specifikt på samma frågor i relation till material i digitala arkiv.

I båda fallen är en viktig aspekt att fråga sig vad som egentligen är nytt i den digitala världen och vad som inte förändrats sedan tidigare. Ett första, ganska givet och allmänt svar på det som ändå är viktigt att framhålla, är att källkritik i grunden alltid är och har varit ett förhållningssätt och inte i första hand en teknik. Däremot kommer självfallet denna hållning till uttryck i form av olika "tekniker" eller handfasta råd, och därför är stora delar av detta och följande avsnitt tämligen konkreta och praktiska till sin karaktär.

Källkritikens grundprinciper

Som en utgångspunkt för det som följer kan det vara en poäng med att kort summera den "klassiska" källkritikens grundprinciper, de begrepp som generationer av historiestuderande i både Norden och andra delar av världen har fått lära sig genom åren. Kort uttryckt är källkritiken, ur ett forskarperspektiv, en samling riktlinjer för att avgöra en källas äkthet, trovärdighet och relevans för den fråga vi söker svar på. Inom historieämnet utgjorde dessa principer länge, och är kanske i vissa avseenden fortfarande, det sammanhållande metodologiska ramverket i avsaknad av mer utvecklade teoretisk-metodologiska modeller; som ett brett ämne med många

subdiscipliner har källkritiken sålunda utgjort ett förenande band mellan utövarna. Åtminstone är det ofta så det framställs i grundutbildningen i historia.

Dessa grundprinciper har äldre rötter men kom att systematiseras i den form vi känner dem idag från 1800-talet och framåt, under en period när många historiker eftersträvade en mer "objektiv" och "exakt" historievetenskap med inspiration från natur- och samhällsvetare. Ambitionen att med ett batteri kritiska frågor avtäckta Sanningen, uttryckt i Leopold von Rankes berömda formulering "Vad har egentligen hänt?", har fått sig många törnar sedan dess och anspråken har skruvats ner något; idag talar vi mer om att målet för vår granskning av källorna är att kunna uttala oss om sannolikheter och "giltighet".⁶²

Källkritikens bud kan sammanfattas och grupperas på olika sätt. Ett är den heltäckande frågan, begriplig utan några som helst fackkunskaper, "Vem säger vad till vem i vilken situation och med vilket syfte?" Ett annat är att tala om "yttre" kritik – en källas tillkomst, bakgrund och eventuella avsikter – respektive "inre" kritik – själva innehållet och dess trovärdighet – samt växelspelen däremellan i en hermeneutisk cirkel eller (i bästa fall) spiral. Den yttre kritiken hjälper oss utföra den inre, vilket ger en bättre förståelse för att utveckla den yttre kritiken osv. Ett tredje sätt, för de flesta kanske det mest bekanta, är att systematisera källkritiken utifrån fyra moment, vilka alla är beroende av varandra i varierande mån:

- **Äkthet:** Är källan vad den utger sig för att vara, vad den tycks vara? Detta är givetvis en första förutsättning för att alls kunna avgöra användbarheten för den fråga vi vill besvara. Även en förfälskning kan vara av relevans beroende på hur och när den tillkommit, men att någorlunda säkert kunna etablera vad en källa faktiskt representerar (eller inte) är en grundförutsättning för resterande steg i värderingen.

⁶² Jfr Rolf Torstendahl, "Källkritik, metod och vetenskap", *Historisk tidskrift* 125:2 (2005), http://www.historisktidskrift.se/fulltext/2005-2/2005-2_209-217.htm (2013-09-07). Artikeln ingår i *Historisk tidskrifts* temanummer om källkritik (125:2) från 2005, som innehåller flera intressanta bidrag och sammantaget ger en god bild av hur historiker idag ser på dessa frågor.

- **Närhet (samtidighet):** Hur nära i tid och rum till det vi vill veta något om ligger källans tillkomst? Ju närmare desto bättre (allt annat lika), men ibland får vi nöja oss med källor tillkomna ganska långt från/efter det de berättar om.

- **Beroende:** Flera källor som berättar om samma sak är alltid bättre än en enda, men om en av dem helt bygger på den andra – är beroende av den – har den givetvis inget självständigt värde som källa. Åtminstone är det så i normala fall; återigen kan det bero på hur relationen mellan den s.k. primärkällan och sekundärkällan ser ut mer i detalj, eftersom tillkomsthistorien också kan ge intressant information som belyser det vi vill veta något om.

- **Tendens:** Här ställs frågan vilka avsikter och intressen som präglar källan, dvs. om det finns skäl att anta att framställningen är skev i en viss riktning för att någon har intresse av att ge en viss bild av en händelse eller ett förlopp. Det är inte nödvändigtvis så att en mer "neutral" framställning alltid är bättre än en mer tendentiös, åtminstone inte om man är medveten om tendensen; tvärtom kan skevheten i källan, särskilt när den lämnar uppgifter som går på tvärs mot tendensen (alltså strider mot upphovspersonens intressen), bidra till att stärka trovärdigheten i just dessa uppgifter.

En viktig aspekt i den källkritiska prövningen är till sist om källan kan betraktas som en kvarleva eller en berättande källa. Är det en direkt kvarleva av ett historiskt förlopp och vi kan fastställa dess äkthet – t.ex. ett kontrakt, en lagsamling eller ett mynt – är den definitionsmässigt en primärkälla som värderas mycket högt eftersom den genom sin blotta existens s.a.s. *är* en bit av historien. Olika former av beskrivningar gjorda av människor – vare sig det är rättegångsprotokoll, brev eller målningar – är däremot s.k. berättande källor där beroende, tendens och närhet blir viktiga kriterier för granskningen. Samtidigt är det alltid den ställda frågan som avgör om en källa är berättande eller kan behandlas som en kvarleva, vilket är en helt avgörande poäng. Om själva syftet exempelvis är att undersöka hur en resenär uppfattade ett främmande land blir hennes eller hans beskrivning en *kvarleva* i relation till frågor om författarens

synsätt, men en *berättande källa* i förhållande till frågor om situationen i det aktuella landet.

Även om alla de moment som berörts här är relevanta i den källkritiska värderingen, handlar denna i praktiken alltid om en helhetsbedömning där olika aspekter vägs mot varandra och där kunskaper om det historiska sammanhanget används för att ytterligare fördjupa förståelsen om källans värde och relevans för frågan. Därmed är den källkritiska granskningen alltid en hermeneutisk process med en ständig växelverkan mellan del och helhet, källan och dess sammanhang. Av den anledningen är det också många som idag hellre talar om källanalys eller källvärdering snarare än källkritik.

Webben som medium

Oavsett vad vi kallar den är ovanstående framställning präglad av de former och uttryckssätt som förknippas med traditionell historisk forskning baserad på tryckt och otryckt material, främst text på papper av olika slag. Hur förändras källkritikens förutsättningar av det nya digitala landskapet i allmänhet och webben i synnerhet? Vad skiljer egentligen dessa två miljöer från varandra och vad är samma? För att kunna säga något om det behöver vi först ringa in vad internet, särskilt webben, är och vad det *inte* är, eftersom det inte alltid är självklart. Några av dessa kännetecken har vi redan berört, andra inte.

För det första bör det påpekas att internet egentligen är summan av en lång rad nätverk och protokoll som numera knyter samman världen: FTP, e-post, Usenet och, förstås, HTTP (webben) för att nu bara nämna några. Idag spelar också olika typer av specialiserade appar, program för både datorer, surfplattor och mobiltelefoner en växande roll för trafiken på internet, men det lämnar vi därhän för nu. I det följande är det just World Wide Web som står i centrum, eftersom det är där de verkligt stora informationsmängderna finns och det är där vi s.a.s. måste navigera utan att någon föregående filtrering egentligen har skett. Det är en stor skillnad mot tidigare, då publicering innebar att ett material hade passerat *något* slags gallrings- eller filtermekanism som innebar att inte riktigt vad som helst kunde slinka igenom.

På webben är det inte så, för även om nya filter börjat växa fram också där, innebär publicering av ett innehåll på webben i sig inget annat än att en person med tillgång till nätet valt att lägga upp det där. Med webben "drabbas" vi alltså dels av ett informationsöverflöd av tidigare oanade proportioner, dels av det faktum att vi själva måste bli än mer kapabla att snabbt kunna granska och värdera uppgifter vi hittar där. Webben är alltså inte en källa utan ett *medium*, en kanal. Materialet där är inte bättre eller sämre än i något annat medium, men förutsättningarna för publicering ser ut på ett visst sätt och det påverkar hur vår källkritiska hållning bör komma till konkret uttryck.

Ett kännetecken är, per definition, att material på webben (vare sig det är text, bilder eller något annat) i praktisk mening inte går att granska fysiskt utan bara som "data" eller "innehåll".⁶³ Den typ av källkritik som tidigare handlade om att studera papperssorter, bläcktyper m.m. är inte möjlig här, och om vi inte anlitar oerhört sofistikerad teknisk expertis med tillgång till serverloggar etc. kan vi oftast inte heller säga något "säkert" om hur en viss fil hamnat i ett visst sammanhang på webben (vare sig rent tekniskt, i form av var den lagras på internet, eller i mer överförd bemärkelse hur den kommit till en viss webbplats). Av samma skäl kan vi ofta inte skilja på original och kopior på samma sätt, då duplicering i princip inte kostar någonting, och det är också relativt lätt att manipulera innehållet i exempelvis en text utan att vi direkt kan se det.

Det leder oss in på ett annat kännetecken för webben och för internet i allmänhet, nämligen att innehållet är föränderligt – vissa skulle säga instabilt – på ett sätt som inte riktigt gäller för pappersbaserat material. Även tryckta verk är visserligen inte fullt så oföränderliga och stabila som vi ibland vill tro, men det är ändå en kvalitativ skillnad, och ofta en del av själva poängen med webben att vi till i princip noll marginalkostnad kan förändra, utveckla, revidera, duplicera och radera innehåll där. Det betyder att beroendeförhållanden kan bli oerhört komplexa att reda ut när till exempel en viss text förekommer i ett stort antal

⁶³ Stig Roland Rask, "Nätet som källa och text" (Stockholm: Delegationen för IT i skolan 1999), http://www.skolverket.se/polopoly_fs/1.151275!/Menu/article/attachment/520185.pdf (PDF, hämtad 2013-09-02). Denna lilla skrift är, som Rask själv skriver i [en artikel från 2010](#), "ett barn av sin tid" och därför delvis föråldrad, men den innehåller också en rad kloka observationer om vad som kännetecknar internet jämfört med tidigare dominerande medier.

varianter efter att ha spritt sig genom sociala medier, på bloggar och andra webbplatser. En annan konsekvens, som egentligen hör till avsnittet om hänvisningar nedan men bör nämnas redan nu, är att det i forskningssammanhang alltid är en god idé att spara undan eller göra en utskrift av webbsidor man tänker använda sig av eftersom de kan vara borta när man väl behöver dem; därför är det också viktigt att i näthänvisningar alltid ange datum för när en sida eller fil har hämtats. En sådan datumangivelse gör det också lättare att kontrollera om en i tid närliggande version av webbplatsen finns arkiverad hos Internet Archive (<http://archive.org>), om originalsidan inte längre finns kvar.⁶⁴

Ett sista karakteristiskt drag för webben är slutligen det som gett den dess namn, nämligen det täta nät av ömsesidiga hänvisningar mellan sidor och resurser som gör att vi talar om hyperlänkning. Dessa nät av länknings är viktiga för att de sätter in sidor i ett sammanhang som på vissa sätt (men inte alla) kan jämföras med ordnandet av material i ett arkiv efter proveniens (härkomst), dvs. genom att studera det kringliggande sammanhanget och källans placering i nätverket får vi viktig information om hur vi ska förstå den och dess tillkomst. De inbördes relationerna av länkar är också betydelsefulla i ett helt annat avseende, nämligen genom den viktiga roll de spelar för hur exempelvis Google rankar sökresultat; en av de faktorer som avgör hur högt en sida kommer i sökningar är just hur många andra sidor som länkar dit (och hur "vällänkade" dessa sidor i sin tur är).

Källkritik på nätet

Vad betyder då allt detta för källkritik på nätet mer konkret? Ja, för det första får vi här, liksom annorstädes, aldrig glömma bort att vilka svar som är mest värdefulla beror på den fråga vi ställer. Det kan låta trivialt, men det är ändå viktigt att påpeka att det inte är samma sak att snabbt kolla upp en enkel

⁶⁴ Det finns också argument för att obeständigheten hos hemsidor gör att man *inte* bör eller behöver ange datum då man hämtat en sida, och dem har Jessica Parland-von Essen skrivit om bl.a. i två blogginlägg (båda hämtade 2013-09-07): "Webbarkiv", *Essetter* 2010-01-01, <http://essetter.blogspot.fi/2010/01/webbarkiv.html>, och "Om att hänvisa till webbsidor", *Essetter* 2011-02-03, <http://essetter.blogspot.fi/2011/02/om-att-hanvisa-till-webbsidor.html>.

faktauppgift som blivit aktuell vid middagsbordet, som att få ett sakligt och allsidigt underlag kring någon laddad fråga som man tänker skriva om i ett vetenskapligt sammanhang; i det förra fallet är konsekvenserna av en felaktig eller missvisande uppgift mindre än i det senare, där det också blir mycket viktigare att bilda sig en uppfattning utifrån genuin och djup sakkunskap från flera experter på ett område.

För det andra är en mer generell förändring, kanske tvärtom mot vad många skulle tro i första rummet, just detta att betydelsen av djup ämneskunskap om något bara blir än viktigare när man använder sig av material från webben. Ja, där finns en ocean av information, tillgången är större än någonsin, men detsamma kan sägas om osäkerheten kring hur all denna rikedom ska värderas. Och det enda sättet att göra sådana värderingar är att kunna utgå från just sakkunskap, antingen sin egen eller andras som man litar på. Inom områden man själv inte behärskar är det därför viktigt att kunna falla tillbaka på en mer generell kompetens att identifiera personer och institutioner man bedömer ha stor trovärdighet.

För att kunna göra det blir, i internetvärlden, för det tredje *sammanhanget* (i vid mening) som en källa återfinns i av central betydelse. Just eftersom det handlar om ett nätverk, ytterst ett socialt system, blir källans plats i detta system ett någorlunda säkert mått på dess trovärdighet – särskilt i de fall där vi själva saknar kunskap för att bedöma den. Dvs. *vem* är det som säger något och vilken *auktoritet* (authority) bedömer vi att den personen har? Där kan vissa tekniska faktorer som domänadress och institutionstillhörighet spela in (t.ex. tillmäts välrenommerade universitet hög trovärdighet i vetenskapliga sammanhang), men också och kanske än mer vilka som länkar till eller behandlar källan ifråga som trovärdig. Samtidigt måste man också hela tiden i detta nätverk vara vaksam på att varje enskild länk kan förflytta en runt halva jorden och ställa en inför en helt ny avsändare med en helt annan avsikt än den vi betraktade för bara en minut sedan; och en enstaka länk betyder inte nödvändigtvis att någon vi litar på gett sitt "godkännande" av någon för oss helt okänd person.

I sista hand är det alltså, i den digitala liksom den fysiska världen, alltid en fråga om en helhetsbedömning där olika faktorer vägs mot varandra. Ingen källa står över kritik, vi har alla våra motiv och intressen (även om det i vissa fall blir mer problematiskt än andra), och även den mest trovärdiga auktoritet kan ha fel. Det gäller alltså att utgå från vad ens fråga är, vilken typ av information och kunskap man är ute efter och väga källans värde utifrån ens samlade erfarenhet. Som Maria Ågren har påpekat i en mycket inflytelserik artikel om källkritik måste vi också alltid komma ihåg att minst lika viktig som källkritik är *självkritiken*, att vara medvetna om de tendenser som präglar vår egen bedömning av källorna och de uppgifter de innehåller.⁶⁵

Det hittills sagda sammanfattas nedan i en mer handfast "checklista" med några av de frågor som man kan eller bör ställa sig när man försöker bedöma en källa på webben, en lista som är sammanställd utifrån några av de oftast förekommande lathundarna för källkritik på webben.⁶⁶ Viktningen av respektive moment kan självfallet variera beroende på vilken fråga man har och vilken typ av material man söker, och i slutändan blir det alltid fråga om jämkning och sammanvägning av dessa (och kanske andra) aspekter.

Vem är det som uttalar sig?

- Var är jag? Tekniskt: Domänadress (www.uu.se osv.)
- Framgår det vem som står bakom/skrivit texten?
- På vilka grunder/utifrån vilka kunskaper uttalar hon/han sig?
- Är sidan publicerad av en välkänd/respekterad institution?

⁶⁵ Maria Ågren, "Synlighet, vikt, trovärdighet – och självkritik. Några synpunkter på källkritikens roll i dagens historieforskning", *Historisk tidskrift* 125:2 (2005), http://www.historisktidskrift.se/fulltext/2005-2/2005-2_249-262.htm (2013-09-02).

⁶⁶ Se t.ex. "Critically Analyzing Information Sources", Cornell University Library, <http://olinuris.library.cornell.edu/ref/research/skill26.htm>; "Källkritik", Umeå universitetsbibliotek, <http://www.ub.umu.se/skriva/kallkritik>; "Kolla källan", Skolverket, <http://www.skolverket.se/skolutveckling/resurser-for-larande/kollakallan>; och Kristina Alexandersson, "Källkritik på Internet", Stiftelsen för Internetinfrastruktur, <https://www.iis.se/lar-dig-mer/guider/kallkritik-pa-internet/> (samtliga hämtade 2013-09-03).

Vad är avsikten?

- Försäljning, ideologiska motiv, (påstått) saklig information...?

När och hur tillkom informationen?

- Är informationen aktuell?
- Anges det när sidan uppdaterades?
- Anges det vid vilken tidpunkt informationen kom till?

Relevans/värde

- Täcker källan in min fråga avseende bredd, djup, period, område?
- Är framställningens nivå rätt för mitt syfte?
- Finns materialet i en mer ursprunglig eller för mig mer relevant version någon annanstans?

Innehållets trovärdighet

- Hur kontrollerbart är innehållet? (Referenser, länkar osv.)
- Anges grunden/underlaget för de påståenden som görs?
- Är antaganden och slutsatser rimliga i förhållande till materialet?

Sammanhanget

- Hur hamnade jag här? Vem har länkat till sidan?
- Vad säger andra om denna källa/sida/upphovsman?

Länkning och hänvisningar

Till sist några ord om hantering av länkning och hänvisningar på webben. Hyperlänkarna är, som sagt, ett av mediets kännetecken, men komplexiteten och föränderligheten kan också göra det svårt att uppnå den stabilitet och permanens i hänvisningarna som traditionellt har varit ett ideal inom forskningen. En annan utmaning är att information, uppgifter, data som hämtats från internet, till och med om vi begränsar oss till just webben, kan komma i så

många olika former och uttryck att det kan vara svårt att hitta format som tydliggör både vad det är för typ av källa och var man hittat den. För ett blogginlägg blir hänvisningen inte likadan som för en tweet, och kanske inte helt identisk med en enskild sida på en "statisk" webbplats heller.⁶⁷

Även om en praxis håller på att växa fram på olika håll – t.ex. har den amerikanska Modern Language Association formulerat en rekommendation för hänvisningar till tweets – råder det därför viss oklarhet om hur vetenskapliga citeringar ska utformas.⁶⁸ I den osäkerheten gäller det att påminna sig om syftet med sådana hänvisningar, *nämligen att göra det möjligt för läsare att lokalisera det åberopade materialet eller en motsvarande källa och granska det själv*. Det gör att hänvisningar bör utformas så att de ger uppgift om:

- *upphov* (person/institution – erkänsla åt skaparen av verket/arbetet),
- *rubrik/titel* på citerat verk/material; vid behov också den
- *publikation* verket är en del av (t.ex. blogg, webbplats, socialt nätverk) samt ev.
- *datum* för tillkomst/publicering,
- *adress* (t.ex. en http-länk i klartext så den följer med vid utskrift/konvertering),
- *datum* för hämtning/nedladdning och (om det behövs för tydlighets skull)
- *typ* av källa/material/verk, t.ex. "blogginlägg", "tweet", "Facebookstatus" etc.

För ett blogginlägg kan en hänvisning enligt ovan t.ex. se ut så här:

Jessica Parland-von Essen och Kenneth Nyberg, "Text: Inledning", *Historia i en digital värld* 2013-05-15, <http://digihi.se/2013/05/15/text-inledning/> (hämtad 2013-09-05). [Blogginlägg.]

⁶⁷ Jfr Jessica Parland-von Essen, "Att hänvisa till en tweet", *Essetter* 2012-03-11, <http://essetter.blogspot.fi/2012/03/att-hanvisa-till-en-tweet.html> (hämtad 2013-09-07).

⁶⁸ "How do I cite a tweet?", *Modern Language Association*, http://www.mla.org/style/handbook_faq/cite_a_tweet (hämtad 2013-09-07).

Och för en tweet:

Barack Obama (BarackObama), "Four more years. [pic.twitter.com/bAJE6Vom](https://twitter.com/bAJE6Vom)", 2012-11-07, <https://twitter.com/BarackObama/statuses/266031293945503744> (hämtad 2013-09-05). [Tweet.]

Observera att denna form inte är den som rekommenderas av MLA, där hela tweeten citeras men länkadressen utelämnas enligt följande:

Obama, Barack (BarackObama). "Four more years. [pic.twitter.com/bAJE6Vom](https://twitter.com/bAJE6Vom)" 7 November 2012, 5.16 a.m. Tweet.

Att utelämna adressen på detta vis anser vi är olämpligt, särskilt eftersom det är notoriskt svårt att få fram just tweets som är mer än några dagar gamla på grund av hur Twitters sökfunktion är upplagd. Liksom för annat webbmateriel tycker vi därför att tweets bör citeras med en länk som leder direkt till den aktuella sidan.

Källkritik och hänvisningar i material i digitalarkiv

Jessica Parland-von Essen

Eftersom man ofta använder digitala material på ett annat sätt än fysiska arkiv finns det en del frågor som är specifika för digitaliserade material. Metadata är viktigt då det gäller digitala material, även digitaliserade, och man borde därför inkludera den i en källkritisk granskning.

Materialets proveniens och kontext

De allra viktigaste frågorna gäller förstås det sammanhang där man finner sin källa. Man måste därför allra först ställa frågor som: Vem är utgivaren av

materialet? Är det en trovärdig instans med tillräcklig sakkunskap, så att man till exempel kan lita på den metadata som finns om objektet? Finns det möjligheter att lätt kontakta utgivaren? Eftersom *alla databaser innehåller fel* är en aktör som tydligt och enkelt erbjuder användarna möjligheter att berika eller korrigera information ofta en pålitligare källa. Det innebär nämligen i regel att fel rättas oftare.

En trovärdig utgivare av digitaliserade material erbjuder också mycket information om sin verksamhet, om digitaliseringsprocesserna, beskrivning av materialen och de digitaliserade helheterna och principerna för hur man skapar metadata. Dessutom används internationella standarder åtminstone för beskrivande och teknisk metadata.

Då det gäller digitalbaserade material bör det finnas bevaringsplaner, inkluderande processbeskrivningar och gärna redovisningar för att man arbetar med OAI, TRAC och till exempel Drambora. Tillräcklig metadata för proveniens bör finnas i form av PREMIS-metadata. (I detta skede räcker det att veta att man ska fråga efter dessa akronymer, vi behandlar innebörden av dessa frågor utförligare i följande kapitel.)

Materialens representativitet och urval

Det är viktigt att beakta hur man tagit fram materialet för sin forskning, det vill säga har man bläddrat eller gjort sökningar i en databas. Vilket material kan finnas som blivit utanför; alltså hur urvalet av de källor som digitaliserats gått till och hurdana sökmöjligheter erbjuds till materialet?

Vilken är alltså mängden material man sökt i? Är den konstant eller växande? Om man gör samma sökningar om ett år eller tio - kommer man då att få samma resultat? Vilken är kvaliteten på metadata? Vem har tillsatt den och hur påverkar den sökresultaten? Svaren på dessa frågor övergår inte sällan i frågor som handlar om själva forskningsmetoden och de är därför ytterst relevanta att ställa.

Då det gäller metadata är det bra om den baserar sig på kontrollerade vokabulärer. Om sökningarna i ett arkiv fungerar bra, det vill säga att man med

några sökningar får fram relevanta resultat, är det vanligen ett tecken på att metadata håller god kvalitet. Det är ändå viktigt att komma ihåg att databaser kräver betydligt mycket mer övning och tålamod att söka i än vad många i dag är vana med. Det är därför helt normalt att man måste göra flera sökningar kring varje ämne. Men man ser enkelt om sökningar på olika namnvarianter till exempel ger olika resultat. Det tyder på att metadata inte är normaliserade och sökningarna är därför opålitliga.

Kvaliteten av själva digitaliseringen

Beskrivande metadata är också här en relevant faktor. Det kan vara av intresse att veta måtten eller papperskvaliteten på objektet man studerar. Vid högklassig bilddigitalisering finns både ett mått och en färgskala *inne* i bilden. (Se ovan i avsnittet om bilddigitalisering.) Dessa syns dock inte alltid i brukskopiorna på webben, även om det vore bra om man vill ta en bild för tryckning. Med hjälp av det lilla färgbandet kan man nämligen kalibrera både datorskärm och tryckning så att färgerna återges korrekt. Vid digitalisering av god kvalitet har dessutom båda sidorna av varje dokument digitaliserats, eller så anges åtminstone eventuell text från baksidan i metadata.

Persistent identifiers

Varje dokument borde vara möjligt att länka till med ett id som är "evigt", det vill säga en webbadress som är oberoende av systembyten eller annat. Dessa igenkänns ofta på adresselement som DOI, Handle eller URN. De bygger på tjänster med register som innehåller webbadresserna till respektive dokument och automatiskt länkar användaren vidare till den adress som är aktuell och fungerande. Om sådana id:n finns ska de alltid användas i första hand i stället för någon annan webbadress.

Hänvisning och länkning

I ett professionellt upprätthållet digitalt arkiv har varje fil ett unikt signum, även om man ibland anser att detta signum inte hör till en viss fil, utan till ett

specifikt "informationsinnehåll", varvid filen exempelvis kan bytas ut eller samma signum kan gälla flera olika versioner. I digitalarkiv finns till exempel i regel en arkivfil som ofta är en TIFF och en användarkopia som är en jpeg-fil. Dessutom kan det finnas minibilder (thumbnails) som visas i sökresultaten och till exempel en utskriven textfil och övrig metadata, som allt sammankopplas med en PI.

Julkaisun pysyvä osoite on <http://urn.fi/URN:NBN:fi-fe201209188104>

Nimeke: The dominions of the Czar of Russia Alba or Great Duke of Moscovia in which are the Dukedoms of Moscow, Wolodimer, Rezan, Worotin, Novogorod-Sewierski, Czernihow, Smolensko, Reschow, Twer, Novogorod-Weliki, Biele Jezora, Wologda, .. /
Tekijä: Berry, William
Päiväys: 1682
Formaatti: Scale [ca. 1:3,992,000]
Kuvaus: 1 kartta :, käsinvär. :, 57 x 88 cm, lehti 62,5 x 92,5 cm
Avainsanat: historialliset kartat; kartat; Venäjä
[Näytä kaikki kuvailutiedot](#)

Pakatut kuvat



Nimi: 1201013795.jpg
Koko: 7.139Mt
Formaatti: JPEG image

[Avaa tiedosto](#)

Alkuperäiset kuvat



Nimi: 1201013795.tif
Koko: 238.8Mt
Formaatti: TIFF image

[Avaa tiedosto](#)

URN pekar här på katalogposten och bildfilerna har egna namn som är synliga. Men kan vi vara säkra på att filen är den samma om 10 år, även om den har samma namn? Det beror mycket på vilket förtroende vi tillmäter utgivaren. Skärmdump från <http://www.doria.fi/handle/10024/79198> (2013-09-07).

I verkligheten finns det förstås unika stigar eller namn på varje fil av tekniska orsaker, men frågan är huruvida en forskare behöver bry sig om det eller inte - räcker det att hänvisa till ett visst dokument i form av en katalogpost eller ett "informationsinnehåll", som av institutionen kan bytas ut till exempel vid en

omdigitalisering? Den frågan är tills vidare öppen och måste också bero på kontexten, alltså på hur materialet används av forskaren. Och hur är det med hanteringen av versioner av digitalbaserade material, där frågan blir ännu mera prekär?

Då det handlar om digitaliserade material finns det förstås både ett fysiskt original, ett (eller flera) "informationsinnehåll" och en eller flera filer att hänvisa till. Varierande numrering av filer eller separata adresser till olika sidor och paginering ställer dessutom till det på ett sådant sätt att man i hänvisningen borde nämna alla dessa explicit.

Idealet vore om forskaren kunde skapa nya permanenta adresser enligt behov till arkivhelheter som man själv plockar ihop. Då kunde också samtidigt en relation registreras i *den andra riktningen*, nämligen från arkivmaterialet till forskaren och publikationen. Detta skulle vara mycket värdefull information, inte minst för arkiven själva, men också med tanke på framtida lösningar med länkad data. Klickbarheten borde finnas från båda hållen, både från forskningstexten till källan och från källan till forskningen.

Detta kräver att man vid arkiven lär sig resonera, på ett ytterligare nytt sätt. Nu tänker man gärna att vi i avsaknad av original som ska bevaras bara ska bevara "information" i många kopior och versioner. Man borde i stället ta ännu ett steg vidare och beakta att den "information" man bevara *de facto inte existerar utan sitt analoga uttryck*, alltså mjuk- och hårdvara.⁶⁹ Vi bör därför fästa uppmärksamhet också vid de tekniska aspekterna av digitala arkiv. Det är skillnad på fil och fil, liksom på dator och dator, och informationen finns inte alls utan dessa.

I slutändan gäller förstås för forskaren att använda sitt sunda förnuft och eget omdöme och ange så mycket information som möjligt för att underlätta för andra att hitta fram till det material man använt. Om bestående id:n eller länkar saknas är det särskilt viktigt att ange så detaljerad information om materialet som möjligt. Det är alltid bra att testa de webbadresser man anger. Till exempel

⁶⁹ Matthew Kirschenbaum: "The .txtual Condition: Digital Humanities, Born-Digital Archives and the Future Library", *Digital Humanities Quarterly*, 2012:6.3, <http://www.digitalhumanities.org/dhq/vol/7/1/000151/000151.html> (2013-09-07).

kan det vara värdefullt att länka till specifika vyer eller sökresultatsidor, men då är det bra att testa att det faktiskt fungerar.



Skärmdump från

[http://topelius.fi/index.php?p=texts&bookId=12#itemId=12_1§ionId=ch3&columns=\[\[3,-1\],\[2,0\]\]](http://topelius.fi/index.php?p=texts&bookId=12#itemId=12_1§ionId=ch3&columns=[[3,-1],[2,0]]) (2013-09-07). En trovärdig utgivare som arbetar långsiktigt kan man förvänta sig att se till att dylika php-baserade adresser fungerar även i framtiden.

Kapitel 5. Metoder inom digital historia

Jessica Parland-von Essen och Kenneth Nyberg

Miljöer, verktyg och arbetssätt

Kenneth Nyberg

Digitala forskningsmetoder kan handla om undersökning, bearbetning, presentation och spridning av material och/eller resultat. Principiellt sett kan vi skilja mellan metoder som hjälper forskaren att snabbare eller mer effektivt göra något som skulle gå även utan digitala verktyg, och sådant som blivit möjligt (eller åtminstone realistiskt) först med hjälp av sådana verktyg. I praktiken är det inte så lätt att göra den gränsdragningen helt exakt, och det går också att diskutera vad som ska ses som något kvalitativt *nytt* eller inte, men distinktionen är ändå viktig.

Strikt talat består metoden i ett visst projekt av ett helt upplägg där olika arbetssätt och resurser relateras till material, teorier och andra faktorer. I mycket av det som följer handlar det alltså inte om digitala metoder i egentlig mening utan om digitala verktyg eller resurser vilka är en *del* av forskares valda metod. Tyngdpunkten kommer att ligga på just forskningsmetoder, men det bör ändå nämnas att digital teknik i ökande takt genomsyrar även forskares vardagsverktyg. Historiker använder, liksom de flesta i vårt samhälle idag, sedan länge datorutrustning för rutinmässiga uppgifter som ordbehandling, informationssökning och kommunikation.

Internet i allmänhet och e-post i synnerhet fick sålunda fäste i den akademiska världen mycket tidigt, långt innan nätet fick mer allmän spridning. Forskare har i flera decennier varit aktiva på webbaserade diskussionslistor och deras föregångare (Usenet osv.), och e-post är sedan länge viktigt för kommunikationen både i närmiljön och mellan forskare som arbetar på olika håll i världen. På senare tid har också mer kraftfulla verktyg för samarbete och

kommunikation dykt upp, där inte minst Google Docs (numera Google Drive) med dess möjlighet att i realtid samarbeta om gemensamma dokument används av många.⁷⁰ Längre fram återkommer vi till betydelsen av bloggar och (andra) sociala medier som kontaktytor både mellan forskarna och mellan dem och det omgivande samhället.

Till kategorin vardagsverktyg hör också dataprogram för att hålla ordning på käll- och litteraturreferenser, där de mest spridda för närvarande är [EndNote](#) och [Zotero](#).⁷¹ Långtifrån alla historiker använder sådana program, men för några har de blivit ett viktigt arbetsredskap.

Resten av detta kapitel kommer dock att behandla metodologiska verktyg och praktiker i mer egentlig mening, där, som redan nämnts, det ofta på olika sätt handlar om "big data", de möjligheter vi har idag att bedriva forskning som bygger på bearbetning av stora datamängder. Kapitlet inleds därför med några allmänna resonemang kring att strukturera information, vilka är tillämpliga för de flesta forskningsprojekt med någon form av digitala metoder. Därefter följer ett avsnitt som behandlar databaser, vilka utgör en infrastruktur som dels är av relevans för många olika typer av digitala projekt och dels för med sig vissa mycket specifika frågor kring just databaskonstruktion och de många olika kompetenser sådant arbete kräver. Kapitlet avslutas med några kortare avsnitt vilka i tur och ordning diskuterar två av de huvudsakliga typer av metoder som nämnts redan tidigare – text mining och visualisering – samt fenomenet "crowdsourcing" eller kollektivt arbete som underlag för forskningsprojekt.

⁷⁰ Jfr Susan Hockey, "The History of Humanities Computing", i Susan Schreibman, Ray Siemens och John Unsworth (red.), *A Companion to Digital Humanities* (Malden, MA: Blackwell 2004), s. 15. Finns även i digital version på adressen <http://www.digitalhumanities.org/companion/> (2013-09-16).

⁷¹ Se <http://endnote.com> och <http://zotero.org>.

Att strukturera information

Jessica Parland-von Essen

En viktig sak att beakta då det gäller forskning och en digital forskningsprocess är att information som finns i ett datasystem *alltid är strukturerad* på något sätt. Den kunde vara strukturerad på olika sätt, men man måste vanligen välja ett. Traditionellt har historiker ofta opererat utifrån ett kronologiskt eller narrativt modus, men dessa faller lätt sönder i den digitala världen eller blir i alla fall endast ett av många tänkbara sätt att organisera och strukturera information. Den struktur man går in för påverkar hur informationen rör sig i systemet och hur den presenteras och också hur vi uppfattar den. Detta måste man först som sist komma ihåg: datorn presenterar information om världen enligt en given tolkning, en modell, som ofta är ett resultat av mängder av medvetna och omedvetna val, ända sedan datorns barndom. Man bör alltså reflektera över hur saker egentligen hänger ihop, då man börjar samla eller arbeta med digitala material. Det vill säga hur saker hänger ihop i verkligheten enligt en själv, och hur de enligt systemet hänger eller borde hänga ihop. Vilka informationsbitar finns och vilka relationer finns det mellan dessa?

Då man under forskningsarbetet samlar på sig digital information, vilket är fallet de flesta gånger idag, är det viktigt att man försöker planera hur man ska organisera och hantera informationen. Ibland kan det finnas behov att samarbeta med it-kunniga personer för att klara av de tekniska utmaningarna och hitta goda lösningar. Det finns framför allt två saker som är avgörande för hur man lyckas med digitala resurser i humanistisk forskning: Den första är tillräcklig planering, den andra är ett kreativt och flexibelt grepp vid förverkligandet av planerna. Trots att det är svårt att förutsäga alla risker och problem måste man verkligen anstränga sig för att planera in i detalj, eftersom det är avgörande för hur man överlag lyckas med forskningen. Det beror på att finansieringen måste vara tillräcklig. Man måste ha en realistisk budget och en långsiktig planering.

Man måste allra först ägna sig åt research vad gäller andra liknande forskningsprojekt och lära sig av andras metoder, misstag och misslyckanden. Viktiga frågor är:

- Hurdan information ska samlas/bearbetas? Detta måste analyseras och struktureras
- Vad skall göras med informationen?
- Kan man beakta och underlätta eventuell senare återanvändning redan vid struktureringen? Kan den länkas till andra resurser? Finns det liknande material på andra håll som handlar om samma saker?
- Hur hantera frågor om autenticitet och proveniens? (Käll)kritik måste kunna göras både på och under den aktuella forskningen och helst senare också
- Hurdan metadata behövs, på vilken nivå och var ska den finnas?
- Vilka tekniska lösningar finns färdiga att få?
- Hur mycket teknisk utveckling behöver göras?
- Hur kan man säkra hänvisningar och reproduktion av processerna på sikt?
- Hur kan materialet och själva mjukvaran bevaras (backup) och återanvändas (tillgängliggöras) – också på lång sikt?
- Hur skall arbetet läggas upp? Ansvar, tidtabeller och arbetsfördelning
- Vad kan gå fel? Vilka risker, brister finns eller kan uppstå?
- Fundera noga över hur olika lösningar påverkar forskningsmetoden och resultaten

Det skulle vara mycket önskvärt att också forskare inom humaniora skulle erbjuda all data de samlat och använt fritt på webben inklusive all dokumentation. Förutom att detta är bra för att man på så sätt ger möjlighet för andra forskare att verifiera resultaten, bidrar man också till annan forskning genom att erbjuda källmaterial. Också av denna orsak måste man ha en bra dokumentation av hela systemet och vad det innehåller.

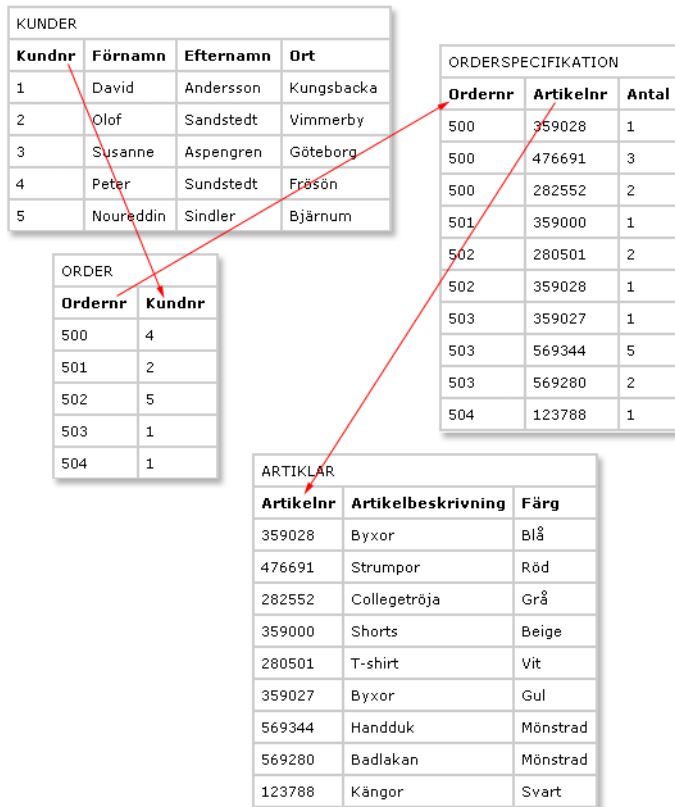
En viktig aspekt, utöver att kartlägga andra liknande projekt och de modeller och standarder som använts i dem, är dessutom att fundera över om man kan utnyttja redan existerande data på annat håll. Går det att berika den egna informationen genom att länka till andra resurser? Sådant kan göras på många sätt. Det kallas *länkad data* (linked data) och kan ge betydande mervärde för den egna eller annan framtida forskning. Kan man till exempel länka i materialet förekommande personnamn till exempel till något auktoritetsregister? För länkning av data finns olika standarder. Geografisk information finns öppet tillgänglig på exempelvis [Google Maps](#), vilket erbjuder möjligheter att med rätt enkel teknik kunna presentera informationen grafiskt på en karta. Sådana lösningar kallas *mash up* och går ut på att man kombinerar olika resurser och tjänster. Särskilt inom arkeologin har man redan länge använt sig av geografisk information och tredimensionella modeller. Det finns program där man kan lägga kartor och annan information i lager på varandra, vilket kan vara till mycket stor hjälp för att gestalta samband eller processer.

Databaser

Jessica Parland-von Essen

Det finns många olika sätt att strukturera information. Renodlade traditionella relationsdatabaser är ett. I dem ordnas informationen i tabeller, där varje värde får ett eget id, som man sedan hänvisar till i andra tabeller. Gemensamt för dem är att man försöker ordna informationen på ett sådant sätt att den kan hanteras rationellt och effektivt, så att samma uppgift inte upprepas flera gånger utan att man bara kan hänvisa till rätt ställe vid behov. På det sättet förhindrar man till exempel att man måste göra mångdubbelt arbete. Man kommer alltså i ett tidigt skede in på frågor om begrepp och klassificering, som mycket snabbt får konsekvenser för både forskningsprocessen och resultaten. Det är därför viktigt att man undviker tunneltänkande och använder sig av olika "etiketter" för saker utan att reflektera och analysera varje begrepp. En viktig princip är att hellre sönderdela informationen i för många typer än för få. Det är

nämligen alltid enklare att slå ihop information än att i efterhand börja dela och sortera i klasser. Sådant kräver ofta mycket manuellt arbete.



Exempel på hur en enkel relationsdatabas kan se ut. Källa: Webdesignskolan, "MySQL och databaser", <http://ramp.hostingsiteforfree.com/Webdesignskolan/mysql/mysql.htm> (2013-09-26).

Om en person till exempel byter namn, kan man då göra ändringen bara på ett ställe, så att det nya namnet sedan syns både i det som ser ut som "Telefonbok" och "Adressbok". Information kan alltså sammanställas till olika typer av "fönster", som plockas ihop av uppgifter från vitt skilda ställen i datasystemet, men visas i en specifik "vy". Ett system med till exempel en sökvy och en resultatvy och visning av enskilda "poster" kan tillsammans utgöra vad man kallar användargränssnitt. Det är ett ställe där ett datasystem möter en

människa. Ett vanligt och konkret exempel är en bibliotekskatalog på webben (ofta kallade OPAC, Online Public Access Catalogue), men i praktiken är ju allt man ser på sin skärm strängt taget användargränssnitt. Gränssnitt finns också ofta mellan olika datasystem, där information kan röra sig mellan de olika systemen.



Användargränssnitt kallas det verktyg som ger användaren tillgång till databasens innehåll. Innehållet i databasen kan presenteras i olika "vyer" där information presenteras enligt på förhand programmerade anvisningar och de sökfrågor användaren via verktyget skickar till databasen. Skärmdump från Databasen Henrik, http://dbgw.finlit.fi/henrik/henrik_svenska.php (2013-09-26).

Många moderna användargränssnitt använder sig ofta av webbkod för att formulera grafiken, men också andra möjligheter finns. För att informationen skall löpa från gränssnittet till det underliggande datasystemet behövs kommandon, sökfrågor eller olika andra skript som berättar för datasystemet vad det ska göra. Ett vanligt språk är SQL, *Structured query language*, som används i relationsdatabaser. Då man kommunicerar via ett gränssnitt i en webbläsare måste man förpacka SQL-koden i en annan kod; är det andra typer av gränssnitt eller databaser behövs andra språk. Då man använder databaser och datasystem vid forskning är det mycket relevant hur dessa kommandon ser

ut. Om man endast sparar rådata räcker det inte för att belägga forskningsresultaten, eftersom exempelvis sökfrågorna ("ta alla enheter som innehåller värdet x och y där värdet för y är mindre än 1790 och räkna dem och visa summan") de facto är en del av forskningsmetoden, om man går ut med siffran man fått som ett forskningsresultat. Om det i denna kod finns felaktigheter eller brister i logiken kan svaren vara helt felaktiga. Eller snarare är frågorna felaktiga och forskaren får svar på andra frågor än han tror sig få besvarade. Det betyder att de informationssystem man använt måste dokumenteras noga antingen de bevaras som helhet eller inte.

För att sökningar skall fungera måste man ofta använda sig av normalisering eller tolkning av källorna. Normalisering betyder att man till exempel ändrar i stavningen, så att ord eller namn alltid stavas på samma sätt. Sådant kan ibland vara försvarbart, men man måste komma ihåg att man samtidigt korrumpierar informationen i källan. Alltså måste man vara mycket tydlig med att man gjort detta. Det kan vara bra att göra sådant ändå ibland. Tidsuppgifter är en sådan typ av information, att det kan vara försvarbart att av ekonomiska skäl helt kallt normalisera datumangivelser. Då gör man en tolkning redan vid skapandet av den digitala resursen som inte kan kontrolleras annat än mot originalet.

Ett bättre men mer resurskrävande sätt är att ange både den ordalydelse och formulering som finns i originalet och den tolkning som behövs för att uppnå god sökbarhet och funktionalitet. Detta kan lösas genom att ge varje värde ett eget id-nummer som man sedan hänvisar till i samband med uppgiften. Problemet är att det ofta finns en hel del osäkerhet vid identifieringen. Till exempel i Helsingfors fanns under slutet av 1700-talet två handelsman Lampa, varför det ofta är helt omöjligt att koppla ihop ett viss omnämnande av "handelsman Lampa" i en källa till en viss person. Ofta kan det vara Clas Lampa lika väl som Carl Lampa. Att hantera denna typ av osäkerhet är svårt i digitala sammanhang. Egentligen bör man hålla de "verkliga fysiska personerna" skilda från förekomsten av ett namn i en källa, och båda borde ha egna id:n. Dessutom har man ofta ett stort antal namnvarianter att hantera.

Om man lyckas länka sina egna resurser till någon annan resurs på webben, till exempel en ontologi är det särskilt bra, eller om man själv lyckas strukturera sin information på ett sådant sätt. Länkning sker genom att man anger ett id eller helst en bestående webbadress till en särskild uppgift i en annan resurs. En ontologi är en resurs där man organiserat begrepp så att relationerna mellan begreppen finns sparade på ett sådant sätt att en maskin kan använda strukturen till exempel vid sökningar. Ett enkelt exempel är geografiska namn. Ta stadsdelen Hagnäs i Helsingfors, säg att den förekommer i relation till en viss uppgift i ditt material. Säg att du sedan har liknande geografisk information om tusentals andra enskilda uppgifter. Du vill kanske i framtiden jämföra uppgifter från Helsingfors och Ekenäs. Nå väl, för att kunna göra det borde du förutom "Hagnäs" i informationen också uppge "Helsingfors" så att uppgiften kan hittas för en jämförelse. Men du vill kanske också jämföra alla uppgifter från Österbotten med alla uppgifter från Nyland. Alltså måste du också ange "Nyland". Det betyder att man är tvungen att upprepa samma textsträngar tusentals gånger.

Det säger sig självt att det inte är särskilt effektivt eller rationellt. I stället kunde man ha en annan resurs, en ontologi, där man räknat upp alla ortnamn och hur de förhåller sig till varandra: "Nyland" = "Helsingfors" + "Esbo" + "Lovisa" etc. Vidare kan man ange att "Helsingfors" = "Hagnäs" + "Sörnäs" + "Kronohagen" etc. Då räcker det att från varje enskild uppgift i ditt material peka på en enda geografisk information. Datasystemet kan själv räkna ut att om du vill ha alla uppgifter från "Nyland" hör också uppgiften från "Hagnäs" dit. Dylika resurser finns och är i många fall tillgängliga på webben. Sådana finns över mängder av olika typer av begrepp på olika språk, vilket ger möjligheter till mycket effektiva sökningar också över språkgränser i vissa fall. När man väljer begreppsontologier måste man analysera dem noga, så att man är säker på att de motsvarar ens världsbild och begreppsapparat. Man bör komma ihåg att ontologierna är tolkningar och modeller av hur världen är konstruerad, inga absoluta sanningar. Det finns kulturella och disciplinära skillnader som kan vara

mycket stora. Väljer man en ontologi som ur ens eget perspektiv sett innehåller tankefel, blir kvaliteten av forskningsresultaten lidande!

Att konstruera databaser för forskningsändamål är ingen enkel konst. Det kräver vana att skapa modeller av information som är logiskt hållbara och rationella. Sådant utvecklingsarbete kräver nära samarbete mellan forskaren och it-programmerare och helst också en informationsspecialist. Men ansvaret för att se till att det finns tillräcklig teknisk dokumentation är i slutändan forskarens, forskaren själv måste kunna fråga efter den och informationsspecialisten kan möjligen hjälpa till med att definiera vilken dokumentation som är mest relevant.

Ofta finns det kommunikationsproblem mellan it-experten och forskare. I synnerhet humanister är ofta omedvetna om vad som ens i teorin är tekniskt möjligt och de kan därför inte ens be om det. Å andra sidan vet it-experterna inte alltid vad humanisterna egentligen vill göra eller är ute efter, varför de inte alltid kommer sig för att erbjuda olika lösningar. Dessutom är informationsteknologin ett mycket vitt område med otaliga olika typer av kompetenser gällande olika system och typer av programmering. Ingen it-kunnig kan allt. En grundregel är ändå för humanisten i svårare förhandlingssituationer att vad som helst är möjligt att göra i teorin, åtminstone med existerande information. Frågan är bara vad man är beredd att betala för olika lösningar. Oftast har man begränsade resurser och då är det mycket viktigt att kunna samarbeta nära och i god anda med it-experten, trots att det kan vara svårt att hitta ett gemensamt språk ibland. Det lönar sig att alltid be om konkreta exempel och om man själv har förebilder eller goda exempel att visa, ska man göra det! Man måste förklara vart man vill komma och vad man behöver göra.

Då man hanterar mycket stora mängder data finns det alltid en större risk för enstaka fel. Om databaser dessutom lever och fylls på gör man också ofta korrigeringar då man hittar felaktigheter. Databaser är alltså ofta genuina digitala texter i det att de inte kan återges meningsfullt på papper och att de ofta lever och förändras. Text som tagits fram ur sådana system är i ovanligt hög grad konstruktioner, resultat av komplicerade tekniska processer som är helt

osynliga för den som bara tittar på skärmen. Bakom den bilden finns många lager av tolkningar som går tillbaka ända till hur man skapat modellen och hur informationen motsvarar den verklighet den avbildar. En viktig fråga är om man skilt på namn och objekt och hur systemet hanterar olika varianter av språkliga begrepp och varianter. Finns dessa också representerade i systemet, eller måste den som använder systemet hantera dessa manuellt?

För att vara trovärdig måste information vara kopplad till annan information som berättar om proveniens och kontext. Detta är mycket viktigt då det gäller digital information. Vad, när, vem och framförallt *hur* är frågor som måste få svar i resursen. Detta måste gälla all information i ett system eller projekt, man måste försäkra sig om att data inte seglar fritt någonstans i systemet utan kontext och historia. Detta kräver i normala fall metadata. Många saker kan också förklaras i vidhängande dokumentation, såsom fältbeskrivningar eller kodningsmanualer, som också måste finnas tillgängliga. Man måste också kunna redogöra för principer vid tolkningar av oklara fall. Det är av största betydelse att sådant dokumenteras under arbetets gång så att man uppnår konsekvens i informationen och ger möjlighet till källkritiska bedömningar.

Ofta händer det dessutom att man använder assistenter vid inmatning eller bearbetning av data. Detta kan ibland vara förrädiskt om man inte följer upp arbetet mycket noga, eftersom det i själva verket många gånger kan vara helt avgörande för forskningens slutresultat hur en enskild assistent resonerat i tolkningsfrågor. Om man sedan dessutom använt sig av flera olika personer för arbetet utan mycket noggrann kollationering eller dokumentation, kan man plötsligt ha ett forskningsmaterial av sämre kvalitet än man tänkt sig. Utgångspunkten måste därför alltid vara att man framskrider *iterativt*, det vill säga stegvis, och i synnerhet i början måste man vara färdig att också ta några steg tillbaka emellanåt och göra om eller komplettera något. Även ett rutinarbete som kodning kan bli mer intressant och givande för den som gör det, vilket ju måste anses som ett plus för alla parter.

Det behövs oändligt mycket kommunikation mellan alla involverade parter och många gånger också teknisk personal. Det är å andra sidan ett minus: räkna

med oändliga möten och diskussioner om olika små detaljer – kom då ihåg att varje detalj kan vara av mycket stor principiell betydelse och att det är viktigt att eftersträva konsekvens. Det är i synnerhet detta som avses med att man måste vara kreativ och flexibel vid genomförandet av arbetet. Trots att man satt mycket tid på planering, måste man vara inställd på att planerings- och utvecklingsarbetet fortsätter under hela projektet. Man måste ständigt ta ställning till nya frågor och kanske till och med revidera sina planer. *Kill your darlings* kan vara den enda lösningen ibland, om något visare sig för dyrt eller ta för lång tid. Då gäller det att vara kreativ.

Big data

Kenneth Nyberg

”Big data” kan, som redan framhållits, ses som en samlande term för mycket av det som är nytt med digitala humaniora – eller för den delen det digitala samhället i stort. Enkelt uttryckt är de nya möjligheterna att utnyttja gigantiska datamängder en följd av dels tillgången på data i digital form (vare sig dessa är digitalbaserade eller digitaliserade), dels de alltmer kraftfulla datorer (*computers*, räknemaskiner) vi har till vårt förfogande för att bearbeta dessa data. Denna uppskalning av beräkningskraften är så omfattande och går så snabbt att man kan tala om en radikal förändring av vilken typ av frågor vi kan ställa oss och rimligen förvänta oss att få svar på.

Följaktligen uppfattas framväxten av digitala humaniora inte sällan som en förskjutning från en tyngdpunkt på kvalitativa metoder till en dominans för kvantitativa sådana. Inte minst inom historieämnet har just uppdelningen kvalitativ kontra kvantitativ metod ofta betraktats som grundläggande, där det stora genombrottet för de sistnämnda under 1960- och 70-talen inte välkomnades av alla ”traditionellt” arbetande historiker. Efter den s.k. kulturella eller språkliga vändningen inom ämnet under 1980- och 90-talen kan det tyckas som att pendeln nu är på väg att svänga ännu en gång, och vissa konfliktlinjer

från den tidigare debatten om kvantitativa metoder kan återigen urskiljas i diskussionen kring digital humaniora.

Det ligger säkert något i dessa positioneringar, vilka avspeglar att forskare har olika prioriteringar och är intresserade av delvis olika saker i sitt studium av det förflutna. Samtidigt ska man inte överdriva motsättningen mellan kvalitativ metod å ena sidan och kvantitativ å den andra; snarare handlar det om en skala utan några skarpa övergångar, ett spektrum där ett givet tillvägagångssätt kan placera sig närmare ena änden och en annan metod hamnar närmare den andra. En gemensam nämnare för nästan all forskning är, trots allt, att hitta mönster i data, att urskilja en signal i bruset, men det kan göras på olika sätt.

Det är också missvisande att kalla mycket av det som för närvarande väcker mest uppmärksamhet inom DH för "kvantitativ metod" i traditionell mening eftersom det ofta handlar om exempelvis text mining (se nedan) snarare än renodlade statistiska analyser. De nya verktygen används dessutom i många fall för att hitta intressanta ingångar i materialet snarare än för att skapa beräkningar vilka i sig ses som forskningens slutresultat. Detta arbetssätt, där man systematiskt växlar mellan empiri och teori, dvs. data och tänkandet kring data, kallas ibland för *abduktion*, vilket skiljer sig från induktion där man drar slutsatser utifrån empiriska data och deduktion där man formulerar hypoteser och teorier vilka sedan testas mot empirin.⁷²

Med allt detta sagt kvarstår det faktum vi började med, att en stor del av möjligheterna med DH – åtminstone som de uppfattas i nuläget – på många sätt är kopplade till användningen av stora datamängder. Dels sker det i form av utveckling av konventionella statistiska metoder som "bara" handlar om mer data och snabbare datorer, dels om helt nya arbetssätt som snarare handlar om kvantifiering av kvaliteter, dvs. att analysera egenskaper och relationer på grundvalen av mycket stora material. Nära förknippad med båda dessa utvecklingslinjer är ytterligare en central företeelse inom DH, nämligen visualiseringar, vilka behandlas i nästa avsnitt. Statistik i sig går vi inte in på i

⁷² Jfr Lev Manovich, "The meaning of statistics and digital humanities", *Software Studies* 2012-11-27, <http://lab.softwarestudies.com/2012/11/the-meaning-of-statistics-and-digital.html> (hämtad 2013-10-01).

detta sammanhang men några ord behöver sägas om det som kallas text mining, vilket också kommer att tas upp i separata fördjupningsartiklar.

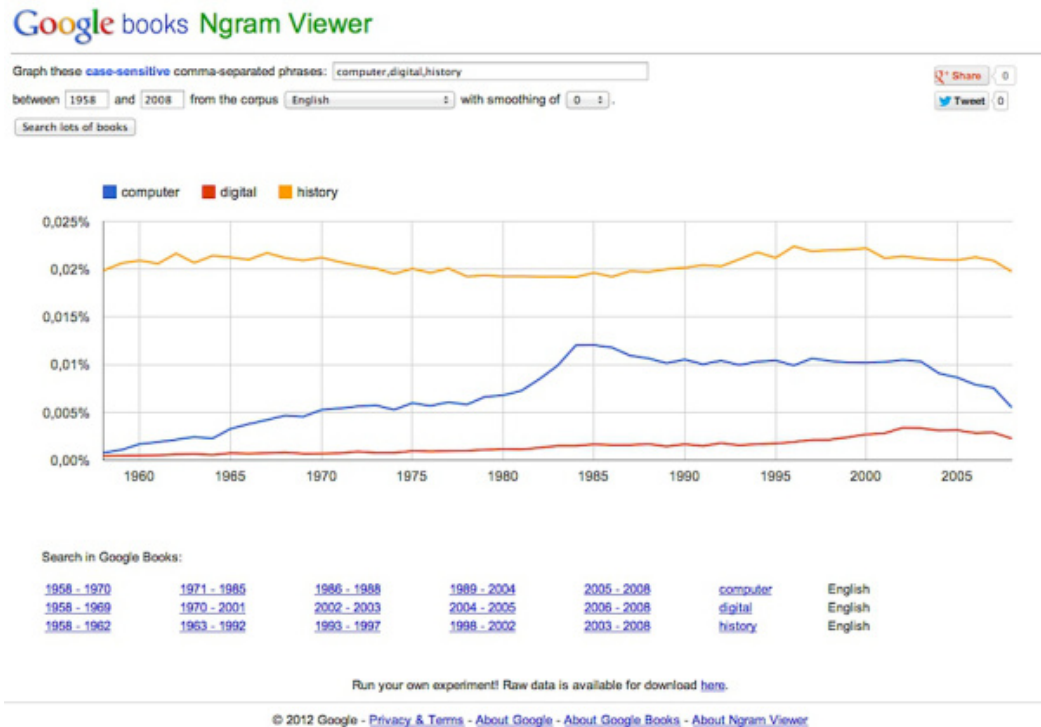
Text mining är en tillämpning av idén om "big data" på stora textmängder, snarare än exempelvis sifferdata. Det handlar om hur man med hjälp av såväl kvalitativ som kvantitativ databehandling kan analysera stora mängder digital(iserad) text, vare sig det är historiska eller litterära källor. En enkel form av text mining är att i stora korpusdatabaser (av korpus, textsamling) söka efter frekvensen av olika ord och hur den har förändrats över tid. Det går också att studera korrelationer av olika slag, i vilka sammanhang begrepp har använts historiskt (som det avspeglas i de analyserade texterna), vilka ord som tenderar att förekomma nära varandra och så vidare. Kvantitativa studier av stora mängder litterära texter kallas *distant reading*, där den italienske forskaren Franco Moretti – nu verksam i USA – är en pionjär, och *topic modelling* är en benämning på analyser av texters tematiska struktur som bygger på studiet av vilka begrepp som används, i vilka sammanhang de förekommer och hur de relaterar till varandra.⁷³

Ett känt och omdiskuterat projekt som bygger på text mining kallas Culturomics och baserar sig på de miljontals böcker som Google digitaliserat, där man genom frekvensmätningar försöker studera kulturella förändringar av olika slag.⁷⁴ Vem som helst kan också göra enkla sådana analyser i databasen genom webbapplikationen Google Books Ngram Viewer (ofta förkortat "Google Ngrams"). Den typen av studier kan vara mycket fruktbara, men som många har påpekat är det viktigt att fundera på vad de egentligen säger om djupare betydelser eller större historiska sammanhang och inte "bara" om orden eller tecknen i sig. Dessutom är det just i Googles fall ofta svårt att veta vilken datamängd det egentligen är man söker i, då den hela tiden förändras i en

⁷³ Om Moretti och distant reading se Kathryn Schulz, "What Is Distant Reading?", *New York Times* 2011-06-26, <http://www.nytimes.com/2011/06/26/books/review/the-mechanic-muse-what-is-distant-reading.html?pagewanted=all&r=1&> (hämtad 2013-10-02). För exempel på hur resultaten av topic modelling kan se ut se Manovich, "The meaning of statistics".

⁷⁴ Projektet introducerades i en omtalad uppsats i *Science*: Jean-Baptiste Michel m.fl., "Quantitative Analysis of Culture Using Millions of Digitized Books", *Science* vol. 331 no. 6014 (2011-01-14), s. 176–182. (Tillgänglig digitalt på <http://www.sciencemag.org/content/331/6014/176.abstract>.)

process som inte är särskilt genomskinlig. Ett exempel på verktyg för topic modelling som fått viss uppmärksamhet är Paper Machines av Jo Guldi, vilket rent tekniskt är en insticksmodul till referenshanteringsprogrammet Zotero.⁷⁵



Google Ngrams. Sökning i Google Ngrams som visar frekvensen av orden computer, digital och history mellan 1958 och 2008 i den engelskspråkiga delen av verktygets korpus, vilken totalt består av flera miljoner böcker på olika språk utgivna mellan 1500 och 2012. (Källa: Google Books Ngram Viewer, <http://books.google.com/ngrams>, hämtad 2013-01-15.)

I Sverige är det än så länge främst språkvetare och litteraturforskare som arbetat med text mining. En viktig resurs i det sammanhanget är Språkbanken vid Göteborgs universitet, en databas med svenska texter som innehåller ca en miljard ord och sträcker sig flera hundra år tillbaka men också innehåller material från ett antal nutida svenska bloggar. Andra svenska korpusdatabaser

⁷⁵ Google Books Ngram Viewer, <http://books.google.com/ngrams>, och Paper Machines, <http://papermachines.org> (båda hämtade 2013-10-02).

är Litteraturbanken och Svensk prosafiktion 1800–1900. Dessa används i första hand av litteraturvetare, som i dem kan undersöka olika frågor om exempelvis sociala nätverk i texterna, hur personerna i dem rör sig i rummet och så vidare. Principiellt sett finns det inga hinder för att utnyttja sådana databaser även för mer historievetenskapliga undersökningar, även om det är viktigt att vara medveten om deras begränsningar.⁷⁶ (Vi kommer att utveckla resonemanget om betydelsen av kritisk granskning i kapitlets sista avsnitt.)

⁷⁶ Språkbanken, <http://spraakbanken.gu.se>; Litteraturbanken, <http://litteraturbanken.se>; och Svensk prosafiktion 1800–1900, <http://spf1800-1900.se> (alla hämtade 2013-01-18).

Fördjupning: Digitala textarkiv och forskningsfrågor

Mats Malm

Digitaliserade och sökbara samlingar av texter har blivit en allt viktigare resurs för forskningen inom en lång rad discipliner. Tidigast byggdes sådana arkiv upp som textkorpusar för lingvistiska studier, men efter hand har både materialen som förs in i dem och teknikerna som gör dem sökbara utvidgats så att de blir användbara för allt fler frågeställningar. Ofta har man märkt upp materialet i korpusarna för att hjälpa analysen, men efter hand som arkiven blivit mer omfattande har behovet uppstått att utveckla metoder som gör det möjligt att utvinna så mycket information som möjligt ur materialet utan att behöva ägna alltför mycket tid åt att märka upp det.

Internationellt har man på senare tid kommit fram till hållningen att så stora mängder material faktiskt har blivit digitaliserade, att utmaningen nu är att utveckla metoder som gör materialen tillgängliga på effektivare sätt. Det är en formidabel uppgift att strukturera en stor samling texter så att den blir begriplig och navigerbar. Metoderna att utvinna information ur och om textmaterial som är större än en människa kan hinna läsa har kallats *distant reading*, fjärrläsning i stället för närläsning.⁷⁷ Den kommer aldrig att ersätta närläsning och textanalys, men den ger nya ingångar till både kultur och historia. Ett av de redskap som har diskuterats flitigt de senaste åren är [topic modeling](#): metoder att urskilja teman på grundval av vilka ord som förekommer i närheten av varandra. På så vis kan man maskinellt fånga upp tematiska sammanhang inom texter och framför allt mellan texter i stora material.

Urval och representativitet

En av de grundläggande frågorna kring användningen av textkorpusar är vad materialet egentligen är representativt för, dvs. vilka slutsatser man faktiskt kan

⁷⁷ Termen myntades av Franco Moretti: se hans *Graphs, Maps, Trees. Abstract Models for Literary History* (London: Verso 2005) och *The Novel. History, Geography, and Culture* 1–2 (Princeton: Princeton UP 2006).

dra av det. I vissa projekt digitaliseras helt enkelt alla böcker i en samling eller ett bibliotek: man får då ett stort material av mycket olikartade texter. De resultat man får i en sökning kommer alltså ur mycket olika kontexter och behöver tolkas utifrån det, och frågan uppstår vilka material som inte är representerade i urvalet. Andra projekt kan fokusera på till exempel ett författarskap: det kan säga mycket om sin samtid, men det är knappast representativt för den. Flera projekt styrs implicit eller explicit av ett klassikertänkande: man samlar i första hand verk och författare som uppfattas som viktiga av en eller annan anledning. Sådana korpusar kan vara mycket givande, men de ger inte en representativ bild av vilka frågor som faktiskt gestaltades i litteraturen vid en bestämd tidpunkt eller i ett bestämt sammanhang. Snarare riskerar de att vidareföra den kulturella självbild som har etablerats över tid. Kulturarvet består ju till sin kärna av sådant som äldre tider velat befästa sin identitet med, och därmed också velat vidareföra till senare tider. I den meningen är kulturarvet avsett att forma vår identitet. Men till vårt litterära kulturarv hör ju också det 'ofrivilliga' arvet: de texter och röster som har glömts bort eller rent av marginaliserats. Kan vi få fram dem, kan vi nå en mer representativ bild av historien och dessutom få bättre överblick över hur kulturarv och kanon fungerar.

Alla slags texter är av potentiellt intresse för humanistiska och samhällsvetenskapliga studier, men fokuserar man på kulturarv och kanon blir skönlitteraturen ett ovärderligt källmaterial i sin egenskap av spegel för samhället. Det är inte en okomplicerad spegling: ibland är litteraturen före sin tid, ibland efter, ibland styrs den av maktens intressen och ofta påverkas den av ekonomiska hänsyn. Men också dessa förhållanden gör att den säger något väsentligt om samhället, så länge den behandlas källkritiskt. Särskilt romanen och novellen tydliggör inte bara kulturen i snäv bemärkelse, utan kan visa vad som rör sig i samhällets inre. Det kan gälla vilka samhällsfrågor som problematiseras, hur frågor kring världsbild, kön, nationell identitet eller det främmande bearbetas, vilka konsumtionsvanor som kommer till uttryck, hur ord och begrepp förändrar sina innebörder eller vilka estetiska föreställningar som

utprovas i teori och praktik. I skönlitteraturen pågår ständiga förhandlingar mellan gamla och nya värderingar, mellan gamla och nya teknologier, mellan gamla och nya möjligheter. Men skönlitteraturen inte bara speglar samhället utan sätter också sitt avtryck på det. Konsumtionsvanor kan etableras i litteraturen på ett sätt som skapar nya levnadsmönster.⁷⁸

Prosafiktionen kan användas som källmaterial inom en mängd olika humanistiska och samhällsvetenskapliga discipliner, men den har det problemet att den är svårtillgänglig för historiskt inriktade studier. Den sociolog, etnolog eller idéhistoriker som vill veta hur rasbiologiska, nationalistiska, politiska etc. föreställningar kom till uttryck kring år 1900 kan inte nöja sig med de romaner som levt kvar i kanon. Det kan finnas många viktiga yttringar hos kända författare som August Strindberg, Selma Lagerlöf eller Hjalmar Bergman, men ofta är de ändå undantag som inte säger mer än en del om vad människor i allmänhet läste och vilka frågor som gestaltades i litteraturen vid en viss tidpunkt. Yttringarna hos mindre kända författare kan vara mer representativa men är betydligt svårare att hitta.

På svenskt område har man tillgång till digitaliserad och sökbar skönlitteratur främst hos [Projekt Runeberg](#) och [Litteraturbanken](#): de arbetar efter helt olika principer både beträffande urval och textetablering, men gemensamt för båda är att de har en rad olika slags material där inte allt, men en stor del, är sådant som brukar betecknas som klassiker. Som ett försök att i stället etablera ett representativt urval finns nu också pilotprojektet [Svensk prosafiktion 1800–1900](#), där man i stället får tillgång till all svenskskriven prosafiktion som utgavs för första gången åren 1800, 1820, 1840, 1860, 1880 och 1900. Det handlar om sammantaget 300 verk, och tanken är att man där skall kunna få en mer representativ bild av vilka olika frågor som faktiskt gestaltades och problematiserades i prosafiktionen – enligt dessa kriterier -- ett visst år, och kunna göra jämförelser över tid. Får man tillgång till motsvarande

⁷⁸ Se *Historier. Arton- och nittonhundratalets skönlitteratur som historisk källa*, utg. Christer Ahlberger et al. (Göteborg: Institutionen för historiska studier, Göteborgs universitet 2009) och *Moderna historier. Skönlitteratur i det moderna samhällets framväxt*, utg. Henric Bagerius och Ulrika Lagerlöf Nilsson (Lund: Nordic Academic Press 2011).

material från andra språkområden kan man också göra jämförelser internationellt. Webbplatsen är öppen för alla och skall göra det möjligt att pröva tekniker och metoder att se kulturarvet och historien från nya aspekter, genom de bortglömda och marginaliserade författarna i stället för att följa den etablerade kanon.

Metoder

Så är frågan vilka metoder man kan använda för att strukturera materialet. I regel kan man göra basala sökningar i sådana textarkiv, men det går också att tillämpa en uppsättning mycket mer sofistikerade verktyg på Svensk prosafiktion 1800–1900 och Litteraturbanken genom att studera dem som självständiga korpusar i [Språkbankens](#) portal för korpusar: [Korp](#). De verktygen är under kontinuerlig utveckling, och därtill arbetar vi med att utveckla metoder för *topic modeling* som ger möjlighet att kartlägga samband i stora material.

De potentiella felkällorna i detta digitaliserade textflöde är förstås många, men de potentiella möjligheterna är också stora. Två amerikanska skandinavister, Peter Leonard och Timothy Tangherlini, har utvecklat en särskilt fruktbar form av *topic modeling* som låter dem definiera "teman" utifrån ett välkänt verk och sedan identifiera sammanhang och påverkan på större material. På så vis kan de till exempel visa hur Darwins teorier spred sig från den första översättningen av *On the Origin of Species* till skönlitterära klassiker som vi redan kände till men också till skönlitteratur som blivit marginaliserad, hur de spred sig till exempelvis kriminalvården och, inte minst, hur de spred sig till historieskrivningen och alltså gav upphov till nya sätt att förstå och tolka den danska historien.⁷⁹ På så vis kan man kontrollera äldre föreställningar, nyansera dem och ibland göra betydelsefulla korrigeringar som preciserar vårt vetande och ger en fullständigare bild av ett historiskt skede.

På så vis kan textarkiven effektivisera traditionella typer av undersökningar och ge upphov till helt nya frågeställningar, gärna på ett sätt som inbegriper

⁷⁹ Peter Leonard och Timothy Tangherlini, "Trawling in the Sea of the Great Unread: Sub-Corpus Topic Modeling and Humanities Research", under publicering i *Poetics*.

tvärvetenskapligt utbyte av perspektiv och metoder. Än mer tankeretande, och svårförutsägbar, är möjligheten att låta maskinerna identifiera tematiska sammanhang som *inte* ingår i vår förförståelse av historien. Det gör det möjligt för oss att låta ordens nya flyktiga natur föreslå helt nya sammanhang för oss, sammanhang som vi inte alls har varit uppmärksamma på. Sådana ansatser kan i bästa fall ge oss möjligheten att komma förbi en del av våra egna begränsningar och blinda fläckar.

De digitala materialen i sig själva erbjuder förstås problem, inte minst genom brister i ocr-läsningen och metadata: går de inte att rätta till, måste de räknas in i metodiken som felkällor. De kvantitativa metoderna kan aldrig ersätta traditionella kvalitativa metoder, och de innebär hela tiden en risk att perspektivet blir för snävt. Men just anläggandet av perspektiv är en av vetenskapens grunder och riskerna är till för att hanteras. Det finns goda skäl att tro på möjligheten till nya fruktbara kombinationer av kvantitativ och kvalitativ analys.

Fördjupning: Kulturomik: Att spana efter språkliga och kulturella förändringar i digitala textarkiv

Lars Borin och Richard Johansson

Vilka vetenskapliga frukter kan vi skörda av de ansträngningar som gjorts att digitalisera text från olika tidsepoker? En av möjligheterna som öppnas är att kvantitativt studera hur språket i materialet förändras över tiden. Detta ger oss förstås en bild av hur själva språket förändras men även av hur händelser i världen och samhället påverkar det som skrivs. I en omtalad artikel i Science (Michel et al. 2011) beskriver ett forskarlag från Google och några universitet hur man kan utnyttja de stora mängder böcker som Google har digitaliserat för att göra storskaliga kvantitativa undersökningar av språklig och kulturell förändring över perioden 1800–2000. Studien lanserades under rubriken *culturomics* – kulturomik – i analogi med *genomics* och *proteomics*, beräkningstunga, massivt databeroende ansatser inom molekylärbiologi.⁸⁰

Kulturomikartikeln ledde till en livlig metoddiskussion, där man bland annat påpekade att författarna verkade helt ovetande om de språkliga aspekter som skulle behöva hanteras när man skalar upp den här typen av undersökning från traditionell mänsklig 'närläsning' till helautomatisk bearbetning av stora textmängder.⁸¹ En människa som behärskar språket har inga problem med att föra samman olika böjningsformer eller stavningsvarianter av samma uppslagsord (t.ex. förstå att *telegrafen* och *telegrafer* hör hemma under *telegraf*), att skilja homonymer åt (t.ex. inse när *friser* anger ett folkslag och när det handlar om ett slags utsmyckningar) eller att tolka när situationer beskrivs från olika utgångspunkter (t.ex. ha klart för sig att när man säger att *misstag har begåtts* vill man förmedla en annan bild av det skedda än om man säger att *vi har begått misstag*).

⁸⁰ Se även <http://www.culturomics.org>.

⁸¹ Se t.ex. Mark Libermans artiklar på Language Log: <http://languagelog.ldc.upenn.edu/nll/?p=2848> och <http://languagelog.ldc.upenn.edu/nll/?p=4456>.

I de metodologiska anmärkningarna döljer sig dock en möjlighet. Den ursprungliga kulturomikundersökningen förfogade över ett enormt material, nästan 5,2 miljoner böcker eller över 500 miljarder ord. Även det minsta delmaterialet, det hebreiska (som inte användes alls i artikeln), omfattade c:a 2 miljarder ord. Detta är viktigt, eftersom låg grad av språklig analys i viss mån kan kompenseras av att man har ett mycket stort material. Omvänt kan man förvänta sig att bra verktyg för automatisk språkanalys kan göra att man uppnår jämförbara resultat även med mindre materialmängder. Detta är en av förutsättningarna för ett svenskt kulturomikprojekt som bedrivs med ett rambidrag från VR (Borin et al. 2013).⁸² Nedan diskuterar vi de möjligheter vi ser när det gäller att utföra den här typen av studier på svenskspråkigt material.

Språkbankens textsamlingar och sökverktyg

Språkbanken (<http://spraakbanken.gu.se>) är en forskningsenhet vid institutionen för svenska språket på Göteborgs universitet. En av Språkbankens viktigaste verksamheter är att samla in svenskspråkiga textsamlingar (*korpusar*) och lexikonresurser, och göra dem tillgängliga för allmänheten. Dessa resurser kommer från en rad olika tidsperioder, från de äldsta medeltida lagtexterna fram till nutida material som nyhetstexter och texter från sociala medier.

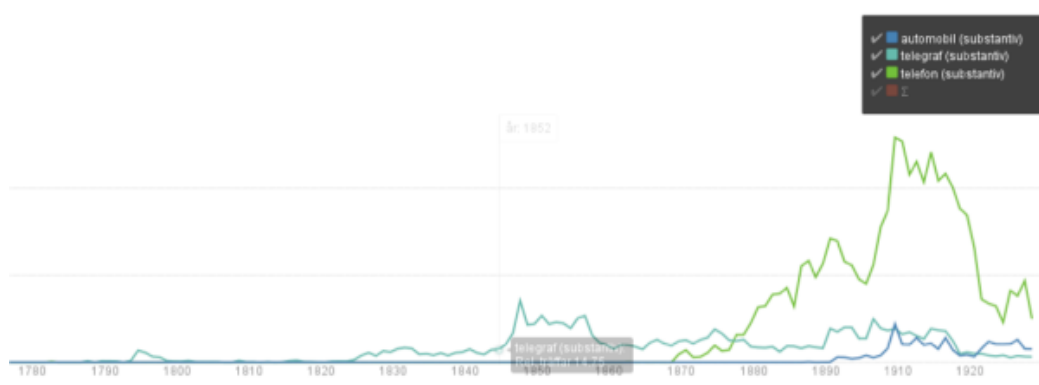
Många av Språkbankens korpusar innehåller information om texternas tillkomsttid. Detta gör det möjligt att söka i materialet och studera hur det förändras över tiden. Ett intressant exempel på detta är *KB-materialet*, ett omfattande textmaterial som kommer ur Kungliga bibliotekets skorskaliga digitalisering av historiska dagstidningar. Hittills har en stor mängd svenskspråkigt tidningsmaterial från sent 1700-tal fram till tidigt 1900-tal digitaliserats. Tidningstexterna kommer från 20 olika tidningar, framför allt landsortstidningar, och huvuddelen av materialet är från andra halvan av 1800-talet. Materialets exakta omfång i ord är svårt att ange på grund av ojämn kvalitet i digitaliseringen (se nedan), men det handlar om knappt 48.000 tidningsnummer omfattande totalt omkring 700 miljoner ord.

⁸² Se även <http://spraakbanken.gu.se/eng/culturomics>.

För att söka i Språkbankens textsamlingar används sökverktyget *Korp* (<http://spraakbanken.gu.se/korp>). Detta verktyg kan användas till exempel för att söka efter enstaka ord eller ordkombinationer och deras sammanhang (*konkordans*) och för att jämföra ords frekvenser. För den som är intresserad av språkbrukets förändring över tiden finns möjligheten att använda *trenddiagram*, som visar ordens förekomstfrekvenser år för år. Nedan visar vi ett antal användningar av trenddiagrammen, framför allt genom sökningar i KB-materialet.

Exempel på enkla tidssökningar

Neologismer (nya ord) kan ge oss intressanta perspektiv på den tid där de uppstår. Under 1800-talet sker det en hel del tekniska förändringar i Sverige, vilket vi ser avspeglas i tidningsmaterialet. Några exempel på detta är orden *telegraf*, *telefon* och *automobil*. Telegrafen uppfanns i slutet av 1700-talet och nämndes i den tidens svenska tidningar, men fick ett medialt genomslag först när den blev praktiskt användbar i och med Morses elektriska telegraf från 1837. Telefonen uppfanns av Bell 1876 och blev därefter snabbt populär i Sverige. Ytterligare en teknisk uppfinning från denna tid är automobilen, som förekommer i slutet av perioden.⁸³ Nedanstående figur visar resultatet av en sökning med Korp i KB-materialet efter dessa tre ord.

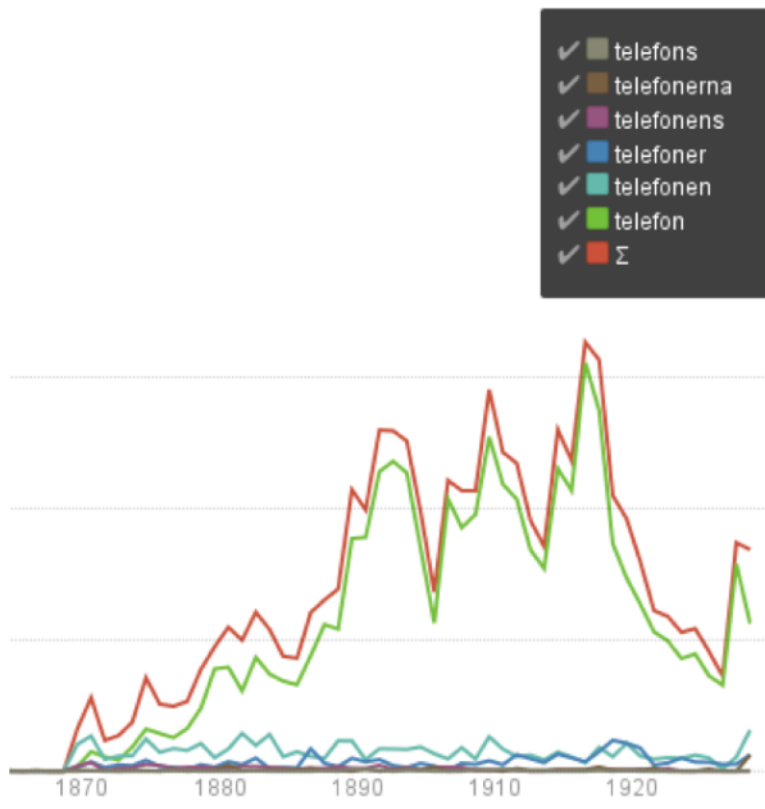


⁸³ De teknikhistoriska detaljerna är hämtade från Tekniska museets webbsidor: <http://www.tekniskamuseet.se>.

Även ur idéhistoriskt perspektiv är 1800-talet intressant att studera på en tidslinje. Till exempel orden *kommunistisk*, *kommunism* och *kommunist* nämns för första gången 1841. Därefter finns det tre perioder då dessa ord förekommer ofta i tidningsmaterialet: under revolutionsperioden runt 1848 (då dessutom *Kommunistiska manifestet* författades), under perioden runt Pariskommunen 1871, samt en kraftig ökning i slutet av tidslinjen, vilket sammanfaller dels med ryska revolutionen och dels med att det svenska kommunistiska partiet bildades genom en utbrytning från socialdemokraterna.



I båda dessa fall har Språkbankens lexikon och språkverktyg använts för att föra ihop textorden till lexikonord. Följande diagram visar hur de olika böjda formerna av ordet *telefon* förekommer i materialet. Även om formen *telefon* är absolut mest förekommande, blir det ändå en märkbar skillnad i det sammanlagda antalet förekomster (den översta, röda kurvan: Σ), som avspeglar summan av de olika formernas frekvens. Särskilt när antalet förekomster är lågt, som till vänster i diagrammet, kan den här typen av språklig bearbetning hjälpa oss att få ut mer av materialet.

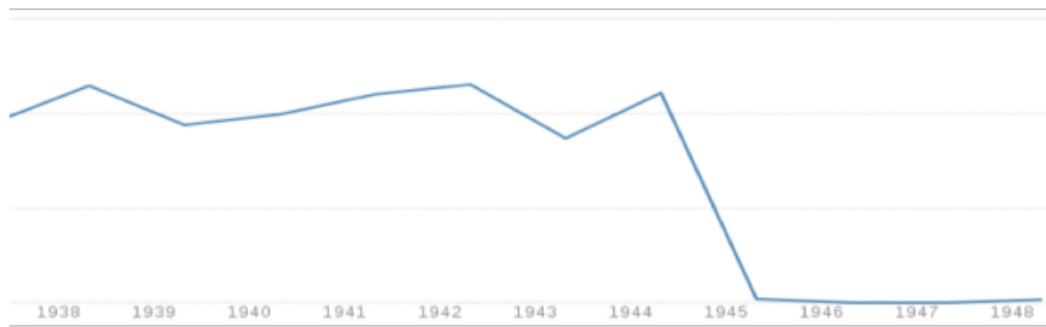


En annan intressant idéhistorisk tendens under 1800-talet är framväxten av den rasbiologiska forskningen, och detta påverkar också det allmänna språkbruket. För att ta ett exempel kan vi söka efter uttryck av typen *rasen* föregånget av ett adjektiv. De två vanligaste uttrycken av denna typ i KB-materialet är *gula rasen* och *hvita rasen*. Vi ser att sådana uttryck kommer i bruk under andra halvan av 1800-talet, vilket också passar bra ur ett idéhistoriskt perspektiv: detta var efter att Retzius metoder för skallmätning presenterats på 1840-talet och inflytelserika verk som Gobineaus *Essai sur l'inégalité des races humaines* (1853) och Darwins *On the origin of species* (1859) publicerats.



Den uppmärksamme kanske undrar hur det kommer sig att vi ser en topp på 1830-talet. Dessa träffar kommer av felaktigheter i arkiven som beror på misslyckad digitalisering, vilket vi kommer att diskutera i mer detalj nedan.

Språket förändras över tiden, inte bara som vi sett ovan ordförrådet utan också grammatiken. Ett exempel på detta är att svenska verb förr böjdes efter *numerus*: de hade en singularform (t.ex. *jag är*) och en pluralform (t.ex. *vi äro*), liksom flertalet europeiska språk fortfarande har. Pluralformerna försvann ur det svenska skriftspråket i början och mitten av 1900-talet. Denna på sin tid kontroversiella förändring skedde gradvis i skönlitteraturen, medan de flesta tidningar övergav pluralformerna någon gång i perioden 1943–1945. Detta ser vi tydligt i korpusen ORDAT, som består av Svenska Dagbladets årsböcker från åren 1923–1945 samt 1948 och 1958: här sker övergången mellan 1944 och 1945. Vi sökte efter sju av de vanligaste distinkta pluralformerna (*äro*, *voro*, *kommo*, *fingo*, *gingo*, *sutto*, *lågo*) och resultatet visas nedan.



Relation till humanistisk (och samhällsvetenskaplig) forskning

Man kan invända att exemplen i föregående avsnitt inte tillför någon ny kunskap, utan enbart ytterligare bekräftar vad vi ändå vet på annat sätt. Det är naturligtvis sant, och våra exempel, liksom de som anfördes av Michel et al. (2011) i den ursprungliga kulturomikartikeln, förstås kanske bäst om man tänker på dem ungefär som kalibrering och verifiering av ett mätinstrument. I och med att metoden faktiskt ger rimliga utslag för kända fakta, kan vi med viss tillförsikt ge oss på att använda den för att söka efter ny kunskap, t.ex. mer

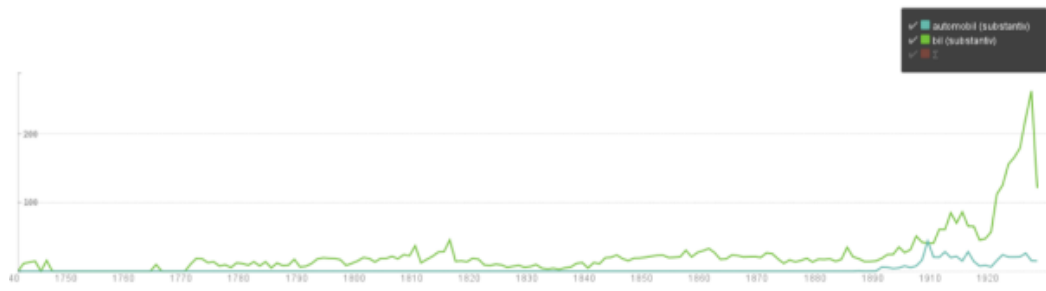
En särskild utmaning ligger här i att utveckla metodologi och verktyg som på ett enkelt sätt låter forskare röra sig mellan kulturomikens storskaliga kvantitativa studier och den traditionella humanistiska forskningens detaljerade närstudium. Som ett litet embryo till detta kan man i Korps trenddiagram klicka på varje datapunkt och i en separat flik få upp en konkordans för just den datapunktens träffar i materialet, som i följande bild, som visar de 47 förekomsterna av *telefonen* för år 1896 i KB-materialet. Det är inte tekniskt svårt att införa möjligheten att gå vidare till den fullständiga texten från varje konkordansrad.

Tekniska utmaningar

132

Teckenigenkänning (OCR)

När vi gjorde de teknikhistoriska undersökningarna ovan, varför sökte vi på *automobil* men inte den moderna varianten *bil*? Kan vi för övrigt säga när *bil* blev vanligare? Nedanstående figur visar resultatet av en sökning efter de två varianterna.



Resultatet tycks paradoxalt: *bil* verkar förekomma under hela 1700- och 1800-talet, trots att denna kortform enligt Svenska Akademiens Ordbok började användas först runt 1900. En inspektion av träffarna visar vad problemet är. Till exempel i Dalpilen 1893 ser vi träffar som *bil jätter* (biljetter) och *bil hörd* (bli hörd), vilka båda är uppenbart felaktiga. Detta beror inte på att tidningsskribenterna var slarviga utan på att överföringen från papper till dator inte är felfri. Det känsligaste steget kallas *teckenigenkänning* eller *teckentolkning*, på engelska *optical character recognition* (OCR), och innebär att datorn ska tolka de inskannade bilderna av de tryckta tidningssidorna och avgöra vilka bokstäver de motsvarar. Detta är svårare för äldre text av flera skäl. Gamla tidningsexemplar är tryckta på tunnare och porösare papper så att trycket flyter ut eller slår igenom till baksidan, tidningsexemplaret kan helt enkelt vara slitet, OCR-programmet kan ha en ordlista som inte är anpassad till det äldre språket, och slutligen avkodar alla OCR-program överlag äldre fraktur betydligt sämre än både nyare fraktur och antikva. Vi ser mycket riktigt i KB-materialet att de äldre texterna ofta innehåller en betydligt högre andel feltolkningar.

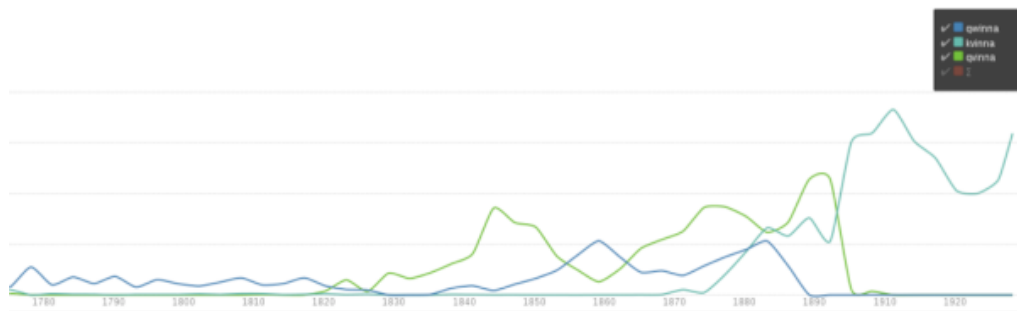
Så varför får vi så många felaktiga *bil* men inga *automobil*? Detta beror helt enkelt på att eftersom *bil* är ett kort ord så är det lätt att ha otur och få detta ord

vid en felläsning. *Automobil* är längre och det krävs därför betydligt mer otur för att få detta ord av misstag.

Stavningsvariation

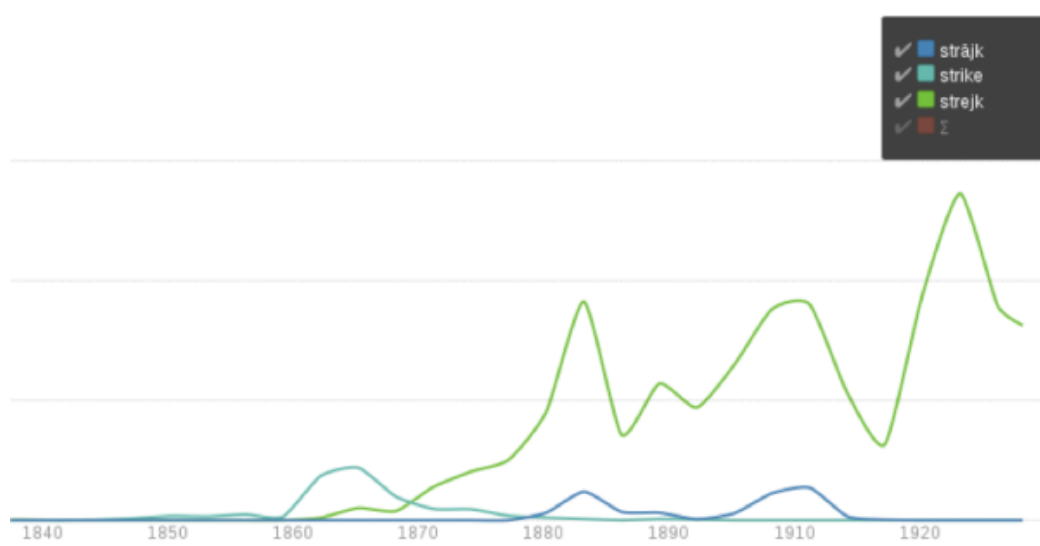
Anta att vi vill studera om ordet *kvinna* nämndes oftare i tidningarna under den tid då frågor som t.ex. kvinnors rösträtt började diskuteras. En sökning på detta ord i KB-materialet verkar vid en första anblick stödja hypotesen: från 1870-talet och en bit in på 1900-talet ökar detta ords frekvens markant. En närmare inspektion visar dock att vi har något problem med vår sökning eftersom det knappt förekommer någon träff alls innan 1870, och de enstaka som förekommer verkar vara sådana där teckenigenkänningen gått fel.

I det här fallet är förklaringen att ordet *kvinna* har stavats på flera olika sätt genom tiderna. På 1600-talet (t.ex. i korpusen *Stockholms stads tänkeböcker*) skrevs det ofta *quinna* eller *qwinna*, och om vi går över till KB-materialet (se figur nedan) så dominerar först stavningen *qwinna*, därefter *qvinna*, och den moderna stavningen *kvinna* först i och med stavningsreformerna i början av 1900-talet.



En viktig social förändring som sker under 1800-talet är framväxten av en organiserad och ideologisk arbetarrörelse, och spåren av detta kan vi också studera i den tidens tidningar. Det leder oss till ett annat ord vars stavning har varierat över tiden, nämligen *strejk*. I detta fall beror svårigheten på att det tagit ett tag innan stavningen av detta lånord stabiliserats. Strejker börjar nämnas i svenska texter under mitten av 1800-talet. I början används den engelska

stavningen *strike*; stavningen *strejk* tar över på 1870-talet. Alternativet *sträjk* levde kvar en bit in på 1900-talet.



Hur kan man göra det möjligt att genomföra sökningar av denna typ för den som inte är språkhistoriskt bevandrad? Språkbanken tillämpar två olika metoder för att hantera detta problem. Den första metoden är att använda *diakroniska lexikon*: ordlistor där vi helt enkelt kan slå upp att ordet *kvinna* tidigare har stavats *quinna*, *qwinna*, och *qvinna* (Borin och Forsberg 2011). Därmed kan vi även automatiskt ta med ordens alla stavningsvarianter i sökningarna. Detta fungerar väl när det gäller tidigmodern text, t.ex. från 1800-talet som i KB-materialet, då det förekommer ett litet antal standardformer. Den andra metoden baseras på *ungefärliga ordjämförelser*: vi kan säga att ett textord *qwenna* förmodligen motsvarar lexikonordet *quinna* eftersom *qw* är en stavningsvariant av *qu*, och ljudet *e* ligger nära *i*. Denna metod kan tillämpas vid analys av äldre text, t.ex. från medeltiden, där det inte ens är meningsfullt att tala om standardformer och antalet varianter är stort (Adesam, Ahlberg och Bouma 2012).

Lingvistiskt komplexa sökningar

De undersökningar vi hittills har visat har haft den begränsningen att de baserats på förekomst av enskilda ord, men det finns många sätt man skulle vilja undersöka frågeställningar som inte så lätt låter sig brytas ned till enkla ordsökningar. För att möjliggöra mer komplexa undersökningar finns en hel del olika lingvistiska analysverktyg.

Om vi till exempel vill undersöka vad man åt på 1800-talet (eller åtminstone vad tidningarna skriver om ätande) kan vi söka på förekomster av verbet *äta* och se vilka substantivobjekt det samförekommer med. För nutida material är detta relativt oproblematiskt: om vi till exempel söker i Göteborgsposten mellan 2001 och 2012 så ser vi att de vanligaste sakerna som man äter är *lunch, middag, kött, frukost* och *fisk*. För att avgöra vilket som är verbets objekt använder vi ett *syntaxanalysverktyg* (på engelska *parser*), och en *ordklassmärkare* kan avgöra om ordet är ett substantiv. Andra möjligheter är till exempel att använda en *namnuppmärkare* för att avgöra vilka typer av person- och ortnamn som omnämns.

Dessa lingvistiska analysverktyg är baserade på moderna ordlistor samt ordstatistik som insamlats genom att observera moderna texter, och att de är byggda för modernt språk gör att de har svårt att hantera äldre texter (Pettersson, Megyesi och Nivre 2012). Detta ser vi när vi söker på *äta* och dess objekt i KB-materialet. De vanligaste korrekta substantiven som vi hittar är *middag, frukost, kött, bröd* och *gräs*, alltså nästan detsamma som i det moderna materialet, men i topplistan finns också en hel del felaktigheter. Till exempel ser vi adverbet *deraf* (därav), vars stavning ställer till problem för ordklassmärkaren, och ett antal OCR-relaterade problem, exempelvis *stall* (från *åter skall*) *stola* (från *åter skola*). Metoder för att hantera språkliga genreskillnader (*domänanpassning*) är ett område som på sistone fått mycket uppmärksamhet inom den språkteknologiska forskningen, och det återstår att se om dessa metoder också kan användas för att hantera språkliga skillnader som beror på språkförändring över tid.

Sammanfattning

Textsamlingar där texterna innehåller information om tillkomsttid öppnar nya möjligheter för kvantitativa studier av språkhistoriska, kulturhistoriska och idéhistoriska frågor, med den nya forskningsmetodologi som kallas kulturomik. Detta ställer dock krav på att det finns användbara sökverktyg för att söka i den typen av samlingar på ett överskådligt sätt, och leder också till en hel del tekniska utmaningar och öppna forskningsproblem inom t.ex. teckenigenkänning, hantering av stavningsvariation, samt anpassning av språkteknologiska verktyg till äldre tiders språk.

Referenser

Yvonne Adesam, Malin Ahlberg och Gerlof Bouma (2012). *bokstaffua, bokstaffwa, bokstafwa, bokstaua, bokstawa...* Towards lexical link-up for a corpus of Old Swedish. *Proceedings of the 11th conference on natural language processing (KONVENS)*, 365–369. Wien: ÖGAI.

Lars Borin, Devdatt Dubhashi, Markus Forsberg, Richard Johansson, Dimitrios Kokkinakis och Pierre Nugues (2013). Mining semantics for culturomics: Towards a knowledge-based approach. *Proceedings of the 2013 international workshop on mining unstructured big data using natural language processing*, 3–10. New York: ACM. <http://dx.doi.org/10.1145/2513549.2513551>.

Lars Borin och Markus Forsberg (2011). A diachronic computational lexical resource for 800 years of Swedish. Caroline Sporleder, Antal van den Bosch och Kalliopi A. Zervanou (red.), *Language technology for cultural heritage*, 41–61. Berlin: Springer.

Jean-Baptiste Michel, Yuan Kui Shen, Aviva Presser Aiden, Adrian Veres, Matthew K. Gray, The Google Books Team, Joseph P. Pickett, Dale Hoiberg, Dan Clancy, Peter Norvig, Jon Orwant, Steven Pinker, Martin A. Nowak och Erez Lieberman Aiden (2011). Quantitative analysis of culture using millions of digitized books. *Science* 331: 176–182.

Eva Pettersson, Beáta Megyesi och Joakim Nivre (2012). Parsing the past – Identification of verb constructions in historical text. *Proceedings of the 6th EACL*

Workshop on Language Technology for Cultural Heritage, Social Sciences, and Humanities, 65–74. Avignon: ACL.

Fördjupning: Open research methods in computational social sciences and humanities: introducing R

Markus Kainu

Introduction – Open Research Methods

The debate on open science in the context of Social Sciences and Humanities (SSH) has been predominantly focusing on open access to research publication and opening up the various types of digital research data (open research data). The openness of research methods has received less attention.

I can think of two main reasons for that. On the one hand, research methods in SSH have predominantly been qualitative where software has played only a supporting role. Such research methods, let's take *discourse analysis*, have always been open, free to use and to modify and redistribute. On the other hand, the quantitative fields of SSH have mostly used statistics or survey and register data, or other, often closed, tailor-made data that custom proprietary data analysis tools such as SPSS, Stata or Excel are well suited for. However, the future of SSH looks somewhat different as the quantity and multiplicity of sources of digital data are challenging both traditional approaches in SSH the field, the purely qualitative approach and custom tools approach in quantitative analysis. The future that Gary King (2014, p. 166), the director of the [Institute for Quantitative Social Science](#) at Harvard University describes as:

An important driver of the change sweeping the field is the enormous quantities of highly informative data inundating almost every area we study. In the last half-century, the information base of social science research has primarily come from three sources: survey research, end-of-period government statistics, and one-off studies of particular people, places, or events. In the next half-century, these sources will still be used and improved, but the number and diversity of other sources of information are increasing exponentially and are already many orders of magnitude more informative than ever before.

In the data rich future of SSH research, as the role of software and computation becomes more central, the questions of licensing, ownership,

modification and distribution of that software will become increasingly important. This chapter will introduce one viable option for analysing your data called R.

What is R?

R is one of the most popular platforms for data analysis and visualization currently available. R is distributed under the terms of the *GNU General Public License* so it is free and open source and it can be distributed under those conditions. R is available from [Comprehensive R Archive Network](#) (CRAN). The name R comes from the first names of two New Zealand statisticians, Ross Ihaka and Robert Gentleman, who created the language in the late 1990s.

R can be regarded as an implementation of the S language which was developed at Bell Laboratories in the 1970s by Rick Becker, John Chambers and Allan Wilks (Venables, Smith, and Team 2013). R is an object-oriented programming language which means that unlike in SPSS or SAS that give you abundant information on a particular model you implement, R creates *objects* in memory that can be used in subsequent analysis. This structure of R directs the user to implement the data-analysis as stepwise process which becomes very useful later on when solving complex research problems using vast and messy data typical for emerging computational SSH research.

R user-interfaces

R runs in Windows, Mac OS X and GNU/Linux operating systems on a local computer, but different server implementations are becoming increasingly popular, such as [R-Fiddle](#) or [rnotebook](#). The most basic user interface for R is *console*, which allows the user to type in commands and outputs the results of the analysis. If the results is a plot a pop-up graphical window is opened. There are several graphical user interfaces (GUI) in R that may be helpful in the beginning, like [RCommander](#) or [Deducer](#). Perhaps the most productive way for using R is through an [integrated development environment](#) (IDE) that provides the user, in addition to console, several useful functionalities for controlling the

whole research project. [RStudio](#) has gained a lot of popularity in the last couple of years and is also my personal favourite IDE. It combines the console with script editor, plot browser, file browser and environment window. If the user uses plain text (latex or markdown) for typesetting the texts, RStudio has a tailored text editor and support for version control either in [git](#) or in [subversion](#). In addition, RStudio has native support for html-based presentation graphics using [reveal.js](#)-framework. All these operations makes it possible to squeeze the whole research process within a single software environments from planning to publishing. Rstudio can also be run on a remote server through a web browser. The RStudio company has another exciting open source tool for R called [shiny](#) that can be used for creating interactive web applications such as this [experimental gadget of mine](#).

Structure of R-project

For someone new to R, the peculiar structure of the language creates a very steep learning curve. The same applies to learning how the whole project is organised.

The official name [The R Project for Statistical Computing](#) refers both to the centrally maintained core as well as R's distributed structure of contributed extensions, called *packages*. Packages in R are collections of functions and/or data that are packaged for convenience. Installing a package broadens the functionality of your R installation. Basic R installation consists of so called *base installation* that includes the core with some 25 additional *packages* for the most basic functionality. The core of the language is maintained by *R Development Core Team*, but the additional packages are developed and maintained by individual developers and research institutes. R users often create packages for themselves, but if one thinks the package could be useful for other users too, the packages can be distributed through repositories.

CRAN is the "official" repository for contributed packages and currently hosts 5,150 packages that can be used to extend R. In the last couple of years various code hosting sites such as [GitHub](#) have become increasingly important resources

especially for collaborative development of new packages. Github currently hosts roughly 1,500 packages for R. [Bioconductor](#) is another separate package repository, but can be regarded as *domain specific* for it hosts packages for *the analysis and comprehension of high-throughput genomic data*. Other such domain specific projects are for example [rOpenSci](#) and the emerging [rOpenGov](#) that provide tools for open science and open government data, respectively.

Learning the language

As the internet has brought together the vast community around R, the internet has become the main channel for delivering instructions for R. The official *Introduction to R* by Venables, Smith, and Team (2013) is an important document to master when getting into the language. Besides this general introduction R-project has also a domain specific structure where you can start learning from so called [task views](#). For SSH researchers the [social sciences](#) and [natural language processing](#) task views are good places to begin with.

Discussions and announcement on R happen mainly through [R official mailing lists](#) that have their own lists for development and user help. [R help](#) is the main list for general help and receives tens of mails per day. Most of the individual packages have their own mailing list for development where anyone can join if wanting to contribute to the packages.

The official mailing lists have recently been challenged by so called Question & Answer -sites like [Stack Overflow](#) in delivering solutions for one-off user questions. Stack Overflow has currently almost 47,000 questions tagged with R. In comparison to proprietary software, there are 2,014 questions tagged with SAS, 616 with Stata and 362 with SPSS. These figures are used as one indicator of the [increasing popularity of R](#). Besides the Question & Answer sites, there are hundreds of blogs discussing specific analytical problems using R and feeds from the blogs are aggregated in [R-bloggers](#)-website.

Another, more formal channel for distributing and communicating R have become the so called massive open online courses (MOOC). MOOCs seem to work well for teaching programming and many courses in [Coursera](#) and [EdX](#)

have become hugely popular, attracting tens of thousands of students each year. The free licensing of R has made it the primary language on these courses as it is basically the only viable alternative for teaching statistical programming for massive crowds.

Aside with vibrant internet community more and more books are being published on R. Books can be put in three categories. First are the general introductions to statistics using R. *Discovering Statistics Using R* by A. Field, Miles, and Field (2012) and *R in Action: Data Analysis and Graphics With R* by Robert Kabacoff (2013) are popular examples of that category. Second there are more and more books addressing how to solve some specific analytical problems using R. A prime examples of books in this category are *Complex Surveys: A Guide to Analysis Using R* by Thomas Lumley (2011), *Text Analysis with R for Students of Literature* by Matthew M. Jockers (in press), *R Graphics Cookbook* by Chang (2012) and *Dynamic documents with R and knitr* by Yihui Xie (2014). A third category are the books that focus on specific theoretical issues in statistics and use R as a primary language to demonstrate this. Such books are for instance *Bayesian Data Analysis* by Gelman et al. (2013) or *Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modeling* by Snijders and Bosker (2011).

Use of R language

Throughout its existence the main use of R has been implementation of new statistical methods. This is still the case and implementations of new statistical methods are usually first available in R. However, various fields of applied statistics have become more active as researchers across disciplines have started to migrate into R. *Bioconductor* was already mentioned as an example of a domain specific initiative to apply R in their analysis, for genome data in this case. Natural sciences in general have been early adopters and for example in Geographical Information Systems (GIS) the R has started to rival proprietary GIS-software. In the case of GIS in R it is possible to combine traditional statistical methods and programming with spatial data and statistics in one

environment. In SSH this is useful as there are a lot of spatial data available and researchers may want to cluster the data thematically, but also visualize it as maps. As for humanities, Matthew M. Jockers (in press) book is one of the first attempts to foster use of R. In the digital humanities blogosphere there are a few others besides [Jockers blog](#) that are worth reading, namely [W. Caleb McDaniel from Rice University](#) and [Quantifying Memory blog by Rolf Fredheim from Cambridge](#).

In addition to academic applications, R has become a major player in business analytics. This is largely due to R's capabilities in visualisation and analysing so called *big data*, but also due to companies like [Revolution Analytics](#) that have started providing consultation and creating tailored application for enterprise needs. The annual [R/Finance 2014](#)-conference gives nice overview of adoption of R in banking and insurance sectors. For example [Google](#) uses R in-house and also provides packages as [r-google-analytics](#) or [rgooglevis](#). One emerging field is so called data journalism where major players like [New York Times](#) or [Guardian](#) use R in data-driven stories [such as this](#).

Conclusions

R is certainly not the only alternative for proprietary data analysis software or for analysis of complex digital data. For example, [Python](#) is another viable option especially for someone looking for a more general purpose language that also masters data analysis. [Whether Python is going to displace R](#) has recently been debated in the data science blogosphere. The data analysis is becoming mainstream in many fields, not just in academic research, but R is still remaining hard to learn and very much research oriented. Programmers rather want to extend the language they already know than learn a new one and python is a lot more common than R. For very intensive computation [Julia](#) is becoming a popular open source option, too. It is still in an early phase of development, but is already a viable option if processing time is important.

But for scientific work I would emphasize the *licensing* of the software more than the name of the particular technology. It is well possible that the recent

buzz around digital data in SSH marks only the beginning of a data intensive research tradition. For someone wanting to gain success in that game it will be equally important to develop the substantial understanding of the research topics as well as technological understanding of the new emerging tools. R is a prime example of this development where academics have taken a major role in software development and created tools that are better suited for their research problems than proprietary software.

This development will go on and therefore it is advisable for someone who is interested in learning these techniques to carefully look at the licensing before investing time and effort in learning the technology. Free and open source tools are great in this respect as once you can pick up the skills to use the technology, you soon will find that it needs to be improved for your purposes. In free and open source technology you can learn how the code works, write improvements and then publish them for the wider research community for use and for further development. In addition, free licensing also allows you to teach the technology, apply it in any purpose, including commercial, and to distribute it. Open source research software is not always the easiest and quickest way to get the job done, but in the long run they are often worth the time invested.

In addition, the openness of the computational research methods is important from the *reproducibility* of your research. Along with demands for open access of research publications there are tendencies that more and more journals in computational sciences will require both the data and algorithms behind the results to be published together with the article. As SSH scholars are moving towards computational analysis this issue of reproducibility should also be taken into account. R is a great tool that fulfills all these conditions, but there are several others out there, too. After all, it is not necessary for all to become software developers, but to have basic understanding and to pair with developers who know more.

Steve Lohr (2013) interviewed some leading digital humanists in the New York Times article *Literary History, Seen Through Big Data's Lens* on the future of SSH and posed a question whether these emerging computational technologies

will undermine the role of qualitative research in the field. Matthew Jockers, whose book *Macroanalysis: Digital Methods and Literary History* (2013) was central in the article, emphasized that finding the right questions and flaws in the analysis still requires deep, both qualitative and quantitative, understanding of the field:

But we're at a moment now when there is much greater acceptance of these methods than in the past. There will come a time when this kind of analysis is just part of the tool kit in the humanities, as in every other discipline.

And that:

Quantitative tools in the humanities and the social sciences, as in other fields, are most powerful when they are controlled by an intelligent human. Experts with deep knowledge of a subject are needed to ask the right questions and to recognize the shortcomings of statistical models.

The quest for new kind of collaboration between scholars and fields of research is also emphasized by professor Gary King (2014). He claims that the analysis of large digital data requires skills that can't be found from traditional fields of social sciences.

Through collaboration across fields, however, we can begin to address the interdisciplinary substantive knowledge needed, along with the engineering, computational, ethical, and informatics challenges before us.

In addition, King (2014) assumes that this collaboration will eventually blur the dichotomy between qualitative and quantitative analysis, and he portrays a future where both traditions have merged into social sciences where the important research problems are solved in collaboration.

Instead of quantitative researchers trying to build fully automated methods and qualitative researchers trying to make do with traditional human-only methods, both now are heading toward, using, or developing computer-assisted methods that empower both groups. This development has the potential to end the divide, to get us working together to solve common problems, and to greatly strengthen the research output of social science as a whole.

This may well be true for humanities as well if we dare to take upon the challenge.

References

- Chang, Winston. 2012. *R Graphics Cookbook*. O'Reilly.
- Field, Andy, Jeremy Miles, and Zoë Field. 2012. *Discovering Statistics Using R*. SAGE.
- Gelman, Andrew, John B. Carlin, Hal S. Stern, David B. Dunson, Aki Vehtari, and Donald B. Rubin. 2013. *Bayesian Data Analysis, Third Edition*. CRC Press.
- Jockers, Matthew. In press. *Text Analysis with R for Students of Literature*. Quantitative Methods in the Humanities and Social Sciences. Springer.
- Jockers, Matthew L. 2013. *Macroanalysis: Digital Methods and Literary History*. University of Illinois Press.
- Kabacoff, Robert. 2013. *R in Action: Data Analysis and Graphics With R*. MANNING PUBN.
- King, Gary. 2014. "Restructuring the Social Sciences: Reflections from Harvards Institute for Quantitative Social Science." *PS: Political Science and Politics* 47: 165–172. http://journals.cambridge.org/repo_A9100Nlq.
- Lohr, Steve. 2013. "Literary History, Seen Through Big Data's Lens." *The New York Times* (January). <http://www.nytimes.com/2013/01/27/technology/literary-history-seen-through-big-datas-lens.html>.
- Lumley, Thomas. 2011. *Complex Surveys: A Guide to Analysis Using R*. John Wiley & Sons.
- Snijders, Tom A. B., and Roel Bosker. 2011. *Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modeling*. Second Edition. Sage Publications Ltd.
- Venables, William N., David M. Smith, and R. Development Core Team. 2013. *An Introduction to R*. Network Theory. <http://www.math.vu.nl/sto/onderwijs/statlearn/R-Binder.pdf>.
- Xie, Yihui. 2014. *Dynamic Documents with R and Knitr*. CRC Press.

Visualiseringar

Jessica Parland-von Essen och Kenneth Nyberg

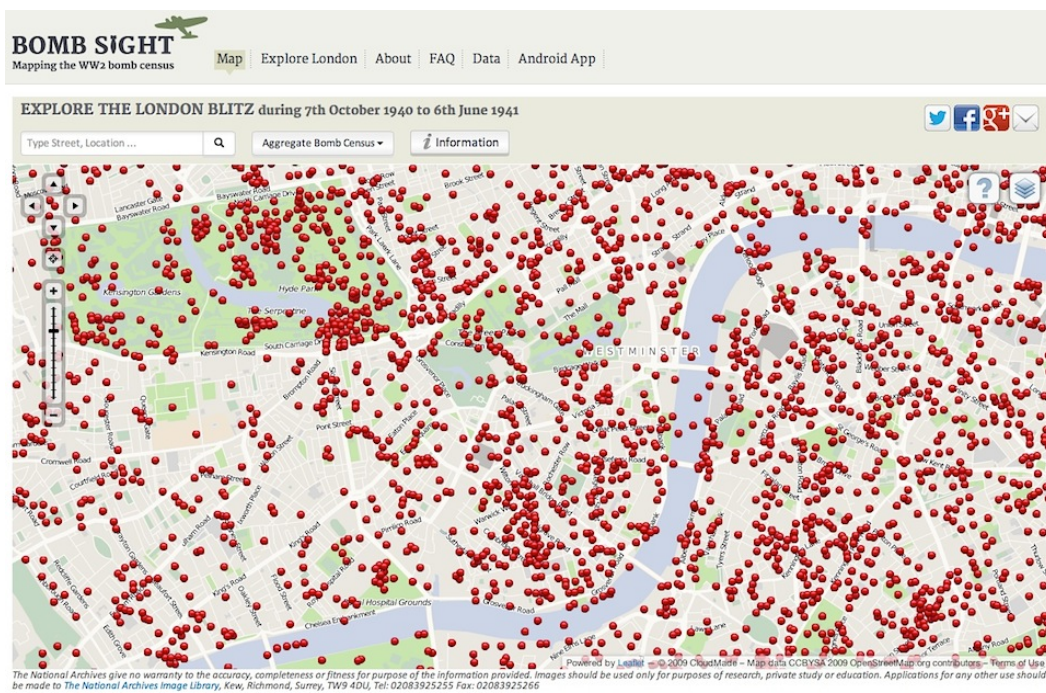
Datavisualisering, som med fördel kan kombineras med text mining, handlar om hur olika data och deras samband kan framställas och analyseras i visuell form. Potentiellt sett är det betydligt bredare än att bara vara ett verktyg eller en metod, eftersom visualiseringar kan ses som ett kvalitativt annorlunda sätt att utforska och gestalta verkligheten än texter i traditionell mening. Vill man något tona ned sådana skarpa gränsdragningar kan man också, som många gör idag, tala om ett "vidgat textbegrepp", där även ljud, bild, filmer och andra media kan uppfattas som en form av texter. Likväl är visuell kommunikation i vissa avseenden ett annat "språk" än ord i skrift, med sina egna möjligheter och begränsningar. Visualiseringar kan också användas för olika syften och för att fylla olika funktioner, och därför blir det något av ett paraplybegrepp för en mängd olika tekniker och arbetssätt där vi här bara kan ta upp några mycket översiktliga resonemang och exempel.

I princip finns det fyra olika saker man kan göra med hjälp av visualiseringar:⁸⁴

- Man kan visa skillnader i storlek eller mängd. I synnerhet stora siffror och skillnader mellan dem är svåra att greppa, men kan väl presenteras visuellt. (Se t.ex. <http://labs.hahmota.fi/veropuu2013/>.) Budgetsiffror är ett exempel på ett område där visualisering stöder både kommunikation och demokrati.
- Med hjälp av visualisering kan man visa relationer. Twitternätverk (<http://twittercensus.se/graph2013/>) har varit populära objekt för denna typ av visualisering.

⁸⁴ Fritt efter Alberto Cairo: *The functional art. An introduction to information graphics and visualization* (New Riders 2013).

- Man kan också visualisera förlopp eller processer antingen statistiskt eller med levande bilder. Detta fungerar bland annat bra på kartor.
(http://www.youtube.com/watch?v=nq0KNfS_M44)
- Visualisering kan också innebära en rekonstruktion. Detta används naturligt nog ofta av arkeologer. (<http://www.digital-archaeology.com/3D.htm>)

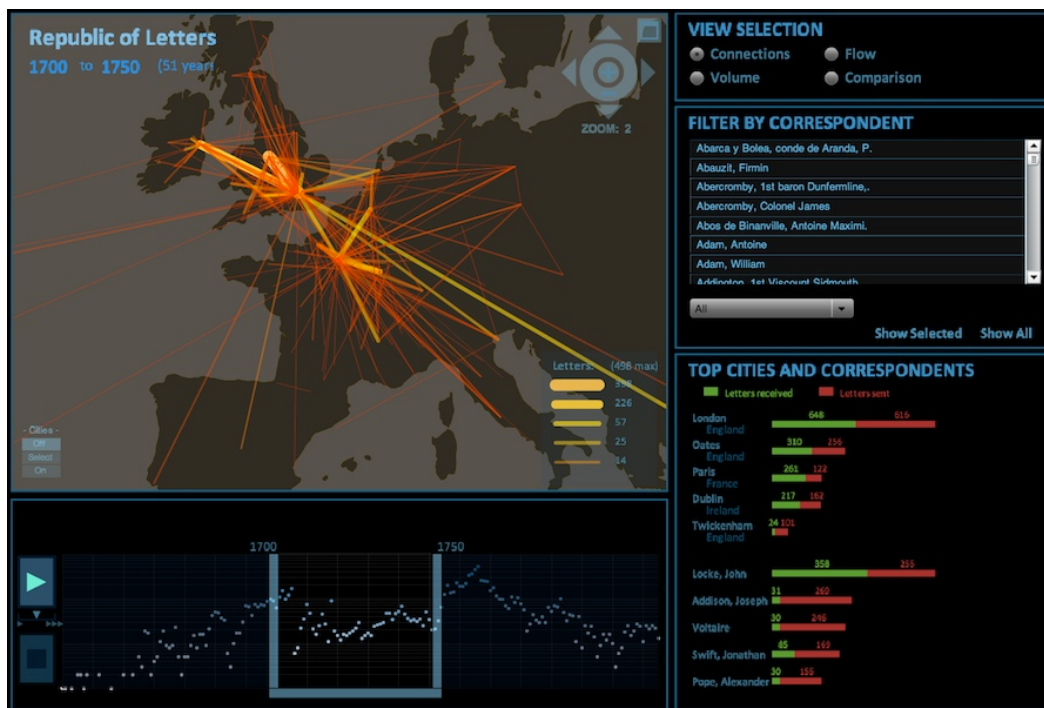


Bomb Sight, brittiska riksarkivets visualisering av tyska bombanfall på London under Blitzten. På den interaktiva webbplatsen kan användaren zooma in och ut i kartan och ta fram data om varje registrerad träff. Bilden visar bomber i centrala London mellan den 7 oktober 1940 och den 6 juni 1941. Källa: National Archives, Bomb Sight: Mapping the WW2 bomb census, <http://www.bombsight.org> (hämtad 2013-01-15).

Två av de enklaste och vanligaste typerna av datavisualisering i historiska sammanhang är dels kartor (se punkt 3 ovan), dels tidslinjer. De förra åskådliggör med visuella medel mönster som kan urskiljas i rummet, de senare demonstrerar tidsrelationer. Detta är dock bara början eftersom visualiseringar

kan användas för att gestalta alla typer av samband, processer och strukturer, även om det ofta rör sig om rumsliga mönster. Sådana grundar sig på spatiala databaser av olika slag, vilka i sin tur ofta skapas med hjälp av GIS (Geographic Information System), en standard för att registrera geografiska data som hittills inte använts så mycket av historiker men länge varit ett viktigt verktyg inom många andra discipliner.

Visualiseringar är särskilt värdefulla för att åskådliggöra och undersöka mycket komplexa studieobjekt med relationer och strukturer i flera dimensioner. Ett tidigt och känt exempel är projektet HyperCities (<http://hypercities.com>), där stora mängder data om befolknings-, sociala och kulturella mönster i storstaden Los Angeles under olika perioder har bearbetats och lagts till grund för avancerade, interaktiva visualiseringar över förändringar i tid och rum. De allra senaste åren har allt fler visualiseringsprojekt av olika typer lanserats; bland dem kan nämnas Bomb Sight (<http://bombsight.org/>), brittiska riksarkivets webbplats om bombanfallen på London under Blitzen 1940–41 och Stanfords Mapping the Republic of Letters (<http://republicofletters.stanford.edu/>) där man kan studera nätverken av tänkare och vetenskapsmän under upplysningstiden i form av interaktiva kartor över deras brevväxling.



Ett av visualiseringsgränssnitten i projektet [Mapping the Republic of Letters](http://www.stanford.edu/group/toolingup/rplviz/) vid Stanford University. Denna bild visar korrespondensnätverken mellan ett antal upplysningstänkare under perioden 1700–1750. Källa: "Electronic Enlightenment Correspondence Visualization", Mapping the Republic of Letters, <http://www.stanford.edu/group/toolingup/rplviz/> (hämtad 2013-10-07).

Som bl.a. digitalkonstnären Jer Thorp har framhållit handlar visualiseringar som de nyss nämnda inte bara om att redovisa eller illustrera "data", utan man borde sträva efter att genom dem också uppnå insikt (*revelation*). (Se t.ex. denna presentation på YouTube från ett seminarium om "Data Visualization from Data to Discovery" den 23 maj 2013: <http://www.youtube.com/watch?v=ivyl-ZWfrDg>.) Det är för övrigt en poäng som borde gälla all digitalisering. Eftersom digitalisering redan i sig innebär reduktion, borde vi inte nöja oss med att skapa nya verktyg och ny tillgänglighet till material genom att omvandla dem till digital form. Vi bör också sträva efter att tillföra ny kunskap, något som bland annat Lev Manovich åtminstone indirekt har påpekat. Som exempel utgår han från en analys av mangaserier. I de digitala bilderna han arbetat med finns över 250 nyanser av grått. De olika nyanserna har inte ett namn, men man kan grafiskt

framställa hur bilderna och exempelvis kontrasterna i dem förändras under berättelsens gång, eller hur "flick-" och "pojkserien" skiljer sig från varandra.⁸⁵

Men den insikt man får genom till exempel visualisering är inte *färdig* kunskap. Den behöver processas för att bli det, inte minst om vi talar om vetenskapligt baserad kunskap. Dan Dixon har skrivit om den kreativa abduktionen som ett sätt att uppnå ny kunskap (jfr avsnittet "Big data", där abduktion också tas upp, och läs mer om "abductive reasoning" på engelska Wikipedia: https://en.wikipedia.org/wiki/Abductive_reasoning). Visualisering och andra digitala metoder kan användas för att hitta mönster i material, som ett hjälpmedel i detta kreativa skede av den kunskapsskapande processen. Vi kan alltså också inom humaniora kanske gå en väg via reduktion och abduktion till ny kunskap.⁸⁶

Just eftersom visualiseringar kan vara ett så kraftfullt verktyg för att gestalta verkligheten gäller det också att förhålla sig lika kritiskt vaksam inför dem som man gör vid läsning av en text. Vi återkommer till detta i kapitlets sista avsnitt, men det bör redan här sägas att visuella representationer lätt kan uppfattas som säkrare, mer "objektiva" och entydiga påståenden än textbaserade resonemang, och det är lätt att bli förförd eller manipulerad av missvisande visualiseringar. Källkritiken blir med andra ord viktigare än någonsin. Hur har visualiseringen framställts? Varifrån kommer de underliggande data? Vad har valts ut, vad har valts bort och vilka parametrar har tillämpats?

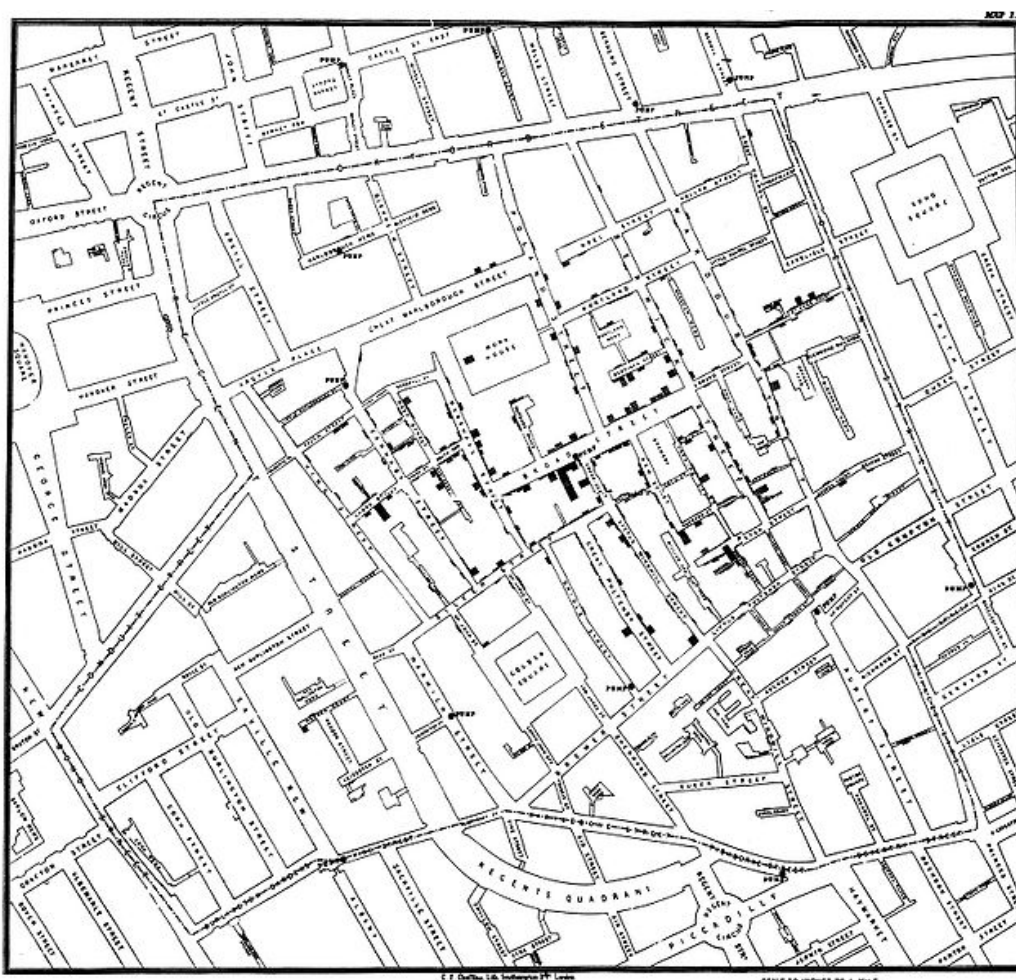
Läsbarhet i infografik

Även visualiseringars läsbarhet i enklare mening är en fråga värd att uppmärksamma här. I massmedier och andra mer offentliga sammanhang stöter vi ofta på visualiseringar i form av så kallad [infografik](#), där olika typer av information sammanfattas och presenteras på ett lättfattligt sätt med olika visuella medel. Idén i sig är ju inte ny, men i dag finns det mängder av enkla

⁸⁵ Lev Manovich, "How to compare one million images", in David M Berry (ed.), *Understanding Digital Humanities* (Palgrave MacMillan 2012).

⁸⁶ Dan Dixon, "Analysis tool or research methodology", in David M Berry (ed.), *Understanding Digital Humanities* (Palgrave MacMillan 2012).

program där vem som helst kan skapa grafiskt snärtiga presentationer av information. Ändå är det ofta mer arbetsdrygt än man först tänker sig, vilket framgår om man provar till exempel Infogram (<http://infogr.am>), Visual.ly (<http://visual.ly>) eller (för de mer tekniskt kunniga) Google Heatmaps (<https://developers.google.com/maps/documentation/javascript/examples/layer-heatmap>).



John Snows klassiska karta över kolerasmitta i London 1854, vilken hjälpte till att spåra smittkällan. Källa: Wikimedia Commons, <http://en.wikipedia.org/wiki/File:Snow-cholera-map-1.jpg#file> (hämtad 2013-10-07).

Visualisering av data har också blivit en [konstform](#), som även [medier använt sig av](#) för att erbjuda nya former av information och erfarenheter. Det fina med det digitala formatet är ju att man kan skapa interaktiva tjänster som ger användare möjlighet att själv forska i datan.

Det är tre viktiga aspekter man måste beakta då man skapar grafik av data: det kommunikativa, det korrekta och det estetiska. En bra grafisk framställning av information har alla dessa element. Inte bara lite, utan på hög nivå. Den erbjuder korrekt information på ett bra sätt och ger både kunskap och en upplevelse för den som tar del av den. Både pedagogiskt och marknadsföringsmässigt är detta en stark kombination. Därför är detta ett mycket krävande område och visualiseringar kräver ofta samarbete av personer med olika kompetenser.

Det kommunikativa och det korrekta är ganska nära besläktade med varandra. Man stöter väldigt ofta på visualiserad information, där det känns som att man inte riktigt haft klart för sig vad det är man vill förmedla. Bra grafik är funktionell, inte bara rolig eller snygg att titta på. Ett vanligt fel är att man vill erbjuda möjligheter att jämföra sifferdata, något som grafisk framställning är oslagbart på, men inte lyckas. Proportionerna mellan 10, 100 och 1 000 000 är till exempel svåra att greppa, vilket vi ganska ofta kan se exempel på i massmedierna, men om man vill att människor ska kunna göra jämförelser måste man välja rätt form för sin grafik. Ofta illustreras den typen av information i form av pajdiagram eller, ännu mer populärt, som en serie cirklar ("bollar") av olika storlek. Jämförelser av summor fungerar dock mycket bättre att återge med stapeldiagram, eftersom det är lättare för flertalet människor att jämföra längd än yta. Även om bollar kan kännas lockande och snärtiga är alltså stapeldiagram ofta det mest funktionella.

Kollektivt arbete (crowdsourcing)

Jessica Parland-von Essen

Webben erbjuder också möjligheter att samarbeta med frivilliga och intresserade lekmän, till exempel i form av det som på engelska kallas crowdsourcing. Det finns många uppgifter som inte kan utföras maskinellt, utan kräver att en människa tolkar något eller manuellt kompletterar material. I vissa fall utbetalas också ersättning för dylikt arbete inom till exempel kodningsarbete.⁸⁷ För exempelvis arkiv är denna typ av gratis crowdsourcing eller på finlandssvenska talkoarbete en värdefull resurs, eftersom behovet av metadata och förbättrade sökningar är skriande samtidigt som resurserna är mycket begränsade. Det kan diskuteras i vilken mån det handlar om en förändring av forskningsmetoden eller bara ett nytt arbetssätt, vilket inte alltid är samma sak. I vart fall kräver också detta eftertanke då man granskar forskningen och dess resultat. Hur har informationen som finns producerats och på vilket sätt har man kontrollerat kvaliteten? Vilken inverkan har enstaka fel på forskningens resultat och kan det finnas strukturella eller systematiska fel, som påverkar kvaliteten?

Artikeln om "Citizen Science" på Wikipedia (http://en.wikipedia.org/wiki/Citizen_science) är påbörjad år 2005 men ett motsvarande ord saknas i den svenska versionen av encyklopedin. Inte desto mindre är förstås fenomenet med lekmäns deltagande i forskningsarbete äldre än så. Till exempel inom biologi och astronomi har crowdsourcing vid insamling av iakttagelser eller uttolkande av bildmaterial redan rätt etablerade roller i en del forskningsprojekt, där Zooniverse (<https://www.zooniverse.org>) tillhör de mest kända. Ett annat, än så länge mer experimentellt projekt som kan bli intressant att följa är en så kallad MOOC (Massive Open Online Course) i bioinformatik där studenterna ska kunna delta i forskningsarbete på massiv

⁸⁷ Se t.ex. Wikipedias artikel om Amazons "Mechanical Turk", http://en.wikipedia.org/wiki/Amazon_Mechanical_Turk (hämtad 2013-10-09).

skala eller MOOR, Massive Open Online Research.⁸⁸ Trots att samarbete med amatörer redan är viktigt inom många områden⁸⁹ har historiker hittills inte varit särskilt öppna för möjligheten att samarbeta med lekmän. Ett exempel med historisk inriktning är dock projektet *Transcribing Bentham* vid University College London (http://www.ucl.ac.uk/Bentham-Project/transcribe_bentham).

Det är först rätt nyligen som man ens från arkivhåll börjat ta samarbetet med släktforskare på allvar, trots ett stort behov av mer utvecklade söktjänster i material. Till exempel i Finland har släktforskare i årtal arbetat med att sammanställa en stor databas över information från kyrkböcker, HisKi (<http://hiski.genealogia.fi/historia/en/ohjeet.htm>). Trots att det Genealogiska samfundet som formellt upprätthåller databasen är ett vetenskapligt sällskap med publikationer som håller hög akademisk nivå, har arbetet med att samla och skriva in informationen inte haft någon koppling till akademisk forskning. Detta beror förstås delvis på de många källkritiska problem som sammanhänger med själva materialet och processen, men man har inte heller från akademiskt håll ansträngt sig för att hitta ett sätt att utveckla kvaliteten så att informationen kunde användas för vetenskaplig forskning.

Samtidigt har man till exempel i Australien genom att låta människor korrekturläsa maskinlästa digitaliserade tidningstexter i Trove-tjänsten (<http://trove.nla.gov.au/newspaper>) engagerat tusentals frivilliga. Genom detta har man fått en mycket bättre kvalitet på materialet. Främst handlar det om förtroende för användarna och deras engagemang och kapacitet. I Finland har man valt att konstruera ett spel för samma syfte med namnet Digitalkoot (http://www.digitalkoot.fi/index_en.html). Frågan är om skillnaden i kvalitet gjorde det värt att investera i relativt kostsam spelkonstruktion och det förarbete som antagligen behövdes. Den största enskilda insatsen i Digitalkoot var 395 timmar för en person, att jämföra med Troves flitigaste medarbetare

⁸⁸ "Is 'massive open online research' (MOOR) the next frontier for education?", *KurzweilAI* 2013-10-03, <http://www.kurzweilai.net/is-massive-open-online-research-moor-the-next-frontier-for-education> (hämtad 2013-10-09).

⁸⁹ Se t.ex. Johan Ohab, "Top Citizen Science Projects of 2012", *PLOS Blogs* 2012-12-31, <http://blogs.plos.org/citizensci/2012/12/31/top-citizen-science-projects-of-2012/> (hämtad 2013-10-09).

som räknar timmarna i tusental – vilket förstås är en rätt orättvis bedömning eftersom Trove varit i gång mycket längre. I Australien räknar man med 270 arbetsår och över 100 miljoner rader korrekturläst text. Vi har heller ingen information om hur många rader text som korrigerats med hjälp av mullvadsspelet, dvs. hur effektivt det är.⁹⁰

I [fördjupningsartikeln av Sakari Katajamäki](#) kan man läsa om andra alternativ för att säkra kvaliteten i dylika projekt. Katajamäki understryker vikten av att låta medarbetarna känna betydelse och meningsfullhet i sitt arbete. I varje projekt måste man också utvärdera hur viktigt det är att arbetet är helt felfritt eller vilken tolerans man eventuellt kan ha för felaktigheter. Då det endast gäller förbättrad sökbarhet torde toleransen för eventuella fel som gjorts av lekmän vara rätt god, i synnerhet om alternativen är ingen sökbarhet alls eller mycket dålig sökbarhet. En viktig poäng i Katajamäkis text är också att det går att göra crowdsourcing utan extra tekniska insatser.

Vi har en mycket stor resurs av människor som är intresserade av släktforskning och historia, men man måste kunna hitta på ett sätt att samarbeta som är givande för alla parter. Att nå dessa människor och värna om deras engagemang kräver att man investerar mycket tid i kommunikation, inte bara att informera utan att reagera på frågor, ge respons och verkligen lyssna och interagera. Om man bygger tekniska system för ändamålet, vilket ofta inte alls är nödvändigt att göra, bör man testa dem noga med många olika försökspersoner, så att man kan minimera antalet fel och problemsituationer som beror på dålig användbarhet eller information i användargränssnittet.

⁹⁰ Marie-Louise Ayres, "Singing for their supper': Trove, Australian newspapers, and the crowd" (IFLA WLC 2013), <http://library.ifla.org/245/1/153-ayres-en.pdf> (PDF, hämtad 2013-10-09).

Fördjupning: Transkribering av manuskript och förstaupplagor med talkokrafter

Sakari Katajamäki

[Wikipedia](#) kan knappast anses vara historieforskarens mest pålitliga källa, men i dag skrivs knappast en enda bok eller artikel inom historia, där denna encyklopedi som uppkommit genom crowdsourcing inte skulle ha utnyttjats. Till och med dess finansiering består till stora delar av små donationer som gjorts av de talrika användarna. Liknande gemensamma ansträngningar kan användas för att underlätta forskarnas arbete också på många andra sätt.

Ett av de mest geniala projekten är reCAPTCHA (<http://www.google.com/recaptcha>), som tjänar både som robotfälla (CAPTCHA, Completely Automated Public Turning test to tell Computers and Humans Apart) och som stöd för automatisk textigenkänning. Bildidentifieringen kan användas som en robotfälla på interaktiva webbplatser för att se till att det inte är ett datorprogram som försöker ändra innehållet på en sida. I reCAPTCHA-tjänsten används två inskannade bilder av ord, där det ena är känt av systemet och mot vilken den inmatade texten matchas, medan det andra ordet är hämtat från en automatiskt inläst digitaliserad text, vars noggrannhet man vill förbättra. re-CAPTCHA-tjänsten används varje dag över 100 miljoner gånger. Det innebär att de granskade ordens antal motsvarar en granskning av mer än tusen romaner av normal längd, ord för ord.

Begreppet crowdsourcing används ofta då man talar om att bygga upp stora system som Wikipedia eller reCAPTCHA. Dessa system används av enorma människomassor över hela världen, men liknande projekt kan också göras i betydligt mindre skala, som i ett litet gårdstälko. Vid Finska litteratursällskapet (SKS) provade vi på ett sådant mindre talko under åren 2011 och 2012, då vi behövde digitala transkriptioner för att kunna göra kritiska utgåvor av dem. Vårt lilla projekt visade att man med rätt litet besvär kan åstadkomma goda resultat till och med på kort tid. Och det bästa är att frukterna av detta arbete också i

framtiden kan producera nya projekt som gynnar hela det vetenskapliga samfundet.

Från digitala bilder till transkriptioner

På hösten 2006 inledde den nya enheten Edith – kritiska utgåvor av finsk litteratur (<http://www.edith.fi/svenska/>) – sin verksamhet vid Finska litteratursällskapet. Enhetens uppgift är att ge ut textkritiska och rikligt kommenterade vetenskapliga utgåvor av den finska litteraturens klassiker. Editionerna ges ut både som tryckta böcker och som öppna utgåvor på webben. Hittills har vi gett ut Aleksis Kivis komedi *Sockenskomakarna* (Nummi et al 2010; Nummi et al. 2011) både som bok och digitalt, samt en tryckt utgåva av författarens brev (Niemi et al. 2012). Editionsarbetet fortsätter med Kivis pjäser och övriga produktion och kommer senare att gå över också till andra finska författares produktion.

På hösten 2011 gjorde Litteratursällskapet en enkät till sina medlemmar. Eftersom svararna i den förra enkäten hade hoppas på möjligheter att stöda sällskapets verksamhet, beslöt vi att denna gång som experiment erbjuda medlemmarna ett tillfälle att delta i några av sällskapets projekt. Som ett alternativ erbjöd vi medlemmarna ett Aleksis Kivi-talko, som gick ut på att skriva in första upplagan av hans roman *Sju bröder* (1870) som digital text. Målsättningen var att en del av den omfångsrika romanen på över 300 sidor som tryckts i fraktur på detta sätt skulle bli transkriberad för fortsatt bearbetning. Intresset för talkot var så stort, över 30 personer anmälde sig och någon till och med från Afrika, att vi redan från början var tvungna att ta med också tidiga upplagor av Kivis andra verk.

Utgångspunkten var att vi inte kunde förvänta oss att medlemmarna kunde eller ville göra transkriptionerna direkt som en xml-fil i den TEI-standard vi använder (Text Encoding Initiative, <http://www.tei-c.org>) och att det inte inom detta projekt skulle vara ändamålsenligt att bygga ett separat webbaserat system för att transkribera materialet. Därför beslöt vi oss för att organisera arbetet så, att var och en som deltog fick använda sig av det

textbehandlingsprogram hen vanligtvis använde och sedan sända oss materialet som en e-postbilaga.

Den andra målsättningen var att vi skulle framskrida med små steg, så att vi under processens gång kunde samla in erfarenheter för att utveckla talkoarbetet så att det blev smidigare. Därför har vi inte heller försökt värva en så stor mängd medhjälpare som möjligt och inte heller gjort reklam för projektet utanför sällskapets medlemskår.

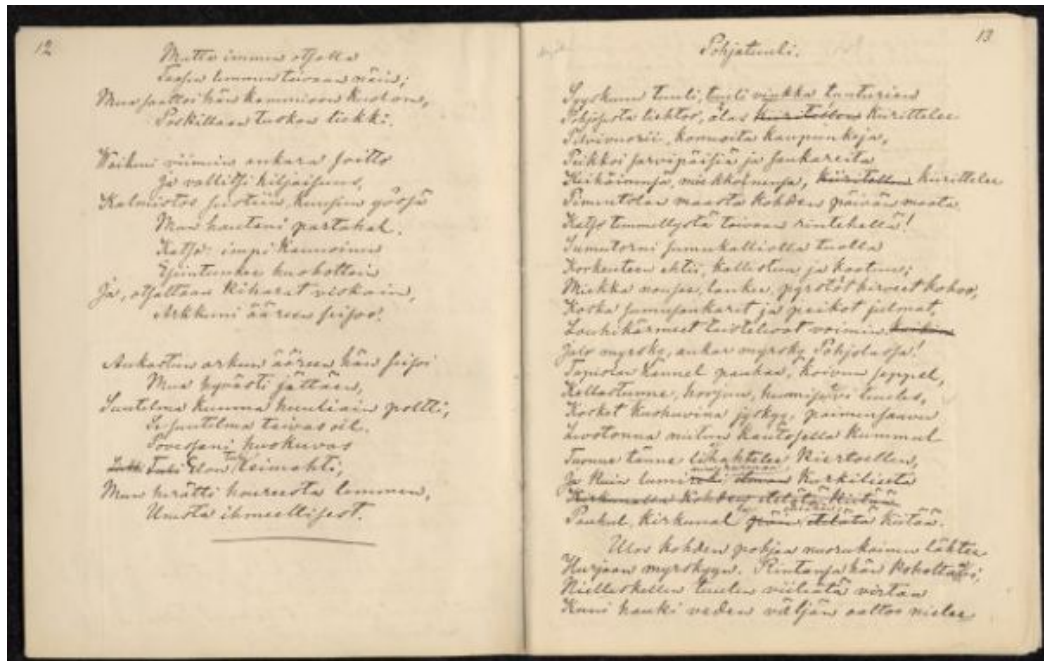
Talkot möjliggjordes av att Finska litteratursällskapet redan 2007 hade publicerat en stor helhet *Tiet lähteisiin – Aleksis Kivi SKS:ssa* på webben (Vägar till källorna – Aleksis Kivi och Finska litteratursällskapet, <http://www.finlit.fi/kivi>), som innehåller bilder på merparten av Kivis förstaupplagor och ett separat digitalarkiv med allt bevarat arkivmaterial, från manuskript till brev och övriga arkivmaterial. Tack vare det digitala arkivet och biblioteket kunde de frivilliga skriva sina transkriptioner från bilderna på webben och vi behövde inte skicka något material separat till någon eller be någon att arbeta i våra utrymmen.

De olika sätten att utnyttja det digitala arkivet och biblioteket var inte påtänkta då de skapades, men de har betjänat arbetet med de kritiska utgåvorna redan under flera år och stöder fortsättningsvis förverkligandet av de textkritiska forskningsprojekten.

Från talkoarbete mot kritisk edition

Då Aleksis Kivi-talkot framskred snabbt och kvaliteten på texterna var god, beslöt vi oss för att använda oss av crowdsourcing också för att transkribera Kivis handskrifter. Processen förlöpte på liknande sätt som med de tryckta texterna, förutom gällande de inre varianterna, det vill säga de ändringar i texten som författaren själv gjort med penna. Eftersom den slutliga målsättningen var att skapa xml-filer som följer den internationella TEI-standard, utformade vi ett enkelt sätt att markera struken eller tillagd text som xml-taggar. Vi bad deltagarna lägga till en klammer och bokstaven P ("P"; på finska *poisto*) och samma beteckning spegelvänd ("P") i slutet för struken text. För tillägg och

förändringar i ordföljden skapade vi liknande annotering. I slutskedet var det sedan enkelt att automatiskt ersätta dessa element med TEI-element.



Handskrift av Aleksis Kivi. SKS KIA (Finska litteratursällskapet, litteraturarkivet), Aleksis Kiven arkisto. Runoelmia 2. Tillgänglig: http://www.finlit.fi/kivi/index.php?pagename=kivendigiaineisto&set=02_runoelmia2&item=4 (2013-10-14).

Syyskuun tuuli, [P tuuli vinkka tunturien P][L vinkka tunturien
tuuli L]
Pohjosesta liehtoo, alas [P kiiritellen P][L kiirittelee L]
Pilvivuorii, komeoita kaupunkoja,
Peikkoi sarvipäisiä ja sankareita
Keihäinensä, miekko[P i P]nensa, [P kiiritellen P][L kiirittelee
L]

Exempel på transkription med annotering av talkoarbetare/frivilliga (ovan)
och TEI-kodad XML (nedan) av några versrader ur början av föregående Aleksis
Kivi-handskrift.

```

<l>Syyskuun tuuli, <subst>
    <del>tuuli vinkka tunturien</del>
    <add>vinkka tunturien tuuli</add>
</subst>
</l>
<l>Pohjosesta liehtoo, alas <subst>
    <del rend="strikethrough">kiiritellen</del>
    <add place="inline">kiirittelee</add>
</subst>
</l>
<l>Pilvivuorii, komeoita kaupunkeja,</l>
<l>Peikkoi sarvipäisiä ja sankareita</l>
<l>Keihäinensä, miekko<del rend="strikethrough">i</del>nensa,
<subst>
    <del rend="strikethrough">kiiritellen</del>
    <add place="margin">kiirittelee</add>
</subst>
</l>

```

De allra svåraste manuskripten gav vi inte till talkoarbetarna, men de klarade av också rätt utmanande texter. I detta skede hade redan en del av de ursprungliga deltagarna hoppat av, men vi hade inte längre något behov att skaffa oss fler medarbetare. Många arbetade mycket dedikerat, som bäst skrev en person in 250 sidor av Kivis manuskript och förstaupplagor.

Genom talkot ville vi ha sådana transkriptioner som innehöll endast enstaka fel. Vi ville ändå inte be talkofolket granska texterna, eftersom vi upplevde att det aktiva producerandet av transkriptionerna var mer inspirerande och belönande. Därför beslöt vi oss för att beställa två eller flera transkriptioner av samma text av olika personer och sedan maskinellt jämföra dem genom maskinell kollationering (jämförelse tecken för tecken), som kunde visa oss på vilka ställen texterna skilde sig från varandra. Genom att kontrollera de ställen där texterna skiljer sig åt kan rätt felfria transkriptioner åstadkommas. Vid kollationering kan man som stöd också använda moderna nätbaserade versioner av Kivis verk, trots att de har fått en moderniserad språkdräkt. Samma metod har använts för bland andra Henrich Heine-portalens material (<http://www.hhp.uni-trier.de>).

Jämförelsen mellan transkriptionerna har vi gjort med ett gratis kollationeringsprogram som heter Juxta (<http://www.juxtasoftware.org>) och som kan laddas ner på webben. Alla filer måste konverteras till ren text (txt). I Juxta kan två textversioner studeras parallellt så att programmet målar de

ställen där texterna avviker från varandra. Man kan också jämföra flera versioner samtidigt, men för att hitta felen visade sig kollationering av två versioner samtidigt vara mest effektivt.

Trots att de texter som producerades genom crowdsourcing redan motsvarade originalet nästan perfekt, bör de granskas ännu flera gånger i senare arbetsskeden, eftersom kollationeringen inte avslöjar sådana fel som människor gjort på samma ställe i texten. Särskilt då originalet innehåller stavfel eller fel satt text, korrigeras de lätt av misstag, trots att texten borde transkriberas som den är. Till exempel upptäckte vi först i ett senare skede att sättningsfelet Äapo (ska vara Aapo) av misstag hade korrigerats i alla transkriptioner.

Med hjälp av det snabbt framskridande talkoarbetet har vi fått transkriptioner av god kvalitet av författarens alla manuskript och tidiga upplagor, vilka vi kan använda som utgångspunkt för de kritiska utgåvorna och som material för att jämföra till exempel hur Kivis ortografi varierat eller leta efter förekomster av enskilda ord och motiv i hans produktion. Att ha texten i digital form redan i ett tidigt skede av redigeringsprojektet har dessutom underlättat möjligheterna att skapa en enhetlig xml-annotering eftersom olika typer av listor av xml-taggar kan göras med hjälp av Oxygen-editorn.

I slutskedet av Aleksis Kivi-talkot hade Finska litteratursällskapet en möjlighet att göra alla transkriptioner till en korpus i XML/TEI-format. Aleksis Kivi-korpusen är den första digitala korpusen av hela hans produktion och som följer den ursprungliga ortografin och informationen om de förändringar som gjorts i texten. Korpusen innehåller transkriptioner av 70 brev, 12 tryckta verk eller delar av sådana, 11 manuskript av pjäser eller prosaverk, drygt 70 diktmanuskript och över tio dikter som publicerats i tidningar.

Vi beslöt att donera Aleksis Kivi-korpusen som skapats med hjälp av talkoarbetet till FIN-CLARIN-projektet (<https://kitwiki.csc.fi/twiki/bin/view/FinCLARIN/KielipankkiFramsida>) för att publiceras i den finska språkbanken Kielipankki (<http://www.csc.fi/english/research/sciences/linguistics/index.html>). (Katajamäki et al. 2013) På så sätt kan även andra forskare få tillgång till

materialet redan innan editionsarbetet är klart. Antagligen kan vi också vid den kritiska editeringen använda det Aleksis Kivi-material som publicerats i Kielipankki.

Varför lyckades projektet?

Att det finns en tradition av frivilligarbete inom till exempel insamling av folkkultursmaterial bidrog sannolikt till att Finska litteratursällskapets Aleksis Kivi-projekt blev så framgångsrikt. Dessutom upplevs Aleksis Kivi som en betydelsefull och intressant författare, varför människor var beredda att sätta tid på att arbeta med hans texter. Vid sidan av dessa immanenta faktorer strävade vi också efter att upprätthålla motivationen på olika sätt.

I början av projektet bedömde vi på vilket sätt arbetet skulle fungera mest smidigt. Ur talkoarbetarnas perspektiv tyckte vi att ganska korta, ungefär 20 sidor långa avsnitt, var mer belönande än att bara skriva så långt man orkar av en längre text. Vi gav en klar deadline som var några månader senare för varje avsnitt. Eftersom alla hade samma deadline, kunde vi skicka påminnelser till samtliga medarbetare samtidigt och sända dem gemensamma mellanrapporter om hur projektet framskred. Vid sidan av dessa "huvudvågor" kunde vi dessutom ge extra uppgifter till de snabbaste och ivrigaste. "Om det ännu kliar i fingrarna, finns det nog mer jobb ..." var en av våra återkommande fraser, då vi skrev till deltagarna och tackade för deras bidrag. Till dem som så önskade kunde vi också ge en mer flexibel deadline. Mellanetapperna gjorde att arbetet var mer motiverande både för oss som organiserade det och för dem som utförde det.

Utöver att projektet som helhet var betydelsefullt och att medarbetarna fick uppmuntrande brev, upplevdes arbetet i sig som belönande. Att noggrannt skriva om handskriven eller i fraktur tryckt text från 1800-talet tvingar en att koncentrera sig på författarens språk, på varje ord och bokstav, vilket leder djupare in i språket, och särskilt att transkribera handskrift är ofta utmanande på ett positivt sätt. Ibland var den iver och det engagemang som människor uttryckte i sina följebrev direkt rörande. Man skrev initierat om Kivis gripande

språk, man återkom till hans minnen eller förundrade sig över någon detalj i något verk. Engagemanget kom till uttryck till exempel i ett brev, där skribenten bad om ursäkt för att denne inte kunde delta fullt ut på ett tag på grund av vård av barnbarnen, balkong- och fönsterrenovering och en större operation!

Aleksis Kivi-talkot visade att också en liten grupp kan ha styrka och att crowdsourcing inte nödvändigtvis kräver några speciella it-system. Kivi-talkots särdrag var att betona ett felfritt resultat framom mängd och snabbhet. Om avsikten är att åstadkomma transkriptioner av ett omfattande material så fort som möjligt, lönar det sig inte att göra kollationering, åtminstone inte i ett tidigt skede av processen. Det är möjligt att få med många frivilliga, om bara arbetet upplevs som meningsfullt.

Referenser

Katajamäki, Sakari – Ossi Kokko – Elina Kela (red.): Aleksis Kivi -korpus (SKS). FIN-CLARIN: Finnish Language Resource Infrastructure / CSC – Tieteen tekniikan keskus: Kielipankki 2013. URL:

<http://www.edith.fi/kivikorpus/index.htm>

Niemi, Juhani (huvudredaktör) – Sakari Katajamäki – Ossi Kokko – Petri Lauerma – Jyrki Nummi (red.): Kivi, Aleksis, *Kirjeet. Kriittinen editio*. Suomalaisen Kirjallisuuden Seuran Toimituksia 1386, Tiede. Helsinki: SKS 2012 (426 s.).

Nummi, Jyrki (huvudredaktör) – Sakari Katajamäki – Ossi Kokko – Petri Lauerma (red.): Kivi, Aleksis, Nummisuutarit. Komedia viidessä näytöksessä. Kriittinen editio. [On-line.] Helsinki: Suomalaisen Kirjallisuuden Seura, 2011. URL: <http://elias.finlit.fi/nummisuutarit/>. URN: NBN:fi:sk-201105261000

Nummi, Jyrki (huvudredaktör) – Sakari Katajamäki – Ossi Kokko – Petri Lauerma (toim.): Kivi, Aleksis, *Nummisuutarit. Komedia viidessä näytöksessä. Kriittinen editio*. Suomalaisen Kirjallisuuden Seuran Toimituksia 1284, Tiede. Helsinki: SKS 2010 (330 s.).

Tiet lähteisiin – Aleksis Kivi SKS:ssa. (Ilkka Välimäki, huvudredaktör; Eeva-Liisa Haanpää; Satu Heikkinen; Irma-Riitta Järvinen; Sakari Katajamäki; Klaus

Krohn ja Tarja Soiniola). Helsinki: SKS 10.10.2007. URL:
<http://www.finlit.fi/kivi/>.

Om kritisk granskning

Jessica Parland-von Essen

Men varför är allt det vi skrivit om i detta kapitel, mycket av det kopplat till "big data" i vid mening, över huvud taget viktigt och intressant? Varför kan man inte bara göra närläsning och analysera, göra kvalitativ analys? Lev Manovichs svar är beklämmande enkelt, men likväl tvingar det till eftertanke: För att man kan. I dag har vi inte möjlighet att bedriva forskning riktigt på samma sätt som förr. Vi har för mycket information. Poängen är ändå att Manovich, liksom många andra, inte alls vill förringa betydelsen av traditionell närläsning och andra etablerade metoder. Men eftersom vi i dag kan teckna även en "fond" av större helheter, vinner forskningen på att göra det också. De kvantitativa metoderna kan, och enligt Manovich också bör, användas även inom humanistisk forskning, som komplement. Det leder till intressanta metodiska frågor.

All datorkod medger inte all sorts modellering av världen. Då vi skapar databaser, visualiseringar eller andra digitala resurser finns det alltid många val i bakgrunden och på vägen. Dessa är dels våra egna gällande vilken data som används eller hur semantiken är uppbyggd dvs vad vi kallar saker, vilka klasser, begrepp och kategorier som användas. Men de är också andras val, deras som har skrivit och konstruerat den hård- och mjukvara vi använder. I praktiken innebär detta att man bör granska hur informationen är strukturerad i en resurs som används vid forskning: Hur är de olika entiteterna klassificerade och hur är deras relationer beskrivna? Ärvs olika egenskaper mellan olika entiteter eller informationsenheter och i så fall på vilket sätt? Hur har man hanterat disambiguering, dvs olika varianter av namn på en och samma sak?

Bernhard Rieder och Theo Röhle skriver i boken *Understanding digital humanities* om de fem utmaningar de digitala forskningsmetoderna medför, och de kan vara en lämplig avslutning på detta kapitel.⁹¹

⁹¹ Bernhard Rieder och Theo Röhle, "Digital Methods: Five Challenges", i David M. Berry (red.), *Understanding Digital Humanities* (Basingstoke: Palgrave Macmillan, 2012).

1. Chimären om objektivitet

På något märkligt sätt tycker vi att siffror är objektiva, neutrala och mer sanna än andra typer av information. Humaniora försöker kanske bli mer tungt och hårt (i Snows tappning) genom att anamma kalkyler som metod? Detta är en tendens som historiskt återkommit med jämna mellanrum under lärdomshistoriens gång, och författarna kanske inte riktigt vederlägger detta antagande med tillräcklig kraft. Alla siffror representerar något, och de är alla delar av en modell, där någon eller några personer gjort en tolkning av något. Rieder och Röhle ser problemet mer i tolkningen av resultatet, medan jag ser redan [uppkomsten av sifferdata](#) som ytterst problematiska förenklingar som man aldrig får glömma att ifrågasätta.

2. Styrkan av visuella bevis

Detta är en mycket intressant poäng, hämtad av [Bettina Heintz](#): Förutom siffror tenderar också visualiseringar att tolkas som "fakta", medan bevis i textform uppfattas som "argument". Detta beror på att argumentationen är transparent i texten, medan siffror och visualiseringar ofta på ett väldigt effektivt sätt kan dölja underliggande tolkningar, antaganden och resonemang. Vi behöver definitivt lära oss källkritik på visualiseringar. Det är svårt.

3. Den svarta lådan

Datorprogrammen, algoritmerna och koden lämnas för ofta utanför ordentlig granskning. Data och kod är delar av forskningsmetoden som kan vara svåra att kontrollera. Förutom att detta kräver tillgång, som kan vara begränsad av många skäl, kan dessa också vara mycket svåra att analysera och förstå sig på. Man bör därför gärna, om möjligt, använda olika verktyg (och således metoder) för samma uppgift, exempelvis för att samla data eller göra sökningar. Då kan man också göra jämförelser och konstatera eventuella skillnader i resultatet.

4. Interdisciplinariteten

Digital humaniora kräver genuint och integrerat samarbete mellan människor från olika discipliner, med olika kompetenser. Hur klarar vi detta och vad kommer det att ha för konsekvenser i längden?

5. Jakten på universalism

Hoppet om att finna en sanning, en universell förklaring på allt, att generalisera modeller inkorrekt, på fel domäner och överföra regelbundenheter på områden där de inte stämmer är en frestelse, som växer sig ännu större med de "allmängiltiga" verktyg datorer till syner erbjuder. Idéer om nätverk, komplexitet, fraktaler, kaos, självstrukturerande system, emergens, entropi och what not kan vilseleda oss och göra att vi missar det som är unikt i olika sammanhang och dölja hur komplex verkligheten de facto är och förleda oss att glömma att modeller alltid bara är modeller och förenklingar.

Kapitel 6. Digitalbaserade material och långsiktigt bevarande

Jessica Parland-von Essen

Den förändrade offentligheten

Jessica Parland-von Essen och Kenneth Nyberg

Offentligheten har länge ansetts vara en grundbult i det demokratiska samhället. Jürgen Habermas är kanske den som tydligast pekat på detta drag i det moderna samhället i sin bok *Strukturwandel der Öffentlichkeit*. *Untersuchungen zu einer Kategorie der bürgerlichen Gesellschaft* (1961, Borgerlig offentlighet. Kategorierna 'privat' och 'offentligt' i det moderna samhället 1962). En fri press och i förlängningen journalistiken har utvecklats till det forum där den offentliga debatten ägde rum under stora delar av förra seklet. Sist och slutligen liknar dagens webbmiljö på många sätt den borgerliga offentlighet som Habermas skrev om, snarare än det journalistvälde som rådde med eller utan samröre med den maktavande eliten under 1900-talet.⁹²

I dag ser medielandskapet annorlunda ut, eftersom medborgare kan publicera sig i offentligheten utan att det finns någon grindvakt. Också att kommunicera direkt med beslutsfattare kan anses vara enklare nu för en vanlig medborgare. Men det handlar inte bara om att sprida sin egen åsikt, utan det som de facto äger rum på webben hela tiden är en, eller egentligen flera, offentliga samtal och diskussioner som påverkar vilka åsikter och argument som omfattas av både enskilda människor och inom myndigheter och organisationer. I dag kan vem som helst i Finland starta en namninsamling för en lagmotion, något som i praktiken ofta förverkligas på webben.⁹³ Det handlar åtminstone delvis om dialog som kan ha stor påverkan på vilka beslut en beslutsfattare tar. Offentligheten handlar inte längre endast om de traditionella medierna och

⁹² Janne Seppänen & Esa Väliaverronen, Mediayhteiskunta (Vastapaino 2012) s. 72–76.

⁹³ En central webbplats för detta är det "öppna ministeriet", <http://www.avoinministerio.fi/>.

“opinionsbildarna”, utan det rör sig om ett intrikat och komplext samspel, där journalister, makthavare och allmänhet i sociala medier och på mediernas och andra webbsidor kommenterar, dementerar eller understöder varandras information och omdömen.

Allt detta är i teorin i dag möjligt att spara och forska i mycket mer noggrant än någonsin förut. Nätverk och kommunikation, som tidigare ofta varit muntliga eller osynliga åtminstone för samtiden, kan nu analyseras i detalj. Samtidigt är situationen den, att en stor del av dessa material i dag inte omfattas av arkiv- eller biblioteksfunktioner. På grund av kommersiella, upphovsrättsliga och historiska orsaker sparas inte denna data särskilt systematiskt eller målmedvetet av de offentliga minnesorganisationerna. Samhällsvetare och framtidens historiker skulle ha mycket stor glädje av material som tidningarnas kommentarsfält eller olika versioner av nyhetssidor, politikers Facebookstatusar m m, men dessa arkiveras nu inte av någon instans.

Ett annat skäl till osäkerheten kring bevarandet av denna typ av material har att göra med en mer allmän fråga som aktualiserats av de senaste årens utveckling av webb och sociala medier: den uppluckring – vissa skulle säga upplösning – av gränsdragningen mellan offentligt och privat som tidigare utgjort en självklar och grundläggande realitet i det moderna samhället. Många av den demokratiska rättsstatens centrala principer och funktioner (allmänintresse, jäv, opartiskhet osv.) bygger till exempel på möjligheten att skilja mellan offentlig ämbetsutövning och rollen som privatperson, något som ofta kan vara ganska svårt i t.ex. sociala medier där man som individ förknippas med både sina offentliga och privata eller personliga roller samtidigt. De sociala mediernas logik är närmast definitionsmässigt den personliga vän- eller bekantskapen, och när denna präglar tilltalet och atmosfären i myndighetspersoners (eller till och med myndigheternas såsom institutioner) kontakter med människor riskerar förtroendet för dessa såsom företrädare för ett allmänintresse eller offentliga organ.

En besläktad fråga rör den om hur det påverkar den journalistiska bevakningen att ledande politiska reportrar och granskare av makten är

“vänner” med politiska makthavare på Facebook eller Twitter med, återigen, den logik av personliga (snarare än professionella) relationer och därav följande intressekonflikter det kan leda till. Samtidigt hävdas ibland att det snarast är positivt att personliga nätverk av det slaget blir synliga i offentligheten och därmed möjliga att granska; tidigare kunde ju mycket nära kontakter av det slaget förekomma mellan politiker och journalister utan att någon kände till det. Frågan är dock om dolda relationer av det slaget har upphört i och med de sociala mediernas framväxt; mer troligt är att de ibland synnerligen personliga band som upprättas i sådana miljöer är något som tillkommit och som ytterligare bidrar till att luckra upp den distans mellan makthavare och granskare som tidigare ansågs självklart eftersträvansvärd.

Om vi återvänder till forskarperspektivet är det dock ganska klart att denna nya typ av material tillför något intressant, men det råder som sagt viss oklarhet om vad som i en etisk mening ska betraktas som privat respektive offentligt material då de sociala medierna verkar i en miljö som löper längs hela skalan från full offentlighet (Twitter) till en gråzon (Facebook) och vidare till rent privata sammanhang (direktmeddelanden och chattar av olika slag, “privata” sociala nätverk som Path, <http://path.com>, etc.). Om exempelvis en politiker (eller någon annan offentlig person) till exempel har en relativt generös hållning till vilka som accepteras som vänner på Facebook och delar en status inte offentligt, utan till vännerna, och en forskare tillhör dessa vänner, är det acceptabelt att använda sig av detta material i forskningssyfte? Mer generellt finns också, som redan antytts, frågor kring vem som egentligen äger innehållet i olika sociala nätverk och därmed vilka möjligheter minnesorganisationer och andra har att systematiskt samla in det för framtida forskning.

Hur vi ska hantera de oerhörda mängder data som numera genereras varje dag på webben och i sociala medier väcker alltså en rad frågor om gränsdragningar, urval, planering och tekniska lösningar m.m., och några av dessa frågor kommer vi att diskutera i de följande avsnitten.

Vad är forskningsdata?

Med forskningsdata menar man ofta data som uppstår i samband med forskning. Forskningsdata som skapas av en historiker kan till exempel utgöras av en databas dit man matat in uppgifter från källmaterial eller de digitala foton som en forskare tagit på ett arkiv. Hanteringen och bevaringen av forskningsdata borde alltid ha en planerad livscykel och tillräcklig dokumentation och struktur så att den kan återanvändas. I praktiken saknas ofta fortfarande kompetens, resurser och infrastruktur för historiker i deras verksamhetsmiljö för att hantera hela denna process med alla dess aspekter.

Men för en historiker kan vilken data som helst i princip utgöra material för forskning. Vi står därför inför en utmaning vad gäller att ta till vara den data som finns i dag med tanke på framtida forskning.

Myndighetsdata

Myndigheter har lagstadgade direktiv för hur de ska hantera data och det finns noggrann reglering för elektroniska dokumentarkiv, som baserar sig på arkivteoretiska modeller om ärendehantering. Myndigheterna besitter ändå en hel del data som inte direkt faller in under dessa kategorier. I Finland har den så kallade JUHTA-delegationen utfärdat direktiv för offentlig informationsförvaltning, där man utgår från Arkivverkets SÄHKE 2-direktiv, som uttryckligen handlar om handlingar, snarare än till exempel dataregister eller andra informationssystem. Ämnet behandlas också i Statsrådets förordning om informationssäkerheten inom statsförvaltningen 1.7.201/681 och till exempel i [Arkivlagen 4 §](#), där det stadgas att "Arkivverket har till uppgift att säkerställa att handlingar som hör till vårt nationella kulturarv bevaras och är tillgängliga, främja forskning samt styra, utveckla och undersöka arkivfunktionen." Generellt är regleringen fokuserad på ärendehantering, "arkivhandlingar" och dokumenthanteringssystem, även om alla register också i princip borde ses som "handlingar".

Myndigheter har i dag en del data som inte direkt faller under denna kategori. Ett exempel, som direkt berör myndigheters offentliga kommunikation, är deras

kommunikation med allmänheten över sociala medier. Denna kommunikation arkiveras inte systematiskt på alla ställen. I vissa fall har ovissheten kring statusen av denna typ av myndighetskommunikation helt förhindrat kommunikationen med medborgarna på detta sätt. Om Facebookstatusar uppfattas som "handlingar" bör de arkiveras, samtidigt som denna tolkning kan ifrågasättas. Liksom inte heller alla telefonsamtal eller möten heller bandas och sparas, kunde denna typ av kommunikation helt enkelt uppfattas ha en status som liknar muntlig kommunikation. Med tanke på att arkivering ändå i teorin är möjlig och med hänsyn till medborgarnas rättsskydd borde man sträva efter att bevara dem.



Ska myndigheter diarieföra sin kommunikation via t.ex. Twitter på samma sätt som (andra) offentliga handlingar? Här väcks den frågan på – var annars? – Twitter.

Till myndigheternas kommunikation med medborgarna hör i dag också olika medborgarinitiativ och projekt (som avoinministerio.fi, otakantaa.fi eller medborgarinitiativ.fi). Dessa är ibland skapade av myndigheter och ibland av

privata aktörer, men de fungerar ändå som viktiga kommunikationskanaler och även reella påverkningssätt mellan medborgarna och beslutsfattarna (i båda riktningarna).

Forskningsdata

Tillvaratagandet av forskningsdata är viktigt eftersom vetenskapen kräver att forskningen ska kunna granskas och upprepas. Vetenskapen kräver också största möjliga transparens, varför publicering av dessa data är att vänta (och kräva) i allt högre grad.

Utmaningarna är ändå mycket stora eftersom datamängderna är enorma och standardisering och metadata är av mycket varierande nivå. Också inom till exempel humaniora uppstår i dag data, som inte alltid ens av forskaren själv uppfattas som data. Nationella infrastrukturer som sträcker sig ända in i universiteten också för humanister vore nödvändiga att åstadkomma snabbt för att trygga forskningens nivå i såväl Finland och Sverige som andra länder.

En delvis ny utmaning i sammanhanget har att göra med att många av de data som används och som genereras i samband med forskning numera är webbaserade. Det betyder att de ofta är mer kortlivade och föränderliga än pappers- och tryckbaserade media, vilka i många fall är mer permanenta men i gengäld har sämre tillgänglighet då de bara finns i ett eller (relativt sett) få exemplar på en eller ett begränsat antal fysiska platser.

Publicerad data

Data eller information publiceras också av privata aktörer. Tidigare tryggade friexemplarslagen att dessa material bevarades för eftervärlden. I Finland utgör [Nationalbibliotekets](#) samlingar ett exceptionellt heltäckande arkiv över den tryckta offentligheten och det kompletteras av [Nationella audiovisuella arkivet](#) och [Yles arkiv](#); i Sverige fyller idag [Kungliga biblioteket](#) samma uppgifter. Sedan den digitala eran börjat, har en ny digital offentlighet uppstått som närmast bevarats i [Internet Archive](#) i USA och sedermera i nationella webbarkiv. Dessa arkiv är endast så kallade skrapningar, ögonblicksbilder som ger en helhetsbild

av webben. Sökbarheten och användbarheten är inte så bra och materialet inte så omfattande som de kunde vara.

Det övriga internet, som i dag omfattar till exempel en del populära spel som har sina egna klienter på användarnas maskiner (Minecraft, Angry Birds etc), ingår inte i webbarkiven, utan arkiveringen är helt beroende av spelarna i en community eller de kommersiella företagens egna arkiverings- och backuprutiner.

En särskild kategori material är det material som tjänstemän och politiker mer eller mindre privat sprider på sociala medier. Hit hör också journalistiska material producerade inom eller utom traditionella medieföretag och material som journalister sprider s.a.s. privat på plattformar som Twitter och Facebook. Dessa material kan kraftigt påverka politiska beslut. Dessa material hör också enligt min preliminära bedömning till de allra mest akuta att få kontroll på.

Även mediehus publicerar i dag en hel del material digitalt som inte nödvändigtvis arkiveras systematiskt, så som digitala specialupplagor, tv-sändningar eller kommentarer på webbsidorna. På grund av att skrapandet av webben är tungt för systemet, brukar inte webbarkivering utifrån göras särskilt djupt i strukturerna. Diskussioner och forum riskerar att ställas utanför systematisk arkivering.

Privata material

E-post, sms och fotografier utgör exempel på material som i dag utgör väsentliga delar av vår kultur och som endast undantagsvis tas om hand för långsiktig bevaring.

Problem

Det finns flera olika typer av utmaningar som gör arkiveringen av material svår att genomföra. Till dessa hör lagstiftning som gäller upphovsrätt, personuppgifter och äganderätten till data. Till exempel äger Twitter den data som deponerats vid [Library of Congress](#) och man kan inte utan lov och bistånd

av företaget Twitter göra körningar i materialet. Det är inte heller möjligt att ladda ner data från en Facebookgrupp eller -sida.

Arkivering av digitalbaserade material

Med digitalbaserade material avses sådana material som uppkommit i digital form och som inte har något "fysiskt original". Det gäller i teorin i stort sett alla material i dag, men i praktiken bara en del, eftersom man fortfarande printar ut en pappersversion i väldigt många fall där det är möjligt och förklarar det som "original". Men bland materialen finns en ökande mängd handlingar och information som man antingen inte kan överföra på pappersmediet, eller som det inte är ändamålsenligt att göra det med, såsom olika versioner av ett och samma dokument. Versionshantering finns i många dokumenthanteringssystem, men ofta printar man ut endast en slutlig version på papper.

Till material som inte ens går att presentera på papper med rimliga ansträngningar eller utan att förlora mycket information hör en del presentationer och kalkyldokument, databaser (som dock kan beskrivas så att de kan rekonstrueras eller användas manuellt åtminstone i teorin), webbsidor, rörlig bild, ljud och en hel del andra saker såsom 3D-modeller, datorspel, mjukvara osv. Situationen innebär att man inom minnesorganisationerna står inför flera stora utmaningar:

- Vad av allt detta ska sparas?
- Vem ska spara det?
- Hur ska det sparas?

Den första frågan av dessa är den svåraste. De två senare frågorna finns det möjligheter att åtgärda med existerande kompetenser och tekniker. Men det förutsätter att klara och övergripande insamlings- och bevarandestrategier görs upp på nationell nivå. En inte heller oviktig aspekt är frågan om vad allt detta får kosta och hur mycket som kan göras inom existerande budgetmedel bara genom att helt lägga om verksamhetsmodeller, och vad som kommer att skapa nya

kostnader. Hur mycket är samhället berett att betala för att det digitala kulturarvet bevaras? Vad har vi råd att förlora eller att göra om från början om information försvunnit? Vilka delar är sådana som inte kan återskapas ens till någon som helst kostnad?

Med tanke på medborgarnas rättsskydd och för att vi skall kunna förstå och undersöka samhällsliga fenomen behöver vi spara en hel del data. Att gamla webbsidor eller delar av dem, till exempel kommentarfält försvunnit, gör att vi kommer att ha svårare att se och beskriva vad som hände i webbens första skeden, eller hur journalistiken påverkats av nya typer av interaktion med läsarna. Det handlar om kritiska utvecklingsskeden, men det handlar också om kunskap som kan hjälpa oss att utveckla teknik och kommunikation på sådana sätt som stöder en sådan samhällsutveckling som vi anser önskvärd. Det gäller frågor som demokrati, utbildning och ekonomi.

Då det gäller digitalbaserade material är det avgörande att arkivfunktionerna och hela processen är ytterligt transparent och väldokumenterad. Eftersom vi saknar ett fysiskt original och ett digitalt original kan vara fullständigt likvärdigt och kan te sig identiskt med sin (förfalskade?) kopia, är den digitala proveniensens viktig. Men lika viktig är trovärdigheten hos den som levererar materialet. Ett trovärdigt digitalt arkiv följer nationella och internationella standarder och har en utförlig och öppen dokumentation.

Vid den tekniska processen att arkivera digitalbaserade material är en tumregel att man borde komma in i processen så tidigt som möjligt för att ha hela livscykeln under kontroll och dokumenterad. Detta är viktigt både av principiella skäl (behovet av grundlig proveniensdokumentation) och av tekniska och i förlängningen ekonomiska orsaker. Då det gäller dokument som är normala handlingar är detta ofta möjligt att lösa med ett bra dokumenthanteringssystem. För att det ska fungera är det mycket viktigt att användbarheten är god och funktionerna motsvarar arbetsprocesserna väl, annars hamnar relevant information lätt utanför informationsförvaltning och arkiv.

Då det gäller arkivering av andra typer av data är utmaningarna större och kräver främst eftertanke och planering på längre sikt, så att man inte plötsligt besitter endast obsolet information eller, ännu värre, ingen information alls. En verklig risk är också den, att man köpt system som tjänster och plötsligt inte mer har tillgång till informationen i sin strukturerade form. Till denna kategori hör till exempel webbsidor. Vill man då förlita sig på de skrapningar som görs av webbarkiven, eller vill man själv arkivera sina webbsidor, så att man alltid vid behov (t ex vid en rättsprocess) kan få fram all data? Detta är något man inom alla organisationer borde ta ställning till. För detta behöver man en tillräckligt omfattande arkivbildningsplan.

Vad gäller information som finns ute på webben i övrigt kan man konstatera att många leverantörer inser att det är ett kriterium för trovärdighet och kundernas förtroende att man alltid kan hämta hem sin data. Detta gäller till exempel Twitter och Google. Samtidigt har man inte på Twitter direkt någon möjlighet att automatiskt ladda ner även *andras* tweets, vilket kanske en myndighet borde göra om man svarat på en fråga, och bestämt att sådan korrespondens ska arkiveras. På Facebook ger man inte heller (för närvarande åtminstone) upprätthållare av sidor eller grupper möjlighet att ladda ner hela sidans innehåll. Orsaken till denna linjedragning kunde tänkas ligga i upphovsrättsfrågor, men detta borde vara en sak som går att stipulera i användaravtalen. Det vore önskvärt att arkivväsenden i olika länder, helst koordinerat, kunde vända sig till företagen (i synnerhet Facebook) och försöka avtala om denna fråga. Det skulle underlätta arkiveringen betydligt för många myndigheter, företag och föreningar, om de enkelt kunde kopiera ner innehållet från sina Facebooksidor.

Urval – vad ska sparas?

Vi granskade ovan vilka olika typer av data som uppstår i dag i digital form. Dessutom finns en typ av data som är essentiell i sammanhanget: själva datorkoden, som på sätt och vis kan vara publicerad som produkt, men vars struktur kan vara hemlig av kommersiella skäl. Ett program eller en applikation

kan alltså vara publicerad, trots att koden inte är öppet publicerad. Denna typ av data hör till den privata sfären som ägs av företag, eftersom själva koden inte har upphovsrätt utan kan ha patent som skyddar den mot direkt kopiering.

Om vi närmar oss problematiken ur kulturarvs- och forskningsperspektiv för att försöka ringa in vilka data som är relevanta att spara, hamnar man snabbt in på väldigt teoretiska och filosofiska frågor, såsom vad kulturarv egentligen är. Det är inte ovanligt att man indelar kulturarv i materiellt och immateriellt arv, vilket i sig blivit problematiserat under 1900-talet i takt med att modern teknologi fått en allt större betydelse i vårt samhälle, samt på grund av utvecklingen inom kultur- och mediehistorisk forskning. Det är i dag uppenbart att mediet inte kan förbises som en bidragande faktor i hur det "immateriella" kulturarvet formas. Den bokhistoriska forskningen, men också hela den moderna ekonomiska historieskrivningen, har visat för oss att det materiella inte är något som kan förbises, och att ekonomiska och mediala faktorer påverkat till exempel de aspekter av kulturarvet som undersöks inom folkloristiken, för att inte tala om lärdomshistoria eller politisk historia. Samtidigt är det ytterst viktigt att man inte sammanblandar de ekonomiska och resursrelaterade aspekterna med frågor om värde och värderingar, varken i dag eller då man talar om det förflutna.

Om vi som utgångspunkt väljer att tänka oss att framtida forskning i kultur, samhälle och kulturarv i stort kommer att ha samma strukturer som i dag, riskerar vi att göra en missberäkning som kommer sig av att mycket vetenskaplig kunskap föds genom interaktion *mellan* olika discipliner. Detta är dessutom en trend som stärks av den tekniska utvecklingen, som ger redskap för att integrera olika typer av metoder och material på radikalt nya sätt.⁹⁴ Den tvärvetenskaplighet och mångdisciplinaritet som det talats om i så många år, utan att de så ofta lett till genuint integrerade typer av ny kunskap, kommer

⁹⁴ Lev Manovich, *Software Takes Command* (Bloomsbury Academic 2013), http://issuu.com/bloomsburypublishing/docs/9781623566722_web (2013-11-19), och Martin Weller, *The Digital Scholar* (Bloomsbury Academic 2011), DOI: <http://dx.doi.org/10.5040/9781849666275>; för avsnittet om interdisciplinaritet se http://www.bloomsburyacademic.com/view/DigitalScholar_9781849666275/chapter-ba-9781849666275-chapter-006.xml (2013-11-19).

sannolikt, i takt med att digitala forskningsmetoder och till exempel länkad data blir vanligare, att faktiskt leda till *ett* allt mer enhetligt fält inom humaniora, där man inte längre kan dra klara disciplinära gränser. Vi kommer troligen allt oftare att arbeta med liknande material och liknande metoder inom traditionellt skilda humanistiska forskningsdiscipliner. Det är förmodligen just i skärningspunkter mellan olika traditionella områden och även i samarbete med helt andra vetenskaper och till och med konstnärer, som den nya kunskapen kommer att födas.

Det är alltså inte ändamålsenligt att tänka sig att man kan definiera kommande behov av forskningskällor utgående från tanken att vi behöver källor för "ekonomisk historia", "politisk historia", "begreppshistoria" eller "förvaltningshistoria". Snarare kunde man försöka se på den pågående utvecklingen i samhället och fråga sig vilken information som kan vara relevant med tanke på för det första rekonstruktion av olika sakförhållanden, och för det andra för att kunna förklara olika större skeenden i samhället. I slutändan kommer vi rätt nära frågan om vilka material journalister, konst- och samhällsforskare i dag använder eller borde använda i sitt arbete.

Det bästa sättet att närma sig frågan är antagligen därför att försöka ställa potentiella forskningsfrågor till samtiden. Dessa kunde vara, förutom "traditionella" forskningsfrågor om utvecklingen inom förvaltning, lagstiftning, ekonomi eller personhistoria, frågor som berör just rekonstruktion av till exempel offentlig eller privat debatt, frågor om maktbruk eller informationens gång, eller om hur själva kulturen och samhället förändras på ett mer principiellt plan. Eftersom data och teknologin är i en nyckelroll, blir det viktigt att spara särskilt den typ av data som nämndes i inledningen, det vill säga att bevara själva mediet, vilket i synnerhet betyder datorkod.

Tyvärr är just datorkoden något som i dag faller utanför den så kallade plikt- eller friexemplarslagstiftningen. Den faller dessutom utanför upphovsrätten, trots att den är proprietär i många fall. Just nu är det ägaren till koden som ansvarar för bevarandet, vilket inte är ett särskilt pålitligt tillvägagångssätt. Företag tenderar i allmänhet inte värdesätta "gammal information" tillräckligt

enligt vad åtminstone historiker anser. Om man sökt patent finns sannolikt en del information om lösningarna hos myndigheterna, men de gäller inte hela program, utan enskilda lösningar. Det bästa sättet att lösa detta är sannolikt att lagstiftarna tar den rätt de har att bestämma över företag för att få ut information för bevarande, vilket kan göras utan att kompromettera konkurrensen. Att forska i själva datorkoden som ett kulturellt fenomen är också ett relevant område.⁹⁵

Andra frågor som nu känns relevanta, är frågor om hur makten och offentligheten förändrats i vårt samhälle i och med de nya mediernas intåg i allmänhetens vardag. För att forska i detta kommer vi att behöva mycket material från webben och sociala medier. Vi behöver också känna till hur fenomen som Wikileaks, Wikipedia eller Open Knowledge-rörelsen utvecklats, både produktionen och användningen av material och tekniker, och hur medierna ompositionerat sig i förhållande inte bara till läsarna, utan också annonsörer, makthavare och journalister i och med den digitala utvecklingen. Också utvecklingen av näthandel och företagen är relevant, företag som Amazon och Google besitter mycket data om penningströmmar och kulturella beteenden i sina databaser.

För att kunna rekonstruera och förstå verkligheten för dagens människor behövs exempel på helheter ur sociala medier, spel, sms, e-postanvändning, bilder, appar, spam ... Eller räcker det att vi har subjektiva beskrivningar av människorna själva? Ser vi själva som individer vad som händer just nu vad gäller långsammare strukturella förändringar? Hur har webben påverkat hur offentlighet, kultur och makt ser ut?

Det förefaller uppenbart att vi inte i dag kan gestalta vår samtid och utvecklingen tillräckligt tydligt.⁹⁶ Historien är per definition något som skrivs efteråt. Den kommer sannolikt delvis att skrivas med hjälp av stora datamängder, som öppnar för strukturell analys av olika fenomen, så som

⁹⁵ Se t.ex. Marc C. Marino, "Critical Code Studies", *Electronic Book Review* 4.12.2006, <http://www.electronicbookreview.com/thread/electropoetics/codology> (2013-11-19).

⁹⁶ Här kan man hänvisa till Paul Ricœurs omfattande produktion kring historia, t.ex. *La mémoire, l'histoire, l'oubli* (Seuil 2000).

kommunikation eller penningtrafik. Det ter sig därför viktigt att faktiskt försöka spara stora datamängder. I dag är mycket av dessa datamängder ändå i huvudsak privat egendom hos stora företag. Microsoft, Nokia eller Apple vet minst lika mycket om oss och vårt samhälle som våra myndigheter. Till exempel Twitter har delvis öppnat sin data för externa forskare, medan många andra företag och medier skyddar sin data av kommersiella och integritetsskäl.

Man kan ändå tänka sig att minnesorganisationer och lagstiftare kunde och borde göra upp systematiska planer för hur man kunde ta tillvara digitalt kulturarv och försöka åtgärda de juridiska behov som finns gällande pliktexemplar och upphovsrätt. Forskningens behov borde ha högsta prioritet, även om man givetvis bör se till att man inte riskerar att störa marknader eller hotar principerna för personlig integritet. Forskare har traditionellt haft en etisk kodex och man har hyst förtroende för deras integritet och goda avsikter, och detta förtroende som nu blivit kringskuret av förlegad eller ensidig lagstiftning borde explicit utsträckas till användningen av digitala material. Det betyder i praktiken att man för akademisk forskning borde kunna få kopiera, spara och behandla vilka data som helst.

Forskare bör vara utbildade i att kunna bedöma hur de presenterar sin forskning så att den inte kränker några rättigheter. Personlagstiftning och upphovsrättslagstiftning reser i dag hinder för verksamhet inom forskning och minnesorganisationer. Dessa hinder borde snarast åtgärdas. Det är inte heller orimligt att till exempel företag, som drar stor nytta av grundforskning och offentliga utbildningssystem, ger sitt bidrag tillbaka också i form av data, som i sin tur efter forskningen sedan kan stöda deras egen utveckling och innovationsverksamhet. Forskarna och minnesorganisationerna bör å sin sida sedan vara mycket noggranna med hur man använder data, så att man inte förvränger konkurrensen eller läcker sådan information som skadar någon.

Långsiktigt digitalt bevarande

Förutom valet om vad som skall sparas, står vi förstå inför ytterligare frågor då det gäller *hur* sparandet ska gå till. Vi måste också fråga oss *vilka egenskaper*

vi vill spara hos det vi vill bevara, eftersom man sannolikt i något skede kommer att stå inför konvertering av material och man måste bestämma hur viktiga olika egenskaper hos ett material är: Behöver vi i framtiden veta vart alla länkar ledde från denna webbsida? Behöver vi ha kvar PowerPoint-presentationens animerade övergångar? Måste vi ha kvar annonserna i YouTube-videon (om nu någon arkiverar YouTube...)? Vi kommer här att kort gå igenom några grundbegrepp, som är viktiga att känna till. Utgångspunkten är att allt digitalt material föråldras och blir det som ibland kallas obsolet också på svenska, det vill säga att en hård- eller mjukvara blir föråldrad eller ett filformat blir gammalt och helt oanvändbart förr eller senare. Dessutom kan både medier, mjukvara och data gå sönder av många olika orsaker. För att eliminera risker behövs systematisk planering av bevarandet.⁹⁷

I huvudsak finns det två olika strategier för att bevara digitala material. Ett vanligt alternativ inom arkivvärlden i dag är att man väljer **migrering**, vilket innebär att man är inställd på att konvertera mellan medier, filformat och program allt efter att tekniken utvecklas. Man migrerar alltså både själva innehållet till nya format och flyttar det till nya medier, eftersom båda anses ha begränsad livstid. Till exempel en CD-skiva anses inte ha så mycket längre livstid än en C-kassett av god kvalitet, det vill säga ett par decennier. Som medium har man inom arkiv ofta i dag så kallade LTO-band, det vill säga magnetband av hög kvalitet. Orsaken till detta är delvis historisk och kommer från en tid då utrymme på datorernas hårddiskor var mycket dyrt. Man har nu investerat mycket i dessa i sig ekonomiska och välbeprövade tekniker, varför andra metoder kanske ibland nedvärderas utan tekniska orsaker. Numera diskuterar man nämligen både bevaring i egna distribuerade system, t ex LOCKSS ("Lots of copies keeps stuff safe", läs mer på <http://www.lockss.org/>) och andra former

⁹⁷ Ett centralt arbete på detta område är C Becker et al, "Systematic planning for digital preservation: Evaluating potential strategies and building preservation plans", *International Journal on Digital Libraries* 10:4 (December 2009), s. 133–157, DOI 10.1007/s00799-009-0057-1, tillgänglig som pdf på adressen: http://publik.tuwien.ac.at/files/PubDat_180752.pdf (2013-11-27).

av användning av hårddiskivor eller material som glas.⁹⁸ Magnetbanden är rätt känsliga för strålning, men å andra sidan drar de ingen ström så länge de vilar. Också data på magnetband bör emellertid kontrolleras och förfriskas, det vill säga skrivas om regelbundet, helst med ett års mellanrum.⁹⁹

Formatmigrering betyder att man konverterar data från ett filformat till ett annat. Eftersom det rör sig om översättning finns alltid en allvarlig risk för att information ska gå förlorad. Frågan är bara hur mycket och vilken information man väljer att tappa. Dessa frågor är mycket tekniska och beror på vad man tänker sig att man i framtiden skall använda informationen till (det kan man trots allt inte med säkerhet veta, men man måste ändå göra medvetna val här och inte låta den tekniska terminologin skrämman en, utan be att få den förklarad för sig genom konkreta exempel). Format föråldras i takt med att programvarorna utvecklas. Har programmet en öppen kod som finns tillgänglig på webben, betyder det att en it-kunnig person alltid kan ta fram information ur en fil, vilket inte alltid är fallet med kommersiella produkter. För att stå på säkrare grund gäller det alltså att fundera på att helst välja:

- filformat som är mycket vanliga (för bilder tiff el jpeg, för text pdf, olika xml-baserade format)
- filformat som har öppen kod (för text odf, presentationer odp)

Och som sagt man måste fundera över vilka egenskaper man vill bevara, vilken funktionalitet som är relevant i synnerhet då det gäller digitalbaserade material. Detta är inte enkelt, men det är viktigt att beakta. Då man använder sig av databaser vid sin forskning blir dessa frågor ännu viktigare för den vetenskapliga trovärdigheten.

⁹⁸ Chris Welsh, "Hitachi invents quartz glass storage capable of preserving data for millions of years", *The Verge* 2012-09-27, <http://www.theverge.com/2012/9/27/3417918/hitachi-quartz-glass-data-preservation> (2013-11-27).

⁹⁹ Mycket bra diskussion förs på bloggar. Bra ställen att börja är Library of Congress blogg *The Signal*, <http://blogs.loc.gov/digitalpreservation>, och David Rosenthals blogg *DSHR*, <http://blog.dshr.org/>.

Den andra strategin för bevarande är **emulering**. Det innebär att man ser till att man bevarar de tekniska miljöerna, eller tillräcklig information om dem, så att man kan skapa virtuella miljöer för att använda gamla filformat eller program. Denna metod verkar vara mer populär bland personer med rent teknisk bakgrund. Den ter sig enklare och mindre komplicerad och riskfylld. I praktiken är metoden ändå en verklig utmaning, eftersom mycket av mjukvaran är proprietär och det inte finns några juridiska möjligheter i dag att arkivera och dokumentera till exempel de i dag så oerhört vanliga programmen från Microsoft eller Adobe. Samtidigt kan man konstatera att konvertering till öppna format och till exempel arkivdugliga pdf-filer innebär stora risker för förlust av relevant information.

Som vi tidigare skrivit är **teknisk och administrativ metadata**, samt annan dokumentation så som systemens loggar, viktiga för både själva bevarandet och för trovärdigheten. Den **tekniska proveniensen** är oerhört viktig för källkritisk bedömning av material; det är en sak som direkt handlar om informationens autenticitet och integritet. **Autenticitet** är nära förknippat med trovärdighet och proveniens. Eftersom digitala objekt lätt kan kopieras och även förändras på ett ofta osynligt sätt kan det vara svårt att fastställa autenticiteten på digitala objekt. Med noggrant dokumenterad proveniens kan man tolka det digitala objektet i förhållande till ett originalobjekt och fastslå graden av autenticitet. Integritet är också viktigt för autenticiteten. **Integritet** innebär att det digitala objektet inte har förändrats inom en viss tidsperiod. Det kan man försäkra sig om genom att räkna ut **kontrollsummor (checksum)** av ett objekt på olika tidpunkter och jämföra dem. Om summorna är samma är objektet oförändrat och dess integritet bevarad. Orsaker till att objekt förändras kan t ex vara medveten förändring eller dataröta.

Dataröta är ett fenomen som förekommer mycket sällan och innebär att koden utan synbar orsak av sig själv förändrats. Har man att göra med mycket komprimerade filer är detta ändå ett mycket stort problem. Det beror på att man vid komprimeringen använder sig av matematiska algoritmer för att krympa ihop data, så att den sedan kan "svällas upp" igen då den ska användas genom en

omvänd komplicerad räkneoperation. Då kan det räcka med att en enda liten nolla blivit en etta eller försvunnit för att hela processen av återskapande skall misslyckas totalt och all information vara förlorad. Detta är den enkla orsaken till att man alltid försöker undvika komprimering då det är möjligt. Dataröta är alltså mycket sällsynt, men ju större mängder data man sparar desto sannolikare är det att man råkar ut för det. Det är också en tungt vägande orsak till att man gärna skall ha många kopior av all data.

För att bekräfta att en fil är oförändrad måste man regelbundet räkna kontrollsummor och varenda förändring eller annan åtgärd måste dokumenteras i metadata. Detta kallas filens proveniens och den måste, liksom då det gäller alla arkivmaterial, vara dokumenterad så bra som möjligt, eftersom den påverkar den källkritiska bedömningen.

I slutändan handlar ändå allt om hur trovärdigt ett arkiv är, för även digital proveniens kan vara bristfällig, för teknisk för en lekman att värdera eller helt enkelt rent förfalskad. För själva bevarandet kan det däremot vara vitalt på lång sikt att veta exakt vilket program som använts och vilka åtgärder som vidtagits under olika skeden av filens livscykel. Genomskinligheten och öppenheten är i sig kanske det allra viktigaste kriteriet för en humanist vid denna bedömning. Ger arkivet ut information om teknisk proveniens och en beskrivning av sina system in i minsta detalj? Använder man sig av internationella standarder som Open Archival Information System (OAIS) eller Preservation Metadata (PREMIS)? I annat fall kan och bör man fråga sig: varför inte? Detta verkar möjligen lite överdrivet tillsvidare, men betänk en situation, som är rätt nära förestående, då stora delar av alla offentliga handlingar bara finns i digital form. Då måste vi kunna vara säkra på att ingen fixat till ett eller annat protokoll eller beslut, oberoende av om det migrerats flera gånger.

Fördjupning: Bevaringssynpunkter på några vanliga filformat

Juha Lehtonen

Att välja filformat då man planerar bevaring är inte en enkel process. I allmänhet rekommenderas att vanliga "kontorsdokument", såsom presentationer, textdokument och kalkyldokument, konverteras till antingen [PDF/A](#) eller något annat öppet xml-baserat format, så som de format som används av [Open Office](#).

Då man väljer filformat ska man alltid tänka på ändamålet: Hurdana egenskaper behöver filen ha? Är det viktigt att bevara utseendet, innehållet eller funktionaliteten? Åtminstone då det gäller filer skapade i vanliga kontorsprogram (ODF, OOXML, PDF) kan alla funktioner vara svåra – om inte omöjliga – att bevara, eftersom filerna kan innehålla väldigt mycket egenskaper och olika komponenter. Dessutom kan olika element infogas i en enda fil nästan hur som helst, till och med innanför varandra.

Och hur går det med att migrera mellan olika filformat? Möjligheten att redigera dokumentet går förlorad, men utseendet kan vanligen bevaras ganska väl, när ett MSOffice-dokument konverteras till en PDF/A. Å andra sidan bevaras redigerbarheten, men inte nödvändigtvis utseendet, om samma dokument konverteras till ODF-format. Innehåll och utseende är inte heller nödvändigtvis samma sak, då man konverterar till ODF. Presentationer kan till exempel innehålla funktioner, då man på samma bild upprepade gånger lägger nya element ovanpå andra genom en animation. När filen konverteras till PDF/A, syns nog de översta objekten, men animationsfunktionen försvinner och därför försvinner också den information som ingick i den (alltså de underliggande elementen). Ibland kan det vara en fördel att man tappar information. Till exempel räcker kanske ett komprimerande ljudfilsformat bra, då det gäller att bevara talande röster, medan ett format utan packning verkligen tar mycket utrymme.

Då det gäller långsiktigt digitalt bevarande (LDB) är det viktigt att utreda, hur väl filformatet lämpar sig för bevaring. En av kärnfrågorna är: Kommer vi att få problem, när vi vill överge det filformat vi nu tar i bruk? Det leder oss in på en del nya frågor: Finns dokumentationen av detta format öppet tillgänglig? Kostar den? Är specifikationen formellt standardiserad? Hur mycket används filformatet vid bevaring eller över huvud taget? Är specifikationerna etablerade eller kommer det ofta nya versioner av filformatet? Beaktar man kompatibiliteten framåt och bakåt mellan de olika versionerna? Har filformatet egenskaper som är beroende av en specifik hårdvara? Innehåller filen länkar eller hänvisningar till externa objekt? (Observera att detta också kan gälla de delar av filen som användaren normalt inte ser, alltså metadatan inne i filen.) Hurdana infogade objekt innehåller filen? Kan filen repareras om den gått sönder? Oberoende av hur bra ett filformat är, kommer det en dag att inte längre vara i allmän användning, och då står vi inför samma frågor med de nya formaten.

Filformatens specifikationer är en noggrann beskrivning av hur filen är uppbyggd och vad den kan innehålla. Då det gäller långsiktigt bevarande är specifikationen mycket viktig och därför bör man föredra sådana filformat, som det går att få tag i specifikationen för. Om den är standardiserad, är det tryggare att lita på formatet, för då uppstår det inte så lätt olika versioner eller varianter av filerna som sedan måste beaktas vid bevarandet. Av en noggrann specifikation får man på ett teoretiskt plan reda på alla egenskaper filen kan ha och man kan direkt bedöma hur väl de lämpar sig för LDB. Specifikationen möjliggör också förverkligandet av sådana verktyg, där de olika egenskaperna är beaktade med tanke på just LDB. Om man inte har specifikationerna för ett format och de program som normalt används skulle försvinna ur bruk, kan en fil i värsta fall bli helt oläsbar.

Hur är det då med oberoende? I kontorsfilerna är en vanligt förekommande hänvisning i filen uppgiften om typsnittet (ofta kallat fonten). I normalt bruk inkluderas inte fonten i själva filen, utan den innehåller ofta endast en hänvisning till en viss bokstavstyp, som mjukvaran sedan hämtar ur den aktuella

datorns typsnittsbibliotek. Om typsnittet inte hittas, väljs någon annan förhandsinställd bokstavstyp, vilket ibland kan förorsaka problem: till exempel kan raderna och sidnumreringen förändras och hänvisningar till texten kanske inte längre stämmer, texten i tabeller som ingår kanske inte längre passar in i sina celler osv. I kontorsfilerna kan man i dag också inkludera hela typsnittet, dvs exakt hur alla bokstäver ska se ut, och då kan texten alltid visas korrekt – förutsatt att mjukvaran man använder klarar av att utnyttja de infogade bokstäverna.

Att baka in olika typer av information i själva filen är inte alltid en bra lösning, eftersom detta kan komplicera innehållet ytterligare och göra bevarandet ännu svårare. Olika funktionella komponenter (t.ex. ActiveX-komponenterna) hör till den typen av egenskaper i en fil som inte är lätta att bevara. Kraven på bevaring kan föda nya typer av filformat, såsom PDF/A, som är en version av PDF som är en ISO-standard, där vissa vanliga egenskaper i PDF-filer är förbjudna.

Målsättningen med PDF/A är att säkra att filerna är oberoende av vilken hårdvara som används, att de inte innehåller några hänvisningar utanför dokumentet självt (t ex ingår hela bokstavstypsnittet) och filen dokumenterar (beskriver) sig själv. Allt detta är egenskaper som är relevanta vid LDB.

Till slut ännu några ord om metadata som sparas inne i själva filen: Metadata kommer att utgöra en mycket viktig del av bevarandet, eftersom man i dem kan ange en beskrivning av innehållet, proveniens, tekniska uppgifter och rättigheter. I en del filformat kan man infoga väldigt mycket metadata och det lönar sig förstås att göra där också, men metadata borde också bevaras någonstans separat i ett för dem lämpligt format, till exempel som xml. Att bevara metadata separat tryggar tekniskt deras läsbarhet, och möjliggör förverkligandet av effektiva sökningar, sortering eller olika statistiska operationer. Dessutom är det inte sagt att man vid en formatmigrering lyckas bevara all metadata som finns inne i filen, varvid en separat hantering av dem ändå blir aktuell.

Kapitel 7. Att förmedla historia i en digital värld

Kenneth Nyberg och Jessica Parland-von Essen

Att arkivera och publicera forskningsdata

Det är inte bara det sammanhang i vilket forskning bedrivs, de frågor som ställs och de metoder som används som påverkas av den digitala utvecklingen, utan också hur resultaten av forskningen bevaras och görs tillgängliga. Om kärnan i den digitala revolutionen är att information kan lagras och spridas snabbare, enklare och billigare än förut, får det betydelse för vilka krav och förväntningar som finns på forskningens öppenhet och tillgänglighet. Transparens, möjligheten att granska de förutsättningar, material och metoder vetenskapliga studier bygger på, är ju en fundamental utgångspunkt i all forskning, och ur den aspekten har den digitala utvecklingen lett till nya möjligheter och utmaningar. Dels handlar det om hur själva resultaten sprids och görs tillgängliga, dels om hur underlaget med digital teknik numera också kan tillhandahållas offentligt och bevaras mer långsiktigt, och slutligen – inte minst – om hur de kunskaper forskningen genererar kan förmedlas i mer populära sammanhang och i undervisning.

Bevarande och tillgängliggörande

Frågan om forskningsdata är rent praktiskt den man måste börja tänka på tidigast i forskningsprocessen, eftersom det måste finnas utvecklade planer kring hur data ska hanteras redan innan arbetet inleds. I takt med att allt större del av det material man använder som underlag är digitalbaserat – och i princip allt det material forskarna själva producerar är ju det redan idag – kommer denna typ av bevarande och tillgängliggörande att bli en allt större fråga. Kraven på även humanistiska forskare att ha genomtänkta strategier kring detta ökar därför för varje år som går, samtidigt som många av dem – särskilt kanske inom ämnen som historia – inte är vana att tänka på sina empiriska material i termer

av "forskningsdata". Även om det är en term som ofta förknippas med kvantitativa forskningsmaterial i digital form är emellertid också kvalitativt material (digitalbaserat eller ej) forskningsdata som kan struktureras och märkas upp på ett standardiserat sätt.

Varför är då detta viktigt? I Finland utkom i november 2013 en rapport om tillgång till forskning vari man också inkluderade öppna data.¹⁰⁰ I rapporten framkommer vilken stark agenda som finns för s.k. Open Access både internationellt och nationellt. På hemsidan för [Svensk Nationell Datatjänst](#) (SND) i Göteborg som vi strax återkommer till, finns en mer utförlig redovisning av skälen till att bevara och tillgängliggöra forskningsdata, men bland annat handlar det om att forskningsdata kan vara användbara i andra sammanhang än de där de ursprungligen samlades in eller skapades. Ofta är forskningen offentligt finansierad och kostar en hel del att genomföra, varför det vore slöseri att inte ta tillvara både resultat och råmaterialet – forskningsdata – för framtida bruk. Mer principiellt är det ju också, som redan påpekats, alltid viktigt med genomsynlighet i forskningen, det vill säga att grundvalarna för resultat och slutsatser kan kontrolleras av andra forskare. Ytterligare ett skäl som lätt glöms bort är att tillgängliga forskningsdata av hög kvalitet kan vara värdefulla material att användas i undervisning på olika nivåer.

Sedan 2012 kräver Vetenskapsrådet att alla ansökningar om projekt där insamling av data utgör en viktig del ska innehålla en datapubliceringsplan.¹⁰¹ Där ska det framgå vilken typ av data som samlas in och hur, hur dessa data lagras och bearbetas under projektets gång samt hur de ska bevaras på längre sikt och göras tillgängliga för andra. En viktig punkt i datapubliceringsplanen är en "metadatadeklaration", där man beskriver vilken typ av metadata som materialet kommer att märkas upp med. Metadata är, som vi skrivit om här tidigare, mycket viktiga för att göra ett material tillgängligt och användbart, och

¹⁰⁰ En länk till rapporten samt en kommentar på svenska finns på bloggen *Essetter* <http://essetter.blogspot.fi/2013/11/open-access-i-praktiken.html> (2013-11-29). I rapporten ingår också en sammanfattning av läget i Finland, tyvärr endast på finska.

¹⁰¹ "Vetenskapsrådets Stora utlysning 2012 har öppnat", <http://www.vr.se/franvetenskapsradet/nyheter/nyhetsarkiv/nyheter2012/nyheter2012/nukanduborjaforberedadinansokan.5.25e86c0f1350e9e545d148c.html> (2013-11-17).

de bör därför följa etablerade standarder och format som gör dem möjliga att enkelt "översätta" i olika sammanhang och för olika användningsområden.

För Riksbankens Jubileumsfond, den andra stora finansiären av svensk humanistisk-samhällsvetenskaplig forskning, är läget mera oklart; i de allmänna anvisningarna för forskare som söker anslag nämns inget om detta, men RJ:s VD Göran Blomqvist skriver i ett kort nyhetsinlägg från april 2012 att stiftelsen "ställer krav på att resultatet av alla infrastrukturella projekt ska göras allmänt tillgängligt". Det är dock oklart om detta syftar på publicering av forskningsresultat i sig eller tillgängliggörande även av forskningens råmaterial i form av de data som samlats in eller producerats under arbetets gång. Blomqvist skriver att det kan finnas "goda skäl" för RJ att liksom VR ställa krav på samtliga projektansökningar, inte bara de infrastrukturella, att innehålla en datapubliceringsplan, men något sådant krav verkar inte finnas i dagsläget.¹⁰²

Arkivering och länkning av data

Som ett svar på dessa behov har, på Vetenskapsrådets initiativ, [Svensk nationell datatjänst](#) bildats. Det är ett serviceorgan som lagrar, dokumenterar och tillhandahåller forskningsdata för svensk forskning inom humaniora, samhällsvetenskap och medicin. Syftet är att göra det lättare för svenska forskare att få tillgång till både svenska och internationella forskningsdata, och det är också möjligt för enskilda forskare att överlämna data till SND för tillgängliggörande. De accepterar bara kopior av material, eftersom lärosäten och andra organisationer har en lagstadgad skyldighet att arkivera material från de projekt de är huvudmän för. Liknande dataarkiv som SND finns i många andra länder och de samarbetar i ett nätverk som utgör "en viktig del av forskningsinfrastrukturen".¹⁰³

I Finland finns [Finlands samhällsvetenskapliga dataarkiv](#) i anslutning till Tammerfors universitet, där informationsvetenskapen är en stark disciplin. Dataarkivet är grundat 1999 och innehåller över tusen dataset av både

¹⁰² Göran Blomqvist, "Ska forskningsdata bli tillgängligt?", <http://www.rj.se/svenska/130/var/newsID/563> (2013-11-17).

¹⁰³ SND:s hemsida, <http://snd.gu.se/> (2013-11-17).

kvantitativ och kvalitativ karaktär. Det mesta materialet är, som namnet säger, samhällsvetenskapligt. Historia finns inte starkt representerat, men Riksarkivet är med i de styrande organen, och arkivet har en mycket tillmötesgående inställning till forskare. Trots detta har inte historiker tydligen kommit sig för att donera sina data till arkivet eller deponera dem där. Dataarkivet kunde eventuellt erbjuda en möjlighet också för digital humaniora, men det kräver målmedvetet arbete. Ett annat problem är att en del forskningsdata som historiker arbetar med de facto utgörs av material som finns hos minnesorganisationerna, varför det kunde vara befogat att fundera om till exempel arkiv borde ta emot och bevara material som direkt har anknytning till deras egna samlingar, till exempel fotografier som forskare tagit av deras annars odigitaliserade material.

Hela detta problem med materialens "hem" är emellertid mycket mindre i praktiken om man kan arbeta med genuint öppen och i synnerhet länkad data,¹⁰⁴ vilket borde vara den självklara målsättningen. Det viktiga är därför att fundera över struktur, dokumentation, bevarande och publicering redan i ett tidigt skede av forskningen, så att man till exempel som forskare ser till att man har möjligheter och rättigheter att publicera så mycket av materialen som möjligt. Det lönar sig att åtminstone konsultera dataarkiven gällande skapandet av dataresurser. Man bör förstås också vara ytterst försiktig då det gäller personuppgifter och till exempel lägga upp det hela så att materialen vid behov enkelt kan anonymiseras.

För att data ska kunna länkas är det av stor vikt att en nationell infrastruktur finns, som erbjuder standarder för hur man beskriver sina material. Vi behöver också känna till och använda oss av internationella vokabulärer och andra resurser för att hänvisa till exempel till ämnesord, format, rättigheter, geodata och mycket annat. Detta kräver en målmedveten administrativ styrning, som för

¹⁰⁴ Länkning av data innebär att man markerar informationen med exakta koder för relevanta begrepp. Tekniken har också kallats semantisk webb eller webb 3.0 och har vunnit allt mera terräng inom webben tack vare att aktörer som [World Wide Web Consortium](http://www.w3.org/), Tim Berners-Lee och Google för fram tekniken. Läs mer t ex på <http://www.w3.org/> eller <http://linkeddata.org/home>.

historikernas del omfattar både forskningsdata och minnesorganisationer och deras material.

Det nya publiceringslandskapet

Frågan om [bevarande och tillgängliggörande av forskningsdata](#) har ännu inte väckt någon större uppmärksamhet bland historiker och andra humanister. Frågan om tillgängliggörande av forskningsresultat har däremot diskuterats alltmer intensivt under senare år och är just nu brännhet i Sverige, delvis på grund av [Vetenskapsrådets beslut nyligen](#) att dra in sitt stöd till humanistiska och samhällsvetenskapliga tidskrifter. (Vi återkommer till det i nästa avsnitt.) Här liksom i andra sammanhang är tillgänglighet dock inte bara, eller kanske ens främst, en fråga om tekniska möjligheter utan också om institutionella förutsättningar, och det är ofta kring dem debatten om "open access" har rört sig. Den diskussionen har i sin tur sammanfallit med flera andra förändringsprocesser som börjat påverka publiceringsformerna inom humanistisk forskning på senare år.

Traditionellt sett har dessa skilt sig ganska mycket från förhållandena inom naturvetenskap och medicin, där publicering av artiklar på engelska i tidskrifter med "peer review" (kollegial förhandsgranskning) länge varit normen. Humanister har oftare skrivit monografier eller artiklar och bokkapitel på sitt eget modersmål, vilket vanligen skett i tidskrifter utan systematisk peer review. Det beror åtminstone delvis på att deras skrifter i högre grad läses även av lekmän och att forskningen av naturliga skäl ofta berör specifikt svenska eller finska förhållanden. Inom humaniora är publiceringen traditionellt också en betydligt mycket mer långsam process än inom andra vetenskaper, inte minst på grund av att textmängderna är så mycket större. Det krävs mycket tid att skriva, bearbeta, redigera och granska de långa texter som humanister behöver för att förklara och belägga sina resultat.

Under det senaste decenniet har emellertid de naturvetenskapliga publiceringstraditionerna alltmer börjat påverka även humanistisk forskning. Skälet till det är enkelt: En växande andel av forskningsmedlen fördelas idag på

grundval av system för att mäta och värdera forskarnas produktivitet ifråga om vetenskapliga publikationer, så kallad [bibliometri](#). Dessa system omfattar i allt högre grad humaniora och samhällsvetenskap, men de är uppbyggda utifrån de normer som traditionellt har gällt inom naturvetenskap, medicin och teknik, vilket betyder att artiklar i internationella peer review-tidskrifter premieras och monografier och artiklar på nordiska språk räknas som mindre värdefulla. Trycket på historiker och andra humanister att skriva för sådana sammanhang har alltså ökat kraftigt, vilket de senaste åren börjat få tydligt genomslag i särskilt yngre forskares sätt att tänka kring sin publicering. Eftersom de internationella tidskrifterna ofta ägs av kommersiella förlag (till skillnad från de nationella, som vanligen drivs av vetenskapliga samfund och sällskap) får dessa förändringar konsekvenser för forskningens tillgänglighet, genom att de ställer frågan om upphovsrätt kontra öppenhet på sin spets.

Upphovsrätt till forskning

Av hävd anses forskare ha upphovsrätt till sin forskning. Upphovsrätten innebär att upphovsmannen under alla omständigheter har rätt att bli omnämnd då man använder sig av hans forskning, vilket också fyller en funktion i den vetenskapliga traditionen där all kunskap måste underbyggas och förhålla sig till annan forskning. Emellertid kan forskaren ge ifrån sig rättigheten att publicera sin text och forskning genom ett avtal med en förläggare. Denne kan sedan låta granska arbetet med hjälp av externa experter, vilket ger det mervärde som akademisk merit och i bästa fall som forskning. Forskaren måste i detta läge avväga hur mycket det är värt att publiceras av ett visst förlag och om det är mera nytta eller skada av att forskningen finns tillgänglig endast för en begränsad publik, vilket också kan minska på spridningen och antalet citeringar.

De kommersiella förlagen, vilka ofta ingår i multinationella mediekoncerner, bevakar omsorgsfullt sina rättigheter genom att se till att deras publikationer bara når dem som betalar för tillgång. Kunderna är till stor del universitetsbibliotek runt om i världen, vilka köper in tidskrifter och böcker för att ge "sina" forskare tillgång till dem. Priset för många tidskrifter har ökat

kraftigt en lång rad år i följd, vilket gjort det allt vanligare att bibliotek – även ganska välbeställda sådana – minskar på antalet tidskrifter de köper in. På många sätt har alltså tillgängligheten till forskningsresultat, ofta producerade med stöd av offentliga medel, minskat under de senaste decennierna samtidigt som de tekniska möjligheterna att tillhandahålla material digitalt via webben har utvecklats dramatiskt. Det har gjort att forskare och deras resultat har hamnat i kläm mellan å ena sidan ökade förväntningar på tillgänglighet, å andra sidan krav på att publicera sig i kanaler som blir allt mindre tillgängliga.

Komna så långt kan det vara värt att påpeka att själva idén om upphovsrätt historiskt sett är relativt ny. Den är en viktig komponent i den moderna ekonomin, särskilt i och med framväxten av det så kallade kunskapssamhället, men den utmanas också på ett fundamentalt sätt av den nya digitala tekniken. Den har gjort att det i all praktisk mening varken kostar pengar eller ansträngning att reproducera och sprida idéer, kunskap och information för den individ som har en "dator" (brett definierat) och en nätuppkoppling. De flesta verksamheter som bygger på sådan reproducering och spridning, och som tidigare skyddats av det faktum att den bara kunnat ske i materiella former som krävt kapitalstarka investeringar, riskerar att undergrävas när transaktionskostnaderna nästan helt försvinner. Företrädare för sådana branscher försvarar därför tanken om upphovsrätt, medan andra menar att det är ett förlegat synsätt som bygger på förutsättningar vilka inte längre existerar. Det finns också många som argumenterar för att få eller inga idéer egentligen är genuint nya eller kan tillskrivas en enskild individ, utan ["everything is a remix"](#) av tidigare idéer; de kan därför definitionsmässigt inte heller ägas av någon. (Jfr avsnittet ["Publicerandets ekonomi"](#) i kapitel 2.)

Inom forskarvärlden har traditionellt tanken om den individuella originaliteten varit mycket viktig, eftersom det ofta är just genom enskilda bidrag – mer eller mindre betydelsefulla – som forskningen går framåt. Samtidigt har det alltid funnits ett medvetande om att det arbete som görs idag med nödvändighet vilar på det som gjordes igår, och framför allt inom medicin och naturvetenskap har forskningen i praktiken länge bedrivits i stora (och

växande) forskargrupper. En stark strömning inom humanistisk och samhällsvetenskaplig forskning de senaste decennierna har också betonat det problematiska i den traditionella bilden av det ensamma vetenskapliga geniet, och visat på hur forskningens framsteg alltid (i någon mening) är resultatet av gemensamma ansträngningar. Att så är fallet betyder givetvis inte att individuella bidrag inte kan ha stor betydelse för den fortsatta kunskapsutvecklingen, utan bara att enskilda människor alltid verkar i ett större sammanhang.

Öppenheten som ideal

Oavsett synen på individuell originalitet – och därmed hur mycket erkänsla enskilda forskare ska få för sina idéer – är öppenhet, transparens och fritt delande av resultat självklara och helt fundamentala värden inom all vetenskaplig verksamhet. Utvecklingen mot alltmer av publicering i låsta, kommersiella tidskrifter går emot sådana traditionella normer, och det är därför inte konstigt att det uppstått en motreaktion inom forskningen och en diskussion om arbetsfördelning, finansiering och distributionsformer inom vetenskaplig publicering. Åtminstone delvis knyter denna debatt an till den större, och ibland rätt radikala, rörelse av nätaktivister som verkar för större informationsfrihet och ett ifrågasättande av upphovsrätten i dess nuvarande form.

Mot den bakgrunden kan "open access" ses som uttryck för en mer principiell hållning rörande alla delar av forskningsprocessen. Det är därmed ett samlingsnamn på en rad olika rörelser och initiativ för att göra både forskningsresultat och forskningsdata så öppet tillgängliga som möjligt, men här och i följande avsnitt är det framför allt publicering av resultat som står i fokus. (Se föregående kapitel för motsvarande diskussion om [arkivering och publicering av forskningsdata](#).) "Tillgänglig" syftar i detta sammanhang egentligen på flera saker, främst kanske att materialet är gratis att ta del av men också att det är lätt att hitta och tekniskt enkelt att få tillgång till samt, inte minst, att det bevaras och arkiveras på ett sätt som gör det tillgängligt även på

längre sikt. Olika modeller för att åstadkomma detta, och hur frågan om open access har utvecklats under de senaste åren på en mer allmän nivå, utgör ämnet för nästa avsnitt.

Open access i teori och praktik

Jessica Parland-von Essen och Kenneth Nyberg

Eftersom frågan om "open access" (OA) fram tills nyligen främst berört forskare inom naturvetenskap och teknik har mycket av diskussionen om den hittills mest kretsat kring publicering av tidskriftsartiklar. I det följande kommer vi att ta upp några aspekter av den diskussionen och några av de OA-modeller som föreslagits och prövats, och även här ligger tyngdpunkten på artikelpublicering. I [Yrsa Neumans fördjupningsartikel](#) behandlas dock lite närmare de frågor som gäller just publicering av böcker med OA, vilket ju har särskilt intresse för historiker och andra humanister.

Öppna publikationsarkiv

Inför millennieskiftet år 2000 utlyste [Clay Mathematics Institute](#) i Cambridge, Massachusetts, det så kallade Millenniumpriset. Man hade valt sju olösta svåra problem, som matematiker kämpat med under 1900-talet. Lösningar skulle belönas med en miljon dollar. En av uppgifterna var beviset för Poincarés förmodan, formulerad av Henri Poincaré år 1904: "Varje slutet, enkelt sammanhängande 3-dimensionell mångfald är homeomorf med 3-sfären."¹⁰⁵ (Det är en sats som handlar om sfärers topografi.)

Hösten 2002 laddade den ryskjudiska matematikern Grigori Perelman upp sina första bevis på Poincarés förmodan, som följdes av två andra artiklar på den öppna publikationsplattformen [arXiv](#). De väckte genast stort intresse bland matematiker runt om i världen som gav sig i kast med att undersöka bevisen,

¹⁰⁵ Clay Mathematics Institute, "Millennium Problems", <http://www.claymath.org/millennium-problems> (2013-12-12). Den svenska problemformuleringen är hämtad från Wikipedia, "Poincarés förmodan", http://sv.wikipedia.org/wiki/Poincarés_förmodan (2013-12-12).

som verkar hålla. Att bevisa matematiska problem (av den kalibern de är i dag) är extremt omfattande och det behövs alltid en granskningsprocess som tar lång tid. I mars 2010 beviljades Perelman det första Millenniepriset. Han vägrade att ta emot det eftersom han ansåg sig ha utnyttjat en annan forskares arbete i sådan utsträckning att denne också borde ha belönats.

ArXiv har under 2000-talet blivit den plattform där nästan alla preprints av artiklar inom matematik, fysik, astronomi, datavetenskap, kvantitativ biologi och statistik publiceras.¹⁰⁶ Artiklarna utsätts inte för någon förhandsgranskning. Perelman valde att publicera sig direkt på arXiv, som uppstått som ersättning för en klumpigare e-postlista, i stället för en traditionell förhandsgranskad tidskrift. Hans och även andra forskares exempel har visat att det hela fungerar mycket bra. Forskare är rätt noggranna med vad de lägger ut på arXiv; kvaliteten är i allmänhet god, eftersom författarna är rädda om sitt forskaransende.¹⁰⁷ Samtidigt går själva publicerandet snabbare och billigare än någonsin tidigare.

Ett digitalt publikationsarkiv är ett ställe där forskarna kan publicera sina artiklar, avhandlingar och annat material i elektronisk form. De digitala publikationsarkiven bygger på OA och har uppkommit som ett steg i tillgängliggörandet av forskningsresultat. En allmän tillgång av forskningsresultat genom ett öppet publikationsarkiv ligger i såväl såväl forskningsfinansiärernas som forskarnas intresse. Forskarna ser många fördelar med dylika öppna arkiv, eftersom de också fungerar som en sorts publikationsförteckningar i fulltext och ökar synligheten för den egna forskningen.

Det öppna publicerandet har också kritiserats.¹⁰⁸ En konservativ syn gällande vetenskapligt publicerande fokuserar på försvar av upphovsrätt och en oro för

¹⁰⁶ Preprints är långt gångna men oftast inte helt slutgiltiga versioner av vetenskapliga artiklar som accepterats för publicering. Ibland kan det avse den version en författare lämnat in för kollegial förhandsgranskning, ibland den helt färdiga texten innan den fått sin grafiska utformning av förlaget inför publicering.

¹⁰⁷ Wikipedia, "arXiv", <http://en.wikipedia.org/wiki/ArXiv> samt <http://arxiv.org/> (båda 2013-12-12).

¹⁰⁸ Kate Worlock: "The pros and cons of Open Access", *Nature* [undated], <http://www.nature.com/nature/focus/accessdebate/34.html> (2013-12-12).

forskarnas och forskningens anseende (jfr [föregående avsnitt](#)). Men det borde vara självklart att risken för plagiat är mycket, mycket mindre om originalpublikationen finns öppet tillgänglig på internet. Dessutom har det visat sig, som på arXiv, att seriösa forskare tar publicerandet på allvar också om det är fritt och inte har förhandsgranskats. Överlag är socialt kapital en valuta som kraftigt underskattats i tidigare teorier: det har visat sig att folk i allmänhet tillmäter sitt anseende mycket stort värde på webben även i mindre seriösa sammanhang.¹⁰⁹ För forskare är det givetvis också frågan om professionalitet och etik.

Digitala publikationsarkiv har vuxit fram sedan 2003 då Berlinkonventionen undertecknades och ett internationellt samarbete inleddes, sedan biblioteken inte på annat sätt lyckats pressa de stora förlagen till rimliga licenskostnader. I mars 2008 rekommenderade European University Association universiteterna att vidta åtgärder för att skapa digitala publikationsarkiv. I de nordiska länderna startade nationalbiblioteken 2007 ett projekt för att koordinera sina arkiv. Det har sagts att 2010 var det år då OA-publicerandet nådde sitt verkliga genombrott.¹¹⁰ Under året vaknade det amerikanska forskarsamfundet till de problem som man redan länge tampats med inom biblioteksvärlden, dvs. de fullständigt överprissatta tidskrifterna som monopoliserat forskningsresultat – som dessutom ofta finansierats med offentliga eller allmännyttiga medel.

Det finns över 2500 öppna digitala publikationsarkiv i världen. Europa har gått i spetsen för utvecklingen, medan amerikanska institutioner står för en knapp femtedel. Endast en knapp femtedel av arkiven har haft en klart definierad publikationspolicy. De flesta arkiven har en betoning på vetenskapliga artiklar och avhandlingar, men över fyrtio procent innehåller

¹⁰⁹ Jfr Clay Shirky, *Here comes everybody: The power of organizing without organizations* (New York: Penguin, 2009).

¹¹⁰ Se till exempel Heather Morrisons blogginlägg där hon gått igenom statistik från olika öppna publikationsarkiv: "Dramatic Growth of Open Access: December 11, 2010 early year-end edition", *The Imaginary Journal of Poetic Economics* 2010-12-11, <http://poeticeconomics.blogspot.com/2010/12/dramatic-growth-of-open-access-december.html> (2013-12-12).

också opublicerade rapporter och "working papers" och i cirka en tredjedel finns konferensmaterial.¹¹¹

Modeller för open access

På Wikipedia finns en [överskådlig genomgång](#) av de olika lösningar och modeller för OA som föreslagits och prövats, men i korthet finns det ett par huvudspår. En modell är att författarna själva betalar för publiceringen i så kallade OA-tidskrifter som tillämpar kollegial förhandsgranskning ("Gold Open Access"); dvs. tidskrifterna finansieras av författaravgifter och därmed behöver de inte låsa materialet bakom betalväggar. Denna modell innebär att penningströmmen för publiceringen i vår nordiska modell inte längre går från staten via universitetsbiblioteken till förlagen. I stället går den från forskningsfinansiärerna, alltså staten eller enskilda fonder, direkt till förlagen via forskarna, som i samtliga modeller är utan annan egentlig ersättning än upphovsrätten och vetenskaplig meritering.

En annan variant ("Green Open Access") är att författarna själva kan göra tidiga eller slutliga versioner av sina texter tillgängliga i öppna arkiv på webben, men att den publicerade versionen med förlagets layout osv. ligger bakom en betalvägg. Denna modell är möjligen svårare att genomföra på grund av förläggarnas motstånd, men är i slutändan förmånligare ur ett samhällsperspektiv. Den kräver inte heller att man ökar mängden forskningsbidrag med pengar som hamnar hos en tredje part med kommersiella intressen. Det är en modell som innebär att idéerna i artiklarna får spridning samtidigt som det för citering krävs att man har tillgång till den betalda tidskriften, som på så vis kan bära sig ekonomiskt. Många kommersiella tidskrifter tillåter i dag, i varierande grad, parallellpublicering men kräver då ofta ett s.k. embargo. Det innebär att man inte får lägga ut artikeln på webben förrän en viss tid efter publicering – vanligen sex, tolv eller tjugofyra månader.

¹¹¹ Siffran från 2013-12-12, se närmare statistiken på *Directory of Open Access Repositories*, <http://www.openaccess.org/find.php?format=charts>.

Vissa tidskrifter som inte räknas som OA publicerar numera själva sina artiklar i fulltext på webben med embargo, oftast sex månader eller ett år; i Sverige gäller det till exempel [Historisk Tidskrift](#) och [Scandia](#). Det finns därtill en mängd andra varianter som faller någonstans på skalan mellan helt "stängd" och helt "låst" publicering, allt eftersom forskare, finansiärer, förlag, lärosäten och andra intressenter försöker navigera i det nya landskapet. Men ansvaret för att rättigheterna hanteras på ett korrekt sätt läggs nästan alltid helt och hållet på forskaren själv, såsom upphovsman och ursprunglig ägare till rättigheterna. Detta innebär att alla forskare måste sätta sig in i dessa frågor om de vill OA-publicera material. Det är något universiteten borde inse kan vara en onödig spärr för spridningen av forskning som producerats i anslutning till dem. Universiteten borde därför stödja OA-publicering av rent och skärt egenintresse på en mycket konkret och praktisk nivå.

Ett belysande exempel

Ett mycket belysande exempel på de friktioner som kan uppstå i det nya publiceringslandskapet är diskussionen om det beslut [Vetenskapsrådet](#) i Sverige nyligen fattade om att dra in sitt stöd till ett antal tryckta tidskrifter inom samhällsvetenskap och humaniora, en bidragsform som det inte finns någon motsvarighet till för naturvetenskapliga tidskrifter.¹¹² Beslutet motiverades med att stödet var för dyrt att administrera i förhållande till dess mycket blygsamma andel av VR:s totala budget. Ett annat skäl var att svenska forskare i högre grad borde publicera sig i internationella peer review-tidskrifter, gärna på nätet. Senare har VR-företrädare också menat att dess uppdrag är att stödja forskning av högsta kvalitet, inte "förutsättningar för forskning" till vilka tidskrifterna anses höra.

Reaktionen från de berörda tidskrifterna och många andra humanister och samhällsvetare blev kraftig och mycket negativ, där kritiker menade att just tidskriftsstödet är något av det viktigaste VR kan göra för forskningen, dessutom

¹¹² På Historisk tidskrifts hemsida finns en samling länkar till olika inlägg i frågan: <http://www.historisktidskrift.se/> (2013-12-12).

till en relativt låg kostnad. Andra anser att det därtill borde vara ett särskilt ansvar att slå vakt om svenskspråkiga publikationer inom humaniora och samhällsvetenskap, eftersom engelskspråkiga tidskrifter inte är lika tillgängliga för svenska läsare. Det gör i förlängningen att den humanistiska forskningens resultat inte får de samhällsnyttiga effekter som enligt den samtida forskningspolitikens nyttotänkande är dess viktigaste syfte.

Vad detta exempel ändå visar är att den nya situation vi befinner oss i, med de konsekvenser den fått för synen på forskningens tillgänglighet och publiceringsformer, sammantaget innebär en mycket svår utmaning mot de traditionella tidskrifterna. Visst är det viktigt att förstå att förutsättningarna inom humaniora och samhällsvetenskap skiljer sig från dem inom naturvetenskap och medicin, som med nödvändighet är mer internationella. Men samtidigt måste man fråga sig om det är ändamålsenligt att helt försöka hålla fast vid den traditionella modellen med regelbundet utkommande papperstidskrifter, ofta med mycket långa publiceringstider, på samma gång som tempot i den vetenskapliga diskussionen blir allt högre liksom i samhället i övrigt. Det finns de som anser att publicering i tidskrifter alltmer håller på att bli slutpunkten i, "arkiveringen" av, vetenskapliga utbyten snarare än startpunkten, och att forskning därför "begravs" där.¹¹³

Det kanske är att gå lite långt, men helt uppenbart är att webbpublicering ger möjligheter till mycket mer varierade och dynamiska former för både spontant tankeutbyte och vetenskaplig publicering i mer traditionell mening, liksom integration i de mer allmänna debatter och diskussioner som förs i samhället i stort. Det ställer frågorna om både tidskrifters och monografiers fortsatta funktion inom forskningen på sin spets i flera avseenden, och vi kommer därför att behandla dessa andra sätt att publicera sig i nästa avsnitt.

¹¹³ Nobelpristagaren Paul Krugman menar till och med att inom hans eget fält, ekonomi, fungerade tidskrifterna redan i början av 1980-talet som "gravvårdar" ("tombstones") för forskningsresultat. Paul Krugman, "Open Science And The Econoblogosphere", *The Conscience of a Liberal* 2012-01-17, <http://krugman.blogs.nytimes.com/2012/01/17/open-science-and-the-econoblogosphere/> (2013-12-12).

Fördjupning: Open Access-böcker – vad säger forskningen?

Yrsa Neuman

Open Access är inte bara beskrivningen av en elektronisk publikationsform, utan också en ideologi vars mål är att så stor del av forskningspublikationerna som möjligt ska göras tillgängliga gratis. Den här rörelsen är en reaktion på [the serials crisis](#), den ofantliga prisstigningen på prenumerationer på akademiska tidskrifter. Men den är också ett vetenskapspolitiskt ställningstagande: vetenskapen produceras för alla människor, i alla samhällen, och därför ska dess resultat inte hållas svårtillgängliga bakom till exempel betalmurar. Idag finns många forskare på universitet världen över vars bibliotek inte har råd med den litteratur de skulle behöva. Samtidigt är mycket av forskningen idag betald av stater, och man kan förvänta sig att också resultaten bör ges över till samhället i bred mening.

De mesta OA-ideologerna önskar att alla akademiska artiklar ska bli Open Access genast, utan några som helst barriärer. En sådan ideologi uttrycker till exempel den definition på Open Access som BOAI, [Budapest Open Access Initiative](#) formulerade 2001. Den här definitionen ligger som grund för DOAJ, [Directory of Open Access Journals](#), en databas över Open Access-tidskrifter som blivit tongivande idag. (Databasen är inte helt konsekvent på den här punkten men torde kunna bli det i framtiden.) DOAJ är ett mycket viktigt verktyg för tidskrifters synlighet. När det gäller tidskrifter, dvs. artiklar, har Open Access-scenen förändrats stort inom de senaste åren, men för böcker ser det annorlunda ut. Det finns en motsvarighet, DOAB, men den är knappast lika känd. OA-forskning om böcker inom humaniora är det egentliga temat för denna fördjupningsartikel.

Stegvis mot Open Access: hybridlösningar

Samtidigt som OA-ideologin frodas undrar man om det verkligen är möjligt att få forskningsvärlden att gå mot Open Access, och mot den mest ideologiska

formen, meddetsamma. En sak som vi som stöder Open Access borde arbeta för är att frigöra de mest prestigefyllda tidskrifterna ur de multinationella förlagens grepp. Men många nya frågor uppstår: hur ska arbetet med tidskrifterna finansieras, om vi egentligen vill ha en värld utan förlag? Vem ska se till att de akademiska texterna är indexerade och lätta att hitta, vem ska stämpla publikationsdatum för texter, och vem ska sköta kvalitetsgranskningen? I ljuset av insikten att världen kanske snarare borde ändra sig lite i taget mot mer Open Access finns nuförtiden en hel del forskning om möjliga alternativa affärsmodeller och publikationsmodeller där förlag ännu har en roll att spela. Hit hör forskning i Gold Open Access, dvs. när förlagen själva publicerar tidskriftsartiklar eller böcker Open Access, och olika hybrider, som man kallar tidskrifter där författaren väljer om artikeln blir Open Access eller inte – vanligen mot betalning. Att författaren å sin sida betalar för Open Access-möjligheten är bara en av flera möjliga modeller, om än en av de mer kända. Inom humaniora är den snarast en ökänd modell.

Inom det EU-finansierade forskningsprojektet [Agora – Scholarly Open Access Research in European Philosophy](#) som vi, en forskargrupp vid ämnet filosofi, Humanistiska fakulteten vid [Åbo Akademi](#) deltagit i, undersöktes några modeller som ligger mittemellan ideologisk OA och kommersiellt gångbar OA. Vi har undersökt både bokpublicering och tidskriftspubliserings inom europeisk filosofi. Våra resultat torde gå att generalisera någorlunda inom sektorerna humaniora och samhällsvetenskaper fastän ämnesvisa skillnader ofta finns.

Open Access-böcker – är det någon vits?

I samarbete med det tyska lilla men internationella filosofiförlaget Ontos publicerade Agoraprojektet den 1 juni 2011 på Ontos webbplats 27 böcker som givits ut i tryck tidigare som Open Access-publikationer. Böckerna hade publicerats mellan 2003 och 2011 och fyra av dem för färre än 18 månader sedan – tiden från tryck varierade alltså mellan 9 år och 6 månader.

Försäljningssiffrorna följdes noggrant upp innan dagen för OA-publicering och också efteråt. Likaså följdes nerladdningssiffrorna för pdf-filerna upp i

denna longitudinella studie, vars mål var att försöka ta reda på risker och möjligheter med en s.k. "delayed OA-modell" för böcker. Open Access för böcker har inte undersökts alls i samma utsträckning som tidskriftsartiklar och man får akta sig för att tänka att OA i dessa två fall fungerar likadant. Artiklar är enkla att skriva ut och många läser dem på skärm. I det fallet är en artikel det intressanta objektet istället för t.ex. ett nummer av en tidskrift som till sin omfattning snarare liknar en bok. Böcker däremot fungerar annorlunda: de är inte så enkla att skriva ut och humanioras invånare verkar inte vara mogna för att läsa böcker på skärm (vilket bl.a. forskningsprojektet [OAPEN-UK:s stora forskarenkät](#) antyder). Böcker recenseras. Det gör sällan artiklar eller tidskriftsnummer. Affärsmässigt finns också stora skillnader: böcker säljs en och en medan tidskrifter säljs som prenumerationer, enskilda eller i paket, s.k. *bundles*.

Nedladdningen av pdf-böckerna i vårt experiment var krångligare än vanligt: man måste registrera sig som kund hos Ontos för att få en e-post med länk till den pdf-bok man ville ha. Däremot fanns en GoogleBooks-version tillgänglig redan i förlagets elektroniska bokkatalog. Det här brukar man ibland kalla för en "Freemium-modell", när en lite sämre version är öppet tillgänglig och en finare version, dvs. Premiumversionen, i pdf-form antingen kostar eller är lite svårare att få fatt i (se till exempel [Open Book Publishers](#) för en mera avancerad Freemium-modell och [OAPEN Library](#) för en aggregator för böcker inom humaniora). Att nedladdningen var lite krånglig var kanske bra för oss forskningsmässigt – även om idealet ju är att de elektroniska böckerna ska vara lätta att hitta så att tillgången blir så stor som möjligt – eftersom tröskeln gör att vi vet att det inte var slappa klickanden utan att en potentiell läsare ligger bakom en nedladdning. En såld bok behöver ju inte heller betyda att någon läst den, kan man ju konstatera. För en författare till en bok kan ökad uppmärksamhet vara till nytta och målet borde vara att få så stor spridning som möjligt för ett alster som man jobbat länge med.

Många förlag är oroad över Open Access, och tänker att nedladdningsbara böcker kommer att hindra försäljningen, som ofta är den enda intäktskällan för akademiska bokförlag. Vår undersökning tyder på att det inte alls är fallet, utan

att Open Access antingen har en neutral inverkan på försäljningen, eller, i några enstaka fall, en mycket positiv inverkan på försäljningen.

Därtill ökade böckernas synlighet väsentligt. I och med pdf-nedladdningarna kom för hela boklasset i medeltal 40 % flera potentiella läsare till utöver de sålda exemplaren. Vi kan konstatera att sådana här modeller i högsta grad är i författarnas intresse.

Försäljningsuppföljningen gav vid handen att 90 % av cirkulationen skedde inom de första 12 månaderna, dvs. även forskningsböcker "kallnar" snabbt på marknaden. Där ger Open Access ju en förlängd tillgång efter att böcker ev. är slut från förlaget.

Delayed OA för böcker är en ofta förbisedd idé som kan gagna både förlag och forskare. För tidskrifter gäller det samma, att delayed OA är en rätt okänd men lovande möjlighet åtminstone i vissa fall. Det verkar vårt andra experiment med den nystartade tidskriften [Nordic Wittgenstein Review](#) antyda. Delayed OA för tidskrifter är inte ovanligt, visar [Mikael Laakso och Bo-Christer Björk](#) vid Svenska Handelshögskolan i Helsingfors, men de modellerna har fått stå i skuggan av modeller med OA-avgifter, som också många forskningsfinansiärer och universitet börjat räkna med efterhand.

Samtidigt erbjöd Ontos en annan OA-möjlighet till de författare som fick böcker antagna för publicering under 2011-12, nämligen en hybridmodell som vi kallar för Ontos Open. Enligt den modellen skulle författarnas bok publiceras OA samtidigt som den kom ut i tryck för en avgift på 1500€, eller 12 månader efter tryck för 750€. Förlagsdirektören hade räknat ut summan utgående från sin förväntade vinst och risken att gå miste om en del av den. Av lite över 100 författare som erbjöds detta hakade endast 9 på, och de ville alla att boken skulle bli Open Access meddetsamma. Taget i betraktande att OA knappast hämmar försäljningen av de tryckta böckerna vore den här modellen en riktig guldgruva för förlaget. Och allt oftare har författare institutionella medel till sitt förfogande för Open Access – och då ofta i storleksklassen 2-3000€ för en artikel, vilket de stora prestigefyllda förlagen ofta kräver.

Dock visar vår enkät till Ontos' alla publicerade författare eller antologiredaktörer under 2011-12 att många författare varit lite yrvakna när det gäller Open Access. De har erbjudits att publicera Open Access, de tycker egentligen att det vore en bra idé, men de har inte riktigt noterat möjligheten när den fanns där. I vissa fall anger de att de aldrig skulle kunna tänka sig att låta sin institution stå för en sådan kostnad – inom humaniora generellt är motståndet mot avgifter på författarsidan rätt stort, även om många förlag sedan tidigare kräver tryckbidrag som man om man är kritisk till förlagens roller kan räkna som en annan summa pengar som strömmar från universitetet eller en fond in i förlagets kassa.

Motståndet är enligt Agoras forskningsinsats inom Open Access befogat: Open Access ska minska universitetens problem att betala i överkant för publikationer, vars innehållsproduktion de redan stått för, och eftersom förlagets risk för förlust på bokförsäljningen pga. OA verkar obetydlig borde ingen avgift behövas för att täcka den. Istället kan en Open Access-version vara till fördel för både författare och förlag.

När förlagen inser detta kommer vi kanske att kunna ta flera mindre steg mot Open Access, i en framtid där förlagen fortsättningsvis har en roll att spela.

Nya publiceringsformer

Kenneth Nyberg och Jessica Parland-von Essen

Den diskussion om open access-publicering som vi behandlat i de föregående avsnitten berör på ett direkt sätt de viktigaste och mest traditionella publiceringsformerna för humanister: monografier, tidskriftsartiklar och antologier. Under de senaste åren har dessa frågor gjort sig mycket konkret påminna genom de ökande kraven på OA-publicering från forskningsfinansiärerna. Medvetenheten är därför i dag rätt hög hos de flesta historiker och deras humanistkolleger om problematiken med OA, pappers- kontra nätpublicering etc. En ännu så länge mindre synlig fråga bland forskarna, åtminstone ur ett nordiskt perspektiv, är den om de helt nya sätt att publicera sig som den digitala utvecklingen och särskilt webben har fört med sig.

Gamla modeller – och nya?

Vad som då egentligen ska räknas som "nytt" och vilken betydelse det har kan givetvis diskuteras, men det "gamla" – den rådande normen eller modellen för publicering av humanistisk forskning – är desto tydligare och kan definieras ganska enkelt. Den består av (huvudsakligen) textbaserade publikationer med en linjär struktur och en viss omfattning – antingen ca 15–30 sidor (artikel/kapitel) eller minst ca 200 sidor (bok). Det som händer i den digitala världen är att hela denna modell utmanas på en lång rad punkter och av olika skäl som har med både teknik och andra delar av samhällsutvecklingen att göra. Vilka av dessa förändringsprocesser som kommer att få störst genomslag och mest långtgående effekter på forskningen är i dagsläget omöjligt att säga, men i det följande tar vi upp några av de aspekter som tycks mest relevanta i sammanhanget.¹¹⁴

¹¹⁴ För två intressanta perspektiv på hur vetenskaplig publicering kan komma att förändras framöver – lämpligt nog förmedlade i form av videoklipp på YouTube – se "Anthony Grafton: The Future of History Books", *YouTube* 2014-01-03, <http://www.youtube.com/watch?v=FCGm2mGz9p0> (2014-01-06), och John Wilbanks, "The Fragmentation and Re-Integration of Scholarly Communication", *YouTube* 2011-05-19, http://www.youtube.com/watch?v=UqYiqzD_L0 (2014-01-06).

Att det inte alltid är givet vari det nya består kan, till att börja med, illustreras av den förändring som är mest uppenbar och för många kanske fortfarande ses som den största: övergången från tryckta till digitalt publicerade böcker och tidskrifter. Självfallet kan sådana i en bemärkelse ses som nya publiceringsformer vilka påverkar både distribution och konsumtion av forskningens resultat, men i sig handlar det främst om själva mediet: från papper till digitalt. Formen i egentlig mening har ju inte förändrats, bara med vilken teknisk lösning vi tar till oss texter – och mest är det än så länge just texter, utan nämnvärda grafiska eller multimediala inslag. Med nätbaserade tidskrifter eller, som det ännu ofta handlar om inom humaniora, nätversionen av existerande tidskrifter, kan publiceringen gå lite snabbare, men processen och formatet i övrigt skiljer sig inte från de tidigare. För många är detta också en självklarhet, eftersom de menar att vetenskapligheten ligger i just processen och de strukturer den är en del av: kollegial förhandsgranskning, redaktionell bearbetning, publicering i ett visst sammanhang som signalerar kvalitet osv.

Potentialen till förändring i både liten och stor skala ligger alltså inte främst i den digitala formen som sådan, utan i det faktum att den undanröjer många av den traditionella modellens inneboende begränsningar. En grundläggande sådan handlar om de normer för akademiska texters längd som nämndes ovan, dvs. att humanistisk forskning antingen publiceras som artiklar/kapitel eller böcker av viss omfattning. Dessa normer är inte järnhårda och de varierar mellan discipliner, i takt med att publiceringsprocessen strömlinjeformats under det senaste århundradet har också konventionerna för texters längd stelnat alltmer. I grunden är de dock kopplade till praktisk-ekonomiska hänsyn som inte längre gäller i digitala sammanhang, och en fråga som ibland väckts bland exempelvis amerikanska digitalhistoriker är därför hur dessa konventioner i sig blir begränsande. De argumenterar för betydelsen att också kunna skriva texter av medellängd, dvs. långa artiklar eller korta monografier om, säg, 50–80 sidor.

Detta kan tyckas vara trivialt, men faktum är att textens längd kan ha stor betydelse för vad vi kan eller inte kan göra i en vetenskaplig publikation.¹¹⁵

Men det finns också många andra, potentiellt sett mer omvälvande följder av den digitala tekniken och webbens utbredning. För att återknyta till kapitlets första avsnitt är det nämligen inte bara befintliga delar av publiceringslandskapet som förändrats, utan helt nya områden har tillkommit som mer fundamentalt kan komma att påverka vad forskning är och uppfattas vara. Dessa nya former, som dels handlar om det som ibland kallas "det vidgade textbegreppet" och dels om nya publiceringssammanhang i form av exempelvis sociala medier, är föremål för mycken diskussion inom digital humaniora-kretsar i vid mening, men bland forskare i övrigt är det (åtminstone i Sverige och Finland) ganska tyst. Återigen handlar det om en för de flesta humanister helt ny terräng, och därför visar de mindre intresse för den än för de OA-frågor som mer direkt påverkar deras redan kända publiceringsvägar.

Ett vidgat textbegrepp

Det vidgade textbegreppet innebär i detta sammanhang att humanistisk forskning inte bara, eller ens främst, behöver ta formen av en huvudsakligen textbaserad publikation med en linjär struktur. Ett första, relativt enkelt steg i den riktningen är att texter på ett annat sätt än tidigare kan varvas med olika former av audiovisuella inslag, inte bara bilder utan också ljud, video, animationer osv. I förlängningen kan man också tänka sig att redovisning av forskning kan ta sig helt nya uttryck där texten har underordnad betydelse och de multimediala inslagen är huvudsaken. Ett forskningsprojekt skulle alltså kunna bestå i att producera en interaktiv webbplats, att utarbeta en databas med dess gränssnitt eller att skriva ett program som utför en viss uppgift.¹¹⁶ (Ett

¹¹⁵ Här kan det kanske påpekas att om vi går ett halvsekel tillbaka varierade artikellängden i t.ex. *Historisk tidskrift* och *Lychnos* betydligt mer än nu; särskilt i den sistnämnda var det inte ovanligt med bidrag om just ca 50–80 sidor.

¹¹⁶ I ett avsnitt av poddsändningen *Digital Campus* spår [Sharon Leon](#) att 2014 blir året då de första "betydande digitala avhandlingsprojekten" (i USA) kommer att slutföras. Se "Digital Campus on 2013 and the Uncertain Future of Amazon's Drones", *Digital Campus* 2013-12-18,

antal exempel på den typen av projekt ges i avsnitt och fördjupningar om databaser, digitala textarkiv, visualiseringar etc. i [kapitel 4](#) och [kapitel 5](#).)

Argumenten för den typen av projekt är främst att synen på vad vetenskaplig verksamhet är har vuxit fram i ett tekniskt sammanhang som nu i grunden är radikalt förändrat, och därmed bör även idén om vad forskning är förändras. Andra skulle hävda att det ligger en viss typ av kritiskt analyserande hållning i texten som medium och att just denna hållning är ett avgörande kriterium på vetenskaplighet, varför de är skeptiska till att betrakta multimediaproduktioner eller datorprogram som forskning i konventionell (enligt deras synsätt "egentlig") mening.¹¹⁷

Andra utmaningar mot den traditionella, linjära texten handlar om både mediet och strukturen. Historieskrivningen och därmed också -vetenskapen är sprungna ur litteraturen som genre, textformatet med dess narrativa struktur är historiens inneboende form och format. Därför kan de också finnas skäl för oss att närmast se på vad som under de senaste decennierna skett inom skönlitteraturen. Det digitala formatet har även där öppnat för interaktivitet och utnyttjandet av rörlig bild och ljud som en del av en "berättelse". Antingen ges möjligheten att följa en viss storyline i egen takt, så som i till exempel [Inanimate Alice](#) eller också erbjuds läsaren att välja och påverka händelseförloppet, varvid man alltmer närmar sig spelvärlden. Ett koncept är också de så kallade "wovels" eller webbromanerna, där författaren skriver sin roman enligt läsarnas önskemål under processens gång (se t.ex. <http://www.underlandpress.com/whatwovel.cfm>).

Den väsentliga poängen här är att även texten kan sönderdelas och upplösas, så att den kan läsas på ett annat sätt än enligt en enda, i förväg bestämd storyline. I detta format faller forskarens argumentationslinje sönder och

<http://digitalcampus.tv/2013/12/18/episode-102-digital-campus-on-2013-and-the-uncertain-future-of-amazons-drones/> (hämtad 2014-01-06), tidsmarkering 48.10.

¹¹⁷ Jfr den delvis annorlunda men delvis parallella fråga som Jessica Parland-von Essen tar upp i ett blogginlägg om en disputation, där respondenten valde att inte redovisa/hänvisa till vissa av sina underliggande kvantitativa studier i textkorpusar för att de enkelt kunde upprepas av vem som helst; man kunde därför anse att de inte var forskning i egentlig mening. Jessica Parland-von Essen, "Om idéer och kvantitativa metoder", *Essetter* 2013-12-14, <http://essetter.blogspot.se/2013/12/om-ideer-och-kvantitativa-metoder.html> (2014-01-06).

materialet utelämnas åt läsarens val och tolkningar. I Sirkka Havus och Sanna Järvinens verk [Kiehtova kirja](#) presenteras bokhistorien som en matris som kan läsas antingen tematiskt eller kronologiskt. Då erbjuder man ett antal olika berättelser, men ger samtidigt läsaren möjlighet att helt bryta berättelserna enligt eget önskemål när som helst eller växla mellan de olika berättelserna. Författaren förlorar på så sätt kontrollen (ännu mer) över läsarens tolkning; tankegången och argumentationen man tidigare kunnat leda läsaren igenom, ersätts av en modulär och evinnerligt varierande struktur.¹¹⁸ Förändringen är förutom strukturell även temporal; den digitala tiden går inte heller linjärt, utan kan gå i olika slag av öglor, där man förflyttas eller förflyttar sig fram och tillbaka i en virtuell tid som kan "snabbspolas" eller bromsas upp. Hur historikern kan få fram sina poänger är något som kräver noggrann planering.

Men kan historien berättas på detta sätt, som en modell eller rekonstruktion utan en klar berättelse, och på vilket sätt är det forskning? Historien och historievetenskapens produkter kan i den digitala världen tolkas och konstrueras modulärt i motsats till det traditionella lineära formatet. Detta är förutom ett potentiellt hot om förlorat tolkningsföreträde också en möjlighet att (re)presentera historien i ett format som mer liknar "verkligheten". Modulariteten och de andra aspekterna av de nya mediernas språk och strukturer ger också möjligheter att presentera alternativa berättelser som likvärdiga eller att pendla mellan mikro- och makroperspektiv, eller källor och olika tolkningar eller relationer. I själva verket ter sig sådana representationer mer verklighetstroga och utmanande för forskaren. De kräver grundliga analyser av semantiska och ontologiska relationer, men så länge de inte presenteras i en narrativ redogörelse platsar de i dagens läge ännu inte ensamma som akademiskt meriterande forskning. En sak som ytterligare gör det hela utmanande är att dylika projekt på grund av sin omfattning och komplexitet kräver mer integrerade mellan- och inomdisciplinära samarbeten än vad historiker är vana med (och utbildade för).

¹¹⁸ Jfr Lev Manovich, *The Language of New Media* (MIT Press, 2001) s. 30–45.

Fragmentering och länkning av information

En aspekt som Lev Manovich egentligen inte behandlar i sina annars mycket grundläggande arbeten om de nya medierna är länkning av data, vilket förstås delvis beror på hans ansats. Men för att förstå effekterna av digitaliseringen för publicerandet bör vi också beakta öppen länkad data. Vi måste se de möjligheter och utmaningar som följer med det faktum att vi publicerar material i ett digitalt nätverk som i princip ompänner hela världen och länkar samman information på olika sätt. Modulariteten begränsar sig ingalunda till en enskild dator, utan omfattar hela internet. Modulariteten innebär också en aldrig tidigare skådad mängd av helt fragmenterad information, det vill säga att sammanhang och kopplingar, kända och okända, helt tappats bort.

Inom naturvetenskaperna har man redan börjat operera med nanopublikationer. Det innebär att varje enskilt antagande eller data betraktas som en skild publikation som man kan hänvisa till. I praktiken betyder det att varje påstående om något sakförhållande består av ett subjekt och ett objekt och dessa två förenas av ett predikat som beskriver deras inbördes relation. I synnerhet det sista saknas alltför ofta i traditionella informationssystem. Dessutom behövs metadata om under vilka förhållanden påståendet gäller, när detta är fastslaget och av vem.¹¹⁹

För historikern ter sig dylikt atomiserande av kunskap olämpligt, eftersom vi ju oftast strävar just till motsatsen, till att se skogen för träden, hitta sammanhang, mönster, påvisa nya samband, de stora helheterna och strukturerna. Vi arbetar med kunskap och inte enskilda små datakorn. Samtidigt erbjuder öppen länkad data nya möjligheter och det handlar ofta bara om att strukturera den data som finns på ett rikare, smartare sätt, för att göra den mer användbar. Till exempel vore det önskvärt att bättre kunna länka alla källor, oberoende var de finns, till forskningen. Att publicera informationen om att ett visst dokument använts för att belägga ett visst påstående i en viss publikation

¹¹⁹ Se t.ex. Sally Chambers presentation "Nano-publications in the arts and humanities?" (2011), <http://www.slideshare.net/schambers3/nanopublications-in-the-arts-and-humanities>. Standardiseringsarbete pågår fortfarande; se närmare Concept Web Alliance http://nanopub.org/guidelines/working_draft/ (båda hämtade 2014-01-05).

kan integreras i publikationen som en enkel länk. Traditionellt har det gjorts som en språklig hänvisning men kan nu kompletteras eller ersättas av en maskinläsbar hyperlänk, dvs. en adress till källan. Men uppgiften om hänvisningen kunde också göras som en separat information som länkas till bägge, varvid informationen blir läsbar så att säga från båda hållen.

Att kunna få en sådan inblick i historievetenskapens strukturer och resonemang som öppen länkad data möjliggör, skulle kunna tillföra vårt vetande en hel del. Samtidigt har vi ännu en lång väg att vandra för att uppnå ens den första varianten: att med ett klick kunna förflytta oss från fulltext till fulltext i en hänvisning. Den bättre varianten skulle också ge möjligheten att genom att klicka på till exempel ett namn få fram information om på olika, definierade sätt relaterade publikationer eller institutioner. Vad som behövs är målmedveten styrning och skapandet av nationella och internationella infrastrukturer för ändamålet. Det internationella forskar-idet ORCID är ett steg i denna riktning.¹²⁰ Samtidigt finns ett visst förståeligt och fullt acceptabelt motstånd bland humanister: om vi erbjuder våra läsare direkt tillgång till oändliga mängder relaterade material – kan någon längre ta till sig vår komplexa forskning? Eller sönderfaller allt i ett ändlöst surfande mellan olika textfragment? Men också: har vi råd att räkna med att vårt vetande sprids *endast* genom läsning av våra texter från början till slut?

Sociala medier

En mycket viktig utvidgning av publiceringslandskapet rör visserligen också teknik och form, men handlar i första hand om de sammanhang där forskare kommunicerar sin forskning och om vem de kommunicerar med: Är det kolleger eller allmänhet och i vilken mån måste dessa skilja sig från varandra? De nya kanaler som idag används kan vara (och är fortfarande ofta) främst textbaserade, liksom traditionell publicering, eller mer multimediala, men deras

¹²⁰ Vetenskapsrådets omvärldsbevakning 31.10.2012, <http://www.vr.se/franvetenskapsradet/omvarldsbevakningavinternationellforskningspolitik/internationellforskningspolitik2012/registerforforskaridanserasorcid.5.5b1cf80213a8e1394f82c2.html> (hämtad 2014-01-05).

kännetecken i detta sammanhang är i första hand snabbheten i kommunikationen och öppenheten mot världen utanför vetenskapssamhället.

Denna form av publicering är något som många digitala humanister propagerar för, då de ser sociala medier av alla de slag (bloggar, Facebook, Twitter, Instagram, YouTube, poddsändningar etc.) som en möjlighet att mer eller mindre radikalt riva de murar forskarsamhället i sin långtgående specialisering tenderar omgiva sig med. De menar att bloggar och (övriga) sociala medier både kan vara ett kraftfullt verktyg i den inomvetenskapliga debatten och för individuella forskares nätverkande. Därtill kan sådana kanaler på ett mycket effektivt sätt bidra till kunskapsspridningen i det omgivande samhället, det som ibland har kallats "den tredje uppgiften" (i dag ingående i "samverkansuppgiften").¹²¹ Tyvärr meriterar denna verksamhet fortfarande litet akademiskt, men de flesta historiker är väl medvetna om sitt samhällsansvar och ser också sambandet mellan framtida finansiering för humaniora och en samhällstillvändhet gällande popularisering och marknadsföring av de egna resultaten.¹²²

Här bör en annan aspekt tilläggas, nämligen att många av de nya publiceringssammanhangen är öppna inte bara i den meningen att vem som helst kan "konsumera" eller ta del av dem, utan också på så sätt att användarna/läsarna kan bidra med nytt och påverka innehållet (som i wovellerna) vilket bör ses som en resurs. (Jämför också avsnitten om crowdsourcing i kapitel 6, <http://digihist.se/6-digitalbaserade-material-och-langsiktigt-bevarande/>). Bloggarna inbjuder till kommentarer, liksom Pinterest eller Facebook och Twitter inte bara erbjuder möjligheter till marknadsföring

¹²¹ Jfr också en intressant studie om betydelsen av "altmetrics" kontra "bibliometrics", alltså vilken betydelse det kan ha för ens forskning att man twittrar om den etc. Se Emily Darling, "Twitter and traditional bibliometrics are separate but complementary aspects of research impact", *Impact of Social Sciences* 2014-01-02, <http://blogs.lse.ac.uk/impactofsocialsciences/2014/01/02/twitter-citations-research-impact-darling/> (2014-01-05).

¹²² Se till exempel boken *Humaniora - till vilken nytta?*, red. Tomas Forser och Thomas Karlsson (Göteborg 2013). En presentation av boken finns på adressen <http://www.hum.gu.se/aktuellt/Nyheter/fulltext/till-vilken-nytta--en-bok-om-humanioras-mojligheter.cid1165277> (hämtad 2014-01-05).

och nätverkande utan också till vetenskaplig diskussion med både kolleger och allmänhet. Det vill säga från att den offentliga delen av den vetenskapliga kommunikationen varit i huvudsak enkelriktad (föreläsningar, tryckta texter) har den nu potential att förvandlas till en ömsesidig kommunikation i större skala, som därmed också har potential att engagera större kretsar. I takt med att historiekunskaperna i samhället generellt försvagats (skolundervisningen har skurits ner) faller därför ett allt större ansvar på de enskilda professionella historikerna att hålla "historien levande" och hålla historieförfalskningar och annat tendentiöst historiebruk på stången. Engagemang uppnås bl.a. just genom dialog varför detta är en aspekt man alltid borde hålla i minnet då man publicerar sig – hur kan jag ta emot feedback, mera information, reaktioner, korrigeringar, kommentarer osv? Visserligen har det ifrågasatts hur breda grupper akademiker i praktiken når till exempel genom sina bloggar, men rent principiellt är det ändå en viktig aspekt att både vi som historiker och den forskning vi bedriver är så tillgängliga och öppna för allmänheten som möjligt.¹²³

Relationen mellan forskare och lekmän

Därmed är vi tillbaka i ett tema som vi berört många gånger i denna bok, nämligen frågan om den förändrade relationen mellan forskare och lekmän som en av de huvudsakliga konsekvenserna för akademien av den digitala utvecklingen. Ett intressant exempel på den diskussionen kommer från Dan Cohen, som har propagerat för att journalistik och humanistisk forskning bör närma sig varandra i det nya publiceringslandskapet, vilket skulle ha den goda effekten att den vetenskap som produceras blir mer tillgänglig för allmänheten och inte bara vänder sig till andra forskare.¹²⁴

Alla delar kanske inte uppfattningen att gränsen mellan forskning och journalistik bör lösas upp, men i ett sammanhang där det blir allt viktigare för

¹²³ "Why do academics blog? It's not for public outreach, research shows", *Higher Education Network*, 2013-12-02, <http://www.theguardian.com/higher-education-network/blog/2013/dec/02/why-do-academics-blog-research> (hämtad 2014-01-05).

¹²⁴ Dan Cohen, "Digital Journalism and Digital Humanities", <http://www.dancohen.org/2012/02/08/digital-journalism-and-digital-humanities> (2013-01-19).

oss att kommunicera det vi gör kan det vara klokt att titta på och inspireras av dem som professionellt arbetar just med samhällskommunikation, dvs. journalister, och de metoder de använder. Kan deras arbetssätt, t.ex. ifråga om datajournalistik eller mediekonvergens, på något vis även användas på historiska material? Mer generellt är det uppenbart att både historiker och andra humanister får helt nya möjligheter att dela med sig av forskningsresultat och annan kunskap genom den nya tekniken. Dels är det ganska tydligt att de visualiseringar som behandlades i kapitel 6 (<http://digi.hist.se/6-digitalbaserade-material-och-langsiktigt-bevarande/>) inte bara har ett värde för forskarna själva, utan ofta uppfattas som mer tillgängliga för lekmän än akademiska texter. Dessutom faller successivt kostnaderna och arbetsinsatsen för att spela in ljud eller video, vilket gör att man kan nå ut till allmänheten med egenproducerade program. Sådana kan man sedan sprida via YouTube eller som poddsändningar (podcasts) vilka lyssnare och tittare kan ladda ner till sina mobiler, surfplattor och datorer.

Återigen ligger den typen av multimediaproduktioner kanske inte alltid så nära till hands för historiker och andra forskare, som genom lång träning ofta är mer bekväma med att kommunicera via text. Åtminstone under en övergångsperiod (och möjligen också på längre sikt) är det därför tänkbart att andra sociala medier, som bloggar och Twitter, kommer att spela en större roll som kanaler för populärvetenskaplig kunskapsspridning. Även här kan det finnas ett motstånd bland historiker som är vana att skriva längre texter och inte gärna gör de förenklingar som begränsat utrymme ofta kräver, men tröskeln är i alla fall lägre än för ljud- och videoinspelningar. Samtidigt blir också denna tröskel hela tiden lägre, och det finns många skäl som talar för det allt vanligare bruket att t ex banda och/eller streama föreläsningar eller seminarier när man ändå håller dem, både som intern dokumentation och för dem som inte kan närvara. Detsamma gäller presentationer som kan delas t ex på [Slideshare](https://www.slideshare.net/), [Academia.edu](https://www.academia.edu/), [LinkedIn](https://www.linkedin.com/) eller [Prezi](https://prezi.com/), eller [Twitter](https://twitter.com/)-rapportering som kan dokumenteras t.ex. på [Storify](https://www.storify.com/).

Framför allt i USA börjar det förekomma, även om det fortfarande är ett ganska marginellt fenomen bortom den "hårda kärnan" av digitala humanister, att forskare också diskuterar vetenskapliga frågor sinsemellan på sina bloggar. En stor fördel är att det är en form där man snabbt och prestigelöst kan ventilerat utkast till texter och preliminära resultat och få synpunkter innan de publiceras på traditionellt vis i en tryckt (eller digital) tidskrift med kollegial förhandsgranskning. Som tidigare påpekats är produktionstiderna för sådana tidskrifter ofta mycket långa, vilket gör att frågor kan hinna bli inaktuella innan en undersökning publiceras. Bloggosfären kan på sätt och vis, i bästa fall, fungera som ett öppet seminarium.

Kanske innebär det, som antydde i [föregående avsnitt](#), att vi kommer se en förskjutning där formell publicering i en tidskrift – vare sig den är tryckt eller digital – blir en slutpunkt och betraktas som arkivering av forskningsresultat, snarare än som utgångspunkt för en fortsatt, levande diskussion. Forskning kommer med andra ord i högre grad att presenteras och debatteras redan medan den pågår, inte först när den är genomförd. Det skulle på sätt och vis innebära en återgång till 1600- och 1700-talens situation, där mycket av de vetenskapliga samtalen om pågående arbete inte fördes i tidskrifter eller böcker utan i brevväxling mellan forskarna. Dessa brev sågs ofta som helt eller halvt offentliga och lästes upp vid olika vetenskapliga sammankomster för att göra innehållet mer allmänt bekant. Som fenomen betraktat är alltså forskarbloggar, liksom så mycket annat med koppling till digital historia, en ny företeelse med gamla rötter.¹²⁵

Nya undervisningsformer

Kenneth Nyberg

Undervisning på både lägre och högre nivåer är en av de viktigaste formerna av förmedling av historisk kunskap, och de flesta verksamma historiker ägnar en

¹²⁵ Kathleen Fitzpatrick, "Blogs as Serialized Scholarship", *Planned Obsolescence* 2012-07-12, <http://www.plannedobsolescence.net/blog/blogs-as-serialized-scholarship/> (2013-01-19).

större del av sin tid åt undervisning än åt forskning. Hur undervisningen påverkas av den digitala utvecklingen är därför en viktig fråga, men den är också så stor att den kan betraktas som ett helt eget fält som det inte är möjligt att gå in på närmare här. Några huvuddrag i diskussionen kan vi ändå skissera mot bakgrund av övriga resonemang om historieförmedling i tidigare avsnitt i detta kapitel.

När det gäller utbildning på både högre och lägre nivåer har den tekniska utvecklingen ännu inte fått några mer fundamentala konsekvenser på de flesta håll, trots att det under årtionden förutspåtts att datorerna, IT eller nätet en dag ska revolutionera undervisningen (och kanske forskningen). Gång på gång har dessa löften utfärdats men utan att infrias i nämnvärd grad. I skolan ersattes 1980-talets datorsalar av 90-talets "IT i skolan" och 2000-talets multimediesatsningar. Under 2010-talets första år har många förhoppningar knutits till en-till-en-projekt – satsningar på att ge alla elever varsin bärbar dator, de senaste åren allt oftare i form av en iPad eller annan surfplatta. Och visst har tekniken efter hand sipprat in i skolor och på universitet och kommit till allt större användning, men sällan som något annat än komplement till befintliga, traditionella arbetssätt och undervisningsformer.

De allra senaste åren har emellertid något hänt som innebär en ny och jättelik utmaning för hela utbildningsväsendet, från förskola till universitet. Förändringen kan sammanfattas med ett ord: förväntningar. Visserligen är studenter och andra ungdomar inte alltid så tekniskt kunniga som ofta antas; "den digitala klyftan" går alltså inte nödvändigtvis mellan yngre och äldre. Men ofta gör den det, och till skillnad från de elever och studenter som lärare mötte för trettio, tjugo eller ens tio år sedan har dagens och morgondagens studenter med sig en helt annan teknikanvändning från sitt övriga liv in i lektionssalen. Tidigare löften om teknikens möjligheter byggde på vad som rent tekniskt var möjligt att göra, inte på vad studenterna faktiskt förväntade sig av undervisningen. Idag har dock tekniken blivit så spridd och genomsyrar i så hög grad studenternas hela tillvaro, att de ser det som något självklart att deras lärare och skolor använder sig av liknande redskap.

Så är ofta inte fallet, eftersom undervisningsformerna till stor del ser ut som de gjort under decennier eller århundraden. Givetvis har olika former av digitalt baserad teknik fått ett genomslag, men här liksom i andra sammanhang är det viktigt att skilja mellan förändringar som "bara" är rent tekniska, och sådana som får mer långtgående konsekvenser. Den digitala utvecklingens följd för akademisk (och annan undervisning) handlar alltså inte främst om att lärare använder PowerPoint för sina föreläsningar eller att varje skolelev har sin egen dator, utan om hur sådana och andra verktyg används för att utveckla lärande. I det perspektivet ställs traditionella modeller för undervisning inför helt nya utmaningar med elever och studenter som är ständigt uppkopplade och där undervisning konkurrerar med sociala medier om uppmärksamheten.

Att använda bildspel på traditionellt sätt, eller att organisera "massiva nätkurser" bestående av videoinspelade föreläsningar utan någon lärar-studentkontakt, är mot den bakgrunden allt annat än pedagogiskt innovativt.¹²⁶ Snarare är det uttryck för att cementera traditionella undervisningsmodeller som uppkommit i en tid med helt andra förutsättningar och som nu måste omprövas. Ett exempel på vad sådana omprövningar kan leda till är den mer genuint nytänkande idén om "flipped education" eller "flipped classroom", där enkel inhämtning av gemensamt stoff inte görs i klassrummet i form av föreläsningar utan av studenten själv medan den dyrbara lärarledda tiden ägnas åt gemensam bearbetning och diskussion av detta stoff, som kan bestå av både läroböcker, videoföreläsningar och annat nätbaserat material. På så vis utnyttjas å ena sidan de närmast gränslösa resurser av hög kvalitet som idag finns relativt enkelt tillgängliga på nätet, och å andra sidan klassrumstiden så att den ger mesta möjliga effekt för varje enskild student i form av interaktion med både läraren och andra studenter på plats.

Den typen av nya modeller aktualiserar också en annan stor fråga, nämligen hur det framväxande digitala landskapet kommer att påverka den

¹²⁶ Sedan hösten 2011 har fenomenet massiva nätkurser (Massively Open Online Courses, MOOCs) ägnats stor uppmärksamhet i media och förutspått innebära en revolution av den högre utbildningen, men de har också kritiserats och ifrågasatts. För en serie blogginlägg med en skeptisk hållning och många länkar till material på båda sidor om debatten se <http://kennethnyberg.org/?s=mooc>.

konventionella lärobokens funktion och betydelse. Allt fler webbplatser med resurser, lektionsupplägg och länksamlingar för lärare etableras, till exempel amerikanska riksarkivets DocsTeach.org och teachinghistory.org eller svenska [Stockholmskällan](http://Stockholmskällan.se). Alltmer primärmaterial blir också tillgängligt direkt på nätet, vilket kan bli en utmaning för läroboken om många lärare väljer att bygga upp sin undervisning på fallstudier eller studentcentrerat lärande istället för på den fasta struktur handböckerna erbjuder. Det finns de (t.ex. Mills Kelly, <http://edwired.org>) som starkt förespråkar och praktiserar den typen av arbetssätt.

Här är det visserligen uppenbart att förändringar är på gång; e-böcker, webbresurser och iPadprogram är på frammarsch och utmanar pappersboken. Men ännu har utvecklingen inte gått så långt som man skulle kunna förvänta sig, åtminstone inte på universitetsnivå; de allra flesta lärare och studenter använder sig fortfarande av traditionella läroböcker, oftast dessutom i pappersform. Möjligen kan det bero på att de fungerar som en trygg referenspunkt för de studentgrupper som idag ofta har ganska svaga förkunskaper – den erbjuder orientering och struktur i en ibland förvirrande massa av information och valmöjligheter. Även för hårt pressade lärare kan läroböcker vara ett stöd, liksom alla de resurser och fördjupningsmaterial etc. som förlagen ofta har utvecklat kring dem; det blir ett inte oviktigt faktum när man talar om att ersätta läroböckerna med något annat.

Användningen av sociala medier i undervisningen är ytterligare en alltmer brännande fråga i undervisningssammanhang. Här finns flera aspekter, dels hur lärarens roll och auktoritet utmanas genom att studenter på nätet kan kontrollera påståenden och resonemang under pågående föreläsning, dels hur bloggar och sociala medier kan fungera som mer eller mindre viktiga verktyg i studenternas lärprocess. Ofta kan kommersiella tjänster som Facebook, Google eller Twitter erbjuda stor flexibilitet och tillgänglighet, men samtidigt uppstår då frågor om gränsdragningen mellan privat och offentligt, hur företagen bakom kan använda studenternas personliga data och så vidare. Som Jeffrey McClurken

har påpekat är det därför viktigt att man vid undervisning med sociala medier har en tydlig policy för deras användning.¹²⁷

Den fundamentala fråga vi givetvis måste utgå från i alla dessa diskussioner, liksom annars, är vad vi egentligen vill att studenterna ska lära sig. Detta är kanske också något som behöver omprövas och omformuleras mot bakgrund av den digitala teknikens, webbens och de sociala mediernas utveckling. Är till exempel – enkelt uttryckt – den tydliga kunskapsstruktur som läroböcker mer än något annat bidrar med ett lärandemål i sig, eller bara ett hjälpmedel för att uppnå andra mål? Måste vi på mer fundamentala sätt omdefiniera vilka kunskaper (eller typer av kunskap) studenter kommer att behöva eller är det i första hand formerna för lärande som ska förändras? Först när vi har något slags svar på sådana frågor kan vi på allvar börja diskutera vilka metoder och verktyg som är ändamålsenliga. Med andra ord är det, här liksom i vår forskning, syftet som leder till frågan och frågan som leder till metoderna. Det är en grundförutsättning för både studenters och lärares kunskapsbildning som inte ens den digitala revolutionen har ändrat på.

¹²⁷ Presentation vid AHA:s årsmöte 2012; länkar till resurser och verktyg som då togs upp finns hos Jeffrey McClurken, "Teaching with Social Media", <http://mcclurken.org/presentations/aha-2012/> (hämtad 2014-01-15).

Kapitel 8. Historikerrollens förändringar

Kenneth Nyberg

Vad har egentligen hänt?

Det har nu blivit dags att summera vad allt det ovan sagda innebär för den historiska kunskapsbildningens betydelse och funktion i en digital värld. Vi kommer att göra det genom att diskutera kontinuitet och förändring ur historikernas perspektiv. Hur påverkas historikerrollen av digital teknik och sociala medier samt dessas genomslag i samhället i stort?

För att besvara den frågan kan vi först ställa oss en annan: Vad menar vi egentligen med "historiker" och hur länge har sådana funnits? Svaret kan å ena sidan sägas vara att de funnits i alla tider, om vi bara menar någon som intresserar sig för det förflutna och förmedlar sina kunskaper och tolkningar av det till det omgivande samhället. Å andra sidan genomgår också historikerrollen ständigt nya förändringar, vilket gör att dagens historiker på många sätt skiljer sig mycket från dem som verkade för bara ett par decennier sedan. Man skulle dock kunna hävda att grundmodellen för det som historiker uppfattas vara idag tog form i samband med 1800-talets professionaliseringsprocess. Det är en modell baserad på just historikers ställning som professionella experter, med en vetenskaplig legitimitet grundad i den källkritiska metoden och specialiserad kunskap om befintliga källmaterial och dessas relation till relevanta frågeställningar. På en rad punkter är det den modellen som nu utmanas till följd av de förändringar den digitala utvecklingen för med sig.

Vari består då dessa förändringar? Hur kan man på några få rader sammanfatta konsekvenserna av den digitala revolutionen? Ja, vad det i grunden handlar om är att transaktionskostnaderna för informationsspridning och kommunikation på kort tid fallit dramatiskt, särskilt i industrialiserade länder. Följden har blivit radikalt utvidgade möjligheter till mänsklig interaktion och åtkomst av information oavsett tid och rum, dvs. tillgänglighet, vilket också fört

med sig kraftigt ökade förväntningar på öppenhet från myndigheter, företag och forskare. Även själva idén om en skarp gräns mellan privat och offentligt har börjat undermineras, något som särskilt de sociala mediernas framväxt de senaste åren har bidragit till där offentlig kommunikation har intimiserats enligt mönster som tidigare bara gällde mer personliga relationer.

En annan följd av den lätthet med vilken information sprids och reproduceras är att verksamheter vars existensberättigande i någon mån bygger på monopol på "data" och information utmanas. I takt med att både text, musik och bild har blivit möjliga att representera och lagra digitalt till låg kostnad har en lång rad mediabranscher och andra industrier omvandlats i snabb takt: bokhandlare, skivbolag, kameratillverkare, dagstidningar och filmbranschen erbjuder bara några exempel av många. Den fråga som tidigare ansågs handla om piratkopiering har särskilt de senaste åren alltmer kommit att handla om själva idén om individuell upphovsrätt. Det är en tanke som av många anses central för att en kunskapsdriven ekonomi över huvud taget ska fungera, samtidigt som andra påpekar att det är en relativt ny idé med några få hundra år på nacken och därför inte en tvingande naturlag. Här stöter vi återigen på öppenheten som ideal, ett ideal som har helt andra förutsättningar att få genomslag i en värld där kostnaderna för att realisera det är så mycket lägre än för bara några decennier sedan.

Över huvud taget förändras hela offentligheten, den arena där samhället blir till just ett samhälle, av den digitala utvecklingen. (Fast om det är något genuint nytt som då uppstår eller om det finns historiska paralleller är en annan fråga, och den återkommer vi till nedan.) Utöver öppenheten, tillgängligheten, den oändliga floden av åtkomlig information, skulle många lägga till det ökade tempot i mänsklig kommunikation. Både via e-post, chattprogram, videosamtal och sms kan vi idag hålla kontakt i realtid med människor var som helst i världen; det viktiga är inte avståndet i sig utan om en plats är uppkopplad på det globala nätet. Det har funnits en tendens att kommunikationen blivit allt snabbare, alltmer uppdelad i allt mindre "paket" av data skickade med allt tätare intervaller – alldeles som de paket som all internettrafik rent tekniskt faktiskt

består av. Aldrig har vi kommunicerat så ofta och så mycket med så många människor över så stora avstånd som nu, och det är en utveckling som bara tycks accelerera.

Sammantaget är det på många sätt en värld av möjligheter som har öppnat sig som människor för bara en generation sedan eller två inte kunnat föreställa sig. Men samtidigt har den också mörkare sidor, effekter av mer tveksamt värde. Till att börja med kan själva ymnigheten i utbudet vara överväldigande, och den oändliga valfriheten kan leda till handlingsförlamning. Känslan av att det finns "för mycket att veta" är visserligen långtifrån ny i historien, något som gärna också påpekas i diskussionen, men sällan har den haft så mycket fog för sig som nu.¹²⁸ Det högt uppdrivna tempot i kommunikationen och i vår mediekonsumtion kan också göra att närsyntheten ökar, att vi blir fångar i vår egen samtid och i ett ständigt uppdaterat "nu" där intresse och värde förintar lika snabbt som det uppstår. Perspektivet krymper och känslan av främlingskap ökar inför det avlägset förflutna, eller över huvud taget för livsvärldar som inte präglas av vår ständiga uppkoppling. En sådan utveckling kan också ses som en förlust ur en humanists perspektiv.

Detsamma kan sägas om den globalisering och kulturella harmonisering som den ökade interaktionen mellan människor också med nödvändighet för med sig. Den nya, globala, digitala offentligheten i allmänhet, och digital humaniora i synnerhet, tenderar att vara anglosaxiskt dominerad och genomsyras i hög grad av den (sub)kultur som förknippas med Silicon Valley, hackers och "nördar". Visserligen är det inte minst därifrån de starka idealen om öppenhet och fri spridning av kunskap kommer, men trots allt är det en specifik miljö med vissa normer som får genomslag på andras bekostnad med en slags digital "monokultur" som följd – och det kan inte (bara) vara av godo.

En helt annan risk som de allra senaste åren hamnat i fokus för en bitvis het debatt är den om övervakningssamhället och den orwellianska framtid vi möjligen går till mötes. Den ökade öppenheten och tillgången till allt större

¹²⁸ Ann M. Blair, *Too Much to Know: Managing Scholarly Information before the Modern Age* (New Haven och London: Yale University Press 2010).

mängder av alltmer personliga data leder till att de med tillräckliga resurser kan kartlägga enskilda människors liv, intressen, åsikter, nätverk, konsumtionsvanor och så vidare med en aldrig tidigare skådad effektivitet. Det är möjligheter som utnyttjas av såväl regeringar – i både demokratier och diktaturer – som privata företag, inte minst de verkliga jättarna inom den nya digitala världen: Google, Facebook, Apple, Microsoft och Amazon. Även när det inte finns anledning att ifrågasätta sådana aktörers egna motiv, är det potentiellt riskabelt att de samlar på sig sådana enorma mängder personliga data eftersom de oavsiktligt kan läcka ut genom bristande hantering eller dataintrång, något vi sett flera exempel på under senare tid.

Forskarens roll och expertens auktoritet

Det är ofrånkomligt att stora samhällsförändringar som de vi just skisserat får konsekvenser för forskarrollen, både historikers och andra. Men vilka? Det är mycket svårt att säga, eftersom det mitt i en historisk omvälvning är vanskligt att avgöra vad som är tillfälliga förskjutningar och vad som kommer bli mer bestående. Just för ögonblicket är det svårt att värja sig för intrycket att själva förändringstakten har ökat med en större kortsiktighet som följd, där vi lever i ett alltmer intensivt "nu"-flöde som ger mindre tid till eftertanke och längre perspektiv. Å andra sidan kan det kanske ses som att historiker och andra humanistiska forskare bara har blivit mindre isolerade från det övriga samhället och nu lever "med sin tid" och följer med i den allmänna utvecklingen mer än tidigare.

Vad som står klart är dock att förhållandet mellan forskare och allmänhet måste förändras till följd av de processer vi talat om här. Även om det kan diskuteras vad som närmare bestämt karakteriserar forskare och deras funktion i samhället, handlar det i någon mening om en expertroll: en forskare är någon som besitter specialiserad kunskap som hen i någon grad har ett monopol på i relation till den omgivande allmänheten. Som vi sett är det just den typen av funktion som påverkas när information "befrias" och blir mer tillgänglig för alla och envar – men hur långtgående blir den påverkan? För kunskap och

förtrogenhet är väl inte detsamma som information eller enkla "data" som kan överföras med en enkel knapptryckning?¹²⁹ Kanske inte, men ändå påverkas forskarrollen när informationsspridningen omvandlas så i grunden och förväntningarna på tillgänglighet och öppenhet ökar så radikalt.

Komna så långt kan det vara värt att påminna om hur den moderna vetenskapen i sig en gång i tiden representerade just en radikal kunskapsdemokratisering och en utmaning mot rådande auktoriteter. Istället för att lita blint på sådana auktoriteter lanserades idén, på sin tid ofattbart subversiv, att skaffa kunskap om världen baserat på egenhändiga, empiriska iakttagelser och observationer där man så långt möjligt sökte undanröja potentiella felkällor och snedvridande förförståelser hos forskaren själv. Det var en modell för kunskapsbildning som visade sig vara oerhört kraftfull, men som också krävde stora resurser i form av tid och arbete för att ge mesta möjliga utdelning. Det gjorde att forskarna utvecklades till ett eget skrä, en grupp i samhället med det specifika uppdraget att utveckla ny kunskap om den värld som kan observeras med våra sinnen. Vem som helst kunde inte bli forskare utan det krävde tid, ansträngningar och pengar; och eftersom inte vem som helst kunde bli det fick forskare också en viss status och auktoritet.

Det som nu händer är, i någon mening, ytterligare en revolution i samma riktning: alltmer av de data som utgör den empiriska grunden för forskning kommer att ha digital form och därmed vara tillgängliga för allmänheten på ett sätt som tryckta och otryckta källor inte har varit. Den del av forskarnas särställning som byggde på (i praktiken) monopol på källorna håller alltså nästan helt på att försvinna, och det i snabb takt. I viss mån gäller detsamma också monopolet över metoderna och verktygen, och det på två sätt: å ena sidan kommer forskningen alltmer att kräva teknisk kompetens, kunskap som humanister ofta inte har, vilket därmed undergräver deras expertstatus, och å andra sidan kommer alla slags verktyg successivt att bli alltmer tillgängliga för

¹²⁹ Jfr Sven-Eric Liedman, *Ett oändligt äventyr. Om människans kunskaper* (Stockholm: Bonnier 2001).

dem som inte är forskare. Gapet minskar alltså från två håll, och den skarpa gränsen mellan Forskare (med stort F) och allmänhet suddas alltmer ut.

Detta kan uppfattas som ett hot mot forskningen som en egen verksamhet med ett visst samhälleligt värde vilket ger utövarna något slags ställning och auktoritet. Men det är också en utveckling som ger möjligheter till närmare samverkan med omvärlden och som, om det görs rätt, kan placera forskarna "mitt i byn" igen på ett sätt som inte varit fallet på mycket länge. I takt med specialisering och utbyggnad av forskning och högre utbildning har vetenskapens utövare blivit alltmer inlåsta i sina egna små fack, vilket riskerar att göra dem obegripliga för omgivningen – något som på sikt verkligen är ett hot då så mycket av verksamheten är skattefinansierad. Om ingen förstår värdet av det historiker och andra forskare gör, varför ska man då stödja deras verksamhet ekonomiskt? Med större öppenhet följer också möjligheten att begripliggöra just värdet av det vi gör, att avdramatisera forskarrollen och samtidigt öka förståelsen för varför den rollen även fortsättningsvis är viktig för samhället och aldrig helt kan ersättas av "crowdsourcing" och "citizen science".

Av många skäl är alltså samverkan mellan forskare och allmänhet av stor och växande betydelse för historiker och andra humanister. Därför kan det inte råda något tvivel om att idealet om historikern som ensam på sin kammare skriver lärda och magistrala böcker alltmer kommer att utmanas och försvagas till förmån för mer lagarbete – både med andra historiker, med personer som har olika former av teknisk specialistkompetens och med lekmän (bl.a. i form av just crowdsourcing).¹³⁰ Den typen av samverkan är inte bara nödvändig för att forskningen kommer kräva flera typer av kompetens och arbetsinsatser, utan den kommer också att göra forskarnas roll och värde tydligare och mer självklara för utomstående.

För att försvara den historiska kunskapen i en tid som blir alltmer nutidsorienterad och där kortsiktigheten ökar, blir det också viktigt för

¹³⁰ Detta är något som t.ex. både John Nerbonne och Jo Guldi framförde vid sina besök på Göteborgs universitet hösten 2012. Se Kenneth Nyberg, "En dag om digital humaniora", 2012-10-31, <http://kennethnyberg.org/2012/10/31/en-dag-om-digital-humaniora/> och "Jo Guldi om digital historia", 2012-12-13, <http://kennethnyberg.org/2012/12/13/jo-guldi-om-digital-historia/> (båda hämtade 2014-01-24).

historikerna att mer systematiskt samarbeta med arkiv, bibliotek, museer och andra minnesinstitutioner så att den kunskap som produceras också sprids och tillgodogörs i samhället. Bland annat i sådana samarbeten har historiskt inriktade forskare en nyckelroll i att "översätta" eller förmedla historisk kunskap till den nya kultur och digitala offentlighet som växer fram. Därför behöver de (vi) vara både aktiva och medvetna aktörer i frågor som gäller digitalisering, vilket material som digitaliseras, hur det görs och så vidare. I den digitala världen är det ju tyvärr något av en sanning att det som inte går att googla finns inte; oavsett vad vi tycker om det är det en realitet vi måste förhålla oss till.

Samtidigt finns det idag, i de historiska forskarnas vardag, ett starkt spänningsfält mellan vad vi kan kalla inre och yttre meritering. Å ena sidan har den offentliga sektorns resurstilldelning till forskare i allt högre grad kommit att styras av modeller där publicering i internationella peer review-tidskrifter premieras, vilket är en starkt pådrivande kraft för en långtgående specialisering som fjärmar forskarvärlden från det omgivande svenska eller finländska samhälle som står för det mesta av dess finansiering. Å andra sidan blir kraven allt högre på samverkan och att forskare ska publicera sig open access och synas utåt, att de ska delta i samhällsdebatten och vara tillgängliga i sociala medier. Ibland talas det i de sammanhangen om hur forskare liksom andra offentliga personer bör bygga och vårda sitt "varumärke". Allt talar för att den typen av spänningar mellan delvis motsägelsefulla krav bara kommer att bli tydligare under de närmaste åren.

Framtidens historiker

Återigen bör det påpekas att det ofta inte är digitaliseringen i sig, övergången från en form till en annan, som är den stora förändringen. Istället är det de nya möjligheter denna form öppnar upp och för med sig på längre sikt, något som vi såg särskilt tydligt illustrerat i diskussionen om [nya publiceringsformer](#) i [kapitel 7](#). Den till synes kaotiska och delvis oreglerade offentlighet vi nu rör oss i, där gamla strukturer brutits ned och auktoriteter utmanats, har exempelvis klara

paralleller till 1700-talet med dess framväxande, länge rätt vildvuxna dagspress och pamflettdebatt. På liknande sätt påminner forskarbloggar, som många anser kan bli en viktig framtida publicerings- och kommunikationskanal för forskare, i flera avseenden om hur tidiga forskningsrön rapporterades i form av brev som publicerades i de då nya vetenskapliga tidskrifterna.

Över huvud taget är det givetvis viktigt att, här liksom annars, ha med sig ett lite längre historiskt perspektiv. Å ena sidan är de nya formerna ofta inte fullt så nya som vi gärna vill tro, å andra sidan är varje tid och plats i någon mening ett unikt sammanhang med sina helt egna förutsättningar och förhållanden.

Vetenskaplig publicering liksom vetenskaplig verksamhet i stort har alltid varit ett system med många olika (rörliga) delar, där varje del fyller sin funktion för helheten: brevväxling, offentliga föreläsningar, konferenser, tidskrifter, monografier, antologier, populärvetenskapliga texter och program i etermedia, läromedel och många andra. Vissa är snabba, andra är långsamma. Vissa vänder sig "inåt" vetenskapen och andra "utåt". Vissa är prövande och utforskande, andra mer slutgiltigt redovisande.

Den digitala tekniken och webbans utveckling i sig förändrar inte detta i grunden, utan tillhandahåller bara ett antal nya former, medier och kanaler och gör (möjligen) att andra uppfattas som mindre ändamålsenliga än förr. Men samtidigt kommer man inte ifrån att den sammanlagda effekten av tekniken både på den vetenskapliga verksamheten som sådan och på forskningens relation till det omgivande samhället på några års sikt kan bli omvälvande. I någon mening är ju en grundläggande förutsättning för forskning och för värdering av den att det är en egen typ av verksamhet som på vissa tydliga sätt skiljer sig från det vilken lekman som helst kan göra; en forskare, oavsett specialområde, är definitionsmässigt en expert och det är just experternas hegemoni som utmanas av den kunskapsdemokratisering webbens utbredning (ur ett perspektiv) utgör. Men det kan och bör också ses som en möjlighet till en förutsättningslös omstart för historisk och annan forskning, en chans att ompröva och befästa grundvalarna för vår verksamhet i en ny situation och att återupprätta en öppnare och närmare relation mellan forskare och lekmän.

Därmed är vi tillbaka vid några av de frågor som ställdes i slutet av kapitel 1 om hur historiker ska förhålla sig till den digitala utvecklingen och dess konsekvenser. Som vi antydde där finns det en tendens att diskussioner av det slaget blir polariserade mellan de som starkt betonar teknikens möjligheter och de som mest ser riskerna. Vi som skrivit denna bok tror att den framkomliga vägen ligger någonstans däremellan, och att vi med både öppet sinne och kritisk blick behöver närma oss den digitala teknik som nu breder ut sig och får allt större genomslag i allt fler delar av samhället. Att den utvecklingen kommer att fortsätta är helt ovedersägligt, även om det är omöjligt att mer i detalj veta vad det betyder eller vilka följder det kommer att få på sikt. Trots den osäkerheten är det inte konstruktivt att ignorera det som händer, eller att tro att det inte kommer att få några konsekvenser för både oss själva och det arbete vi ägnar oss åt. Vi måste därför, som också framhållits flera gånger i tidigare kapitel, aktivt förhålla oss till utvecklingen för att inte överflyglas av den.

Den enskilt viktigaste utgångspunkten för en sådan reflektion kring var vi står och vart vi är på väg är påminnelsen om det grundläggande syftet med historisk och annan forskning: att utveckla och förmedla kunskap av värde för det samhälle vi är en del av. Med "värde" åsyftas då inte ett nyttotänkande av det kortsiktiga, snäva slag som ofta förfäktas av exempelvis politiker och på längre sikt ofta är kontraproduktivt. Snarare är det en markering av att forskningen i någon bemärkelse givetvis måste tillföra samhället någon form av mervärde – hur svårt det än må vara att enas om vari detta består – med tanke på de resurser den tilldelas. Utifrån vår egen kunskap om hur samhällen fungerar förstår vi också att vår egen verksamhet inte existerar i ett vakuum utan är en del av omvärlden. Hur mycket vi än drivs av vår egen nyfikenhet i det vardagliga vetenskapliga arbetet, och hur viktig en sådan individuell frihet än är för att forskningen som kollektiv kunskapsbildning ska fungera, är syftet med verksamheten som helhet ytterst att den ska komma samhället till godo.

Å ena sidan måste vi alltså vara beredda att ompröva vad vi gör och hur vi arbetar när det omgivande samhället och dess förväntningar på oss förändras. Å andra sidan måste vi utåt samtidigt försöka argumentera just för en bredare och

mer långsiktig uppfattning om vad "nyttig" och "värdefull" kunskap egentligen är. Detta är i själva verket en av våra viktigaste uppgifter som forskare: att slå vakt om det öppna, prövande och kritiska förhållningssätt som – åtminstone idealt sett – kännetecknar vetenskaplig verksamhet, men som det kan vara svårt att få gehör för i ett medialt klimat präglad av benägenheten att söka enkla svar på komplicerade frågor. Hur vanskligt det än må vara att förutspå framtiden tycks alltså så mycket vara säkert som att vi måste bli mer synliga i den offentlighet där människor rör sig – dvs. i allt högre grad på webben och i sociala medier – både för att förmedla värdet av det vi gör och påverka den allmänna debatten om forskningens funktion och betydelse. Dels behöver vi bidra med våra egna perspektiv på sådana frågor, dels måste vi själva få en större förståelse för omgivningens uppfattningar om och förväntningar på historisk forskning så att resultaten av vårt arbete i högre grad ska kunna komma samhället till godo.

Både av detta skäl och som ett led i själva forskningen kommer historiker mycket snart att ställas inför helt andra krav på digital kompetens än tidigare. Hur dessa krav kommer att se ut på sikt är svårt, nästintill omöjligt, att säga, men dels handlar det om en generell förståelse för digitala material, miljöer och verktyg (vilka lär bli allt vanligare oavsett forskningsinriktning), dels om behovet i vissa projekt att använda mer specialiserade och komplexa tekniska lösningar. Det förra är något som berör alla historiker, medan det senare beror mer på forskningens karaktär och dessutom kan mötas på olika sätt. Som framgick i [föregående avsnitt](#) är det många som tror att det framöver blir vanligare att historiker arbetar i tvärdisciplinära grupper där teknisk expertis är ett inslag. Även i sådana fall kommer dock de humanistiska forskarna att behöva mer omfattande kunskap om och förståelse för digital teknik än idag, eftersom de måste kunna kommunicera med programmerare och andra specialister på ett sådant sätt att de tekniska lösningar som byggs blir ändamålsenliga.

Vad de nya förutsättningarna och förväntningarna innebär mer konkret för oss som historiker kommer alltså att variera efter vilken typ av forskning vi arbetar med och hur den är organiserad. För vissa kan det handla om att bättre förstå de möjligheter och problem som är förknippade med datavisualiseringar,

för andra att lära sig TEI-kodning och topic modelling och för några kanske rentav att sätta sig in i programmering på olika nivåer. Oavsett inriktning behöver vi dock alla bli bättre på att förstå sociala medier och vilken betydelse de kan ha för både själva forskningsprocessen och för att kommunicera dess resultat. Och alla kommer vi att verka i ett helt annat medialt och samhälleligt landskap än det de flesta av dagens historiker har vuxit upp i och utbildats för, ett landskap där forskningen bedrivs i och utgör en del av en sammanhängande digital offentlighet präglad av en annan öppenhet för omvärlden än tidigare.

Det är denna öppenhet som utgör den verkliga utmaningen, men också oanade möjligheter, för den historiska forskningen idag och imorgon. Framtiden är, som alltid, redan här.