



Vetenskapsrådet
Database Infrastructure Committee
(DISC)

Databaser och digitalisering inom humaniora - existerande resurser och framtida behov

Juni 2007

Förord

Styrelsen för Database Infrastructure Committee (DISC) beslutade vid sammanträde 2006-05-10 att uppdra åt undertecknad att initiera arbetet med en kartläggning av förhållandena kring databaser för forskning inom humaniora. Vid följande möte 2006-09-13 preciserades uppdraget till att gälla

- En inventering av befintliga stora databaser inom humaniora.
- En kartläggning av framtida behov av stora databaser inom humanistiska ämnesområden.
- En rimlig avgränsning av datamaterial vid myndigheter och organisationer utanför universiteten som bör ligga inom DISC:s ansvarsområde.

Denna kartläggning av forskningsinfrastruktur inom hela det humanistiska fältet delades upp i två delar, en med inriktning på språkteknologi och en med inriktning på humaniora i övrigt. Den första delen har tidigare redovisats i rapporten *Svensk språkteknologi – existerande forskningsinfrastruktur och framtida behov* från februari 2007. Den andra delen, som redovisas i föreliggande rapport, beskriver läget vad gäller existerande resurser och framtida behov inom ett brett humanistiskt fält.

Många personer har bidragit till kartläggningen. Det gäller främst de direkt involverade – de ämnesrepresentanter som stått för det huvudsakliga underlaget för kartläggningen: Inge Bartning, Stockholm, Magnus Bergquist, Göteborg, Phil Buckland, Umeå, Lars-Erik Edlund, Umeå, Olle Ferm, Stockholm, Britt-Louise Gunnarsson, Uppsala, Lena Johannesson, Göteborg, Kristian Kristiansen, Göteborg, Merja Kytö, Uppsala, Mats Malm, Göteborg, Marie C. Nelson, Linköping, Lars Nilsson, Stockholm, Ester Pollack, Stockholm, Jørgen Straarup, Gävle. Men det gäller också många av deras kolleger som bidragit med uppgifter och därmed gjort framställningen mer fullständig än den annars varit.

Ytterligare andra har i samtal eller på annat sätt varit mig till hjälp i att hitta lämpliga personer för kartläggningen och att i övrigt överblicka, och göra rimliga överväganden inom så varierande delar av humaniora som det här är fråga om. Till dem som delat med sig av sin tid hör Tom Ericsson, Johan Nordlander, Pat Poussa, samtliga Umeå, Johan Svedjedal, Uppsala, Anders Steinvall och Patrik Svensson, Umeå och Maria Ågren, Uppsala. Av Börje Justrell, fick jag vid besök på Riksarkivet ingående information om samlingarna där, läget i fråga om digitaliseringen och utnyttjandet av materialet för forskningen.

Slutligen har Jan Larsson i egenskap av forskningssekreterare i DISC, men också som historiker bidragit på många sätt till kartläggningen – i planeringen och uppläggningsen och inte minst i det heldagsseminarium som ingick som en del i kartläggningen.

Till alla dessa, ett varmt tack!

Umeå 18 juni 2007



Eva Strangert

Uppdraget redovisades den 12 september 2007 inför DISC:s styrelse.

INNEHÅLL

INLEDNING	3
SAMMANFATTNING OCH SLUTSATSER.....	5
BAKGRUND	7
Humaniora – Frågeställningar, metoder och relationer till andra vetenskaper	7
Humaniora i ett samhällsperspektiv	8
Digitalisering av kulturarvet	10
ÖVERVÄGANDEN INFÖR OCH UPPLÄGGNINGEN AV KARTLÄGGNINGEN	12
SITUATIONEN IDAG	14
Befintliga resurser.....	14
<i>Språkvetenskap – svenska/nordiska språk, engelska, romanska språk</i>	14
<i>Litteraturvetenskap</i>	15
<i>Konst- och bildvetenskap</i>	16
<i>Medie- och kommunikationsvetenskap</i>	16
<i>Etnologi</i>	16
<i>Historia – modern historia och medeltidshistoria</i>	17
<i>Arkeologi, inkl miljöarkeologi</i>	18
<i>Religionsvetenskap</i>	18
Nyttjande	19
Erfarenhet av samverkan vid uppbyggnad av databaser.....	19
Betydelsen av befintliga databaser vid arkiv, bibliotek och museer, mm	20
FRAMTIDA BEHOV	21
Digitalisering av material vid arkiv, bibliotek, museer, mm	21
<i>Olika ämnesområden – likartade behov och likartad finansiell problematik</i>	22
Nya databaser	23
<i>Språkvetenskap</i>	23
<i>Litteraturvetenskap</i>	23
<i>Konst- och bildvetenskap</i>	24
<i>Medie- och kommunikationsvetenskap</i>	24
<i>Etnologi</i>	25
<i>Historia – modern historia och medeltidshistoria</i>	25
<i>Arkeologi, inkl miljöarkeologi</i>	25
<i>Religionsvetenskap</i>	26
<i>Olika ämnesområden – likartade behov</i>	26
UTVECKLING OCH NYTÄNKANDE.....	27
Vetenskapligt intressanta utvecklingslinjer	27
Samverkan för att utveckla nya databaser med bred användning	27
Databaser och synlighet i internationellt perspektiv	28
GENERELLA PROBLEM SOM KRÄVER EN LÖSNING	30
Tillgänglighet.....	30
<i>Upphovsrättsliga begränsningar</i>	30
<i>Etiska hinder</i>	30
<i>Forskningsinterna hinder</i>	30
<i>Kostnadsaspekter</i>	30
Teknik, metoder och teknisk samordning.....	31
<i>Teknikutveckling, dokumentation, standardisering och verktyg</i>	31
<i>Samverkan och samordning av insatser</i>	31
SLUTSATSER OCH FÖRSLAG TILL ÅTGÄRDER	32

BILAGOR (i separat dokument)

1. Inbjudan inklusive bakgrund till kartläggningen samt, i separat bilaga, frågeställningar att besvara)
2. Svenska/nordiska språk, nyare inriktning (Britt-Louise Gunnarsson, Uppsala universitet)
3. Svenska/nordiska språk, äldre inriktning (Lars-Erik Edlund, Umeå universitet)
4. Engelska (Merja Kytö, Uppsala universitet)
5. Romanska språk (Inge Bartning, Stockholms universitet)
6. Litteraturvetenskap (Mats Malm, Göteborgs universitet)
7. Konstvetenskap (Lena Johannesson, Göteborgs universitet)
8. Medie- och kommunikationsvetenskap (Ester Pollack, Stockholms universitet)
9. Etnologi (Magnus Bergquist, Göteborgs universitet)
10. Historia, modern inriktning (Lars Nilsson, Stockholms universitet)
11. Historia, modern inriktning (Marie C. Nelson, Linköpings universitet)
12. Historia, äldre inriktning (Olle Ferm, Stockholms universitet)
13. Arkeologi, inklusive miljöarkeologi (Kristian Kristiansen, Göteborgs universitet, samt Phil Buckland, Umeå universitet)
14. Religionsvetenskap (Jørgen Straarup, Umeå universitet/Högskolan i Gävle)
15. Inventering av språkliga korpusar och datoriserade lexikon vid Språkvetenskapliga fakulteten, Uppsala universitet (Anna Sågvall-Hein, Uppsala universitet)

INLEDNING

Uppdraget från Database Infrastructure Committee (DISC) var att beskriva läget – befintlig forskningsinfrastruktur såväl som behovet av infrastrukturella resurser i ett femårsperspektiv – inom det humanistiska fältet. Med forskningsinfrastruktur avses här databaser, i första hand digitala datasamlingar, men också befintligt icke-digitalt material, av stort värde för forskningen nationellt och internationellt. I uppdraget ingick också att göra en rimlig avgränsning av datamaterial vid myndigheter och organisationer utanför universiteten som bör ligga inom det område som DISC ansvarar för.

En utgångspunkt var en tidigare rapport från april 2005¹ som sammanfattar resultatet av ett regeringsuppdrag till Vetenskapsrådet att "... som underlag för sitt arbete med vetenskaplig infrastruktur ... kartlägga den nationella infrastrukturen inom humaniora och samhällsvetenskap..." Denna kartläggning begränsades till befintlig infrastruktur, främst digitala forskningsdatabaser, men även icke-digitaliserade materialsamlingar ingick. Den nu aktuella kartläggningen är att se som en komplettering och utvidgning av denna tidigare inventering, men nu med fokus på humaniora. Den förväntas också ge också underlag för prioriteringar som måste till inför kommande satsningar och anknäyer därmed till Vetenskapsrådets långsiktiga plan för forskningsinfrastruktur från juni 2006² utarbetad inom Kommittén för forskningens infrastrukturer (KFI).

I infrastrukturplanen (s. 20) tas digitalisering inom humaniora upp som ett område under rubriken "Infrastrukturer eller områden med stor potential".³

Under rubriken "Digitalisering av det svenska kulturarvet" (s. 106) står:

Många humanister arbetar direkt med material från arkiv, forskningsbibliotek, museer och andra samlingar. Det finns värdefulla material av stor vikt för humanistisk forskning som ännu inte är digitaliserade. Om de vore digitaliserade kunde de ingå i nationella bild-, ljud-, språk- och andra forskningsdatabaser. Omfattande resurser kommer att krävas för digitalisering, dokumentation, lagring och åtkomst av dessa data. Långsiktiga prioriteringar blir nödvändiga och bör baseras på noggranna förstudier om till exempel källmaterialets innehåll och struktur, krav på sök- och analysmetoder, var och hur det digitaliserade materialet ska lagras, hur det ska underhållas och tillgängliggöras för forskarsamhället. Prioriteringarna bör göras i dialog mellan forskarsamhället, Vetenskapsrådets ämnesråd och KFI, samt andra forskningsfinansiärer.

I de sammanfattande rekommendationerna återfinns bland att-satserna (s. 112). "Att KFI utreder ramarna kring digitalisering av delar av det svenska kulturarvet." I anslutning härtill görs också en hänvisning till den då av DISC planerade (och nu, genom denna rapport, föreliggande) kartläggning av bl a digitalisering inom humaniora.

¹ *Om forskningens infrastruktur inom humaniora och samhällsvetenskap i Sverige*, http://www.vr.se/download/18.320a86de108dcd98cbc80004975/Om+forskningsens+infrastrukturer_KFI-HS_slutrappport.pdf

² Se kommentar s. 20 i *Vetenskapsrådets guide till infrastrukturen*, <http://www.vr.se/download/18.7bea596910e36c19cbc80001735/Rapport+14.2006.pdf>

³ Planen tar också upp humanioras roll inom ramen för större europeiska infrastruktursatsningar som ESFRI (European Strategy Forum on Research Infrastructures).

I infrastrukturplanen finns således en fokusering på digitalisering av ännu icke digitalt material. Det är helt uppenbart att många ämnen inom humaniora utnyttjar sådana data ur olika källor och att omfattningen och behovet av digitalisering av detta material måste kartläggas. Men det är samtidigt ett faktum att det också finns omfattande digitala databaser och syftet med kartläggningen är därför att också täcka sådana existerandesurser liksom behovet av helt ny databasuppbyggnad.

Begreppet digitalisering som det används här är inte entydigt, även om ”omvandling från analog till digital form” kan anses vara en gemensam nämnare. Men inom detta ryms allt från avfotografering av serier av föremål, bilder och texter lagrade på dator till avancerade bearbetningar (annotering, andra analyser) av olika originalmaterial (text, ljud, bild) inom ramen för välstrukturerade och sökbara databaser. Digitaliseringsbehovet inom humaniora kan förväntas inkludera såväl dessa extremformer som olika mellanformer.

Att fullständigt täcka ett så brett fält som det humanistiska med alla dess olika ämnesområden är inte görlikt, men syftet har likväl varit att ge en så bred och representativ bild som möjligt. Sammantaget kom kartläggningen att omfatta 11 olika områden – något beroende på hur man räknar – representerande humaniora i dess hela bredd: svenska/nordiska språk, engelska, romanska språk, litteraturvetenskap, konst- och bildvetenskap, medie- och kommunikationsvetenskap, etnologi, modern historia och medeltidshistora, arkeologi och religionsvetenskap.

Övervägandena gällde också hur insamlingen av uppgifter skulle ske. Den slutgiltiga uppläggningsen förelades och diskuterades i DISC och byggde även på omfattande sonderingar med olika företrädare för humaniora. Den kom att bestå av dels ett heldagsseminarium där inbjudna ämnesrepresentanter beskrev situationen inom sina respektive områden, dels skriftliga sammanställningar som byggde vidare på seminariepresentationerna. Dessa skriftliga sammanställningar utgör underlaget för kartläggningen. Därutöver har annan tillgänglig dokumentation utnyttjats i den mån den ansetts ge ytterligare belysning av området.

Rapporten i preliminär form har gått ut på remiss till de olika ämnesrepresentanterna och inkomna synpunkter har inarbetats i den nu slutgiltiga versionen.

I det följande presenteras inledningsvis en sammanfattning av arbetet och de viktigaste slutsatserna. Därefter ges en kort bakgrund – en karaktärisering av det humanistiska fältet med dess relationer till andra vetenskaper, humanioras roll i samhällsperspektiv samt en diskussion kring begreppet ”digitalisering av kulturarvet” i dess relation till humanistisk forskning. I följande avsnitt redovisas överväganden inför och uppläggningsen av kartläggningen samt resultatet med beskrivningar av den forskning som bedrivs inom de olika ämnesområdena, de tillgängliga resurserna och de framtida behoven av digital infrastruktur. Avslutningsvis framläggs slutsatserna av arbetet och ges förslag till åtgärder.

SAMMANFATTNING OCH SLUTSATSER

Humanistisk forskning befinner sig i ett utvecklingsskede, där samverkan såväl inom humaniora som med andra discipliner får allt större utrymme. För det behövs uppdatering och vidareutveckling av tillgängliga databaser och uppbyggnad av nya stora databaser för bred användning. Det behövs också digitalisering – med särskilt fokus på just forskningens behov – av icke-digitalt material vid arkiv, bibliotek och museer. I stort sett alla ämnen i kartläggningen utnyttjar – och är beroende – av sådant material ur de samlingar vid arkiv, bibliotek och museer mm vi räknar till ”kulturarvet”. Förändringen speglar också utvecklingen inom informationsteknologin. Det är lätt att se den potential som finns i tillgängliggörandet av stora mängder data som kan delas mellan många.

De slutsatser som i övrigt kan dras av kartläggningen kan sammanfattas i följande punkter:

- De befintliga resurserna i form av digitala databaser är ojämnt fördelade över de olika ämnena. Här finns etnologi och religionsvetenskap, där databaser spelar en undanskymd roll, andra, som litteraturvetenskap, konstvetenskap, medie- och kommunikationsforskning och arkeologi där forskningen i stor utsträckning är *beroende* av sådana resurser och så språkvetenskap och historia (den moderna historieforskningen) där forskningen i stor utsträckning *kräver* databasmaterial.
- Storleken på materialen varierar inom vida gränser, från stora datamängder till långt mindre samlingar uppbyggda av enskilda forskare. En del av databaserna är utvecklade inom landet – det gäller i de flesta fall och generellt över ämnena – men databaser uppbyggda på andra håll i världen eller i samarbete med svenska forskare utnyttjas också.
- I stort sett alla ämnen i kartläggningen utnyttjar – och är beroende – av material som finns vid arkiv, bibliotek och museer, främst vid de större statliga institutionerna men också vid bl a kommunala myndigheter och andra organisationer. En del av dessa samlingar föreligger som välstrukturerade och sökbara databaser, men en mycket stor del utgörs av icke-digitalt material som behövs – i digital form – för forskningen.
- Behoven av databasuppbyggnad är omfattande och mångskiftande. Utöver de behov som gäller enskilda områden, framkommer sådana där man tydligt kan se att det finns sammanfallande intressen och därmed potential till bred användning av resurserna. Hit hör: (A) Språkdatabaser med tal- och skriftspråk, (B) det samnordiska projektet *Menota* inriktat på nordiska medeltidstexter, (C) Internetdatabaser med information i form av text, tal, bild, grafik etc samt (D) databaser med bilder och bildrepresentationer av skilda slag som kartor, ritningar, modeller och instrumentella, inklusive rörliga, bilder.
- Inom de flesta ämnesområden framkommer ett starkt behov av dokumentation av dels den forskning som pågår – bibliografiska databaser – liksom av databaser över databaser. Avsaknaden av överblick ses som starkt hämmande för forskningen. Det finns en vilja till större samverkan än idag, men en slutsats som kan dras är att bristen på överblick bidrar till att den samverkanspotential som framkommer i kartläggningen inte uppmärksammas av de berörda och därför inte heller utnyttjas.
- Svensk Nationell Datatjänst (SND) kan i framtiden förväntas fylla en ovärderlig funktion när det gäller att möta behov som ovanstående på dokumentation, överblick och samordning av databasresurser. SND har också en viktig uppgift i fråga om vägledning och stöd samt vad gäller standardisering, användbara gränssnitt, sökfunktioner och uppdatering till dagens teknik av äldre databaser inom humaniora.
- Forskningen hindras av upphovsrättsliga begränsningar för text och bilder, regler kring digital access samt kostnadspålagor för utnyttjande av tillgängligt material. Det är problem som humaniora till viss del har gemensamt med andra vetenskaper.

- Omfattande ekonomiska hinder står i vägen för nödvändig digitalisering och registrering av äldre material inom ramen för sedan länge påbörjade projekt vid myndigheter och organisationer utanför universiteten. Den tilldelning som tidigare gavs i form av anslag direkt till projekten, via bl a statliga forskningsfinansiärer, går nu direkt till myndigheterna inom ramen för totalanslag till verksamheten. Det gör att många projekt, som tidigare haft mer omfattande resurser, i konkurrensen med andra behov inom verksamheten, nu går på sparlåga.
- Kartläggningen visar sammantaget en vilja till nationell – och internationell – samverkan när det gäller utveckling i fråga om databaser, inom ämnen och över ämnesgränser. Man efterlyser t ex bättre samordning av registreringen i demografiska databaser, en portal för arkeologiska databaser för att bättre överblicka området etc. Det gäller nu alltså att gå vidare för att skapa de resurser som den humanistiska forskningen kräver för att utvecklas, inte minst i samverkan med andra vetenskapsområden, där humaniora kan bidra med sina perspektiv. Initiativen här bör tas av forskarna själva, men det krävs också satsningar och stödinsatser för att komma vidare.
- Till de åtgärder som bör övervägas i ett första skede hör utbyggnad av nya databaser på de områden (A)-(D) ovan där man kan se en bred användning med utnyttjande från många olika ämnen. Här krävs dock för vart och ett en mer ingående kartläggning av behoven innan vidare steg tas. Forskningens behov av digitalisering av kulturarvet behöver också utredas vidare. Omfattande digitaliseringsinsatser behövs, men informationen är otillräcklig för att i dagens läge se möjligheten till prioriteringar och att göra kostnadsberäkningar. Här föreslås en planläggning i flera steg som inbegriper såväl forskarna själva som de myndigheter och organisationer som idag har hand om digitaliseringen liksom Vetenskapsrådet (och andra forskningsfinansiärer).

BAKGRUND

Humaniora - Frågeställningar, metoder och relationer till andra vetenskaper

Fokus inom humaniora är människan/det mänskliga – i språket, i historien och i det estetiska uttrycket. Denna beskrivning reflekteras i den traditionella sektionsindelningen inom humaniora i språkämnen, historisk-filosofiska ämnen och estetiska ämnen. Även religionsvetenskap utgör en del av humaniora vid flera svenska universitet. Nationalencyklopedin beskriver humanioras särart – och relationer till andra vetenskaper – på följande sätt:

.... Metodiskt kännetecknas de humanistiska disciplinerna av att tolkning – av texter, bilder, musik m.m. – spelar en stor roll i forskningen. Traditionellt har också studiet av det enskilda, i motsats till det generella och lagbundna, brukat ha stort utrymme inom humaniora. En nyare term, utan den estetiska association som ligger i humaniora, är *humanvetenskaper*, som betyder 'de vetenskaper som studerar människan och hennes situation'. Hit räknas förutom humaniora även discipliner såsom psykologi och sociologi, vilka tillhör de samhällsvetenskapliga fakulteterna.

Frågeställningarna varierar självfallet beroende på ämnesområde, liksom i viss mån metoderna, men generellt gäller således att angreppssättet är beskrivande och analytiskt snarare än deduktivt, dock med inslag av såväl hypotestestande som teoribyggande. Viktiga skeenden i historien kartläggs och analyseras och ger perspektiv på nutiden. Det gäller alla delar av vårt kulturarv – historia, arkeologi, litteratur, konst, religion och språk. Språket i dess olika former av text och tal beskrivs i dess regionala, sociala, ålders- och könsvariationer. Analyser av språkinläringens olika stadier hos det lilla barnet såväl som hos den vuxna inläraren bidrar till förståelse av den komplexa process som språktillägnande utgör. Förutom att vi kommunicerar genom språket kommunicerar vi också via bilder och andra konstformer och detta, liksom andra former av kulturyttringar studeras inom medievetenskap, etnologi, antropologi, konst- och litteraturvetenskap. Värderingar och deras förändringar över tid är ett växande forskningsområde inom religionsvetenskap, filosofi och idéhistoria. Listan kan göras lång och detta är bara några exempel.

Metoderna för datainsamling varierar också. Arkivstudier spelar stor roll främst inom historievetenskaperna, men också inom den historiskt inriktade språkvetenskapen, i synnerhet vad gäller dialekt- och namnforskning, samt inom olika delar av religionsvetenskapen, inom etnologi och inom konst och litteratur. Inom de senare har också samlingar vid bibliotek och museer stor betydelse. Den modernt inriktade språkvetenskapen baserar sig på analys av texter och inspelat tal inom olika genrer och starkt fokus riktas idag på dialog- och samtalsforskning. Text, bild och ljud i olika media utnyttjas inom medie- och kommunikationsforskning, inom språk- och litteraturforskning, för kulturstudier och för att studera värderingar. I studier av värderingar används ofta survey-undersökningar och enkätförfaranden. Receptionsstudier liksom rena perceptionsstudier förekommer också inom vissa ämnen

Informationsteknologin har inte bara förändrat förutsättningarna för utan också hur forskningen bedrivs. Tillgången till teknologi, bl a möjligheten att söka digital information, har medfört nya möjligheter. Det gör också att nya behov uppstår. Arkivforskaren t ex, som tidigare var helt hänvisad till att fysiskt finnas på plats, kan nu via sin dator få tillgång till det önskade materialet över Internet – i den mån det verkligen är digitaliserat. Det öppnar för nya behov av digitalisering av tidigare icke-digitaliserat material. Digitala språkdbaser med text, tal och även tecken byggs upp, samtidigt som det behövs digitalisering av insamlat icke-digitaliserat material. På samma sätt skapar historiker och arkeologer nya databaser. Föränd-

ringen påverkar inte bara tillgången på material, teknologin och den breda tillgången på information öppnar också nya möjligheter att vidga perspektiven på den egna verksamheten. Allt detta gör att den humanistiska forskningen av idag ser avsevärt annorlunda ut än den gjorde tidigare.

Förändringen påverkas också av ett allt större inslag av tvärvetenskap. Samarbeten inom humaniora förekommer idag i växande utsträckning, bl a i form av samverkan där valda tematiska områden utforskas ur olika aspekter. Det ger underlag för större och mer samordnade forskningsprojekt än vad som tidigare förekommit. Det finns också gemensamma forskningsfrågor och samverkan med samhällsvetenskapliga ämnen på det humanvetenskapliga fältet vad gäller språk, media, historia och kulturvetenskaper som t ex etnologi. Det finns en nära anknytning till utbildningsvetenskap ifråga om ämnesdidaktik. Överbryggningsprojekt mot såväl medicin som naturvetenskap förekommer också i viss utsträckning. Här finns exempel inom arkeologin med specialiseringar som miljöarkeologi där rent naturvetenskapliga metoder utnyttjas i analyser av äldre tidsskeden. Ett tydligt exempel på denna tendens till samverkan är det utkast till storprojekt, *What it means to be human*, som skisseras i ett EU-dokument från 2005 inom ramen för NEST (New and Emerging Science and Technology)⁴. Expertgruppen bakom detta, där Sverige är representerad med historikern Arne Jarrick, ser projektets potential i de landvinningar för forskningen som kan väntas genom att föra samman områden som genetik, neurobiologi, kognitions- och beteendevetenskap m fl med humanistiska vetenskaper som historia och filosofi. Syftet är således att se olika aspekter av "det mänskliga" i relation till den kulturella kontexten.

De tvärvetenskapliga aspekterna gör att vissa ämnen till och med är svårplacerade och därmed kan ha olika fakultetstillhörighet vid olika svenska universitet. Till dessa ämnen hör t ex medie- och kommunikationsvetenskap som på en del orter räknas in under samhällsvetenskap och på andra under humaniora. Hit hör också religionsvetenskap som vid vissa universitet ingår under humaniora, vid andra ingår i en större teologisk fakultet.

Humaniora kan därför sammantaget ses som ett område ståt i stark omvandling. Även om det är en överdrift att påstå att den ensamme forskaren böjd över böckerna i arkiven är på väg att försvinna – för den typen av studier behövs i viss utsträckning – så blir den humanistiska forskaren i stor utsträckning alltmer en samverkande medlem i laget som utforskar alltmer komplexa processer, oftast med datorn som arbets- och analysredskap. Forskningen med inriktning på en-personsprojekt var för övrigt snarare en produkt av finansieringssystemet (under den tid när man inte kunde få medel för mer än en forskare i taget) än av forskarnas egna önskemål.

Humaniora i ett samhällsperspektiv

Denna omvandling ger humaniora en tydligare roll i ett samhällsperspektiv. I ett alltmer komplext och mångkulturellt samhälle med snabb förändringstakt och skiftande omvärldsvillkor är en uppenbar uppgift för humaniora att göra människor delaktiga i vad som sker.

Här spelar språket en nyckelroll. Språket, i talad eller skriven form eller som tecken, är vårt främsta uttrycksmedel i det privata, i arbetslivet och i samhället i stort. Dess betydelse i alla dessa aspekter men främst i samhällsperspektiv var utgångspunkten för utredningen *Mål i*

⁴ ftp://ftp.cordis.europa.eu/pub/nect/docs/whatitmeanstobehuman_b5_eur21795_en.pdf

mun (2002)⁵. Propositionen som följde, *Bästa språket – en samlad svensk språkpolitik*, 2005⁶ formulerade de språkpolitiska målen. Dessa innebär att alla ska ha rätt till ett väl fungerande språk för att inte hamna utanför språkliga gemenskaper. Och även om svenskan ska vara huvudspråk i Sverige, så ska det också vara möjligt att kommunicera på minoritetsspråk och på främmande språk. I denna samhällsbärande roll, finns en uppenbar uppgift för språkforskningen, inte minst som bas för kunskap om hur språk tillägnas och hur språkundervisning ska bedrivas effektivt.

I en alltmer internationaliserad omvärld måste överhuvudtaget språkliga barriärer rivas. Samtidigt som det är viktigt att slå vakt om språklig mångfald – i alla länder värnas det egna språket – så måste vi kunna förstå varandra. Om den ekvationen ska gå ihop måste satsningar ske på mångspråkighet. Här utgör språkpolitiken inom EU som går ut på att varje lands språk ska vara EU-språk en stor utmaning för inte minst översättningsforskningen⁷.

En annan roll för humaniora är att bidra med perspektiv och samhällskritisk analys⁸. Hur påverkas människor av att leva i en alltmer internationell och mångkulturell omvärld? Vad sker när människor tvingas bryta upp ur invanda miljöer när arbeten inte längre finns på hemorten? Hur förändras inställningen till livet – från födelse till död – i dagens samhälle? Vilka etiska frågor uppstår när övervakningen av medborgarna ökar genom utvecklingen av nya säkerhetssystem och alltmer avancerad avlyssning av telefonsamtal och loggning av datorer? Hur speglas samhällsfenomen, t ex nationalistiska strömningar, genom litteraturen vid olika skeden i vår historia? Detta är frågeställningar där den humanistiska forskningen kan bidra med analyser utifrån de perspektiv – kulturella, historiska, filosofiska – som humaniora arbetar med.

Den framskjutna ställning som kulturarvet fått idag och inte minst behovet av att göra det digitalt tillgängligt får här sin betydelse. Det är inte minst genom att se samtiden i historisk belysning som möjlighet ges till perspektivskapande. Och det handlar inte bara om det nationella arvet utan i lika hög grad om det nordiska och det europeiska. Men det stannar inte med Europa. Vår egen historia måste sättas in i ett vidare mer globalt perspektiv. Seminarier i Vetenskapsrådets regi under senare tid ger flera exempel på denna globala orientering, ansatser till att skriva en litteraturhistoria som vidgar sig långt utanför en snävt västerländsk sfär, arkeologer som sätter den svenska historien i relation till globala företeelser med exempel från Afrika, Asien och Latinamerika. Följande citat⁹ anger sålunda den färdriktning som idag i hög grad präglar humaniora:

Både nationell och internationell samverkan är nödvändig för att utveckla också humanistisk forskning. Att många humanister forskar om det svenska kulturarvet utesluter inte internationella nätverk eller internationell publicering. Tvärtom, den humanistiska kulturanalysen är nödvändig i vår tid då det stora europeiska projektet går ut på att söka en identitet både inom och bortom nationalstaten i en postkolonial värld.

⁵ www.regeringen.se/sb/d/108/a/1443

⁶ www.regeringen.se/sb/d/5359/a/50761

⁷ Europeiska Ministerrådets rapport *En ny ramstrategi för flerspråkighet*, 2005

Eurllex.europa.eu/smartapi/cgi/sga_doc?smartapi!celexplus!prod!DocNumber&lg=sv&type_doc=COMfinal&_doc=2005&nu_doc=596

⁸ Se regeringens proposition *Forskning och förnyelse* 2000/2001:3, avsnitt 6.3.4.

⁹ Ur *För en offensiv humaniora*, krönika av Margaretha Fahlgren, Forska 2005, nr 3.

Digitalisering av kulturarvet

Att göra det svenska kulturarvet tillgängligt i digital form på nätet är idag en högt prioriterad verksamhet. Det är också en fråga av största angelägenhet för den humanistiska forskningen i dess hela bredd och Vetenskapsrådet ser som nämnts en stor potential i att digitalisera kulturarvet. Det sker också stora satsningar i Sverige, bl a inom ramen för det sk Accessprojektet¹⁰ via Kulturrådet. Liknande och mycket stora satsningar sker för övrigt över stora delar av världen och i vår närhet sker samverkan inom ramen för EU för att värna ”det europeiska kulturarvet”. Hit hör initiativ som ECHO¹¹, där fokus ligger på att samla redan existerande material och göra det tillgängligt på Internet. Det pågående DAM-LR¹² och det planerade DARIAH¹³ inom det större ESFRI-programmet har på liknande sätt till uppgift att tillgängliggöra material som rör kulturarvet.

I flera av dessa initiativ finns svensk medverkan. Språk- och litteraturcentrum/Humanistlaboratoriet vid Lunds universitet är engagerat i ECHO såväl som i DAM-LR¹⁴, det senare främst inriktat på tillgänglighet till språkresurser¹⁵. HUMLab i Umeå är vidare koordinator för projektet QVIZ (Query and Context Based Visualizations of Time-Spatial Cultural Dynamics)¹⁶ inom sjätte ramprogrammet. Syftet är ”att utveckla ett ramverk där digitalt arkivmaterial kan göras tillgängligt via kartor, och där användarna kan bidra till och samarbeta kring materialet.”

Men tillgängligheten till digitala kulturarvsdata gäller inte bara forskningen utan perspektivet är vidare och gäller samhället i stort. Syftet är att sprida kännedom om kulturarvet till breda grupper varav forskarna är en bland andra. Detta framgår tydligt i de statliga dokument som styr bl a den svenska kulturpolitiken. Ett av målen som de beskrivs i kulturpropositionen 1996/97 är att ”jämma ut de skillnader som finns mellan olika grupper av människor när det gäller bevarandet av och tillgången till kulturarvet”¹⁷. I andra måldokument inom arkiv, bibliotek och museer, dvs ABM-sektorn i stort, gäller detsamma – kulturarvet ska göras tillgängligt för forskningen, men också för skolan och olika utbildningar och för medborgaren/allmänheten. En studie av de styrande ideologierna – valet av berörda målgrupper – i tre digitaliseringsprojekt, Sueciaprojektet vid Kungliga Biblioteket (KB), André museet vid Stiftelsen Grännamuseerna och Ostindieprojektet vid Göteborgs universitet, visar just hur man vänder sig till breda grupper av nyttjare¹⁸.

Detta är i sig ytterst positivt, men när fokus som i det här aktuella fallet är forskningen, kan det breda greppet också innebära vissa problem. Även om samma material kan ha betydelse för breda användargrupper, ställer forskningen speciella krav på materialet, inte minst vad gäller dokumentation och metadata. Det gör att de stora ekonomiska satsningar på digitali-

¹⁰ http://www.kulturradet.se/templates/KR_Page.aspx?id=813&epslanguage=SV

¹¹ European Cultural Heritage Online, <http://echo.mpiwg-berlin.mpg.de/home>

¹² <http://www.mpi.nl/DAM-LR/>

¹³ <http://europa.eu/rapid/pressReleasesAction.do?reference=MEMO/06/384&format=HTML&aged=1&language=EN&guiLanguage=en>

¹⁴ Projektmedverkan genom Sven Strömquist, chef vid Humanistlaboratoriet.

¹⁵ För en beskrivning, se <http://www.sol.lu.se/forskning/projekt.html?projectid=65>

¹⁶ Med HUMLabs föreståndare Patrik Svensson som projektledare. För en beskrivning se,

<http://www.qviz.eu/svenska.php>

¹⁷ Prop 1996/97:3, s 126.

¹⁸ Gumælius, M. 2004. *Vad innebär digitalisering av kulturarvet? En ideologianalys av tre svenska digitaliseringsprojekt*. Magisteruppsats, 2004: 53, Bibliotekshögskolan, <http://bada.hb.se/dspace/handle/2320/1088>

sering av kulturarvet (bl a Access-projektet) långt ifrån alltid ger utdelning för forskningen. En nyligen genomförd magisteruppsats vid Högskolan i Borås¹⁹ visar hur kunskapsbrister när det gäller standarder och brist på samverkan mellan olika digitaliseringsprojekt men också olika fokus på projekten kan inverka negativt på verksamheten.

Brister av detta slag föreligger inte när det gäller den digitalisering som bedrivs vid de större nationella myndigheterna inom ABM-sektorn. Här finns kompetensen och även många och omfattande projekt som tillgodoser forskningens behov - men endast till en del. Finansieringen är långt ifrån tillräcklig. Det drabbar för övrigt inte bara olika digitaliseringsprojekt utan också andra projekt som ordboksutgivning etc som bedrivits långsiktigt sedan kanske 30-40 år tillbaka i tiden och för vilka anslagen stegvis urholkats så att de nu endast går på mycket svag sparlåga.

I detta läge behövs sannolikt inte bara mera pengar. Ur forskningens synvinkel kan det också finnas behov av speciella prioriteringar när det gäller att digitalisera ännu inte digitalt tillgängligt material, t ex att vissa insatser bedöms som mer angelägna än andra.

¹⁹ Emanuelsson, C. 2006. *Digitalisering av kulturarvet - En studie av digitalisering vid två museer*. Magisteruppsats 2006: 65, Bibliotekshögskolan, <http://dspace.bib.hb.se:8080/dspace/handle/2320/1480>

ÖVERVÄGANDEN INFÖR OCH UPPLÄGGNINGEN AV KARTLÄGGNINGEN

De överväganden som inledningsvis måste göras gällde hur det humanistiska området skulle täckas så att resultat skulle ge en representativ och tillräckligt omfattande bild av verksamheten. Viss hjälp gavs här genom den tidigare nämnda kartläggningen från 2005 (se inledningen) som byggde på enkätsvar utsända till landets universitet och högskolor. Den bild som framkom där var en stark representation av språkdata-baser – ca hälften av de humanistiska databaserna gällde språk. I den andra hälften fanns de flesta databaserna inom historieämnet samt i viss mån arkeologi. Andra områden var representerade i begränsad omfattning och vissa saknades helt. I det samlade materialet ingick såväl databaser i digital form, uppskattningsvis drygt hälften, som ännu icke digitaliserat material.

Att språk- och historieämnena skulle vara väl representerade i kartläggningen var härigenom naturligt. Men, eftersom mycket talade för att den tidigare kartläggningen, som gjorts under stark tidspress, inte nått ut i tillräcklig omfattning, var det rimligt att även få med så många ämnen som möjligt utöver språk och historia, och även sådana som inte alls fanns med tidigare. De ämnen som slutligen kom att ingå var språk (engelska, romanska språk, svenska/nordiska språk), litteraturvetenskap, medie- och kommunikationsvetenskap, konstvetenskap, etnologi, historia (modern och medeltidshistoria), arkeologi, inklusive miljöarkeologi och religionsvetenskap.

En hel del av den digitalisering och databasuppbyggnad som har betydelse för humanistisk forskning sker utanför universiteten, vid statliga arkiv, bibliotek och museer mm. Trots detta har uppgiftslämnarna i kartläggningen i samtliga fall varit forskare vid universiteten. Utgångspunkten har varit *forskningens behov* av digitala resurser. Det innebär självfallet inte att det behov av digitalisering av "kulturarvet" som är så betydelsefull för humanistisk forskning inte ska ske i dialog med arkiven, biblioteken och museerna. En sådan dialog i ett nästa steg bör också inkludera såväl Vetenskapsrådet som andra forskningsfinansiärer.

Övervägandena gällde också hur insamlingen av uppgifter skulle ske. Den slutgiltiga uppläggningsen, som förelades och diskuterades i DISC och som föregicks av omfattande sonderingar med olika företrädare för humaniora, blev följande:

Kontakt togs med representanter för de ämnesområden som valts ut för kartläggningen. De som tillfrågades och kom att bidra med underlag var personer med överblick och väl ansedda inom sina respektive områden och som därtill accepterade att se sig som *nationella* representanter för sitt ämne. De skulle sålunda svara inte bara för sig själva och sin egen institution utan för sitt *ämne*.

Ämnesrepresentanterna inbjöds till ett heldagsseminarium vid Vetenskapsrådet. Den skriftliga inbjudan (Bilaga 1) gav en bakgrund till uppgiften och vad som förväntades av var och en (nationell förankring av uppgiften, förberedelser inför och presentation vid seminariet och därefter eventuell komplettering samt slutligen en skriftlig sammanställning om 5-7 sidor). Som stöd i sitt arbete fick var och en ett antal frågeställningar att utgå ifrån (se bilaga till Bilaga 1).

Vid seminariet presenterade var och en sitt område och beskrev läget vad gäller tillgång och behov av databaser och digitalisering med utgångspunkt i de givna frågeställningarna.²⁰ Det

²⁰ Den vidtalade representanten för arkeologi fick förhinder och deltog inte i seminariet. Underlaget för kartläggningen ställdes till förfogande genom samverkan av representanter för institutionerna i Göteborg och Umeå.

nationella perspektivet var tydligt. Kolleger vid andra lärosäten än det egna hade på olika sätt bidragit till redovisningen. Uppdraget hade gjorts känt nationellt, frågorna hade skickats vidare och svar inhämtats från kollegerna runt om i landet.

Efter seminariet gavs tid för uppföljning och komplettering inför utformningen av den skriftliga sammanställning som utgjorde den slutgiltiga avrapporteringen av läget inom respektive ämnesområde. Dessa skriftliga sammanställningar utgör det huvudsakliga underlaget för kartläggningen. Därutöver har annan tillgänglig dokumentation utnyttjats i den mån den ansetts ge ytterligare bidrag till beskrivningen av det humanistiska området.

I följande avsnitt ges en översiktlig framställning av svaren på de frågeställningar (se bilaga till Bilaga 1) som ligger till grund för kartläggningen. För mer detaljerade uppgifter om de olika ämnesområdena hänvisas till Bilaga 2-14 som innehåller de ingivna sammanställningarna från områdesrepresentanterna.

SITUATIONEN IDAG

Befintliga resurser

Svar på frågeställningen: *Hur ser situationen ut idag inom ämnet? Finns digitala databaser? Andra typer av material och datasamlingar (tal, text, bilder etc)?*

Bilaga 2-14 innehåller omfattande dokumentation av befintliga databaser såväl inom Sverige som i andra länder. Det gäller material i digital såväl som i annan form, vid universitetsinstitutioner samt vid arkiv, bibliotek och museer som används inom svensk humanistisk forskning. Även om endast delar av dessa resurser täcks in i följande avsnitt, där de får tjäna som exempel, har uppgifterna stort värde som inventering av de samlade resurserna.

Språkvetenskap – svenska/nordiska språk, engelska, romanska språk

Inom språk finns stora mängder material i form av digitala text- och talkorpusar. Många av dessa utgörs av strukturerade korpusar speciellt utvecklade för lingvistisk analys. Hit hör merparten av det digitala materialet. Inom svenska, med fokus på modern svenska (inklusive finlandssvenska), finns när det gäller skrivet språk stora material (tidningstext, romaner och brev mm) i *Språkbanken* vid Göteborgs universitet. Omfattande digitala korpusar finns också vid bl a FUMS²¹ i Uppsala (fackspråk). Härutöver finns också stora talspråksmaterial, den mycket omfattande och endast delvis digitaliserade *Göteborg Spoken Language Corpus* (GSLC), *Talbanken 05* med inspelningar av dels autentiska, dels arrangerade talsituationer samt ett stort och varierat talmaterial vid FUMS.

Vid Institutet för språk och folkminnen (SOFI)²² finns omfattande samlingar av ortnamn, personnamn, dialekter och folkminnen (dvs beskrivningar av händelser och företeelser av olika slag) delvis digital form. *DIABAS*, *Nordisk dialektgeografisk databas* innehåller en stor mängd digitaliserade svenska och nordiska växtnamn. Digitala data om namn (ortnamn) finns också vid Lantmäteriverket och i bl a *Svenskt Diplomatarium* vid Riksarkivet. Till övriga historiskt inriktade datasamlingar hör *Fornsvenska textbanken* i Lund, samt Uppsalas *Samnordisk rundatabas*. Den utgör inte bara en källa för den språkvetenskapliga utan även den historiska forskningen. Detsamma gäller det material som bearbetas inom ramen för *Medieval Nordic Text Archive (Menota)*. *Menota* är en samnordisk organisation med idag omfattande arbete (och även samarbete med USA) i syfte att utveckla digitala redskap för att hantera nordiska medeltidstexter.

Liknande resurser finns för engelska och romanska språk, men här är databaserna i stor utsträckning utvecklade i samarbete mellan svenska och utländska institutioner. I engelska kan nämnas t ex *English-Swedish Parallel Corpus* (ESPC)²³, ett samarbete mellan Göteborgs och Lunds universitet. *A Corpus of English Dialogues 1560 – 1760* har fokus på talspråket i gångna tider. Utöver detta utnyttjas i stor utsträckning databaser som helt och hållet utvecklats på andra håll i världen. Just inom engelska finns här stora mängder material med stor genrevariation att tillgå .

²¹ Avdelningen för forskning och utbildning i modern svenska vid Institutionen för nordiska språk.

²² SOFI är en statlig arkivmyndighet med avdelningar i Uppsala, Göteborg, Lund och Umeå. Verksamheten inkluderar forskning.

²³ En parallellkorpus innehåller samma text (eller tal) på två språk. Vissa parallellkorpusar inkluderar tre eller flera språk. Parallellkorpusar har generellt stor betydelse inom språkvetenskapen och inte minst för översättningsforskning.

Situationen är likartad inom de romanska språken – franska, spanska och italienska – även om de allra största digitala samlingarna finns inom franska. Det gäller såväl det i Sverige utvecklade materialet som det som byggts upp utanför Sverige. Till korpusar med skriven franska hör flera tidningskorpusar bl a *Le Corpus de Stockholm* (COSTO I och II) och här finns också *Français parlé des médias* (FPM) med talad mediefranska. En spansk motsvarighet (vid Uppsala universitet) är *Corpus des periodicas hispanos*. Precis som inom engelska finns parallellkorpusar och andra översättningskorpusar, t ex *Fransk-svensk parallellkorpus* i Uppsala. Och precis som i engelska utnyttjas också en mängd internationellt tillgängliga databaser.

Inom engelska och romanska språk (liksom i andra språk) finns sk inlärarkorpusar med vars hjälp språkinlärningsprocessen (svenska vad gäller invandrare och andra främmande språk vad gäller t ex svensktalande) kan följas och där resultaten kan utnyttjas i undervisningen. Hit hör de svenska delarna av *International Corpus of Learner English* (ICLE). *InterFra*-korpusen och *InterIta* i Stockholm och *Corpus Lund* (i Lund) är exempel på franska och italienska inlärarkorpusar.

Valet av engelska och romanska språk, förutom svenska/nordiska språk, i kartläggningen utgör självfallet en stark begränsning för täckningen av databaser inom språkområdet. Läget inom engelska och romanska språk kan dock sägas vara representativt för en stor del av de språk, förutom svenska, som studeras vid svenska universitet, även om mängden och variationen av material är större än inom i synnerhet engelska och franska än inom många andra språk. En inventering gjord vid Språkvetenskapliga fakulteten i Uppsala, med en rik representation av olika språk, ger en översiktlig bild av förekomsten av databaser också inom bl a asiatiska, slaviska och klassiska språk (Bilaga 15)²⁴.

Eftersom språkvetenskapen inom olika enskilda språk i hög grad – och inte minst vad gäller metodiken – har mycket gemensamt med allmänlingvistikens finns också ett stort överlapp dem emellan vad gäller databaser. Samma typer av material används inom båda områdena och i viss mån också samma databaser. En del av överlappet gäller också gentemot de språkteknologiska databaserna som tillsammans med de allmänlingvistiska varit föremål för en nyligen genomförd kartläggning²⁵.

Litteraturvetenskap

Litteraturvetenskapens bas är litterära texter. Sådana ingår som en del i de språkdatabaser som byggts upp och utnyttjas inom lingvistik och språkteknologi. Dessa databaser utnyttjas till viss del inom litteraturvetenskapen. Men litteraturvetare behöver utöver själva texterna upplysningar om texterna, deras tillkomst, eventuella omarbetningar, upplagor, sökmöjligheter för att göra parallellsökningar mot andra material mm, något som inte är av lika vitalt intresse inom språkvetenskapen. Dessutom måste texterna vara trogna mot originalet och korrekturlästa.

Bristen på sådana databaser är mycket stor inom litteraturvetenskapen. Till det som ändå finns hör *Litteraturbanken* (Inom ramen för Språkbanken i Göteborg) som är ett påbörjat projekt för att samla den litterära delen av det svenska kulturarvet. Här finns också projekt som *Almqvist*, *Strindberg* och *Lagerlöf* för att åstadkomma vetenskapliga utgåvor, utgivningsprojekt som

²⁴ Bifogad med tillstånd av dekanus Anna Sägval Hein.

²⁵ *Svensk språkteknologi – existerande forskningsinfrastruktur och framtida behov*. Database Infrastructure Committee (DISC), Vetenskapsrådet, 2007.

dock skulle behöva kopplas till generell databasuppbyggnad vad gäller editionsfilologi²⁶. Även om ovanstående inte ger en uttömmande beskrivning av de tillgängliga resurserna, är det uppenbart att bristen på databasresurser (se avsnittet ”Framtida behov”) är långt större än tillgången inom det litteraturvetenskapliga området.

Konst- och bildvetenskap

Sedan man i mitten av 1990-talet utvecklat tekniker för att samordna bildåtergivning med textdata om proveniens, tillkomstår, teknik etc²⁷, har digitala medier spelat stor roll inom konst- och bildvetenskap.

Inom de svenska konst- och bildvetenskapliga institutionerna finns emellertid inga bildbanker, trots mångåriga diskussioner om att bygga sådana som en gemensam tillgång för disciplinen. Planerna har fått läggas på is på grund av de olösliga problemen med upphovsmannarättsliga kostnader och BUS-lagstiftningens²⁸ krav. De bildbanker som utnyttjas, och som fyller ett stort behov, finns i stället vid arkiv, museer och gallerier i Sverige (t ex vid Nationalmuseum) och annorstädes (t ex vid National Gallery i London). De förekommer som länkar i de resursbanker som byggs upp vid universitetsinstitutioner världen över. Vid t ex institutionen i Göteborg finns under länken *Konsthistoriska resurser* vidarehänvisningar till bl a ’*Art History Resources on the Web*’. Forskarna bygger också för eget bruk upp egna bildförråd, hämtade från bilddatabanker, webbsajter, egen fotografering och egen inscanning av material.

Medie- och kommunikationsvetenskap

Forskningen här har som tidigare nämnts sina rötter inom såväl samhällsvetenskap som humaniora och inbegriper inte bara samtiden utan också historiska perspektiv. Tvärvetenskapligheten och de många perspektiven gör att många olika slag av databaser används, t ex lingvistiska databaser när offentligt tal och samtal studeras och databaser uppbyggda av historiker och samhällsvetare när fokus ligger på den politiska historien.

Till databaser och arkiv som är väsentliga för medieforskare nämns material vid Statens Ljud- och Bildarkiv (SLBA), KB:s tidningsenhet samt andra press- och bildarkiv, liksom de interna arkiven vid Sveriges Television (SVT) och Sveriges Radio (SR).

Etnologi

Databearbetning i form av kvantitativa metoder och klassificering som tidigare användes inom traditionell folkminnesforskning har idag hamnat i bakgrunden till förmån för mer kvalitativt analytiska angreppssätt och man bygger inte upp egna databaser.

Däremot utnyttjas tillgängliga samlingar och digitaliserade databaser från olika arkiv, främst från de olika folkminnessamlingarna inom SOFI (med sina avdelningar i Lund, Göteborg, Uppsala och Umeå). Det är fråga om texter (frågelistor, intervjuer, dagböcker) och bilder, bl a foton. Nordiska museet har också ett omfattande material, men också samlingar från andra museer och arkiv används i forskningen.

²⁶ Behandlar textvarianter och deras förhållanden till tryckta böcker.

²⁷ Se *Arbetsrapporter från Tema K, 1995:24, Bilddigitalisering vid bibliotek, arkiv och museer. En lägesrapport för Nationella Arkivdatarådet (NAD) och Humanistisk-Samhällsvetenskapliga Forskningsrådet (HSFR)* utarbetad under ledning av Lena Johannesson.

²⁸ BUS (Bildkonst Upphovsrätt i Sverige), se också avsnittet ”Generella problem som kräver en lösning”

Mycket av det material som finns i nämnda samlingar speglar tidigare förhållanden och har betydelse när man vill belysa gångna tider, t ex i relation till nutiden. Men inom etnologin idag, där det finns en stark trend att utforska företeelser i nuet och de snabba förändringar som sker i dagens samhälle, erbjuder Internet en viktig källa till information. Med etnologins intresse för människans olika kulturella uttrycksformer utgör Internet en gigantisk världsomspännande ”databas”, där etnologer kan bedriva fältarbete. Ett svenskt exempel här är en studie av hur programvara gemensamt utvecklas av olika individer på Internet. Begreppet ”databas” kommer i den kontexten att stå för något annat än vi normalt avser, dvs en strukturerad samling data, men har likväl betydelse för de företeelser som etnologin studerar.

Framtida utveckling inom etnologin inkluderar en mer utbredd samverkan och kommunikation inom forskningen genom IT. En del av detta arbete sker i samverkan med forskningsplattformar som t ex HUMLab och i andra samarbetsformer.

Historia – modern historia och medeltidshistoria

Det var först under 1970-talet som större databaser utvecklades inom historieämnet, som då hade en orientering mot samhällsvetenskaperna med inriktningar som demografisk och ekonomisk historia samt socialhistoria. Databaserna byggdes i första hand med olika slags folkbokföringsuppgifter och med bistånd från arkiven. Senare har bilddatabaser tillkommit genom engagemang från arkiv, museer och bibliotek.

Det finns ett antal demografiska databaser i varierande storlek i Sverige, *Demografiska databasen (DDB)* i Umeå, *Stockholms historiska databas (SHD)* och *Demografisk databas södra Sverige (DDSS)*, den senare med nära knytning till ekonomisk historia. *Svenskt arkivregister (SVAR)* och lokalhistoriska och medicinhistoriska databaser samt databaser över kulturmiljöer är andra källor som utnyttjas, liksom de historiska kartor som finns hos Lantmäteriet. Det historiska bildmaterialet finns samlat i bilddatabaser, åtkomliga på nätet. Ett stort antal dataset med många olika slags uppgifter (folkmängd sedan 1570-talet, yrkesfördelningar, näringslivsprofiler mm) om totalt 133 orter i Sverige ingår i *CyberCity* utvecklad vid Stads- och kommunhistoriska institutet. De nämnda historiska databaserna har en bred avnämning som inkluderar geografer, sociologer, arkitekter, medicinare, tekniker och naturvetare utöver historiker. Inom demografiforskningen finns vidare ett utvecklat internationellt samarbete och utbyte med olika centra, t ex IPUMS (Integrated Public Use Microdata Series, University of Minnesota, Minneapolis, USA) and NIDI (Netherlands Interdisciplinary Demographic Institute).

Databaser som utnyttjas i den tidigare-historiska forskningen finns vid Statens historiska museer, bl a *Medeltidens bildvärld*. Vid Riksarkivet finns bl a *Svenskt diplomatarium*, som används inom flera humanistiska ämnen (historia, historiskt inriktad språkvetenskap, litteraturvetenskap) och *Nationalutgåvan av äldre geometriska kartor*, där kartmaterialet bl läggs in i en GIS-databas. Detta och andra kartmaterial har under senare tid fått alltmer ökad och bred användning inom olika ämnen. Vid Stockholms universitets arkeologiska institution registreras vidare myntfynd i Sverige i flera databaser.

Nationell ArkivDatabas (NAD) vid Riksarkivet, som innehåller uppgifter om arkiv och deras förteckningar, är av stort värde som ingång till all historiskt inriktad forskning.

Arkeologi, inklusive miljöarkeologi

Nutida utgrävningar dokumenteras direkt digitalt i form av geografiska databaser och andra digitala dokument, men standarder för lagring mm saknas i stor utsträckning.

Källmaterialet från tidigare grävningar och inventeringar inom arkeologi hämtas ur olika digitala databaser och andra samlingar av fornlämningar i landskapet. *Fornminnesregistret* knutet till Riksantikvarieämbetet (som också har betydelse för den historiska forskningen) innehåller dokumentation om hällristningar runstenar mm. Register över fornlämningar finns vidare vid museer, men det saknas en nationell databas. *Antikvariskt topografiskt arkiv* (ännu inte digitaliserat) vid Vitterhetsakademien skulle, digitaliserat, kunna utgöra en länk mellan *Fornminnesregistret*, museernas samlingar och dokumentation från utgrävningar.

Kvantitativa data från naturvetenskapliga/miljöarkeologiska laboratorier – miljöarkeologiska analyser, C14-dateringar, kemiska analyser, osteologiska bestämningar – registreras också och lagras i databaser. Den arkeologiska/miljöarkeologiska forskningen utnyttjar geologiska och geografiska data inom ramen för Geografiska Informationssystem (GIS) och miljöarkeologi är sammantaget ett tvärvetenskapligt område med sin anknytning till dels andra humanistiska ämnen, dels miljö- och naturvetenskap. Ett exempel här är den miljöarkeologiska databasen *Strategic Environmental Archeology Database (SEAD)* som planeras i Umeå, som dels kommer att innehålla rådata, dels olika redskap, bl a GIS-baserade, för att analysera dessa data.

SEAD, som nu är under utveckling är tänkt att vara en internationellt tillgänglig databas. Det samma gäller *BugsCEP*, också utvecklad i Umeå, som innehåller data över fossila insekter i Europa. Dessa är endast två exempel på internationella databaser som används inom arkeologi tillsammans med databaser på nationell, regional, och individuell nivå.

Religionsvetenskap

Religionsvetenskapens har som det uttrycks i Bilaga 14²⁹ ”bildat bindestrecksvetenskap” med en hel rad ämnen med resultat att det finns specialiseringar som religionsdidaktik, -filosofi, -historia, -pedagogik, -sociologi mm. Vidare finns inom religionsvetenskapen en stark knytning till områden som utomeuropeiska språk och kulturer och antikforskning som till viss del också täcks in i den inlämnade sammanställningen. Denna ämnesbredd gör att variationen i de data som utnyttjas, liksom metoderna, varierar starkt.

En mängd religionsvetenskapliga forskningsfrågor kräver tillgång till stora internationella och väl uppbyggda databaser. Det kan gälla frågor om värderingar i nationellt och internationellt perspektiv. För sådana undersökningar och för andra med anknytning till samhällsvetenskap utnyttjas bl a Svensk Samhällsvetenskaplig Datatjänst (SSD) i Göteborg och liknande dokumentationscentra i andra länder, bl a Norsk Samfunnsvitenskapelig Datatjeneste i Bergen. Generellt gäller emellertid att befintliga digitala databaser används i måttlig omfattning – även om sammanställningen tar upp ett antal såväl svenska som internationella databaser inom olika specialområden. Databaser vid bl a arkiv används emellertid av religionsvetare som arbetar med historiska perspektiv och det finns också icke-närmare specificerade önskemål om digitalisering av material som förvaras vid arkiv, bibliotek och museer. Det gäller dels inscanning av material, dels överföring av material till länkade databaser.

²⁹ Sammanställningen inleds med en läsvärd historia om tidigare former av datainsamling och lagring, men med bäring på nutiden.

Det vanliga är att forskare bygger sina egna databaser ur olika källor. Det kan handla om bl a tidningsmaterial och bilder. En ökande mängd av undersökningar bygger på intervjuer, där materialet sammanställs men sedan stannar hos den enskilda forskaren eller forskargruppen. Ett skäl till detta är de etiska problem man har att handskas med då det gäller uppgifter om personer.

Nyttjande

Svar på frågeställningen: *Hur utnyttjas tillgängliga databasresurser? Av enskild forskare? Grupper?*

Svaren här varierar mellan de olika ämnesområdena. De stora och välutvecklade digitala, ofta internationella, korpusmaterialen inom språken har bred användning. I många fall har svenska forskare bidragit med delmängder (svenska) i fråga om flerspråkiga material. Korpusarna är i stor utsträckning tillgängliga via Internet. Det finns också svenska korpusar som nyttjas av breda användargrupper, t ex material i *Språkbanken*. Inom historia finns bl a de demografiska databaserna i denna kategori. Mycket av det material som finns vid arkiv, men också bibliotek och museer används på liknande sätt av många. Historiskt material vid bl a Riksantikvarieämbetet och Riksarkivet utnyttjas av bl a arkeologer, konstvetare och språkvetare med historisk inriktning, förutom av historiker. Etnologer och språkvetare nyttjar material vid Institutet för språk och folkminnen (SOFI). Det är alltså i huvudsak databaser och annat material inom ABM-sektorn som har bred användning inom den humanistiska forskningen, förutom de ovan nämnda språkliga material som är tillgängliga via *Språkbanken* eller som byggts upp i internationell samverkan och finns tillgängliga via bl a LDC (Linguistic Data Consortium).

På liknande sätt finns bl a när det gäller historia ett brett utnyttjande av databaser. *Stockholms historiska databas* utnyttjas bl a inom ramen för internationella nätverk inom stadshistoria och det finns ett rikt forskarutbyte också kring liknande databaser i andra länder. Tillgång till stora longitudinella databaser av hög kvalitet har varit avgörande för sådana projekt

Vissa databaser uppbyggda inom ramen för projekt har av naturliga skäl ett mer begränsat nyttjande, företrädesvis och åtminstone inledningsvis, inom den egna forskargruppen. Det gäller t ex inom språkvetenskapen. Men sådan databasuppbyggnad sker inte sällan inom stora grupper med representanter för flera universitet, vilket medverkar till ökad spridning av resurserna.

Databaser i större eller mindre skala byggs också av enskilda forskare, oftast för eget bruk. Det kan gälla utdrag och sammanställningar av material ur t ex arkiv – vanligt bland historiker och religionsvetare – eller större språkliga korpusar men också uppbyggnad av egna originaldata.

Erfarenhet av samverkan vid uppbyggnad av databaser

Svar på frågeställningen: *Finns erfarenhet av samverkan och forskningssamarbete gruppvis/nationellt kring uppbyggnad av nu existerande databaser och har det då även handlat om att samarbeta tvär- eller mångdisciplinellt?*

Sådan samverkan förekommer inom delar av humaniora – företrädesvis inom språkvetenskap. Här finns stora likheter med språkteknologin och den allmänna språkvetenskapen, där databaserna oftast tillkommer inom ramen för breda forskargrupper, involverande flera olika universitet. De språkvetenskapliga databaserna kan också sägas vara mångdisciplinära i den meningen att de utnyttjas över de gängse ämnesgränserna. Databaser uppbyggda inom språkteknologi kan utnyttjas av t ex forskare i svenska/nordiska språk och omvänt och på liknande sätt kan databaser med engelskt material nyttjas i språkteknologisk forskning.

Omfattande och bred samverkan kring uppbyggnad av databaser utmärker inte i övrigt humaniora – idag. Men det finns tendenser som pekar mot en framtida större samverkan som kan komma att kräva nya infrastrukturella resurser, en stark trend till tvärvetenskap och mångdisciplinärt samarbete såväl inom humaniora som mellan humaniora och andra vetenskaper. Som läget är idag utnyttjas då redan tidigare befintligt material och eller skapas de data som behövs för projektet utan byggande av databaser. Men ett exempel, där samverkan redan förekommer och fyller ett stort behov, är arbetet kring digitaliseringen av medeltidstexter inom *Menota (Medieval Nordic Text Archive)*. Intresset för att tillgängliggöra och skapa standarder för det medeltida materialet delas av såväl språk- och litteraturvetare som historiker som samverkar för att få detta till stånd.

Betydelsen av befintliga databaser vid arkiv, bibliotek och museer, mm

Svar på frågeställningen: *Vilken roll spelar befintliga databaser vid arkiv, bibliotek och museer, mm för forskningen inom ämnet?*

Betydelsen av denna typ av material för i stort sett alla ämnen inom humaniora kan inte överskattas. Det gäller många olika slag material – historiskt (t ex vid Riksantikvarieämbetet och Riksarkivet) såväl som samtida (tidningsmaterial ur pressarkiv, annat mediematerial (radio, TV) från Statens ljud- och bildarkiv (SLBA etc). Bild- och videomaterial från museer, andra konstinstitutioner och gallerier är väsentliga inom konst- och bildvetenskapen. Det gäller databaser och andra samlingar vid svenska institutioner (bl a Nationalmuseum), men i lika hög grad utländska, som t ex National Gallery i London, Metropolitan Museum i New York. Det nyttjade materialet inkluderar såväl digitala och välstrukturerade databaser som digitaliserat material i form av inscannade bilder av föremål och källtexter liksom icke-digitaliserat material.

Betydelsen av det äldre materialet framgår väl av medeltidshistorikern Olle Fermes sammanställning i Bilaga 12, som illustrativt visar hur materialet vid Riksantikvarieämbetet, Riksarkivet, Statens historiska museum och Stockholms universitet utnyttjas inom arkeologi, språk, konstvetenskap, historia och kulturgeografi. En stor del efterfrågas också av utländska forskare.

FRAMTIDA BEHOV

Digitalisering av material vid arkiv, bibliotek, museer, mm

Svar på frågeställningen: *Finns behov för forskningen av digitalisering av material som förvaras vid arkiv, bibliotek, museer, mm?*

Som nämnts betyder existerande digitala databaser vid arkiv, bibliotek, museer och andra myndigheter en mycket väsentlig roll inom humaniora. Men mycket av materialet inom dessa institutioner – och i synnerhet det äldre – föreligger i icke-digital form. Den historiskt inriktade forskaren är av tradition visserligen van vid att själv gå direkt till källorna, dvs ta sig till arkivet, begära fram materialet, skriva av eller kopiera och sedan återvända hem för att bearbeta och utnyttja materialet i sin forskning. Men det är en krånglig process som dessutom är kostsam (kan innebära dyra rese- och vistelsekostnader). Det riktigt gamla materialet är dessutom mycket ömtåligt och även en del som är skrivet på papper från 1800-talet är på väg att vittra sönder. Forskningen av idag om vår historia i bred bemärkelse – vårt kulturarv – kräver således av många skäl digital åtkomst till materialet. Efter digitalisering i form av inscanning av texter och andra handlingar, bilder, kartor och föremål har man tillgång till det åtråvärda originalet och med registreringar och bearbetningar av olika uppgifter öppnas de samlingar man annars inte har tillgång till. Behovet finns naturligtvis inom historia, men också inom språkvetenskap, i synnerhet den historiskt inriktade, litteraturvetenskap, konstvetenskap, och vissa delar av religionsvetenskapen samt etnologin. Mycket av materialet som behöver digitaliseras har också intresse inte bara inom en ett enskilt ämnesområde utan har ofta mångvetenskaplig betydelse.

En hel del äldre språkligt/ material – det gäller framförallt dialekter och person- och ortnamn – är ännu inte digitaliserat, även om större digitaliseringsprojekt pågår, t ex inom SOFI. Det är frågan om mycket betydelsefullt material för forskningen som alltså kräver omfattande arbetsinsatser. Äldre textsamlingar som *Fornsvenska textbanken* behöver också rustas upp. *Menota*, som kan ses som ett övergripande paraplyprojekt och som bl a syftar till standardisering när det gäller hur materialet ska föras ut till användarna, kräver likaledes omfattande utvecklingsresurser. *Menota* har inte bara betydelse för språkvetenskapen utan har också starka intressenter inom såväl litteraturvetenskap som historia. En ytterligare styrka är den internationella referensramen. Här finns behov av finansiering för att skapa permanenta strukturer för att upprätthålla användbarheten och att utöka verksamheten med nya projekt.

Tillgång till det litterära kulturarvet kräver omfattande digitalisering, bl a inscanning av original- och andra utgåvor, som dock måste kombineras med rik dokumentation om utgivningen och om verken. *Litteraturbanken* (inom ramen för Språkbanken i Göteborg) är ett exempel på ett sådant digitaliseringsprojekt. Det drivs nu som en ideell förening med finansiering från Svenska Akademien³⁰, men resurserna är högst otillräckliga.

Inom konstvetenskap finns stor enighet kring behovet av digitalisering av den del av kulturarvet som svenska konstsamlingar utgör. Det gäller samlingarna vid Nationalmuseum (som borde uppdateras och utvidgas vad gäller digitaliseringen) samt vid Göteborgs konstmuseum, Norrköpings konstmuseum och Malmö museum. Det behövs för att nå ut i forskarsamhället och för att göra det nationella kulturarvet synligt i internationella översikter. Ritningar vid Arkitekturmuseum och teckningar, ritningar, kartor, brev och excerpter ur Kungliga veten-

³⁰ *Litteraturbanken* finansierades tidigare (2004-05) av Riksbankens Jubileumsfond.

skapsakademiens samlingar samt Carolina Redivivas och KB:s omfattande bild- och handskriftssamlingar borde också digitaliseras.

För konst- och bildvetenskapen är det dessutom oerhört väsentligt att rädda de omfattande diabildsarkiv som byggs upp vid institutionerna i Sverige. Det gäller bilder som samlats in under lång tid, totalt över 200 000 bilder, och där det finns en nationell samling kring behovet att säkerställa dessa resurser för forskning och undervisning framtiden.

De stora investeringar som gjorts i de befintliga demografiska databaserna och deras värde för den historiska forskningen talar för en fortsatt uppbyggnad med tidigare material som grund, t ex 1800-talets kyrkböcker datoriserade för hela Sverige. Framtida satsningar beror också på de forskningsuppgifter man vill ta sig an. Här ser man idag en trend mot större tvärvetenskaplighet där befolkningsdata får betydelse i relation till andra data, epidemiologiska etc, i t ex studier av hälsa i övergången mellan olika historiska tidsskeden. Men det finns också ett behov av att arbeta mer "kvalitativt" med utgångspunkt i kulturarvsdatabaser, bilddatabaser och liknande. För framtiden behövs också bättre samordning av pågående registreringar, inte minst vad gäller de demografiska databaserna, men också samordning mot lokalhistoriska databaser och SVAR.

De historiker som koncentrerar sig på äldre tider (medeltid och tidig modern tid) vill se mer källor på nätet. När man bygger sina egna databaser – vilket är vanligt inom historia – är man helt beroende av de databaser eller andra källmaterial som finns vid Riksantikvarieämbetet, Riksarkivet och andra arkiv- och museiinstitutioner. Det materialet är helt oundgängligt för medeltidsforskningen inom historia - liksom inom konstvetenskap, arkeologi och historiskt orienterad språkvetenskap och religionsvetenskap.

Ett problem här är att flera ambitiösa och angelägna projekt inte kunnat fullföljas pga medelsbrist. Till exempel är inte inskrifterna i den *Samnordiska rundatabasen* i Uppsala (knuten till Runverket) översatta till modernt språk, något som gör det svårt för icke språkvetare (som historiker och arkeologer) att utnyttja informationen. *Det medeltida Sverige*, en topografisk historisk handbok, anger grunddata för bebyggelsen i medeltidens Sverige initierades av Vitterhetsakademien.

Inom arkeologi behövs digitalisering av äldre arkivmaterial, liksom av miljöarkeologiska undersökningar.

Etnologer beskrivs inte som storanvändare av digitala forskningsdatabaser, men man är i hög grad beroende av det källmaterial som finns vid arkiv, bl a SOFI, Folklivsarkivet i Lund, samt vid Nordiska museet. Ökad digitalisering antas vidare ha en forskningsinitierande effekt. Det gäller för övrigt inte bara etnologi, utan kan sägas gälla mer allmänt inom humaniora.

Olika ämnesområden – likartade behov - och likartad finansiell problematik

Det är alldeles uppenbart att vissa av de redovisade behoven är gemensamma för olika ämnen. Hit hör det flera gånger nämnda *Menota* som är av stor betydelse inom såväl, språk- litteratur- som historievetenskaperna. Den *Samnordiska rundatabasen* är betydelsefull för historiker och arkeologer utöver de historiskt inriktade språkforskarna. Mycket av materialet vid Riksarkivet och Riksantikvarieämbetet, t ex *Svenskt diplomatarium* ger underlag för studier inom historia, namnforskning mm. Samlingarna av bl a namn vid SOFI är ytterligare ett exempel, liksom *Det medeltida Sverige* (av intresse för såväl humaniora som samhällsvetenskap).

Flera av dessa projekt av vilka de flesta pågått under lång tid saknar idag de medel som behövs för att de ska kunna fullföljas. *Det medeltida Sverige* är ett exempel. Det stöddes från början ekonomiskt av HSFR och medlen för projektet överfördes senare genom ett särskilt riksdagsbeslut till Riksantikvarieämbetet inom ramen för det totalanslag som gavs till myndigheten. I dagens läge med många konkurrerande projekt inom verksamheten, räcker inte medlen till att driva *Det medeltida Sverige* annat än på sparlåga. Projektet bygger på ett internationellt sett unikt källmaterial och borde därför enligt historieforskarna fullföljas, då det har relevans för många ämnen också vid sidan om historia (kulturgeografi, konst, bebyggelse- och socialhistoria). Men situationen är alltså inte unik för detta projekt. Det gäller många likartade verksamheter där materialet måste dokumenteras med stor noggrann- och idoghet under vetenskapligt tillfredsställande former och där slutfasen därför måste ses i långtidsperspektiv. Många större ordboksprojekt hör t ex hit.

Nya databaser

Svar på frågeställningen: *Vilka framtida behov finns i övrigt vad gäller databasuppbyggnad för att upprätthålla hög kvalitet och bidra till att forskningen inom ämnet kan vidareutvecklas?*

Språkvetenskap

Till de framtida behoven inom svenska/nordiska språk hör talspråksdata, inte minst offentligt talad svenska (i form av text och tal lagrade parallellt). Annat som behövs är korpusar med multimodalt skriftspråk (text, bild, grafik) för att fånga dagens textvärld både när det gäller text i tidningar och böcker och den text som förekommer inom datormedierade genrer. E-mejl, bloggar och hemsidor innebär nya sätt att kommunicera där ord, bild grafik, ljud och skrift samverkar på ett sätt som inte var möjligt tidigare.

Behovet av tal- och textdata som reflekterar 2000-talets språk (med genrevariation) går igen också när det gäller engelska och romanska språk (och man kan förmoda att det också gäller för många av de språk som inte täcks i kartläggningen). Också databaser som täcker språkets förändring genom århundradena och geografiska variationer nämns liksom inlärdatabaser för andraspråksforskningen. Till detta kommer behov av tekniska resurser för komplettering av taggning och lagring av befintligt digitalt material och för nya databaser. Nya och bättre verktyg behövs för annotering och sökning och utveckling av gemensamma standarder.

Behovet av stora talspråksmaterial är i stora delar likartat inom språkteknologin, allmänlingvistik och den språkvetenskap som bedrivs inom de här beskrivna och andra skärspråk.

Litteraturvetenskap

Litteraturvetenskapen är generellt sett i behov av att utveckla nya databaser (och att utöka påbörjad digitalisering, se föregående avsnitt). Vissa av dessa behov skulle kunna täckas av myndigheter utanför universiteten, främst Kungliga Biblioteket (KB). Det gäller t ex bibliografiska databaser, som gör det möjligt att söka och finna forskning om svensk/skandinavisk litteratur, något som idag saknas efter nedläggningen av *Svensk litteraturhistorisk bibliografi*. Vältäckande bibliografier är för övrigt ett behov inom flera av områdena i kartläggningen. Det utgör ett stort hinder för forskningen, eftersom forskningsresultaten vidareförs främst genom böcker och informationen idag är svår att få ut. Det saknas också databaser som tillgängliggör

den forskning som förekommer. Kulturartiklar tillgängliggörs genom *Presstext*, men ett textarkiv för svensk litteraturvetenskaplig forskning saknas.

Det behövs också utökade resurser för databaser (som *Libris*) som gör det möjligt att finna källtexter, dvs katalogisering av manuskript och andra material. Dessutom behövs databaser som tillgängliggör sådana källtexter. Till det som ändå finns hör *Litteraturbanken*.

Databaser som kan utveckla och utvecklas av editionsfilologin – för att åstadkomma vetenskapliga utgåvor – behöver också byggas. I våra grannländer finns stora databaser som stöd för större utgivningsprojekt, bl a utgivningen av *Henrik Ibsens Samlade Skrifter* i Oslo, medan Sverige saknar liknande resurser. Den närmaste motsvarigheten inom editionsfilologin i Sverige är det samnordiska *Menota* med sin inriktning på nordiska medeltidstexter och att utveckla en standard för uppmärkning av texterna (se också Språkvetenskap, ovan). *Menota* skulle kunna vara grunden också för utgivningen av svensk litteratur i allmänhet, inte minst med koppling till pågående utgivningsprojekt som *Almqvist*, *Strindberg* och *Lagerlöf*.

Ytterligare ett behov – som litteraturvetenskapen långt ifrån är ensam om – är databaser som erbjuder en arbetsmiljö med utnyttjande av Internet, en arbetsmiljö som uppmuntrar till gränsöverskridande och samverkan. (Det gäller också bl a etnologi.) Ett av flera exempel inom litteraturvetenskapen är *The Rosetti Archive*³¹. *COST*, Action A32, *Open Scholarly Communities on the Web* är ett initiativ för att skapa just sådan samverkan. Överhuvudtaget gäller att Sverige bör vara med när *COST* utvecklar digitala forskningsgemenskaper och Litteraturbanken byggs också upp i fortlöpande kontakt med *COST*.

Litteraturvetenskapens databaser har långt vidare användning än inom det egna forskningsområdet. Litteraturen med dess gestaltning av människor och miljöer i samtiden eller i historien har betydelse också för historiker, inklusive idéhistoriker, inom kultur- och medievetenskap, religionsvetenskap mm. Det litterära kulturarvet i form av texter från olika tider i vår historia utgör således ett centralt forskningsmaterial inom humaniora.

Konst- och bildvetenskap

Ett behov som är gemensamt för samtliga konst- och bildvetenskapliga institutioner i Sverige är att ta tillvara de omfattande samlingar av främst diabilder, men också storbildsdiagram och planscher, som byggts upp vid varje institution och som skulle behöva digitaliseras för användning i forskning och undervisning. Sammantaget uppgår diabilderna i samlingarna till långt över 300 000, även om man bara räknar bilderna vid de större universiteten.

För utvecklingen inom området finns ett intresse av att skapa digitala resurser för studier av nya konstformer som performances, video, film, dans och multimediala konstrepresentationer och likaledes ett intresse för samverkan med vetenskaper som t ex medicin och meteorologi kring instrumentella bilder.

Medie- och kommunikationsvetenskap

Inom den språkvetenskapligt inriktade medieforskningen behövs databaser med transkriberade mediesamtal som t ex intervjuer. Det är ett behov som delas med den allmänna språk-

³¹ <http://www.rosettiarchive.org>

vetenskapen och språkteknologin, som ser det offentliga talet som en genre av talspråk där mycket stora datasamlingar behöver tillskapas (se Språkvetenskap, ovan.)

Inom medieforskningen efterlyses också ökad satsning på press- och bildarkiv, som nu bara är delvis användbara i forskningen. T ex behövs material som går längre tillbaka i tiden, något som gäller bl a *Pressklipp* och *Mediearkivet*.

Det man i övrigt i synnerhet vill värna om och utveckla är det nordiska informationscentret *Nordicom* som ses en ytterst värdefull resurs. Verksamheten tjäna flera syften; den ger nödvändig information för de direkt berörda forskarna inom Norden, men bidrar också till internationell synlighet genom olika publiceringssatsningar. Den engelskspråkiga *Nordicom Review* är högt rankad internationellt och går ut till 125 länder utanför Norden.

Centret, som är en institution inom Nordiska Ministerrådet, inrättades för drygt 30 år sedan och har sitt säte i Göteborg. Verksamheten bygger på kontrakt och överenskommelser med, utöver Nordiska Ministerrådet, UNESCO, Kulturdepartementet och mer tillfälliga uppdragsgivare. Verksamheten är uppbyggd omkring tre huvudsakliga arbetsområden: dokumentation av medieforskningen dess resultat i Sverige och de nordiska länderna, framtagandet och förmedlingen av kunskap om medieutvecklingen inom Norden samt forskning om barn, unga och medier i världen. Trots den lyckade satsningen är finansieringen ett problem, ett problem som blir än större i och med att det också finns ett behov av att utveckla tjänsterna.

Liknande finansieringsproblem och farhågor inför framtiden framhålls också vad gäller upprätthållande och utveckling av SLBA.

Etnologi

Databaser inom etnologin, som den idag studeras, har som nämnts ingen stark ställning. Trots att man kan se fördelar med utnyttjandet av databaser finns det heller ingen uttalad ambition att utveckla sådana. Men man ser Internet som en viktig resurs och det finns också idéer om och behov av att utnyttja program som möjliggör kvalitativa analyser av data i form av intervjuer liksom diskussions- och chattmaterial.

Historia – modern historia och medeltidshistoria

Av den rundfrågning som gjorts bland svenska historiker framgår (Se Bilaga 10 och 11) – kanske inte oväntat – att man inför framtiden dels vill göra nysatsningar, dels satsa på vidareutveckling av de existerande databaserna, och främst då de demografiska. Man skulle gärna se hela Sveriges kyrkböcker från 1800-talet i digitala databaser. En möjlig nysatsning skulle kunna vara en stor databas med dataset över det nordiska och europeiska stadsväsendet. En sådan databas uppbyggd i samverkan mellan forskare inom Europa skulle kunna vara en satsning med stor potential.

Arkeologi, inklusive miljöarkeologi

Utöver det tidigare nämnda behovet av en nationell databas över arkeologiska undersökningar, finns andra högt prioriterade mål, bl a en nationell C14-databas. Det finns vidare enligt uppgiftslämnarna ett starkt behov av en nationell handlingsplan för arkeologiska databaser, eftersom arkeologin i så hög grad är avhängig av sådana.

Till det som behövs hör också en policy för lagring och tillgång till data från nutida utgrävningar. I dagens läge dokumenteras de digitalt, men bristen på standardisering vad gäller lagring och problem med tillgängligheten kräver en lösning. Till övriga teknikutvecklingsbehov hör användbara gränssnitt (GIS-baserade) och tillgänglighet till all form av data via Internet. För att utveckla den tvärvetenskapliga orienteringen med samkörning och länkning av olika databaser krävs samarbete mellan forskarna som producerar data och mjukvaruutvecklare.

Religionsvetenskap

Allt fler forskningsuppgifter bygger idag på uppgifter hämtade via Internet, bl a från hemsidor. Här är det faktum att informationen på nätet i stor utsträckning är "färskvara" ett problem. Det skulle för vissa syften behövas en databas med dag-för-dag-variationer, dels för att underlätta kontroll, dels för diakrona studier av information på nätet. Man skulle då kunna spåra hur olika företeelser, t ex religiöst inspirerat våld, lutar sig mot olika organisationers hemsidor etc.

Religionsvetenskapens utforskande av tendenser i det moderna samhället skulle också vara betjänt av data från radio och TV. Tillgång till den information som t ex via parabol kommuniceras till minoriteter i det svenska samhället skulle kunna belysa den religiösa och kulturella verklighet som många av våra invandrare lever i. En databas som på stickprovsbasis lagrar sådant material skulle vara en stor tillgång.

Det är således material för studier som rör dags- och gårdagsaktuella frågor som utgör det största behovet inom religionsvetenskapen. Vad gäller urkunder och annat historiskt material finns i större utsträckning tillfredsställande lösningar.

Olika ämnesområden – likartade behov

Genomgången ovan visar att det finns vissa typer av databaser som skulle kunna tillfredsställa behov som är gemensamma för flera olika ämnesområden. Hit hör databaser med samtidsspråk – text och tal – och speciellt talare i interaktion, dvs samtalsdata samt korpusar som dessutom inkluderar bilder och grafik. I detta behov förenas såväl språkvetare som medie- och kommunikationsforskare.

Ett annat gemensamt intresseområde är Internet och dess utforskande ur olika perspektiv. Kombinationen av text, bild, grafik och tal är av högsta intresse för medie- och kommunikationsforskaren, liksom för många språkvetare. Etnologerna vill likaledes analysera olika företeelser via nätet. Religionsvetarna ser nätet som en möjlighet att studera värderingar, inställningar till våld etc. Hit hör sannolikt också idéhistoriker. Det gemensamma behovet som det framkommer i kartläggningen är strukturerade samlingar av stickprov (dag-för-dag) över viss tid som utgångspunkt för vidare analys.

Slutligen hör databaser över databaser något som efterlyses, förutom inom medie- och kommunikationsvetenskap, där dokumentationen som sker inom *Nordicom* fyller denna funktion.

UTVECKLING OCH NYTÄNKANDE

Vetenskapligt intressanta utvecklingslinjer

Svar på frågeställningen: *Finns vetenskapligt intressanta utvecklingslinjer och förnyelse-mönster som inte skulle kunna förverkligas utan tillgång till nya databasresurser?*

Samtidsspråk/samtalsdata samt samtidsföreteelser reflekterade genom Internet (se ovan) är två av de områden där uppbyggnad av databasresurser med största sannolikhet skulle bidra till nyskapande forskning.

Nutidens språk är ett gemensamt intresse inom såväl svenska/nordiska språk som andra språk (engelska, romanska språk mm) samt inom medie- och kommunikationsvetenskap. Vidare kräver den didaktiskt inriktade språkforskningen nya inlärar databaser, där vi med underlag i forskningen kan förvänta t ex utveckling av språkinlärningsmetodik och nya IT-tekniska hjälpmedel i språkträningen.

Andra nyare media, förutom Internet, ses också som intressanta källor till vetenskapligt intressanta studier – e-post bland språkvetare t ex – liksom representationsformer som video-installationer, performances etc inom konst- och bildvetenskap, där det också finns ett intresse för instrumentella bilder, bilder i anslutning till t ex medicinska och tekniska undersökningar.

Kartor och kartor länkade till annan dokumentation är en annan typ av data som fyller ett tvärvetenskapligt behov inte bara inom humaniora utan också i tvärvetenskaplig samverkan mellan humaniora och andra vetenskapsområden, t ex naturgeografi.

Det finns också ett intresse – bl a inom etnologi – att experimentera med nya gestaltungsformer. Konceptet här inkluderar databaser men är vidare till sin karaktär. Följande är ett utdrag ur Magnus Bergquists sammanställning (Bilaga 9):

Från Göteborg rapporteras om internationella projekt som organiseras via elektroniska kommunikationsplattformar. I Lund finns HEX - humanistisk experiment-grupp (www.hex.lu.se) där humanister, med etnologen Robert Willim som koordinator, utvecklar nya former för att bedriva tvärvetenskapliga studier och experimentera med nya gestaltungsformer för humanistisk forskning. Etnologen Charlotte Hagström deltar i delprojektet *Historia på plats* som handlar om platser - symboliska, virtuella och reella. Ett av de mer uppmärksammade HEX-projekten är *Surreal Scania* (www.surrealscania.se) som är ett samarbete mellan Willim och filmaren Anders Weberg. Digitala videofilmer kombineras med geografisk positioneringsteknik (GPS) och Google Earth för att skapa nya upplevelser av platser, samtidigt som dessa problematiseras med hjälp av tankar om platsmarknadsföring och branding i förening med konstnärlig gestaltning. Projektet har uppmärksamats på en rad internationella seminarier och festivaler.

Samverkan för att utveckla nya databaser med bred användning

Svar på frågeställningen: *Vilka tänkbara möjligheter finns att samverka i forskargrupper eller i andra konstellationer nationellt och/eller internationellt för att bygga upp och utveckla nya databaser med bred användning i forskningen?*

Forskare inom humaniora har traditionellt sett arbetat individuellt inom enskilda projekt och med egna data. Det gäller inte rakt av för all verksamhet – inom t ex språkforskningen finns sedan länge samverkan kring uppbyggnad av databaser och samverkan och gemensamt nyttjande av databaser sker inom den historisk-demografiska forskningen. Men det har ändå varit en tydlig tendens. Detta är nu på väg att ändras till förmån för forskning som involverar grupper, ofta med tvärvetenskaplig inriktning. Det har också lett till viss ökad samverkan kring data och ett ökande behov av ny databasuppbyggnad, standardisering av tidigare insamlade data och samling kring olika digitaliseringsbehov. Ett bra exempel är *Menota*, inom vars ram det uppenbarligen finns ett starkt intresse från många områden och där arbetet redan är i gång, om än utan tillräcklig finansiering. Ytterligare exempel finns inom konstvetenskap, där man har intresse av samarbete med ”de instrumentella bildpraktikerna inom naturvetenskap (t ex meteorologi) och medicin. Lena Johannesson (Bilaga 7) skriver också:

Inom arkitekturforskningen skulle större databaser vara av vital betydelse, då större cluster av satellitdata, demografiska, geologiska och arkeologiska data etc. exempelvis kunde samköras med långa serier av stadsplanedata, ritningsprojektioner, modeller etc. (Ett liknande projekt utvecklas nu i Rom vid Bibliotheca Herziana, där 1700-talsarkitekten Juvaras samtliga verk lagras i digital form för att kunna underlätta restaureringsarkitekternas och konservatoreernas analyser av hans byggnader.)

Trots denna pågående förändring till samverkan över ämnesgränser är det inte alltid uppenbart för direkt berörda att det ibland finns likartade behov inom olika ämnesområden. Tydliga exempel på detta är samtidsspråkliga text- och taldata, inklusive multimodala data, som efterfrågas från flera håll. Lagring i strukturerad och sökbar form av information på Internet är en annan typ av data som kan förväntas få bred användning om satsningar görs för att skapa sådana resurser.

Något som också framhålls är att databaserna, redan i det skick de föreligger, kan verka forskningsfrämjande genom att de blir kända och görs tillgängliga för andra än de som varit direkt involverade i deras utveckling. Ett exempel är hur historiska data idag i allt större utsträckning kommer till användning inom bl a epidemiologi och genetik, ett annat hur miljöarkeologiska databaser kan utnyttjas brett inom andra verksamhetsområden. Sådant tvärvetenskapligt utnyttjande kan förväntas öka i framtiden i den mån dokumentation och metadata om databaser utvecklas. Det kan sannolikt också skapa behov av nya för den tvärvetenskapliga forskningen betydelsefulla databaser.

Kartläggningen visar således på den potential för samverkan som finns och insikten om de gemensamma behoven bör kunna tas till utgångspunkt för initiativ för samverkan.

Databaser och synlighet i internationellt perspektiv

Svar på frågeställningen: *Kan tillgång till och användning av databaser påverka forskningens synlighet och närvaro i ett internationellt perspektiv?*

En genomgående tendens i svaren är att man ser tillgången till välstrukturerade och sökbara databaser som en viktig faktor när det gäller att göra svensk forskning mer nationellt och internationellt synbar och för att öka samarbetet mellan forskare såväl i Sverige som i världen i övrigt. Så har också skett på en del områden, inte minst när det gäller en del befolkningshistoriska och språkvetenskapliga databaser. Även inom språk och historia finns dock brister

– ett exempel är behovet av digitalisering av namnforskningsmaterial som en del i uppbyggnaden av en större europeisk namndatabas inom ramen för projektet ”EuroGeonames” – men andra humanioraområden är än mer eftersatta.

Det finns intresse i omvärlden för svenska data och svensk forskning. De svenska arkiven, biblioteken och museerna drar idag till sig uppmärksamhet från utländska forskare, en uppmärksamhet som kan väntas öka ju mer av materialet som finns digitaliserat och tillgängligt på Internet. Det gäller alla ämnen och ifråga om bilder gäller detta tveklöst – de är för kapacitetskrävande för utbyte forskare emellan på mer traditionellt vis. Brister här resulterar i internationellt osynliggörande av den nationella konsten och hämmar den konst- och bildvetenskapliga forskningen.

GENERELLA PROBLEM SOM KRÄVER EN LÖSNING

Tillgänglighet

Upphovsrättsliga begränsningar

Upphovsrättsliga begränsningar lägger hämsko på forskningen och tillgången på material. Det gäller språkvetenskap, svenska såväl som övriga språk, liksom bl a litteraturvetenskap. Att få rättighet att inkludera upphovsrättsligt skyddat textmaterial i en databas är inte bara svårt och arbetskrävande utan medför också mycket stora kostnader. Liknande begränsningar finns inom konst- och bildvetenskap där de sk BUS-reglerna³² gäller. Hit hör också frågan om digital access. Det går att få tillgång – via vanlig postgång – till digitalt material från Statens Ljud- och Bildarkiv, t ex vissa TV-program, men det får inte sändas över nätet. Det hindrar idag bl a de stora TV-projekt som man vill starta inom medieforskningen. Sammantaget utgör dessa begränsningar ett allvarligt hinder i utvecklingen av den humanistiska forskningen.

Etiska hinder

Många material är insamlade med medgivande från uppgiftslämnarna att använda data för en viss uppgift, något som är långt ifrån ovanligt och hindrar utnyttjande i annan forskning. I vissa fall, när det gäller speciellt känsligt material kan detta vara befogat (och här finns paralleller till Personuppgiftslagen (PUL)), men i andra fall skulle materialet sannolikt kunna utnyttjas också av andra än de som ursprungligen samlade in uppgifterna, om bara tillstånd givits för ett bredare utnyttjande.

Forskningsinterna hinder

Revirtänkande är inte sällan ett hinder. Forskarna som byggt databasen vill själv utnyttja den. Här behövs ett system – i varje fall när allmänna medel erhållits – för att efter viss tid göra materialet tillgängligt även för andra.

Kostnadsaspekter

Tillgängligheten begränsas också av kostnadspålagor. I många fall regleras tillgången genom avgiftssystem. Det gäller inom historia, där bearbetade data ur t ex *Demografiska databasen* i Umeå liksom kartmaterial från Lantmäteriet kan vara mycket kostsamma att få fram. Processen när det gäller att få tillgång till vissa data för medieforskning från Kungliga Biblioteket beskrivs vidare så här av Ester Pollack (Bilaga 8):

... stora forskningsprojekt som ska samla mycket stora mängder microfilmstexter och där vi själva har behov av att kopiera dessa på papper, arbetar under samma villkor som exempelvis privata släktforskare. Kopieringskort ska betalas kontant, dessa kort får endast laddas med mindre summor åt gången, man måste sätta upp sig i fullteknade köer till kopieringsapparater, de fungerande maskinerna är få och så vidare. Det gör att själva insamlingsprocessen är onödigt komplicerad och långsam.

Andra kostnader uppkommer genom arkivresor, när materialet inte finns tillgängligt på nätet (eller när det är andra skäl är nödvändigt att hämta material direkt ur källorna.

³² Bildkonst Upphovsrätt i Sverige (BUS)

Teknik, metoder och teknisk samordning

Teknikutveckling, dokumentation, standardisering, och verktyg

Många frågor som har med teknik att göra har också effekter på tillgängligheten till data. Den snabba utvecklingen på dataområdet gör att tidigare insamlat material idag kan vara mer eller mindre oanvändbart. Systematik och databearbetning har gjorts med den teknik som var rådande när databasen skapades och databasen sjunker i värde för användarna vartefter tiden går. Anpassning till dagens teknik krävs därför när det gäller många av humanioras databaser.

Otillräcklig dokumentation är ett annat – och stort – problem, liksom utnyttjande av olika standarder för liknande typer av material. Här finns idag en stor medvetenhet om betydelsen av procedurer som ökar användbarheten av insamlade data för breda grupper. Och i vissa projekt, t ex *Menota* är just standardiseringen ett centralt mål för verksamheten. Samarbetet inom EU på flera områden, t ex medverkan i COST (bl a litteraturvetenskap) syftar också till samverkan kring liknande frågor. Trots detta är dokumentation och standarder ett problem inom stora delar av humaniora.

Verktyg av olika slag, bl a för annotering av material, efterlyses, liksom sökverktyg och inte minst användarvänliga gränssnitt.

Samverkan och samordning av insatser

Behovet av samordning framhålls – för bättre överblick av den forskning som pågår liksom för att undvika kostsamma dubbelarbeten. Detta är ett uppenbart försummat område inom humaniora, även om medie- och kommunikationsforskningen och *Nordicom* utgör ett undantag. Behovet av bättre samordning av pågående registreringar tydliggörs, t ex när det gäller demografiska data. Men detta samordningsbehov gäller också andra områden, vilket tydligt uttrycks i följande utdrag ur Lars-Erik Edlunds sammanställning i Bilaga 3):

En nationell strategi bör skapas för att möjliggöra att digitaliserat material, som äger intresse för en större grupp forskare eller för allmänheten, underhålls också efter projektperioden och att adekvata, uppdaterade användargränssnitt skapas för dessa databaser. Här blir Svensk Nationell Datatjänst (SND)³³ självfallet en strategisk resurs.

³³ SND ska ersätta tidigare Svensk Samhällsvetenskaplig Datatjänst (SSD) och få utökade resurser.

SLUTSATSER OCH FÖRSLAG TILL ÅTGÄRDER

En kartläggning med det begränsade uppdrag som det här varit fråga om kan inte ge en heltäckande bild av humaniora och situationen i fråga om infrastrukturresurser. Svårigheten att överblicka ett så stort och diversifierat område och den begränsade tid som stått till förfogande gör det nödvändigt att inte dra alltför vittgående slutsatser. Förhoppningsvis kan kartläggningen ändå sägas ge en rimligt rättvisande bild utifrån det underlag som funnits och här redovisas de slutsatser som kunnat dras ur materialet.

Databaser i digital form har tidigare varit relativt sällsynta inom humaniora, även om forskningen inom speciellt språk och historia kan sägas utgöra undantag. Detta synes nu vara på väg att ändras. Kartläggningen ger ett tydligt vittnesbörd om betydelsen av digitala data – existerande databaser, digitalisering av ännu inte digitala samlingar samt uppbyggnad av nya databaser – liksom behovet av att sådana data tillgängliggörs över Internet.

Humanistisk forskning befinner sig i ett utvecklingsskede, där samverkan såväl inom humaniora som med andra discipliner får allt större utrymme. Även om humaniora i många avseenden också tidigare varit en del av humanvetenskapen med analyser av olika slag av samtidsföreteelser, så har detta nu blivit än tydligare. Det historiska perspektivet bidrar till belysning av den samhällliga och kulturella kontext – i vidaste bemärkelse – som vi är en del av. För att gå vidare behövs stora mängder data, inte minst data som finns vid arkiv, bibliotek och museer och som idag föreligger som icke-digitaliserade samlingar. Förändringen speglar också utvecklingen inom informationsteknologin. Det är lätt att se den potential som finns i tillgängliggörandet av stora mängder data som kan delas mellan många och Internet ses inte bara som ett redskap för att göra forskningen tillgänglig utan också som en datakälla i sig.

De slutsatser som i övrigt kan dras av kartläggningen kan sammanfattas i följande punkter:

- De befintliga resurserna i form av digitala databaser är ojämnt fördelade över de olika ämnena. Här finns etnologi och religionsvetenskap, där databaser spelar en undanskymd roll, andra, som litteraturvetenskap, konstvetenskap, medie- och kommunikationsforskning och arkeologi där forskningen i stor utsträckning är *beroende* av sådana resurser och så språkvetenskap och historia (speciellt modern historia) där forskningen i stor utsträckning *kräver* databasmaterial. Tyngdpunkten inom språk och historia är därmed densamma som i rapporten från 2005 om forskningens infrastruktur³⁴, men bilden blir mer komplex när man som i den här aktuella kartläggningen också inkluderar behoven för framtiden.
- Storleken på materialen varierar inom vida gränser, från stora datasamlingar, t ex större demografiska databaser och databaser som *Språkbanken*, som har många användare, till långt mindre samlingar uppbyggda av enskilda forskare och utnyttjade huvudsakligen av dem själva.
- En del av databaserna är utvecklade inom landet – det gäller i de flesta fall och generellt över ämnena – men databaser uppbyggda på andra håll i världen eller i samarbete med svenska forskare utnyttjas också. Det senare gäller speciellt språkdatabaser.
- I stort sett alla ämnen i kartläggningen utnyttjar – och är beroende – av material som finns vid arkiv, bibliotek och museer, främst vid de större statliga institutionerna men också vid bl a kommunala myndigheter och andra organisationer.
- En del av dessa samlingar föreligger som välstrukturerade och sökbara databaser, men en mycket stor del utgörs av icke-digitalt material som behövs – i digital form – för forskningen. Hit hör speciellt äldre språkligt material, framförallt dialekter och person- och ortnamn, samt

³⁴ ”Om forskningens infrastruktur inom humaniora och samhällsvetenskap i Sverige”, se inledningen

folkminnesmaterial. Behoven av digitalisering är dock långt större än vad som hittills åstadkommit, t ex krävs omfattande digitalisering av andra delar av kulturarvet, bl a inscanning av litterära originalverk, och digitalisering av svenska konstsamlingar samt de omfattande bild- och handskriftssamlingar som finns vid bl a Kungliga Biblioteket. Teckningar, ritningar, kartor vid arkitekturmuseum behöver digitaliseras samt äldre arkivmaterial inom arkeologi. Pågående digitalisering av historiskt källmaterial vid Riksarkivet och Riksantikvarieämbetet behöver vidare öka i omfattning. Bland annat krävs uppdatering (översättning) till modernt språk av inskrifterna i den *Samnordiska rundatabasen* och fortsatt uppbyggnad inom projektet *Det medeltida Sverige*. För den demografiska forskningen slutligen skulle en datorisering av 1800-talets kyrkböcker över hela Sverige utgöra en rik källa för forskningen.

- Behoven av ny databasuppbyggnad är omfattande och mångskiftande. Utöver de behov som gäller de enskilda områdena, framkommer sådana där man tydligt kan se att det finns sammanfallande intressen och därmed potential till bred användning av resurserna: (A) Det behövs annoterade korpusar med tal (i synnerhet samtal och intervjuer) och skriftspråk, inklusive multimodala text- och taldatabaser. Data ska alltså inkludera såväl text och tal som bild och grafik. Fokus är 2000-talets språk med stor variation av genre. Behovet är likartat vare sig det gäller forskningen i svenska, andra språk, eller medie- och kommunikationsvetenskap och sammanfaller också med vad som framkommer i den nyligen gjorda kartläggningen av databaser i språkteknologi. Ett liknande sammanfall med språkteknologins behov gäller inlärdatabaser. Stora delar av detta skulle kunna tillgodoses inom ramen för den planering för databasresurser inom språkteknologi som nu pågår³⁵. (B) Litteraturvetenskapen behöver åstadkomma vetenskapliga utgåvor och behöver databaser som kan utveckla, men också utvecklas av, editionsfilologin. Det samnordiska projektet *Menota* inriktat på nordiska medeltidstexter och utveckling av standarder för uppmärkning av texterna har betydelse för editionsfilologin, men också för alla vetenskaper med medeltiden i fokus – historiskt inriktad språkvetenskap, historia, arkeologi. (C) Internet ses som en viktig källa till kunskap för samtidsforskning inom etnologi, religionsvetenskap, språk, medie- och kommunikationsforskning. Informationen där kan utnyttjas för att studera värderingar och utbyte av åsikter om skilda företeelser, t ex via dag-för-dag-information under utvalda tidsperioder. Intresset för Internet handlar också om mediet i sig och hur kommunikation i dagens samhälle sker i form av e-post, bloggar etc. (D) Bildrepresentationer av skilda slag får alltmer ökad betydelse och bör samlas i databaser. Kartor länkade till annan dokumentation fyller såväl inom- som tvärvetenskapliga behov inom humaniora och inom humaniora i samverkan med andra vetenskapsområden, t ex för arkeologi. Databaser med ritningsprojektioner, modeller etc inom arkitekturforskningen skulle kunna samköras med demografiska och arkeologiska data. Möjlig samverkan med medicin och meteorologi kring databaser med instrumentella bilder kan inte realiserats utan nya databaser.
- Inom de flesta ämnesområden (utom medie- och kommunikationsvetenskap som har ett välfungerande dokumentationscenter) framkommer ett starkt behov av dokumentation av dels den forskning som pågår – bibliografiska databaser – liksom av databaser över databaser. Avsaknaden av överblick ses som starkt hämmande för forskningen. En slutsats som kan dras är att detta bidrar till att den potential till samverkan som framkommer i kartläggningen inte uppmärksammas av de berörda och därför inte heller utnyttjas. Här kan kartläggningen i sig vara ett stöd i att se möjligheterna, men framförallt kan här den utbyggda nationella data-tjänsten SND komma att fylla en ovärderlig funktion i framtiden.
- SND har också en viktig uppgift i frågor som rör dokumentation av databaserna och standardiseringsnormer, användbara gränssnitt och sökfunktioner samt uppdatering till dagens teknik av äldre databaser. Här finns mycket att göra, liksom när det gäller vägledning och stöd i

³⁵ Planeringsanslag från KFI för utveckling av svenska språkresurser (Basic Language Research Kit, BLARK, och Svensk Nationell Korpus, SNK).

övrigt. För språken finns speciella behov ifråga om komplettering av annotering och andra verktyg för äldre digitala databaser som bör ses i anslutning till liknande frågor kring databasuppdatering inom språkteknologin.

- Till de mer generella hinder som finns hör upphovsrättsliga begränsningar för text och bilder samt regler kring digital access, liksom etiska hinder med paralleller till PUL mm. Också hinder i form av revirtänkande förekommer. Allt detta hör sammantaget till problem som den humanistiska forskningen delar med andra vetenskaper och som därför bör diskuteras och lösas i ett vidare sammanhang (bl a inom ramen för verksamheten inom DISC och SND).
- Tillgängligheten till data hindras av kostnadspålagor och praktiska hinder i fråga om material i såväl väl utbyggda databaser, t ex demografiska databaser (här finns paralleller till det avgiftssystem som finns vid SCB) som källmaterial från bl arkiv.
- Ekonomiska hinder står i vägen för nödvändig digitalisering och registrering av äldre material inom ramen för sedan länge påbörjade projekt vid myndigheter och organisationer utanför universiteten. Här spelar det anslagssystem som regeringen tillämpar en väsentlig roll. Den tilldelning som tidigare gavs i form av anslag direkt till projekten via bl a statliga forskningsfinansierare går nu direkt till myndigheterna – t ex arkiven – inom ramen för totalanslag till verksamheten. Det gör att många av dessa projekt, som tidigare haft mer omfattande resurser, i konkurrensen med andra behov inom verksamheten, nu går på sparlåga.
- Kartläggningen visar sammantaget att det finns en vilja till nationell – och internationell – samverkan när det gäller utveckling i fråga om databaser, inom och mellan ämnen. Man efterlyser bättre samordning av registreringen i demografiska databaser, en portal för arkeologiska databaser för att bättre överblicka området etc. Det finns en stark potential i just sådan samverkan, vilket forskarna själva är väl medvetna om. Det gäller nu alltså att gå vidare för att skapa de resurser som den humanistiska forskningen kräver för att utvecklas, inte minst i samverkan med andra vetenskapsområden, där humaniora kan bidra med sina perspektiv. Initiativen här bör tas av forskarna själva, men det krävs också initiativ, satsningar och stödsatser från forskningens finansierare för att komma vidare. Organ med överblick över den humanistiska forskningen, inte minst Ämnesrådet för humaniora och samhällsvetenskap vid Vetenskapsrådet, har här också en central roll. Likaså kan vägledning och stöd från SND förväntas få stor betydelse i den framtida utvecklingen av humanioras databaser. Den satsning på SND som nu sker förefaller sammantaget välmotiverad utifrån vad som framkommer i kartläggningen .

Det har inte legat i uppdraget att föreslå åtgärder. Trots detta är det rimligt att utifrån vad som framkommer i underlaget föreslå visa insatser i en möjlig bredare åtgärdsplan.

- Ifråga om uppbyggnad av nya databaser borde fokus i nuläget inriktas på de områden (A)-(D) ovan, där man kan se en bred användning med utnyttjande från många olika ämnesområden. Här behövs för vart och ett en mer ingående kartläggning av behoven och olika prioriteringar, som måste göras i samverkan mellan Vetenskapsrådet och forskarna på området. Ifråga om editionsfilologin och utgivningen av medeltida texter relaterade till *Menota* under (B) finns redan existerande samarbete att bygga på, vilket borde kunna bidra till en kort startsträcka. I fråga om de språkdata som inkluderas under (A) bör samverkan ske med den nämnda, pågående planeringen för språkresurser inom språkteknologi.
- Vidare utredning krävs i fråga om digitaliseringen av kulturarvsmaterial. Kartläggningen har visat på stora behov uttryckt i relativt generella termer. Dock framställs vissa behov som mer trängande än andra, men informationen är otillräcklig för att se möjligheten till eventuella prioriteringar. Finansieringen är ett problem, men det är svårt att få en uppfattning om kostnadsbehoven. Allt detta talar för att det behövs en bred samling kring dessa frågor. I ett första steg måste prioriteringar utifrån forskningens behov göras av representanter för forskarna själva i berörda ämnen. Här skulle Vetenskapsrådet kunna bidra genom t ex konferensmedel.

Ämnesrådet för humaniora och samhällsvetenskap och andra organ med bred överblick över humanioraområdet bör också vara involverade. Därefter borde fortsatta diskussioner ske i samverkan med de myndigheter och organisationer som idag huvudsakligen är de som utför digitaliseringen (ofta inom ramen för olika Access-projekt) samt med forskningsfinansiärerna (inklusive staten).



Vetenskapsrådet
Database Infrastructure Committee
(DISC)

Bilagor:

**Databaser och digitalisering inom humaniora -
existerande resurser och framtida behov**

Juni 2007



Vetenskapsrådet

Database Infra-Structure Committee (DISC) 2006-10-25

Inbjudan

att medverka vid ett seminarium om pågående humanistisk forskning och behovet av databaser som forskningsinfrastruktur inom humaniora

Database Infra-Structure Committee (DISC) har skapats av Vetenskapsrådet för att arbeta framåtsyftande med utveckling, samordning och finansiering av existerande och nya forskningsdatabaser samt tillhandahålla denna gemensamma resurs för användning inom svensk och internationell humanistisk och samhällsvetenskaplig forskning.

Humaniora har hittills haft mycket begränsade möjligheter att få stöd från Vetenskapsrådet för digitalisering och uppbyggnad av databaser. Tillgängliga medel för infrastruktur har så gott som uteslutande gått till naturvetenskap, medicin och samhällsvetenskap. Med tillkomsten av DISC inleds emellertid en omfattande och samlad satsning på forskningsinfrastruktur, där även humanioras behov skall uppmärksammas.

I arbetet med att göra humaniora väl synligt i detta sammanhang har nu DISC beslutat att göra en kartläggning av existerande databaser och relevanta behov av att bygga upp och utveckla nya databaser för användning inom humanistisk forskning. Undertecknad har, som representant för humaniora i DISC, fått uppdraget att genomföra kartläggningen.

Kartläggningen ligger i linje med rekommendationer som skrivits in i Vetenskapsrådets strategidokument "Vetenskapsrådets guide till infrastrukturen" (se ytterligare info i bilaga), där *digitalisering inom humaniora* tagits upp bland de områden som ska utredas. De insamlade uppgifterna ska utgöra en del av underlaget för långsiktig planering och uppbyggnad av en forskningsinfrastruktur, som kan stärka svensk forsknings position i den internationella konkurrensen.

Inom ramen för detta arbete inbjuder jag dig att medverka vid rubricerade *seminarium*. Tanken är att sätta fokus på humanioras beroende och behov av databaser och digitalisering. Genom att avsätta utrymme för diskussion ska seminariet också ge möjligheter att se nya infallsvinklar som kan leda till satsningar inom det egna ämnesområdet.

Din uppgift skulle alltså vara: *att* inför seminariet tänka igenom situationen i ditt ämne med utgångspunkt i din egen forskning men också den som du *i ett nationellt perspektiv* överblickar, *att* vid seminariet under 15-20 minuter ge en muntlig presentation av ditt ämne med fokus på infrastrukturfrågor och *att* efter seminariet bidra med ett mer genomarbetat kunskapsunderlag. Det handlar om en *sammanställning* om 3-5 sidor.

VETENSKAPSRÅDET
SWEDISH RESEARCH COUNCIL

Postadress/Postal address
SE-103 78 Stockholm
Sweden

Besöksadress/Visiting address
Regeringsgatan 56

Tel: +46-(0)8-546 44 000
Fax: +46-(0)8-546 44 180

Org.nr.
202100-5208

vetenskapsradet@vr.se
www.vr.se





Vetenskapsrådet

Du är alltså en av ett antal inbjudna humanistiska forskare (från såväl språk- som hist-fil-sidan) med erfarenhet och överblick och som kan tala för era respektive områden. I den egenskapen anser jag dig, och övriga inbjudna, sitta inne med värdefulla synpunkter vad gäller läget idag och framtida behov av databaser och digitalisering inom den ämnesinriktning du representerar. Dessa lägesbeskrivningar är helt nödvändiga om DISC ska kunna föra humanioras talan ifråga om infrastruktur för forskningen.

Det är oerhört angeläget att kartläggningen ger en så rättvisande och representativ bild som möjligt. Ett kraftfullt synliggörande av humaniora behövs för att hävda området i konkurrensen om de infrastrukturella resurserna. Du skulle kunna bidra till detta. Min förhoppning är därför att du har möjlighet att ta dig an uppgiften.

Mötet planeras kunna ske i Stockholm **den 12 december klockan 10-18**. Trots det korta varslet finns en stark förhoppning om att kunna samla alla inbjudna till detta tillfälle. Utöver resekostnader utgår arvode för din insats vid seminariet och den skriftliga sammanställningen, som jag behöver **senast den 22 januari 2007**.

Jag hoppas på (positivt) svar från dig senast **den 3 november** med besked om huruvida du har möjlighet att ta på dig uppgiften samt delta i seminariet. Därefter meddelas ytterligare detaljer inför mötet.

Antingen du kan eller är förhindrad att medverka ber jag dig meddela ditt beslut per telefon eller e-post till mig alternativt forskningssekreterare, docent Jan Larsson, Vetenskapsrådet (tel: 08-546 44 335; e-post: jan.larsson@vr.se).

Med vänliga hälsningar
Eva Strangert

Tel: 090-7865680
e-post: eva.strangert@nord.umu.se

Bilaga:

1. Bakgrund och kompletterande information

Bilaga 1:

Bakgrund och kompletterande information

Bakgrunden till uppdraget

I ”Vetenskapsrådets guide till infrastrukturen”¹ framhålls behovet av att utreda frågan om *digitalisering inom humaniora* i syfte att skapa underlag som en del av en långsiktig planering av svensk forskningsinfrastruktur. Det generella syftet med infrastrukturplanen är att *främja och stödja uppbyggnad och utnyttjande av infrastruktur för svensk forskning av högsta vetenskapliga kvalitet*.

Inom humaniora finns behov av tillgång till material som finns vid arkiven, biblioteken och museerna (ABM-sektorn), men där finns också krav på digitalisering av sådant material och därutöver annan infrastruktur. Det finns t ex behov av att skapa helt nya databaser som bygger på egen datainsamling och det finns databaser under uppbyggnad, som behöver kompletteras.

Den kartläggning som nu ska göras ska därför inte begränsas till en inventering av behoven av *digitalisering* av ännu icke-digitaliserade datasamlingar. Den ska även kartlägga behovet av *databasuppbyggnad* i vidare mening och därtill ge en bild av vad som nu föreligger i form av *tillgängliga forskningsdatabaser* (digitala såväl som icke-digitala). Fokus ligger på databaser och datasamlingar för direkt användning *av forskare och för forskning*. Existerande databaser och framtida behov av resurser som kan *utnyttjas i stor skala* (delas av många parter) och har *nationellt intresse* (behovet är uppenbart för en stor mängd forskare) är av särskild betydelse.

Den aktuella kartläggningen ska ses som ett komplement till en tidigare inventering av befintliga databasresurser inom humaniora och samhällsvetenskap i Sverige². Rapporten från april 2005 utgör en redovisning av regeringens uppdrag till Vetenskapsrådet (regleringsbrev för budgetåret 2005). Där konstaterades bl a:

”Befintliga infrastrukturer inom humaniora och samhällsvetenskap är av delvis olika karaktär. Inom samhällsvetenskaperna finns en lång tradition av att bygga register av olika slag. Där finns longitudinella databaser och material baserade på tvärsnittsstudier av individer eller andra sociala fenomen. Många humanister arbetar direkt med material från arkiv, forskningsbibliotek, museer och andra samlingar. Språkdatabaserna vid universiteten är ett undantag. Inom det humanistiska fältet finns också värdefulla material som inte är digitaliserade. En intressant utveckling representeras av humanistlaboratorierna” (s 3).

I rapporten står också att:

”...resultatet av kartläggningen av infrastrukturer visar på en utveckling som går mot att materialen allt oftare utformas i samverkan mellan olika aktörer, såväl nationellt som

¹ <http://www.vr.se/download/18.7bea596910e36c19cbc80001735/Rapport+14.2006.pdf>

² http://www.vr.se/download/18.320a86de108dcd98cbc80004975/Om+forskningens+infrastrukturer_KFI-HS_slutrapport.pdf

internationellt, att forskningsdatabaserna uppvisar en allt större komplexitet och att behovet av tillgänglighet till materialen ökar”.

En del av det material som används inom humaniora föreligger alltså icke-digitaliserat; i vissa fall pågår digitalisering. Det gäller såväl material insamlat av forskare vid universitet och som bevaras där som forskningsmaterial som hämtats från myndigheter och ABM-sektorn. De digitala databaserna är dock till huvuddelen språk- och litteraturdatabaser, men här finns även databaser inom historia och arkeologi.

De allra flesta databaser – och de mest omfattande – finns inom språken, främst lingvistik (inklusive datalingvistik samt språk – och talteknologi), nordiska språk/svenska och moderna språk. Det gäller såväl det digitaliserade som det icke-digitaliserade materialet. Språkdatabaserna har nära knytning till litteraturdatabaserna med sina textdata (som också utnyttjas för språkteknologiska syften). Här finns stora material, ofta som resultat av stora nationella samverkansprojekt. Det beror främst på att stora tal- och textdatabaser inklusive dialektdatabaser är helt nödvändiga för forskningen inom nämnda områden.

Inom humaniora finns också andra slag av databaser. Som exempel kan nämnas arkeologi med dess olika användningsområden, t ex specialiteten miljöarkeologi. Till de större områdena i databashänseende hör också historieämnena med digitala databaser inom främst historisk demografi, men också lokal- och nutidshistoria. Inom svensk medeltidsforskning brukar den icke digitala datasamlingen ”Det medeltida Sverige” med Riksantikvarieämbetet som huvudman nämnas.

Det är sannolikt att denna bild av tillståndet inom humaniora, d v s en koncentration av databaser (och behov av digitalisering) till vissa ämnesområden, bör breddas och revideras. Syftet nu är därför att även inkludera ämnen och forskningsområden, som inte fanns representerade i den tidigare kartläggningen.

Uppgifter för kartläggningen

Din uppgift skulle alltså, som nämnts, vara att inför seminariet tänka igenom situationen – med fokus på infrastrukturfrågor – i ditt ämne, med utgångspunkt i din egen forskning men också övrig forskning *i ett nationellt perspektiv*, att under 15-20 minuter ge en presentation vid seminariet samt att därefter bidra med en skriftlig sammanställning som kan belysa frågeställningar som:

- Hur ser situationen ut idag inom ämnet? Finns digitala databaser? Andra typer av material- och datasamlingar (tal, text, bilder etc)?
- Hur utnyttjas tillgängliga databasresurser? Av enskild forskare? Grupper?
- Finns erfarenhet av samverkan och forskningsarbete gruppvis/nationellt kring uppbyggnad av nu existerande databaser och har det då även handlat om att samarbeta tvär- eller mång-disciplinellt?

- Vilken roll spelar befintliga databaser vid arkiv, bibliotek, museer, mm för forskningen inom ämnet?
- Finns behov för forskningen av digitalisering av material som förvaras vid arkiv, bibliotek, museer, mm?
- Vilka framtida behov finns i övrigt vad gäller databasuppbyggnad för att upprätthålla hög kvalitet och bidra till att forskningen inom ämnet kan vidareutvecklas?
- Finns vetenskapligt intressanta utvecklingslinjer och förnyelsemönster, som inte skulle kunna förverkligas utan tillgång till nya databasresurser?
- Vilka tänkbara möjligheter finns att samverka i forskargrupper eller i andra konstellationer nationellt och/eller internationellt för att bygga upp och utveckla nya databaser med bred användning i forskningen?
- Kan tillgång till och användning av databaser påverka forskningens synlighet och närvaro i ett internationellt perspektiv?

Databaser och digitalisering: modern svenska

Professor Britt-Louise Gunnarsson

FUMS, Institutionen för nordiska språk, Uppsala universitet.

Forskning om svenska språket bedrivs vid institutioner för nordiska språk huvudsakligen i Sverige och i Finland. Man studerar där svenskans utveckling under tidigare perioder och under det senaste århundradet. Professor Lars-Erik Edlund och jag har delat upp svenskämnet mellan oss, och jag kommer att behandla studier av modern svenska, ibland kallad nusvenska, medan han tar hand om studier av äldre svenska.

För att samla in underlag för den här genomgången sände jag ut ett rundbrev med en informell enkät till kolleger i Sverige och Finland. Jag fick glädjande många svar, vilket förstås visar på intresset för databaser och digitalisering bland nordister. Jag har också gått in på olika relevanta hemsidor och därigenom försökt skaffa mig en översikt över området. Någon heltäckande bild av databasutbudet tror jag mig dock inte ha fått fram. Studier av nutida talad och skriven svenska bedrivs ju inte bara vid nordiska språk institutioner utan också vid många andra institutioner, till exempel vid institutioner för lingvistik, datalingvistik och journalistik och vid Tema kommunikation i Linköping.

Min egen bakgrund är den att jag varit med om att samla in stora material inom olika projekt som jag lett. Materialet föreligger i digitaliserad form och ingår i de databaser som finns vid FUMS i Uppsala. I min egen forskning har jag använt de databaser jag själv varit med om att bygga upp. Däremot har jag själv nästan aldrig använt andra forskares databaser.

1 Stora material- och datasamlingar

Min beskrivning nedan av befintliga digitaliserade material- och datasamlingar utgår från en uppdelning på skrivet och talat språk. Jag anger också om materialen är tillgängliga utan eller med särskilt tillstånd och anger adressen till hemsidan som beskriver samlingen.

1.1 Skrivet språk

Språkbanken i Göteborg

På Språkbankens hemsida, <http://spraakbanken.gu.se/konk/down.phtml>, återfinns ett stort antal databaser av skriven svenska, som är tillgängliga för alla. Här finner man till exempel följande skriftspråkskorpuser:

Tidnings- och tidskriftsmaterial:

Press 65

Press 76

DN 1987

Press 95
Press 96
Press 97
Press 98
SvD 00
GP 01
GP 02
GP 03
GP 04
Forskning och framsteg

Romaner och brev
Äldre svenska romaner
Bonniers romaner
Strindbergs brev
Strindbergs romaner

Övriga korpusar
SAOL 11
Psalmboken
Svensk författningssamling 1978-81
SAOB

Skrivsyntaxmaterialet (Lund och Växjö)

Ett annat stort, skriftspråkligt material samlades in under 1970-talet inom projektet Skrivsyntax i Lund. Två av de då insamlade korpusarna är digitaliserade och finns lagrade i *Talbanken 05*, som är en moderniserad version av *Talbanken 76*. Följande skrivna material kan man få tillgång till för forskning efter tillstånd:

Professionell prosa: broschyrer, tidskriftsartiklar, läroböcker, debattböcker (87 000 ord)
Gymnasistprosa (89 000 ord)

Den aktuella hemsidan är knuten till Växjö universitet och återfinns på följande adress:
<http://w3.msi.vxu.se/~nivre/research/Talbanken05.html>

Skriftliga fackspråkskorpusar vid FUMS, Uppsala universitet

Vid FUMS, Avdelningen för forskning och utbildning i modern svenska, vid Institutionen för nordiska språk i Uppsala har två stora skriftkorpusar byggts upp och digitaliserats.

Inom projekten ”Facktexter under 1900-talet” och ”Fackspråkens framväxt” in samlades 360 vetenskapliga och populärvetenskapliga artiklar inom medicin, ekonomi och teknik från 1730-1985.

Inom projektet ”Texter i europeiska skrivsamhällen” insamlades ett stort antal texter inom olika genrer (broschyrer, brev, internmeddelanden, protokoll, årsberättelser, rapporter etc.) tillkomna under 1990-talet i banker, ingenjörbyråer, yrkesmedicinska och historiska institutioner i Sverige, Tyskland och England.

Materialen, som är tillgängliga för forskare efter tillstånd, beskrivs på FUMS hemsida på länken: <http://www.nordiska.uu.se/fums/forskning/skriftligmtrl>

FISC-korpusen vid Helsingfors universitet

Vid Nordica, Helsingfors universitet, har man samlat in och digitaliserat texter skrivna på 1990-talsfinlandssvenska.

Tidningsspråk (1 milj ord)
Skönlitteratur (0,5 m),
Sakprosa (0,4 m)
Myndighetstest (0,4 m.).

Materialen distribueras fritt mot användaravtal. Institutionen kan nås på adressen: www.helsinki.fi/nordica

1.2 Talat språk

Göteborg Spoken Language Corpus (GSLC)

Talspråkskorpusen vid Institutionen för lingvistik i Göteborg är tillgänglig efter visst tillstånd. Korpusen är baserad på 371 olika inspelningar från: auktioner, domstolar, konsultationer, formella möten, hotell, fabriksgolvet, föreläsningar, politiska debatter. Sammanlagt 182 timmar. Hemsida: <http://www.ling.gu.se/projekt/tal/>

Talbanken 05 (Lund och Växjö)

Som nämnts ovan innehåller *Talbanken 05* en moderniserad, digitaliserad version av *Talbanken 76*. Den digitaliserade samlingen av talad svenska, som beskrivs på hemsidadressen <http://w3.msi.vxu.se/~nivre/research/Talbanken05.html>, omfattar:

Inspelningar på arbetsplatser av autentiska talsituationer:
16 inspelningar om totalt cirka 25 timmar.

Studieinspelningar av arrangerade talsituationer: samtal/debatt.
15 timmars inspelat tal.

Talspråksmaterialet vid FUMS

Vid FUMS, Inst. för nordiska språk i Uppsala, har sedan 1970-talet ett stort antal forskningsprojekt samlat in och lagrat inspelningar av talat språk. Dessa har de senaste åren digitaliserats och kan efter medgivande av den som ansvarar för materialet utnyttjas även av forskare utanför FUMS-gruppen. På hemsidan www.nordiska.uu.se/fums/forskning förtecknas 16 forskningsprojekt som bidragit med material till den s.k. talspråksbasen (talspraksbas.pdf). Följande projekt finns där med:

Svenska stadsmål 1. Eskilstuna

Språk, roll och social relationer (Burträsk)
Barnets språkliga identifikation
Stad och omland
Språkanvändning och språkmiljöer i staden och på landet
Ungdomars samtalsstil
Läkarens och lekmannens begreppsvärldar
Kommunikation vid en larmcentral
Talutveckling
Sverigefinnars två språk
Interaktionen vid seminarier
Texter i europeiska skrivsamhällen
Samtal i nödsituation (giftinformationscentral)
Kontinuitet och förändring i nutida talspråk
Samtal, åldrande och identitet.
Samtalsspråkets grammatik

Talat språk vid Nordica i Helsingfors

Vid Nordica, Helsingfors universitet, har man också samlat in helsingforssvenska. Enligt prefekten Jan Lindström rör det sig om 80.000 ord av transkriberad helsingforssvenska.

2 Problem med utnyttjandet av befintliga databaser

Språkbankens material utnyttjas nog ganska allmänt. Materialet är välkänt och framför allt är det lättillgängligt. Ofta hör och ser man hur nordister sökt svar på olika frågor i Språkbankens databaser. I övrigt är det nog primärt forskare direkt knutna till ett visst projekt som utnyttjar det material gruppen samlat in. Detta faktum, som förstås kan anses som ett slöseri med forskningsresurser, skulle jag här vilja knyta till fem olika typer av problem.

2.1 Etiska problem

Det är ett välbekant faktum att mycket av det talpråkmaterial som spelats in av etiska skäl är svårt att få tillgång till av andra än projektets forskare. En anledning kan vara att forskarna, då de genomförde undersökningen, inte begärde in de tillstånd av de medverkande som behövs för att andra än projektgruppen ska få utnyttja materialet i ett senare skede. En annan anledning kan vara materialets allmänna känslighet, till exempel gäller detta för en del samtal inom sjukvården.

2.2 Upphovsrättsliga problem

Ett problem med aktuellt skriftligt material kan knytas till den upphovsrättsliga lagstiftningen. Delar av den korpus som projektet "Svensk dramadialog under tre sekler" vid Institutionen för nordiska språk i Uppsala byggt upp kan till exempel inte läggas in i en databas av denna anledning. Författarna till dramerna i 1900-talsdelen av materialet lever fortfarande eller har nyligen dött.

2.3 Forskningsinterna problem

Inte sällan händer det att forskarna som lagt upp korpusen vill vänta med att lämna ut materialet tills de själva forskat färdigt inom området. Om man så vill, kan man säga att forskarna här vaktar på sitt revir.

2.4 Datatekniska problem

Många stora material har insamlats i enlighet med vetenskaplig systematik och lästs in för datorbearbetning med beaktande av vid tiden aktuell teknik. Vad som sedan hänt är att utvecklingen inom datatekniken gått framåt och att lagringstekniken blivit omodern. För att korpusen ska kunna användas av andra än forskarna själva – och kanske också av dem själva – behöver materialet uppdateras och mallanpassas till aktuell teknik. Ett avsevärt problem gäller avsaknaden av uniformitet över tid.

2.5 Materialbeskrivningsproblem

Ett annat problem som infinner sig när någon utanför den ursprungliga gruppen ska använda ett material är att detta inte är beskrivet och kodat på ett tillräckligt noggrant sätt. Om man till exempel i en sociolingvistisk studie ska analysera ett inspelat material bör man veta vem som säger vad i vilken situation. Det räcker då inte att få tillgång till ett inspelat talspråksmaterial där det helt saknas uppgifter om i vilket sammanhang yttrandet fällt. En följd av en sådan här avsaknad av adekvat materialbeskrivning blir att varje forskargrupp måste göra sina egna inspelningar.

3 Erfarenheter av samverkan och forskningssamarbete

Inom nordistiken och ämnet svenska språket finns det goda erfarenheter av samverkan och forskningssamarbete. Samtliga ovan nämnda databaser har byggts upp i samverkan lokalt och nationellt. De olika korpusarna vid FUMS i Uppsala, Skriv-och talsyntaxmaterialet i Lund, Språkbanken i Göteborg och FISC-korpusen i Helsingfors har kommit till genom samarbete mellan olika forskargrupper. Också många internationella samverkansprojekt har givit resultat i form av stora materialsamlingar. Här kan till exempel nämnas projektet ”Samtalspråkets grammatik”, som omfattat forskare i Göteborg, Helsingfors, Linköping och Uppsala.

4 Framtida behov

Vad gäller studier av modern svenska vill jag peka på följande tre behov av databaser:

4.1 Talspråk

För det första finns det behov av en talspråkskorpus som kan användas i allmänt språkbeskrivande syfte. Det finns mycket vi **inte** kan säga om den talade svenskan utan tillgång till ett mycket större material än vad den enskilda forskaren kan samla in. Vad som behöver läggas kraft på är (a) tydliga etiska riktlinjer, både med hänsyn till de inspelade och

med hänsyn till den/de som har gjort grovjobbet med att samla in materialet, (b) standarder för format, inkl. sökrutiner. Det skulle till exempel vara möjligt att samla in och bygga upp en talspråkskorpus över offentligt talad svenska, det vill säga en talspråkets motsvarighet till Språkbankens tidningskorporusar, så att den inte var integritetsskyddad utan allmänt tillgänglig för studier av olika slag. Förmodligen vore det lämpligt att också länka textremsan till en ljudfil. Med utgångspunkt i en större talspråkskorpus skulle man kunna studera ordförråd, morfologi och syntax på ett sätt som inte är möjligt i dagsläget.

4.2 Multimodalt skriftspråk

För det andra finns det behov av korpusar som kan användas för mer genreinriktade analyser av skrivet språk. Tidningsspråk är förhållandevis väl beskrivet och också vetenskapligt språk fram till 1980, men för andra genrer saknas det stora, lättillgängliga korpusar. Befintliga korpusar har också lagts upp med inriktning på det skrivna ordet enbart. För att fånga dagens textvärld och de förändringar som sker inom olika genrer behöver vi därför lägga upp multimodala korpusar, dvs. sådana som även inkluderar bilder och grafik.

4.3 Datormedierade genrer

För det tredje finns det behov av systematisk lagring av olika typer av datormedierade genrer. I dagens kommunikationsverklighet samverkar ord, bild och grafik och också ljud och skrift, och nya, annorlunda genrer skapas. Ett brev ser inte längre likadant ut som det gjorde för tjugo år sedan, och det kommer säkerligen inte att se likadant ut i nästa decennium. Förändringarna går snabbt och alla stadier hinner inte med att dokumenteras; förra veckans hemsidor finns inte alltid att tillgå också denna veckan. För forskaren inriktad på analysen av modern svenska skulle det därför behöva byggas upp också databaser också för dagens datormedierade genrer.

5 Nya databasers värde för framtida forskning

De behov jag skissat ovan skulle på olika sätt ge bättre underlag för analyser av de förändringar som svenska språket med sina olika genresystem för närvarande genomgår. Att förändringar ofta först uppträder i talspråket innan vi ser dem i den offentligt skrivna svenskan är en allmänt omfattad sociolingvistisk uppfattning. Att språkbruket förändras genom modern teknologi spekuleras det mycket om, men för att få underlag för en mer djupgående diskussion av i vilken grad och på vilket sätt dessa förändringar sker behöver forskaren ha tillgång till systematiskt uppbyggda korpusar. Generellt gäller att de korpusar som enskilda forskare kan bygga upp är för små för att man ska kunna dra några mer generella slutsatser om hur svenska språket förändras.

6 Framtida möjligheter till forskningssamverkan

Instutionerna för nordiska språk är vana vid att samarbeta såväl nationellt som nordiskt med andra nordister, dels med lingvister och datalingvister. Det finns alltså all anledning att se hoppfullt på framtida möjligheter inriktade på att bygga upp stora, gemensamma databaser och ge dessa en bred användning i forskningen. Säkerligen skulle det vara möjligt och fördelaktigt om korpusuppbyggandet samlades till ett centrum.

7 Forskningens synlighet och närvaro i ett internationellt perspektiv

Förmodligen skulle det vara av värde också i ett internationellt perspektiv med stora databaser med modern svenska. I och för sig finns det många sätt för svenska forskare att bli synliga internationellt, men att ingå i större, internationella projekt är en möjlighet. För forskarvärlden utanför Sverige blir det säkerligen extra intressant när svenska språket belyses kontrastivt.

Lars-Erik Edlund: Nordiska språk

Jag har hedrats med uppdraget att kartlägga digitalisering och databasuppbyggnad inom ämnet nordiska språk, i synnerhet inom delområdena ordförråd, dialekter och namn, men även inom områdena runologi och äldre texter. Andra delar av ämnet har kartlagts av Britt-Louise Gunnarsson, Uppsala. En enkät har gått ut till institutioner som bedriver forskning inom ämnet och dessutom till de relevanta enheterna inom Institutet för språk och folkminnen (SOFI). Även relevant material som digitaliserats inom andra enheter nämns nedan, men här är kartläggningen mer sporadisk. Detta senare kompletteras av Olle Ferm i dennes avsnitt, till vilket en generell hänvisning lämnas.

1. Resultatet av kartläggningen

Uppsala universitet, Institutionen för nordiska språk

Här har Samnordisk runtextdatabas, <http://www.nordiska.uu.se/forskn/samnord.htm>, byggts upp, tillgänglig i både svensk och engelsk version. Databasen omfattar omkring 6 500 runinskrifter i translitterering och i normalisering till rundanska, runsvenska respektive fornvästnordiska. Inskrifterna översätts till engelska, undantagsvis till norska. Till varje inskrift kopplas informationer, i den mån sådan är tillgänglig, om fyndplats och nuvarande placering, runtyper, korsformer, datering, stilgruppering, ristare, material och föremålskategori, referenser samt bildlänkar. Databasen uppdateras och utvidgas dels med nya tolkningar och nya inskrifter, dels med nya informationskategorier.

Samnordisk runtextdatabas är utgångspunkten för forskning kring framför allt de vikingatida runstenarna som språkliga och kulturella monument i relation till det forntida samhället och landskapet. Databasen har resulterat i flera ordböcker: Svenskt runordsregister (<http://web.telia.com/~u13403587/rundata/Runordsregister.pdf>), An English Dictionary of Runic Inscriptions in the Younger Futhark (<http://runic-dictionary.nottingham.ac.uk/index.php>) och Nordiskt runnamnslexikon (http://www.sofi.se/servlet/GetDoc?meta_id=1472). Basen drivs nu på ideell bas. Behov finns av ytterligare databasuppbyggnad.

Lunds universitet, Institutionen för nordiska språk

Här finns Fornsvenska textbanken, som innehåller stora mängder fornsvenska texter, och en del nysvenska. Denna nås på: <http://www.nordlund.lu.se/Fornsvenska/Fsv%20Folder/index.html>.

Databasen befinner sig, enligt institutionen, i ett utvecklingskede, och texterna har varierande kvalitet, bl.a. krävs genomgående en ordentlig korrekturläsning. Även sökrutinerna och fakta som redovisas om texter och handskrifter måste kompletteras. Vissa genrer i banken är sämre företrädade. Dessutom finns i de genrer som är representerade tidsmässiga eller "undergenremässiga" luckor.

I nuvarande skick får korpusen enligt institutionens bedömning betraktas som "ålderdomlig". Den bygger inte på en databas, utan varje text finns dels som en html-fil, dels som en rtf-fil. Vid korrigeringar måste ändringar göras i båda. Tanken är nu att skapa en databas, där texterna indexerar och på ett enkelt sätt blir sökbara. Dessutom tänker man sig att forskare skall kunna avgränsa korpusen efter tidsperiod och genre för att sedan kunna göra en mer begränsad sökning. För att öka tillgängligheten finns behov av att översätta alla metatexter till engelska.

Umeå universitet, Institutionen för litteraturvetenskap och nordiska språk

Här finns två databaser som har sin bakgrund i nu avslutade projekt. Svensk etymologisk databas baseras på Elof Hellquists svensk etymologisk ordbok (senaste upplagan), till vilken omfattande, nyskrivna revideringar fogats. Etymologier till en mängd ord som inte finns hos Hellquist har dessutom tillförts. Konkret finns supplement till följande bokstäver: F (av Börje Tjäder), K (av Hans Jonsson), O, P, Q, R (av Gösta Holm), S-T (av Sven Ekbo), U, V, X, Y, Z, Å, Ä, Ö (av Gösta Holm). Arbetet med ytterligare bokstäver pågår. De nyskrivna partierna innebär en avsevärd utökning av Hellquists texter; "grundtexten" byggs ut med cirka 70 %. Databasen föreligger tillgänglig vid institutionen, och underhålls ideellt. Den är (ännu) inte sökbar på nätet. En del frågor från forskare och allmänhet besvaras med hjälp av databasen.

Den andra databasen, Diabas – Nordisk dialektgeografisk databas – byggdes upp inom ramen för ett externfinansierat projekt. I databasen finns drygt 300 000 ordbelägg, däribland samtliga i Sverige upptecknade dialektala växtnamn liksom en del växtnamn från övriga Norden. Just växtnamnen har utgjort utgångspunkten för ett utvecklingsprojekt vid HUMlab, Umeå universitet. Utifrån detta material har olika teorier för språklig och kulturell spridning testats genom modellering, simulering och visualisering. Diabas underhålls ideellt, men efterfrågas både av andra forskare och av allmänheten. Det är uppenbart att intresse för basens material finns, men att institutionen själv inte kan fortsätta uppbyggnaden eller prioritera de uttag som forskare och allmänhet önskar.

Göteborgs universitet, Institutionen för svenska språket

De stora databasresurser som finns vid denna institution redovisas på andra ställen i denna rapport, varför jag här går förbi denna institution i min kartläggning.

Institutet för språk och folkminnen (SOFI)

Institutet finns på fyra orter: Uppsala, Umeå, Göteborg och Lund. Här förvaras stora samlingar av ortnamn, personnamn, dialekter och folkminnen, vilka till vissa delar digitaliserats. Det pågår nu ett projekt för att samordna myndighetens registerkataloger till ett enhetligt system och tillgängliggöra dem över nätet.

Vid Namnavdelningen i Uppsala finns Svensk ortnamnsdatabas. Digitaliseringen av SOFI:s ortnamnssamlingar har pågått sedan mitten av 1990-talet. Redan vid utgången av 1998 var cirka 3 700 000 blad skannade och indexerade, vilket innebar att det topografiska registret i Uppsala var färdigt. Under senare år har bättre sökbarhet i registret eftersträvats. Nu är totalt strax under 3 000 000 poster granskade, och sammanlagt 20 län utlagda på institutets webbplats. Denna digitalisering har gjort det möjligt för forskare att ta del av ortnamnssamlingarna. Avdelning för namnegransking, Oslo, har kopierat etymologiregistret och digitaliserat det. Detta finns nu tillgängligt via nätet. De hittills tryckta häftena av *Sveriges medeltida personnamn* finns i nätversion.

Vid Dialektavdelningen i Uppsala finns stora behov, och här har digitaliseringen inte alls kommit så långt som inom namnavdelningen, trots de rika dialektordsamlingarna. Ett register över senareleder i sammansättningar, suffix i avledningarna och betonade artiklar i verbförbindelser omfattande cirka 260 000 poster har dock digitaliserats. Detta register bygger på de sedessamlingar som utgör OSDs, alltså Samlingar till Ordbok över Sveriges dialekter. Man har dessutom inom institutet bl.a. arbetat med att ta fram en databas för substantivens böjningsmorfologi och dialektal syntax. Detta material skall vara sökbar på olika sätt och kunna kopplas till karteringsprogram, MapInfo. En del resurser har lagts på digitalisering av fonogram. Det finns idag drygt 80 000 accessioner av registrerade inspelningar i databasen 4D, men kommer att föras över till en FileMaker-databas. Nu påbörjas ett arbete med mp3-filer tillgängliga via nätet. I samarbete med projektet Scandiasynt, som undersöker dialektal syntax, tillförs dialektavdelningen transkriptionen länkade till ljudfiler.

Databaserna vid Dialekt-, ortnamns- och folkminnesarkivet i Göteborg, DAG, utgörs till största delen av register över DAG:s samlingar. Det föreligger enligt DAG ett generellt behov av digitalisering av arkivets material för att öka tillgängligheten för forskare och allmänhet. Sålunda finns det en bohuslänsk ordsamling som omfattar cirka 15 000 uppslagsord med uppgifter om uttal och betydelser för varje socken i Bohuslän. Det är en mycket användbar källa, som i ett första skede skall inskannas.

Dialekt- och ortnamnsarkivet i Lund (DAL) har byggt upp namndatabaser och erbjuder dem på nätet: register över skånska bebyggelsenamn, register över skånska naturnamn, Skånes bebyggelsenamn (som färdigställs 2007) och DAL:s handskriftsregister. Planer finns att digitalisera naturnamnssamlingar samt dialektbeläggen från Skåne, Blekinge, Halland, Småland och Öland.

Vid Dialekt-, ortnamns- och folkminnesarkivet i Umeå (DAUM) finns en accessdatabas över samlingarna med för närvarande uppemot 50 000 poster i de olika registren. Digitalisering av materialet i samlingarna har prioriterats under de senaste fem åren.

Lantmäteriverket i Gävle

Vid Lantmäteriverket finns förutom en ortnamnsdatabas arkivaliesamlingar som man kommer åt på www.lantmateriet.se, t.ex. historiska kartor. Om man arbetar med landskaps- och bebyggelseforskning är tillgången till detta material mycket värdefull.

Riksantikvarieämbetet

Riksantikvarieämbetets digitala fornlämningsregister FMIS och den publika delen Fornsök är viktiga databaser. En fortsatt digitalisering av ATA:s samlingar är prioriterad.

Riksarkivet

Uppgifter om svenska brev eller brev rörande Sverige finns samlade i diplomatariets huvudkartotek. Detta omfattar uppgifter om över 40 000 brev och medger sökning på personnamn, ortnamn, datum och viktiga sakuppgifter. Dessutom kan man söka i samtliga brevtexter som hittills tryckts i Svenskt Diplomatariums huvudserie och i serien för 1401-1420. Huvudkartoteket finns tillgänglig

<http://www.ra.se/ra/diplomat.html>. Söksida:

<http://62.20.57.212/ra/medeltid/soksida.html>><http://62.20.57.212/ra/medeltid/soksida.html>.

Medieval Nordic Text Archive (Menota)

Menota tillkom 2001 som samnordisk organisation för utvecklingen av elektroniska redskap för arkivering, behandling och visning av nordiska medeltidstexter, se

<http://gandalf.aksis.uib.no/menota/>. En engelsk handbok för elektronisk kodning av nordiska medeltidstexter har sammanställts, se för en (preliminär) version 2.0

http://gandalf.aksis.uib.no/menota/guidelines-2/contents/contents_2-0-b.page. Problem med att visa medeltida texter via internet eller sända dem mellan olika datorer är välkända. Inom Menota har därför initiativ tagits för att utveckla användbara fonter och att påverka Unicode att anta så många tecken som möjligt i sin standard. En organisation under beteckningen Medieval Unicode Font Initiative (MUFI) samlar deltagare från hela Europa och USA, se <http://gandalf.aksis.uib.no/mufi/>.

Finansieringen av verksamheten har hittills skett med hjälp av externa projektmedel. I Bergen är projektet Tekst- og verktøybase for Medieval Nordic Text Archive. Kodingsstandard, ordbankbygging og Heimskringla-edisjon inne i slutfasen. Man har vidare arbetat med att göra texter från Gammelnorsk ordboksverk tillgängliga via internet. I Oslo har Vadstenaklostrets fornsvenska handskrifter elektroniskt tillgängliga (VadstenaDigital) arbetat med att transkribera och elektroniskt koda handskriftstext ur samlingen av handskrifter från Vadstena i Kungliga biblioteket i Stockholm, se

<http://www.hf.uio.no/iln/forskning/forskergrupper/norron/prosjekter/vadstena/index.htm>. I Reykjavík har Stofnun Árna Magnússonar ett projekt för att producera en elektronisk utgåva av Codex Regius av den äldre Eddan enligt Menotas standard med fullständig lemmatisering och morfosyntaktisk analys. Den elektroniska texten blir tillgänglig via Menota.

Under 2005 inleddes på Menotas initiativ ett samarbete mellan de tre ordboksredaktionerna för Ordbok over det norrøne prosasprog (København), Gammeldansk ordbog (København) och Gammelnorsk ordboksverk (Oslo) och institutioner för språkvetenskaplig behandling av elektroniskt material, Språkdata (Göteborgs universitet) och Enhet for Digital Dokumentasjon (Oslo). I samarbetet ingår även svenska lexikografer. Se för detta <http://www.edd.uio.no/menota/>.

I nuläget inriktas arbetet inom Menota främst på att få ut fler texter i läsbar och sökbar form på nätet. För att Menota skall kunna fortsätta sitt arbete är det avgörande att dels sätta i gång nya projekt, dels permanenta strukturer för att upprätthålla användbarhet för det material som etableras.

2. Reflektioner

I anslutning till kartläggningen tillåter jag mig nu några avslutande reflektioner, som anknyter till de frågor som Vetenskapsrådet önskat få belysta. För det första vill jag lyfta fram det berättigade intresse denna digitalisering och uppbyggnaden av dessa baser har för allmänheten. Baserna innehåller en viktig del av vårt kulturarv och det är därför viktigt att materialen så långt som möjligt görs tillgängliga med väl anpassade användargränssnitt.

Unesco:s konvention rörande skydd av det immateriella kulturarvet (se <http://www.google.com/search?hl=sv&client=safari&rls=sv-se&sa=X&oi=spell&resnum=0&ct=result&cd=1&q=konvention+immateriellt+kulturarv&spell=1>) innebär, om den ratificeras, att språkliga kulturarv i alla former kommer att behöva värderas, sorteras och, kanske, rangordnas och övervakas. I de läget krävs att man har kännedom om, tillgång till och kontroll över ”materien”, och då blir databaser av den typ som här beskrivits av avgörande betydelse.

Det skall dessutom understrykas att databaserna har stort värde för forskarna, inte minst ur ett internationellt perspektiv. De nationella ortnamnsdatabaserna har exempelvis betydelse i det internationella samarbetet. Sålunda pågår i Europa ett stort projekt EuroGeoNames som avser att bygga upp en europeisk namndatabas. Resultatet kommer att bli att namnforskarna får helt andra möjligheter att studera namn, namngivning, namnstrukturer, namnmotiv, och att diskutera namnvård och namnplanering etc. i ett komparativt och globalt perspektiv och i internationella och/eller globala nätverk. Det är i detta sammanhang viktigt att våra rika material digitaliserats, kvalitetssäkrats och gjorts tillgängliga i databaser.

Man måste utifrån den kartläggning jag här gjort konstatera att digitaliseringen och databasuppbyggnaden i flera fall skett inom ramen för externfinansierade och tidsbegränsade projekt. Detta innebär att stora material hanterats med teknik som varit aktuell just vid tiden för projektet. När projektet avslutats har ingen kunnat ansvara för uppdateringar (och det har heller inte funnits pengar), och därför kan också mycket viktiga forskningsmaterial komma att bli outnyttjade eller i varje fall klart underutnyttjade. En nationell strategi bör skapas för att möjliggöra att digitaliserat material, som äger intresse för en större grupp forskare eller för allmänheten, underhålls också efter projektperioden och att adekvata, uppdaterade användargränssnitt skapas för dessa databaser. Här blir Svensk nationell datatjänst (SND) självfallet en strategisk resurs. Viktigt är emellertid också att de forskare som skapat baserna tydligt involveras i arbetet, eftersom de besitter specifika kunskaper som måste tas tillvara vid den fortsatta utvecklingen av databaserna. Här har Vetenskapsrådet en betydelsefull roll att spela.

Behoven av samordning mellan olika aktörer framstår i kartläggningen som uppenbara. De tycks sålunda vara uppenbart att den kompetens som byggts inom Menota bör tas tillvara då det gäller uppbyggnaden av baser över äldre texter. Överhuvudtaget bör, som jag ser det, Menota kunna tjäna som förebild och inspiratör för likartade projekt, varvid det inte minst är av betydelse att Menota har en unik internationell referensram och tar ett pannordiskt grepp. En satsning på Menota från de nordiska finansiärernas sida synes mig vara en viktig forskningsstrategisk satsning. SweDia-materialet, som ger rika prov på det nutida regionala talspråket, bör om möjligt samordnas med de samlingar som digitaliseras vid språkarkiven, eftersom vi därigenom får material till longitudinella studier av svenskt talspråk. En samordning mellan den etymologiska databasen vid Umeå universitet och redaktionen för Svenska Akademiens Ordbok i Lund bör, för att nu ta ytterligare exempel, övervägas. Det finns dessutom god anledning att söka länka samman material från olika ämnesområden och inom olika myndigheter, t.ex. fornminnesmaterial med ortnamnsmaterial. Initiativ av detta sistnämnda slag skulle underlätta och uppmuntra spännande mång- och tvärvetenskapliga undersökningar.

Flera mer konkreta behov vad gäller inskanning, digitalisering och databasuppbyggnad har lyfts fram vid min inventering. Mest akut synes det mig vara att dialektalt material inom SOFI inskannas och digitaliseras, så att det blir någorlunda paritet mellan ortnamnsmaterialet och dialektmaterialet inom myndigheten. Detta måste vara högt prioriterat. Det hindrar dock inte att man också har konkreta digitaliseringsbehov vad gäller namnen, t.ex. beträffande medeltida personnamn och ortnamn på uppteckningskartor.

Vidare finns ett stort och viktigt ordmaterial vid Svenska Akademiens Ordboksredaktion. Detta är användbart för olika forskare, eftersom de excerperade källorna från 1521 och framåt täcker så många olika genrer och sakområden. (Endast cirka 10 % av de 8 000 000 excerpterna kommer med i de tryckta ordboksartiklarna.)

Ett mångfasetterat och betydelsefullt textmaterial från skilda tider utgör Svenskt Visarkivs samling av visböcker. Viktigt är detta material inte minst då det speglar ett folkligt skrivande inom olika samhällsgrupper och av den anledningen tjänar som utgångspunkt för studier på område literacy. Här finns stort behov av inskanning och senare digitalisering.

Med detta anser jag mig ha fullgjort mitt uppdrag.

Lars-Erik Edlund

Merja Kytö
Uppsala universitet

Behovet av databaser som forskningsinfrastruktur inom humaniora

Engelska

Engelskan anses vara det bäst representerade språket vad avser mängden av allmänt tillgängliga strukturerade korpusar kompillerade för lingvistiska analyser. Man har också lagrat engelska i andra former. Nedan behandlas de till seminariedeltagarna utsända frågorna i ljuset av korpusforskning inom engelsk språkvetenskap (frågorna är numrerade inom hakparentes).

[1] Hur ser situation ut idag inom ämnet? Finns digitala databaser? Andra typer av material- och datasamlingar (tal, text, bilder etc)?

Vanligen skiljer man de s.k. strukturerade korpusarna från andra textsamlingar. Det finns också andra databastyper, t.ex. elektroniska ordböcker som kan användas som korpusar. Den allra senaste upptäckten är World Wide Web, den största korpusen i världen, till stor del på engelska men också på många andra språk. Nedan skildras dessa resurser i all korthet.

(1) Strukturerade korpusar ägnade åt lingvistisk analys

En strukturerad korpus kan anses ha följande egenskaper:

- en representativ samling av texter (eller delar av texter)
- maskinellt läsbara texter
- skriftligt eller/och talat språk
- autentiskt material
- en begränsad mängd ord/texter
- en standard referenskälla (undersökningar jämförbara)
- texterna förberedda för datasökningar
- lingvistisk annotering kan ingå (satsdelar, morfologi, prosodi m.m.)

Dessa korpusar kan vara generella med ett antal representerade genrer eller specialiserade med enbart en representerad genre (eller ett genreområde). De är vanligen utvecklade av arbetsgrupper intresserade av en särskild forskningsfråga. Syftet är att lägga fram en textsamling som skulle kunna representera en viss typ språkbruk, t.ex. autentiska texter eller talspråk som kan sökas med dataprogram. De kan vara synkrona eller diakrona, dvs. de kan spegla språket under ett särskilt tidevarv i historien eller i modern tid eller de kan löpa genom historien och erbjuda material från olika tidevarv.

Dessa korpusar är ofta ganska begränsade vad avser storleken (från några tio- eller hundratusentals ord till en eller två till tre miljoner ord) men de kan också bestå av t.o.m. 100 miljoner ord, så som British National Corpus. De borde kunna användas som en referenskälla för jämförbara undersökningar av andra språkvetare.

Dessa korpusar kan bestå av text som man har matat direkt från källan så nära den ursprungliga formen som möjligt. Men det går också att annotera texten med lingvistisk och externlingvistisk information som berättar om satsdelar, morfologi, semantiska aspekter, genrer, språkbrukarnas egenskaper osv. Dessa koder kan sedan användas av olika

datasökningsprogram för att utvinna information avseende språkets bruk, struktur och betydelser. På detta sätt underlättas datainsamling. Manuella excerperingsuppdrag som kunde ta årtal ännu på 1980-talet kan nu åtgärdas manuellt inom några sekunder eller minuter. Det går förstås inte att använda korpusar till alla möjliga områden inom engelsk språkvetenskap, men en hel del intressanta forskningsfrågor kan behandlas inom den korpuslingvistiska analysramen. Två exempel på korpusdata ges nedan. I det första exemplet har man sökt ordet "London" i London-Lund korpus som består av annoterat talspråk under en 30-årig period fr.o.m. 1950-talet. I det andra exemplet ser vi samma ord i Lancaster/Oslo-Bergen korpusen som innehåller skriftlig brittisk engelska från början av 1960-talet.

1 4:Heading

<441 B> ^I !th\ink#

<442 B> *it`s ^not in L\ondon#*

<443 A> *what a ^cl/ever* !thing to **!d\o#**

A:Press:reportage A09:133

who_WPR started_VBD his_PP\$ career_NN *'_*' picking_VBG up_RP pins_NNS in_IN a_AT Paris_NP salon_NN **'_**' and_CC is_BEZ now_RN *London's*_NP\$ leading_JJ couturier_NN ,_, has_HVZ been_BEN chosen_VBN by_IN Katharine_NP Worsley_NP

Det finns ett antal organisationer som distribuerar engelskspråkiga datoriserade korpusar. De bäst kända nämns nedan med information avseende deras utbud.

a) Oxford Text Archive (<http://ota.ahds.ac.uk/>) är en av de viktigaste källorna för korpusar inom engelsk språkvetenskap. Detta textarkiv innehåller flera tusen elektroniska texter och språkkorpusar på ett flertal språk.

b) Aksis Centre (Avdelning för kultur, språk og informasjonsteknologi), Bergen universitet (<http://www.aksis.uib.no/>). Denna enhet distribuerar korpusar som innehåller engelska, norska eller minoritetsspråk i Norge. Dess bäst kända materialkälla är ICAME CD-ROM (<http://icame.uib.no/newcd.htm>). ICAME CD-ROM (1991/2, 1999) innehåller följande material (de asterisk-markerade korpusarna var tillgängliga på den första 1991/2 versionen av skivan).

Skriftligt språk

- Brown Corpus *icke-tagga / tagga (1 miljon ord) = AmE
- *Lancaster/Oslo-Bergen Corpus (LOB) icke-tagga / tagga (1 miljon ord) = BrE
- Freiburg-LOB (FLOB) (1 miljon ord) = BrE
- Freiburg-Brown (Frown) (1 miljon ord) = AmE
- *Kolhapur Corpus (India) (1 miljon ord)
- Australian Corpus of English (ACE) (1 miljon ord)
- Wellington Corpus of Written New Zealand English (WWC) (1 miljon ord)
- The International Corpus of English – East African Component (500 000 ord)

Talspråk

- *London-Lund Corpus (LLC) (500 000 ord)
- Lancaster/IBM Spoken English Corpus (SEC) (80 000 ord)
- Corpus of London Teenage Language (COLT) (500 000 ord)
- Wellington Corpus of Spoken New Zealand English (WCS) (1 miljon ord)
- The International Corpus of English – East African Component (500 000 ord)

Historisk engelska

- *The Helsinki Corpus of English Texts (1.5 miljon ord)
- The Helsinki Corpus of Older Scots (800 000 ord)
- Corpus of Early English Correspondence, Sampler (450 000 ord)
- The Newdigate Newsletters (750 000 ord)
- Lampeter Corpus of Early Modern English Tracts (1.1 miljon ord)
- Innsbruck Computer-Archive of Machine-Readable English Texts (ICAMET)

Satsdelkodad engelska

- Polytechnic of Wales Corpus (barnspråk, 61 000 ord)
- Lancaster Parsed Corpus (LOB)

Bland andra korpusar finns följande:

- British National Corpus, 100 miljon ord (BNC) (90 skriftligt / 10 talat språk)
- British Component of the International Corpus of English (ICE-GB) (1 miljon ord)
- New Zealand Component of the International Corpus of English (ICE-NZ) (1 miljon ord)
- International Corpus of Learner English (ICLE) (2 miljon ord, 14 olika modersmål)
- Uppsala Student English Corpus (USE) (1.2 miljon ord)
- Dictionary of Old English Corpus (3.5 miljon ord)
- York-Helsinki Parsed Corpus of Old English Poetry (79 000 ord)
- York-Toronto-Helsinki Parsed Corpus of Old English Prose (1.5 miljon ord)
- Penn-Helsinki Parsed Corpus of Middle English (två versioner med 510 000 ord och 1.2 miljon ord)
- Penn-Helsinki Parsed Corpus of Early Modern English (1.8 miljon ord)
- Middle English Medical Texts (MEMT) (500 000 ord)
- Zurich English Newspaper Corpus (ZEN) (1.6 miljon ord)
- Corpus of Early English Correspondence (CEEC) (2.7 miljon ord)
- A Corpus of English Dialogues 1560–1760 (CED) (1.3 miljon ord)

Ett antal korpusar finns tillgängliga on-line via internet, t.ex.

- The English-Norwegian Parallel Corpus (ENPC) (<http://www.hf.uio.no/iba/prosjekt/>) (lösenord krävs)
- The COLT Corpus (Corpus of London Teenage Speech): <http://kh.hd.uib.no/tactweb/colta.htm>
- CobuildDirect demo: <http://titania.cobuild.collins.co.uk/form.html>

c) Linguistic Data Consortium (<http://www ldc.upenn.edu/>) erhåller över 40 000 korpusar på flera språk. De tio mest populära korpusarna (december 2006) behandlar engelska, två av dem i ett multilingvalt sammanhang:

[LDC93S1](#) / TIMIT Acoustic-Phonetic Continuous Speech Corpus = talspråk / engelska

[LDC96L14](#) / CELEX2 = lexikon / holländska, tyska, engelska

[LDC93S10](#) / TIDIGITS = talspråk / engelska

[LDC94T5](#) / ECI Multilingual Text = skriftligt språk / albanska, bulgariska, kinesiska, engelska, m.m.

[LDC93S2](#) / NTIMIT = talspråk / engelska

[LDC93T3A](#) / TIPSTER Complete = skriftligt språk / engelska

[LDC99T42](#) / Treebank-3 = skriftligt språk / engelska

[LDC94S16](#) / YOHO Speaker Verification = talspråk / engelska

[LDC2001T02](#) / Message Understanding Conference (MUC) 7 = skriftligt språk / engelska

[LDC2000S85](#) / Santa Barbara Corpus of Spoken American English Part-I = talspråk / engelska

Lingvistisk information kan hämtas från korpusarna med hjälp av sökprogram. Bland populära programspaket för PC datorer finns WordSmith

(<http://www.lexically.net/wordsmith/>) och MonoConc (<http://www.ruf.rice.edu/~barlow/mono.html#mono>). Inom unix-systemet kan man använda flera programmeringsspråk och verktyg för att lägga fram ordlistor, kondordanser m.m.

Ytterligare och mer detaljerad information avseende de olika engelskspråkiga och andra språkkorpusarna kommer att finnas i *Corpus Linguistics: An International Handbook* (HSK, Handbücher zur Sprach- und Kommunikationswissenschaft / Handbooks of Linguistics and Communication Science) redigerad av Anke Lüdeling och Merja Kytö (Berlin och New York, Mouton de Gruyter, under tryck).

(2) Andra strukturerade textsamlingar

Dessa textsamlingar består av texter som man har matat in i datorminnet utan att begränsa antalet ord eller att särskilt tänka på vad som skulle vara relevant för lingvistiska analyser. Vi kan ta några exempel på dessa.

- EEBO = Early English Books on Line består av de över 125 000 böcker som trycktes i England 1475–1700 dvs. från Caxton genom Spencer, Shakespeare och inbördeskriget till slutet av 1600-talet. Böckerna har skannats in som facsimile-bilder vars text inte kan sökas med datasökningsprogram. Men många av dessa böcker skulle kunna användas som sekundärkällor, t.ex. tidiga grammatikor. Lyckligtvis har 25 000 böcker i samlingen matats in som sökbara texter med hjälp av ett pågående massivt Text Creation Partnership samarbetsprojekt. Dessa är mycket värdefulla källor trots att forskaren måste själv beakta externlingvistiska parametrar så som publikationsår, författarna m.m. Det är också en viss begränsning att man inte angivit antalet ord per text i textlistan. EEBO erbjuder ett konkordansprogram som kan användas för att hämta exempel på ord och uttryck i materialet.

EEBO har en fortsättningsdel,

- ECCO = Eighteenth Century Collections Online. Ändamålet är att lägga fram en datoriserad version av alla betydelsefulla publikationer som tryckts i England på engelska eller på ett annat språk på 1700-talet, med tusentals viktiga arbeten tryckta i Amerika. Det gäller arbeten av välkända eller mindre kända författare, prediktor, reklamblad, musik, allt möjligt.
- Lion = Literature Online (Chadwyck-Healey) erbjuder över 750 000 litterära verk bestående av poesi, skönlitterär prosa och drama med en referensbibliotek av litteraturkritik. Denna är den största sökbara litterära och litteraturkritiska databasen i världen. Samlingen har över 16 500 författare med över 100 olika nationaliteter. Engelska är representerad av över 350 000 poesi-, drama- och prosaverk från 700 till 2000-talet. Lion erbjuder 14 textsamlingar och korshänvisningar finns till över 8000 andra elektroniska texter på internet sidor. Bland problemen med den bifogade sökmotorn är att den inte kan behandla stavningsvarianter på ett bekvämt sätt och att som sammanhanget får man bara några ord runt sökordet. Beläggen listas på ett mycket opraktiskt sätt så att det inte går att processera mer än ett tiotal exempel på en skärm.

- The Proceedings of the Old Bailey, London 1674 to 1834 (<http://www.oldbaileyonline.org/>). I denna sökbara samling (över 40 miljoner ord) finns 101 102 rättegångsprotokoll avseende åren 1674–1834 från Old Bailey domstolen i London. Men sökbarheten är begränsad för lingvistiska ändamål (jmf Lion). Professor Magnus Hubert (Giessen universitet, Tyskland) håller på att bearbeta en version annoterad för talare, deras kön, ålder m.m.
- Tidningar (<http://www.chadwyck.co.uk>): *The Guardian*, *The Times* (och *The Sunday Times*), *The Independent* (och *The Sunday Independent*), *The Daily Telegraph* (och *The Sunday Telegraph*), *The Financial Times* och *The Economist*. Problemet med tidningssamlingar är ofta att vissa frekventa ord som t.ex. modala hjälpverb, konjunktioner och negationsord har uteslutits från sökningar. Dessa skulle vara av stort intresse för många lingvistiska undersökningar.

(3) Elektroniska ordböcker

Som ett exempel kan vi ta den för engelska språket centrala ordboken *Oxford English Dictionary* (OED) som numera vanligen finns on-line vid universitets- och andra biblioteken. Sökningar kan utföras genom hela ordboken, dvs. man kan använda exempelaterialet som text. Detta har givit utrymme till ett nytt område, ”dictionary linguistics”, som har sina för- och nackdelar. Vi vet t.ex. att vissa perioder är magert representerade i OED vilket kommer att påverka sökresultat.

(4) Internet / WWW

Bland de senaste upptäckterna finns internet eller WWW (World Wide Web) som framstår som en mer eller mindre obegränsad och konstant växande källa av språkmaterial. Denna källa har både hyllats och kritiserats av språkforskarna. Det är svårt (eller oftast omöjligt) att få information avseende språkbrukarnas bakgrund. Lika svårt är det att bemästra den heterogena genrevariationen som är typisk för internettexter. Ett område som har väckt intresse är om det använda språket är mer likt talspråket eller skriftligt språk. Engelska dominerar på nätet (med 70% i jämförelse med de två nästa språken, japanska och tyska, både med närmare 6%). Men oftast är det inte möjligt att veta vilken geografisk variant som det gäller, dvs. brittiskengelska, irländskengelska, amerikanskengelska, eller varianter som används i Australien, Nya Zeeland, Syd-Afrika, Indien osv.

Två saker bör betonas avseende ett fruktbart och smidigt bruk av alla ovannämnda källor. För det första räcker det inte med korpuser och andra textsamlingar. Vanligen behöver man hjälp från IT-experter för att utnyttja materialet. De traditionella sökprogrammen är ganska begränsade i sina funktioner. Ofta behöver man programmeringshjälp för att kunna begränsa sökningarna på ett sätt som inte är möjligt med paketprogram. För det andra kräver det också tid och medel att bygga upp och underhålla en miljö där man kan använda korpuser. Vissa korpuser och sökprogram lämpar sig för nätbruket men inte alla. Vissa korpuser kan användas av alla, andra finns enbart i in-house bruk. Vissa korpuser kräver uppdateringar, andra inte, osv.

[2] Hur utnyttjas tillgängliga databasresurser? Av enskild forskare? Grupper?

Vanligen av enskilda forskare men också av grupper. T.ex. i Växjö finns det en aktiv GramTime projektgrupp som bearbetar frågor inom modern grammatik. Korpusar och textsamlingar används både i forskning och undervisning.

[3] Finns erfarenhet av samverkan och forskningssamarbete gruppvis/nationellt kring uppbyggnad av nu existerande databaser och har det då även handlat om att samarbeta tvär- eller mång-disciplinellt?

I nationellt perspektiv kan man notera samverkan mellan Göteborgs och Lunds universitet där man har arbetat fram ett antal viktiga korpusar, t.ex. English-Swedish Parallel Corpus (ESPC) kompilerat av Bengt Altenberg (Lund) och Karin Aijmer (Göteborg). Denna textsamling är en systerkorpus till English-Norwegian Parallel Corpus (ENPC), en subkorpus av Oslo Multilingual Corpus (OMC) som har kompilerats av Stig Johansson och hans projektgrupp vid Oslo universitet.

I internationellt perspektiv kan nämnas t.ex. Swedish Component of the International Corpus of Learner English corpus (ICLE) kompilerad av Bengt Altenberg och Swedish Component of the Louvain International Database of Spoken English (LINDSEI) kompilerad av Karin Aijmer. Inom engelsk språkhistoria har man arbetat på ett antal internationella projekt. A Corpus of English Dialogues 1560–1760 (CED) som kompilerades av Merja Kytö (Uppsala universitet) och Jonathan Culpeper (Lancaster University) består av 1.3 miljoner ord och har som syfte att ge material för forskning inom variationsstudier och historisk pragmatik. Det tidiga talsspråket är i fokus: korpusen innehåller domstolsprotokoll, vittnesmål, dramatexter, fiktionsprosa och manualer i handboksformat. Som ett ytterligare exempel kan nämnas A Corpus of Nineteenth-century English (CONCE) som består av ett miljon ord och täcker sju genrer: brev, fiktionsprosa, historieskrivande, naturvetenskapliga texter, dramatexter, domstolsprotokoll och parlamentariska debatter. Korpusen är strukturerad i tre subperioder för att spegla språket i början, mitten och slutet på århundradet; den kompilerades av Merja Kytö (Uppsala universitet) och Juhani Rudanko (Tammerfors universitet). I arbetsgruppen tillhör numera också Erik Smitterberg (Stockholms universitet).

[3] Vilken roll spelar befintliga databaser vid arkiv, bibliotek, museer, mm för forskningen inom ämnet?

Engelska finns som huvudämne vid 24 lärosäten i Sverige; forskarutbildning ges vid 10 lärosäten. Engelskspråkiga korpusar används av professorer, forskarassistenter, lektorer, doktorander och studenter inom engelsk språkvetenskap men också av allmänna lingvister och språkteknologer.

[4] Finns behov för forskningen av digitalisering av material som förvaras vid arkiv, bibliotek, museer, mm?

Det finns stora mängder av mycket intressant material på arkiv och biblioteken i England, ofta i manuskriptform eller i tidiga imprints. Dessa dokument är synnerligen värdefulla texter, inte bara för språkvetare men också för historiker. Som exempel på ett pågående digitaliseringsprojekt kan nämnas ”English witness depositions 1560–1760: An electronic text edition” som syftar mot att transkribera och digitalisera vittnesmål från den nymoderna England (projektet finansieras av Vetenskapsrådet och är aktivt vid Uppsala universitet). Att dylika texter inte har använts i forskning till en stor grad beror på att det inte finns

elektroniska utgåvor som skulle göra olika datasökningar möjliga. Dessa dokument ingår inte alls i Old Bailey, EEBO eller andra datoriserade textsamlingar.

[5] Vilka framtida behov finns i övrigt vad gäller databasuppbyggnad för att upprätthålla hög kvalitet och bidra till att forskningen inom ämnet kan vidareutvecklas?

Följande kärnområden kan nämnas:

- 2000-tals engelska (tal och skrift)
- Inlärarengelska
- Omfattande strukturerade korpusar genom århundradena täckande engelskans olika regionala former
- Annotering av de befintliga och nya korpusarna för effektivare sökningar: lingvistiska och externlingvistiska särdrag
- Utveckling av sökprogram för att kunna behandla stavningsvariation i historiska texter

[6] Finns vetenskapligt intressanta utvecklingslinjer och förnyelsemönster, som inte skulle kunna förverkligas utan tillgång till nya databasresurser?

Följande exempel kan ges:

- Variation och pågående förändringar i engelska språket och i dess regionala former
- Språkinlärningsproblematik
- Engelskans historia genom århundradena, t.ex. talspråket i det förflutna (utveckling av artighetsformulärer, konversationsrutiner, talkaktssärdrag m.m.)

[7] Vilka tänkbare möjligheter finns att samverka i forskargrupper eller i andra konstellationer nationellt och/eller internationellt för att bygga upp och utveckla nya databaser med bred användning i forskningen?

Ett internationellt nätverk av forskare inom engelsk korpuslingvistik har vuxit fram runt ICAME som anordnar årliga ICAME konferenser och publicerar *ICAME Journal* tidskriften. Dessa ger utmärkta möjligheter för nationellt och internationellt växelverkan som i sin tur kan leda till samarbetsprojekt.

[8] Kan tillgång till och användning av databaser påverka forskningens synlighet och närvaro i ett internationellt perspektiv?

Svaret på frågan är 'ja' då redan namnet på databasen kan användas för att synliggöra arbetet, om man så vill. Ett antal forskningscentra har blivit välkända för korpusarna som har utvecklats vid institutionerna (t.ex. Lancaster/Oslo-Bergen Corpus, Freiburg LOB, Helsinki Corpus of English Texts, Uppsala Student English Corpus). Kompilationsprojekt kräver ofta flera år, vilket bidrar till att associationerna blir kraftiga. Man kan också utveckla teoribasen genom att ta fram korpusar och databaser, t.ex. med termen "Helsinki School" hänvisar man till variationsforskning som använder empiriskt material och försöker lägga fram utvecklingstendenser i språket i ljuset av de parametrar som har använts i byggandet av databaserna.

Inge Bartning
 Institutionen för franska, italienska och klassiska språk
 Stockholms universitet
 (januari 2007)

Databaser i romanska språk

I det följande görs en kort översikt av de databaser som används inom forskningen i de tre romanska språken franska, italienska och spanska i Sverige och internationellt. Rapporten bygger på en informell enkät som besvarades av ett antal kolleger i romanska språk ute i landet inför det s.k. DISC-seminariet på Vetenskapsrådet den 12 december 2006. Översikten är inte uttömmande.

Dispositionen är följande:

- A. Databaser som har etablerats av svenska forskare och används i Sverige i franska, italienska och spanska vid Stockholms universitet, Lunds universitet och Uppsala universitet,
- B. Internationella korpusar i de tre språken.
- C. Problemområden

A. Databaser som etablerats och används i Sverige

1. Franska

Stockholms universitet

Institutionen för franska, italienska och klassiska språk

Skrivet språk

- Tidningskorpusen **COSTO I** och **II** (*Le Corpus de Stockholm*): 26 nummer av *Le Monde*, 26 nummer av *L'Express* från 1987 och tio år senare, 1997, ca 2 miljoner ord, etablerad av Gunnel Engwall och Inge Bartning. Korpusen används för avhandlingar, C- och D-uppsatser.

- **Romankorpus**: 25 bestseller-romaner från 1960-talet för vokabulär- och frekvensstudier, t. ex. *Dictionnaire des fréquences*, 1984, Gunnel Engwall, ca 500 000 ord.

- **EU-texter**: *Journal officiel des Communautés européennes, Traité de Maastricht* (använd för fil.lic.-uppsats av Gunilla Allstig).

- **Textutgivning, filologi:**

Vid institutionen finns en textkorpus med medelfranska texter från 1300- och 1400-talen som bygger på texter som forskare vid institutionen gett ut (Anders Bengtsson, Per Förnegård). Denna korpus skall kopplas till den franska *Base de français médiéval* (BFM), Lyon (www.bfm.ens-lhf.fr). Eftersom filologi och textutgivning i franska numera endast finns vid Stockholms universitet, har man inte något formellt nationellt samarbete, däremot finns tvärinstitutionella vid SU och internationella kontakter.

Talspråk

● Korpusen **Talad mediafranska** (*Français parlé des médias, FPM*), etablerad av Mats Forsgren, Matthias Broth; innehåller ca 30 timmar talat material, bl. a. nyhetsprogram, rundabordsamtal, polisserier.

Videoinspelningar: TV-program, inspelningar av TV-produktioner (Broth).

Behov: resurser för taggning, inköp av utrustning för inspelningar och lagring av digitala video- och ljudfiler.

● **InterFra-korpusen** (*Interlangue française – développement, interaction et variation*) har etablerats under 1990-2006 av Inge Bartning samt medarbetare. Den innehåller sex delkorpusar varav större delen är longitudinella: 10-20 nybörjare (inspelade under 2 terminer), 13-20 gymnasister (inspelade 2 ggr under 2 år), 38 universitetsstuderande (varav 8 följts under 4 terminer), ämneslärare (inspelade 4 ggr under 2 terminer), 9 doktorander; 20 infödda. Uppgifter: intervjuer samt återberättelser av bildserier och videofilmer. Korpusen är transkriberad och har etiketterats med grammatiska annotationer (ännu inte fullständigt). Man har hittills använt PCBeta/Tagger-programmen av Benny Brodda och konverterar nu till xml-format och prövas i ITG-miljö av Språkbanken i Göteborg (Lars Borin); ca 500.000 ord; se vidare www.fraita.su.se. Tanken är att korpusen skall bli tillgänglig via lösenord. Uppbyggnaden av denna korpus har skett med hjälp av medel från dåvarande HSFR och senare, Vetenskapsrådet samt av Humanistiska fakulteten vid Stockholms universitet.

Andra korpusar inom InterFra-gruppen:

● **Corpus Paris** (2006): inspelningar med 10 tvåspråkiga informanter bosatta i Frankrike sedan 20-25 år, samma uppgifter som i InterFra-korpusen (Inge Bartning, Nathalie Kirchmeyer, Fanny Forsberg). Finansiering av Vetenskapsrådet.

● **Corpus Forsberg** (2005-2006): 11 'naturliga' svensktalande inlärare samt 6 infödda, 23-33 år; intervjuer och narrationer (Fanny Forsberg, avhandlingsarbete).

● **Corpus Hancock** (2004): 14 svenska Erasmusstudenter i Paris, 20-29 år, franska sedan 4-14 år, intervjuer och narrationer (Victorine Hancock, för forskning som post-doc).

Dessa franska inlärarkorpusar finns som grund för forskningen inom programmet *Avancerad andraspråksanvändning (High Level Proficiency in SLA)*, finansierat av Riksbankens Jubileumsfond, 2006-2009) tillsammans med Kenneth Hyltenstam, Centrum för tvåspråkighet och Lars Fant, Spanska, SU; se vidare www.biling.su.se/~AAA.

Behov: resurser för fortsatt datorarbete med taggning, konvertering och sökverktyg kommer att behövas i framtiden.

Lunds universitet

Institutionen för romanska språk

Talspråk

● **Corpus Lund**: 8 vuxna inlärare, s.k. naturliga inlärare i Paris, samt 7 universitetsstuderande och 5 vuxna nybörjare; longitudinell uppläggning; korpusen är etablerad av Suzanne Schlyter under 1980-1990-talen och den finns i CHILDES/CLAN.

● **Barnspråskorpus**: tvåspråkiga fransk-svenska barn (Suzanne Schlyter); materialet finns i d-base, 1980-1990; DUFDE-projektet, HSFR, Vetenskapsrådet.

● **Direkt Profil:** (Jonas Granfeldt, Suzanne Schlyter, Pierre Nugues, 2005 - 2006), diagnostiskt dataprogram för stadieplacering i franska L2, tal och skrift, självvärderingsinstrument; potentiellt verktyg även för andra språk för vilka man har inlärningsgångar (se vidare: <http://project.sol.lu.se/DirektProfil>); finansierat av Vetenskapsrådet.

Dessutom har följande korpusar använts vid avhandlingsarbeten i franska i Lund:

Talspråk

- **Corpus Westin** (2003, Eva Westin, avhandlingsarbete): samtal mellan avancerade inlärare och infödda (CLAN).
- **Corpus Thomas** (2006, Anita Thomas, avhandlingsarbete): muntligt material från bl. a. vuxna nybörjare i franska (CLAN).

Skrivet språk

- **Corpus Gunnarsson** (2006, Cecilia Gunnarsson, avhandlingsarbete): gymnasisters skriftliga uppgifter inskrivna i Scriptlog samt s.k. tänka-högt-protokoll (CLAN).
- **CEFLE** (*Corpus écrit de français langue étrangère*): etablerad av Malin Ågren, 2005, för avhandlingsarbete; ca 400 skriftliga återberättelser av gymnasiestuderande i franska (CLAN).

Uppsala universitet

Institutionen för moderna språk

Skrivet språk

- **CUCF** - *Le Corpus d'Uppsala du Conditionnel français* (Hans Kronning och Coco Norén): 5000 konditionalisförekomster i romaner, teaterpjäser, dagstidningar, vetenskaplig prosa; 1700-, 1800-, 1900- och 2000-tal, motsvarande ca 30 böcker.
- **CCUCFI** - *Le Corpus Contrastif d'Uppsala du Conditionnel français et italien* (Hans Kronning): 1000 konditionalisförekomster i franska och italienska romaner från 1900-talet; motsvarar ca 8 böcker.
- **Corpus de journaux francophones** (Mats Forsgren): dagstidningar från Belgien och Schweiz, 49 nummer, 1980-90-talen.
- **Fransk-svensk parallellkorpus** (Carina Andersson): fransk-svensk, 2 miljoner ord, romaner från 1900-talet.
- Två meningslänkade **översättningskorpusar** franska – svenska och en korpus med parallelltexter (franska – svenska); en av dessa i samarbete med Romanska institutionen, Göteborgs universitet och Institutionen för lingvistik vid Uppsala universitet.

Uttryckt behov: att etablera en panromansk korpus bestående dels av meningslänkade originaltexter och översättningar, dels av jämförbara originaltexter. Studier baserade på denna typ av material är nästan obefintliga i romanska språk.

Talspråk

- **Dialogues entre jeunes Français** (Coco Norén): informella tvåpartssamtal på franska, 5,5 timmar, 1990-tal.

- **The Uppsala French TAP-corpus:** s.k. tänka-högt-protokoll, etablerad 1993-95 av Kerstin Jonasson, ca 50 sidor

2. Italienska

Stockholms universitet

Institutionen för franska, italienska och klassiska språk

Skrivet språk

- **Primo Levis** hela prosaproduktion (Jane Nystedt): ca 750 000 ord.
- **EU-dokument** från 1988-1992 (Jane Nystedt).

Talspråk

- **InterIta**, italiensk inlärningskorpus med början 2001- (Camilla Bardel): 21 svensktalande studenter på universitetsnivå, från 3-6 tillfällen, sammanlagt 65 inspelningar (intervjuer, återberättelser); 10 infödda talare (Erasmus), sammanlagt 30 inspelningar (CLAN). Planer på skriftligt material som skall digitaliseras, mer talspråksmaterial samt parallellkorpus tal och skrift; finansiering av Vetenskapsrådet. Forskningsprojekt om inläring av 'lexikon' på avancerad nivå ingår i RJ-programmet *Avancerad Andraspråksanvändning* (2006-2009); samarbete med det franska InterFra-programmet vid samma institution. Resursbehov: finansiering av uppläggning av skriftligt parallellmaterial.

Uppsala universitet

- **CCUCFI** (konditionalis) se Franska, Uppsala Universitet, ovan.

3. Spanska

Stockholms universitet

Institutionen för spanska, portugisiska och latin-amerikanska studier

Talspråk

- **AKSAM-korpusen** (Aktivitetstyper och samtalsstruktur hos infödda och icke-infödda talare av spanska), ansvariga: Lars Fant, Johan Gille, AKSAM-SWEDEN: enbart svenskspråkiga samtal, 5 timmar, AKSAM-SPAIN: spanskspråkiga samtal, 6 timmar, AKSAM-CHILE: spanskspråkiga samtal, 10 timmar, AKSAM-BILING: samtal på spanska mellan svenskar och spanjorer, 13 timmar, AKSAM-LUNCHTALK: både spansk- och svenskspråkiga samtal, 2 timmar, AKSAM-NEGOTIATION TALK: innehåller förutom spanskspråkigt material även förhandlingar på svenska och danska, 8 timmar, AKSAM-INTERCULTURAL INTERVIEWS: innehåller förutom spansk-språkigt material även intervjuer på svenska och danska, 4 timmar; totalt omfång, 48 timmar.
- **Korpus Österberg:** longitudinell korpus av ca 20 gymnasister, under 3 år (20>9>5), intervjuer, flerpartssamtal, narration (ca 10 timmar); djupintervjuer och enkäter, aptitudtest och motivationella variabler (2005-2006, Rakel Österberg, avhandlingsarbete).

Uppsala universitet

- **Corpus des periodicas hispanos** (Mats Forsgren): spanska dagstidningar från 1980- och 1990-talen, 116 nummer.

Sammanfattningsvis kan sägas att dessa databaser spelar en mycket stor roll för forskningen i våra språk. De utgör kärnan i forskningen och används av professorer, forskarassistenter, lektorer, doktorander och studenter inom romansk språkvetenskap vid alla nämnda institutioner.

4. Övrigt i Sverige

Adresser som kan vara av intresse:

- **ECHO** (European Cultural Heritage On line), Sven Strömqvist, Lunds universitet (<http://www.mpi.nl/echo/>).
- **DAM LR** (Distributed access management for language resources) Sven Strömqvist, Lunds universitet (<http://www.mpi.nl/DAM-LR/>); Dessa projekt syftar till att underlätta distribution, rättigheter mm kring elektroniskt material för forskning och annat.

Vid **Språkbanken**, Göteborgs universitet (Lars Borin) (<http://spraakbanken.gu.se/>) finns bl.a. följande databaser av intresse för våra språkinstitutioner:

- Spansk korpus: Spanish Online
- **ASU**-korpusen (Allmän språkutveckling), Lingvistik, SU, Björn Hammarberg, longitudinell korpus av talad svenska L2, 10 informanter med olika L1, inspelade 10 ggr under 30 mån.
- **SUC** (Stockholm-Umeå korpusen)
- **SSM** (Svenska som målspråk, Björn Hammarberg, Åke Viberg), skriftligt inläromaterial
- **IT-baserat kollaborativt lärande i grammatik** (ITG): samarbetsprojekt mellan Uppsala (lingvistik och filologi, informationsteknologi), Stockholm (lingvistik) och Göteborg (svenska språket). Projektet syftar till utveckling av ett datorbaserat stöd för undervisning i grammatik på universitetsnivå. Projektledare Lars Borin. Projektid: 2002–2004.

B. Internationella databaser

1. Franska

Skrivet språk

- **FRANTEXT** är den stora franska databasen för franska texter på 210 milj. ord från 3637 olika texter alltifrån den klassiska litteraturen till icke-fiktiv prosa och facklitteratur. Fem sekel (1500-1900-tal) och en mängd olika genrer finns representerade, t. ex. romaner, poesi, teaterpjäser, tidningstext, essäer och brevkorrespondans. Delar av korpusen är grammatiskt kodade (1940 texter). Korpusen är tillgänglig på nätet, men kräver abonnemang (se vidare www.atilf.fr , www.frantext.fr).

- **BTMF** (*Base textuelle du Moyen français*), delkorpus av FRANTEXT över medeltidens litteratur.

● **TLF** (*Trésor de la langue française*), elektroniskt lexikon, som egentligen var målet för FRANTEXT-korpusen; kan bl. a. användas tematiskt.

Andra elektroniska lexikon

● **Dictionnaire de l'Académie française** (4^e, 8^e och 9^e uppl.), Franska Akademiens lexikon tillgänglig på nätet i olika upplagor.

● **FRIDA** (*French Interlanguage Data Base*), skriftlig korpus av intermediära till avancerade inlärare etablerad av Sylviane Granger, Université de Louvain. Vid samma universitet har den välkända korpusen **ICLE** (*International Corpus of Learner English*) också etablerats av Sylviane Granger, (obs! engelska).

Talspråk

● **C-ORAL-ROM** (2005) DVD med talspråk i de romanska språken (franska, italienska, spanska, portugisiska) som modersmål (se vidare Cresto & Moneglia (utg.) *Integrated Reference Corpora for Spoken Romance Languages*. Amsterdam: Benjamins).

● **FLLOC** (*French Learner Language Oral Corpora*) (S. Rules 2004), inlärarkorpusar i franskt talspråk, fri för användning på webben: (<http://www.flloc.soton.ac.uk>), 1.400 ljudfiler och transkriptioner, från 11-åriga nybörjare till universitetsstuderande; flera korpusar etablerade i Storbritannien, allt i CHILDES/CLAN.

● **ELICOP**, databas vid Louvains universitet, Belgien, franskt inlärarspråk och talad infödd franska i Belgien. (<http://bach.arts.kuleuven.ac.be/elicop>). För många andra internationella inlärarkorpusar med franskt talspråk finns en förtjänstfull översikt i *Journal of French Language Studies* 14:3, 2004, översiktsartikel av Sarah Rule, *French interlanguage oral corpora*.

● Corpus **GARS** (*Groupe aixois de Recherche en syntaxe*, Université de Provence), referenskorpus för franskt talspråk (*Corpus de référence du français parlé*, CRFP), tidigare under ledning av Claire Blanche Benveniste; numera heter databasen och forskargruppen ● **DELIC** (*Description Linguistique Informatisée sur Corpus*) och leds av Jean Véronis (<http://www.up.univ-mrs.fr/delic/corpus/index.html>).

● Corpus **PFC** (*Phonologie du français contemporain*), internationellt projekt med talspråkskorpus från stora delar av den frankofona världen; ansvariga B. Laks, C. Lyche och J. Durand vid universitet i Paris X och Toulouse (<http://infolang.u-paris10.fr/pfc/>).

● **CLAPI** (*Corpus de langues parlées en interaction*), projekt om talspråk och interaktion under ledning av Lukas Balthasar, Université de Lyon 2. Databasen CLAPI.1 on line (<http://corpusgric.univ-lyon2.fr:8000/corpus/>).

På hemsidan till *École Normale Supérieure* (ENS-LSH) finns flera andra franska korpusar (<http://www.ens-lsh.fr/>).

Oxford Text Archive (<http://ota.ahds.ac.uk/>): detta arkiv innehåller, utöver alla engelska korpusar, 34 franska.

2. Italienska och spanska

Italienska

Skrivet språk

- **Biblioteca Italiana Telematica (CiBIT)**(<http://www.bibliotecaitaliana.it>); litterära texter med gratis tillgång på webben on line.
- **Corpus di Italiano Televisivo (CiTDemo)**(<http://www.sspina.it/>), transkriberat TV-språk av olika slag.
- **LIZ - Letteratura Italiana Zanichelli** (LIZ 4.0, cd-rom med sökmotor DBT), (Stoppelli Pasquale & Picchi Eugenio, utg.), äldre litterärt språk.

Talspråk

- **LIP** (*Lessico di frequenza dell'italiano parlato*, http://languageserver.uni-graz.at/badip/badip/20_corpusLip.php), italienskt frekvenslexikon över talad italienska, även med regional variation..
- **Corpus di Italiano Parlato** (a cura di E. Cresti), 2000 (versione cartacea+cd-rom), [http://lablita.dit.unifi.it.](http://lablita.dit.unifi.it;); se under Franska, C-ORAL-ROM.

Spanska

- Spanska: Mark Davies, Brigham Young University (<http://www.corpusdelespanol.org/>).
- **CREA** (*Corpus de Referencia del Español Actual, RAE*): Spanska akademiens stora korpus med litterära, fackspråkliga och talade delkorpusar (<http://corpus.rae.es/creanet>).
- **CORDE** (*Corpus diacrónico del español*); diakronisk korpus.
- **NORMA CULTA** (*Norma lingüística culta de la lengua española*); korpus över talad spanska i sex huvudstäder: Madrid, Bogotá, Buenos Aires, Caracas, Santiago de Chile, Mexico City; ansvarig: Umberto Lopez Morales.

C. Problemområden, samverkan och behov

Problem

Generellt kan sägas om de korpusar som etablerats och används i Sverige (se sektion A. ovan) att det är forskare knutna till ett visst projekt som utnyttjar datamaterialen. Ytterst få korpusar är tillgängliga på nätet. Intresserade utifrån får vända sig till projektledarna. Ett annat problemområde är copyright-frågor och medgivande från de inspelade personerna. Utvecklingen har gått så fort att man inte alltid hunnit med i den nya tekniken, vilket innebär att medgivande av medverkande personer från tidigare år för traditionell vetenskaplig publicering inte gäller t.ex. Internetpublicering. Idag väljer flera forskare att använda system med lösenord för åtkomst av databaser. I många forskargrupper väntar man tills forskningen inom gruppen är klar innan man ger tillstånd till utnyttjande.

Ett annat problem är att materialen ofta behöver uppdateras p.g.a. föråldrad lagringsteknik och man söker moderna standardiserade textformat. Metodologin för taggningsprogram har också utvecklats. Materialet och kodningstekniken kanske inte är beskrivet på ett begripligt sätt så att senare forskare lätt kan hantera korpusen. Det visar sig att det inte räcker med att språkforskare etablerar korpusar. Man behöver också hjälp av IT-experten för sökverktyg, standardisering och programmeringshjälp.

Beträffande skriftligt material, t. ex. skönlitteratur krävs villkor som har med den upphovsrättsliga lagstiftningen att göra. Lagstiftningen tycks oklar på många punkter.

Samverkan

Samverkan finns och utvecklas. Inlärningsforskarna i franska i Lund och Stockholm har samarbete i olika former och publicerar gemensamma resultat. I Stockholm har man samarbete mellan fyra institutioner (särskilt franska, italienska och spanska), tack vare RJ-projektet *Avancerad andraspråksanvändning* som nämnts ovan. Franskan både i Lund och Stockholm har täta kontakter och väl utarbetade nätverk med inlärningsforskare ute i Europa, men än så länge har vi inte lagt ut våra korpusar på nätet. Man kan säga att vår forskning blivit uppmärksammas internationellt mycket tack vare inlärarkorpusarna. Man har t. ex. erbjudit oss plats i FLLOC-nätverket (se ovan, under Franska, Internationellt). Gruppen kring *Talad mediafranska* i Stockholm har ett internationellt nätverk med bl. a. Canada. Projektet anordnade ett internationellt Wallenberg-symposium i Stockholm i juni 2006. Språkinlärningsgrupperna i franska och italienska i Stockholm har kommit med i ett nyligen påbörjat europeiskt samverkansprojekt initierat av Jan Hulstijn, Amsterdam. Projektet, kallat SLATE (*Second language acquisition and Testing in Europe*), har som syfte att finna lingvistiska kriterier som karakteriserar varje steg på den skala som Europarådet föreslagit i *The Common European Framework* för nivåbestämning av studenter och kurser. Den befintliga skalan bygger på funktionella och kommunikativa kriterier. Här kommer databaserna med inlärarspråk och utarbetade stadiindelningar väl till pass.

Behov i framtiden

Tekniska

- Mer resurser till de redan existerande korpusarna för komplettering av taggning och lagring av digitala video- och ljudfiler
- Annotering av de befintliga och nya korpusar för effektivare sökningar
- Standardisering av befintliga korpusar och beredskap för nya korpusar
- Utveckling av sökverktyg

Ämnesmässiga behov vad gäller utvecklingslinjer och förnyelsemönster som inte skulle kunna förverkligas utan tillgång till databasresurser:

- Franskans utveckling genom århundradena, i tal och skrift
- Variations- och genreforskning inom tal och skrift i franskan
- Andraspråksforskning
- Att etablera en panromansk korpus bestående dels av meningslänkade originaltexter och översättningar, dels av jämförbara originaltexter. Studier baserade på denna typ av material är nästan obefintliga i romanska språk.

Behovet av databaser för svensk litteraturvetenskap

Mats Malm

Fyra övergripande behov

De litteraturvetenskapliga behoven av databaser utgår främst från fyra funktioner: bibliografering, bevarande, tillgängliggörande, utgivningsverksamhet.

- *Bibliografering.* Varje forskningsområde är beroende av sina bibliografiska resurser. I litteraturvetenskap följer bibliograferingen i hög grad de språk vilka forskningen gäller – anglisterna har sina bibliografier, germanisterna sin, etc.
- *Bevarande* av artefakterna, i den meningen att man skapar digitala faksimil av äldre tryck, är en uppgift som främst ligger hos Kungl. biblioteket. Det är rätt plats för uppgiften även om det är en fördel om KBs verksamhet kompletteras på andra håll.
- *Tillgänglighet.* Ordet ”digitalisering” har en olycklig vaghet. Den digitalisering KB främst ägnar sig åt består i att göra digitala faksimil, men annars avses i regel OCR-bearbetning och uppmärkning av text. Båda formerna av digitalisering kräver stora resurser. Om vi skiljer mellan digitalisering som ger faksimil och digitalisering som ger e-text, är faksimilen av stor nytta eftersom de ger (om än begränsad) tillgång till den fysiska gestaltning en text haft. Men det är tillgång till e-text som är av störst allmän nytta för litteraturvetare.

Den litteraturvetenskapliga forskningen inbegriper en hel del filologi och kan inte alltid avgränsas tydligt mot språkämnen och historieämnena. Det finns några speciella inriktningar som har ett mycket konkret behov av databaser, men den litteraturvetenskapliga forskningens utvecklingspotential ligger inte minst i att dess material tillgängliggörs genom databaser.

- *Utgivningsverksamhet* har blivit en allt större del av litteraturvetenskapen på senare år. Den teoretiska diskussionen har blivit allt mer betydelsefull och problematiserat litteraturvetenskapliga grundantaganden om vad ett verk är. Den tekniska utvecklingen av digitala miljöer för editionsfilologi är en central fråga.

Därtill bör man lyfta fram databasens och webbplatsens potential som miljö för utbyte forskare emellan. Också databaser för texter främst på svenska språket har en stor potential för internationalisering. När vi har ett sökbart material som är så omfattande att man kan uppfatta det som representativt i viktiga avseenden, då kan vi göra parallella undersökningar på svenskt och utländskt material. Vi kan till och med gå i dialog med internationell forskning, utveckla metoder och frågeställningar med anspråket att bidra till forskningen i andra länder. Sådana undersökningar kan ha olika estetiska infallsvinklar eller luta mot filologi: i hög grad tror jag att man här skulle kunna använda litteraturen som en spegel av samhället. Man skulle t.ex. kunna se mentalitetsförändringar kring krigsutbrotten, följa nationalistiska stråk i olika tider, ställa samman olika sorters källor för att dels klargöra specifikt svenska frågor, dels relatera dem till internationella resultat.

Avgränsningar

Punkten *Bevarande* lämnar jag därhän som en fråga för biblioteken snarare än universiteten. Litteraturvetenskapens arbetsmaterial består, förstås, till stor del av skönlitterära texter, i regel på svenska språket. Sådana texter finns i språkdata-baser på ett antal olika håll i landet, men det är viktigt att notera att de ändå inte fyller många litteraturvetenskapliga behov. Språkkorpusar är främst tänkta för språkteknologisk och lingvistisk användning, medan litteraturvetare behöver se materialet i andra kontexter. Texterna behöver vara helt korrekta, dvs korrekturlästa, med precis information om författare, tillkomst- och utgivningssituation, upplaga, eventuella omarbetningar och varianter, med mera. Sökfunktionerna skall inte bara erbjuda konkordanser och index, utan i första hand kunna leda vidare in i hela texter. Därtill skall sökfunktionerna vara sådana att de, tillsammans med hela databasens upplägg och presentation, ger möjlighet till kontextualisering av den enskilda texten. Man skall kunna söka parallellställen både genom strängsökning och sökning efter t.ex. namn, platser, tidpunkter och helst ytterligare uppmärkning. Urvalet behöver i regel ske på helt andra grunder än i språkdata-baser: det handlar inte om att etablera en kanon, utan om att åstadkomma ett material som ger möjlighet till nya ingångar och undersökningar från nya perspektiv. Det litteraturvetenskapliga behovet av data-baser är alltså något helt annat än klassiker till allmänhetens nytta.

Litteraturvetenskapen har ibland utbyte av språkteknologisk forskning och frågeställningar som görs i anslutning till språkdata-baser, men hittills har de litteraturvetenskapliga intressena legat ganska långt i utkanten av sådana projekt. Språkteknologisk forskning och utveckling, kvalificerad sådan, finns på många håll (<http://www.sprakteknologi.se>). Humanistlaboratorierna i Umeå (<http://www.humlab.umu.se>) och Lund (<http://www.sol.lu.se/humlab>) är fruktbara platser för korsningar mellan humaniora och teknik. Så vitt jag vet har man inte hittills tänkt på textdata-baser i en litteraturvetenskaplig mening där. Det finns vissa inriktningar av litteraturvetenskap med teknologiska frågeställningar som går rätt in i den sortens miljöer, men det är undantag snarare än regel.

I det följande går jag alltså inte in på språkteknologi eller datalingvistik, utan försöker hålla mig till textdata-baser som direkt syftar till att erbjuda material åt en vidare uppsättning forskare.

Sex specifika behov

Som jag ser det, har svensk litteraturvetenskaplig forskning i första hand behov av:

- 1 Databaser som gör det möjligt att finna forskning.
- 2 Databaser som gör det möjligt att finna källtexter.
- 3 Databaser som tillgängliggör forskning.
- 4 Databaser som tillgängliggör källtexter.

5 Databaser som kan utveckla och utvecklas av editionsfilologin.

6 Databaser som erbjuder en arbetsmiljö över gränser.

Flera av dessa funktioner kan med fördel förenas i en databas. Situationen idag ser ut:

1 Databaser som gör det möjligt att finna forskning.

Någon internationell skandinavistisk bibliografi existerar inte, och numera inte heller någon vältäckande bibliografi över svensk litteraturvetenskaplig forskning eller forskning om svensk litteratur. Den uppgiften sköttes till för några år sedan genom *Svensk litteraturhistoriska bibliografi*, utgiven årligen av Svenska litteratursällskapet (ett lärt sällskap som också ger ut årsboken *Sammlaren*), en bibliografi som numera tyvärr är nedlagd av kostnadsskäl. Det säger sig självt att bristen på en adekvat bibliografi är ett allvarligt hinder för forskare och studenter.

Vissa resurser finns naturligtvis. Biblioteksdatan Libris förtecknar monografier i svenska forskningsbibliotek (och vissa folkbibliotek), medan databasen ArtikelSök förtecknar artiklar i tidningar och tidskrifter. (I bägge fallen sker också indexering, dvs de bibliografiska posterna förses med ämnesord.) En del universitet har också egna databaser med uppgifter om publicerade arbeten, men deras värde begränsas radikalt dels av att de bara gäller verksamheten vid det egna lärosätet, dels att de bygger på forskarnas egen rapportering, med allt vad det innebär av opålitlighet, luckor, eftersläpning och felinmatningar.

Enkelt uttryckt innebär detta bibliografiska läge att viss periodica och monografier är väl täckta, men att åtskilligt centralt material faller utanför. Forskning om svensk litteratur i internationella tidskrifter kommer inte med i Libris (viktigast och enklast att täcka är den handfull välrenommerade skandinavistiska tidskrifter som kommer ut i England, USA och Tyskland). Inte heller täcks en annan viktig kategori, nämligen enskilda bidrag i konferensvolym, forskningsantologier och andra samlingsverk där flera ämnen/verk/författare behandlas. Bristen är så mycket värre som litteraturvetenskapen i hög grad kommunicerar sina resultat via böcker, snarare än via de tidskriftsartiklar som snarast är normen inom andra discipliner. En litteraturvetenskaplig bibliografering som enbart riktar in sig på periodica fyller inte sitt uppdrag på ett adekvat sätt.

Bibliotekstekniskt kan läget beskrivas som att dessa publikationer behöver fördjupad bibliografering och indexering, dvs en beskrivning av varje bidrag/kapitel. Detta kunde med fördel skötas inom ramen för KB:s uppdrag som vetenskapligt nationalbibliotek och posterna skulle göras tillgängliga via Libris (där *Svensk litteraturhistorisk bibliografi* publicerades som digitalt sökbar resurs från 1993 fram till nedläggningen 1999.) Naturligtvis skulle detta kräva vissa extra resurser för KB, uppskattningsvis motsvarande 1/4 eller 1/3 bibliotekarietjänst.

2 Databaser som gör det möjligt att finna källtexter.

Biblioteksdatan, särskilt Libris, har byggts ut substantiellt, men där finns mycket kvar att göra. I synnerhet finns behov av kataloger för ett stort antal arkiv och

samlingar. I Libris har man, inom det projekt som förbereder utgivningen av Selma Lagerlöfs Samlade verk, börjat katalogisera manuskript av Selma Lagerlöf, och på samma sätt skulle väldigt mängder material behöva registreras. Arbetet skulle lämpligen utföras lokalt, men enligt Libris direktiv.

3 Databaser som tillgängliggör forskning.

I Sverige tillgängliggör Presstext kulturartiklar, men inget textarkiv för svensk litteraturvetenskaplig forskning finns. Ett antal internationella digitaliserade facktidskrifter är tillgängliga genom universitetsbiblioteken.

4 Databaser som tillgängliggör källtexter.

Här ligger Sverige långt efter utlandet. Vid Det Kongelige Bibliotek i Köpenhamn finns Arkiv för dansk litteratur (<http://www.adl.dk>) som tillgängliggjort en imponerande mängd danska ”klassiker”. Runt om i världen tillkommer fler och fler databaser för olika slags litteratur – flera av ansevärda dimensioner. Bland de mest förebildliga är Electronic Text Center (etext.lib.virginia.edu). Ett stort EU-projekt, i2010 Digital Libraries, handlar om att tillgängliggöra allt material på biblioteken i EU-området (europa.eu.int/information_society/activities/digital_libraries/index_en.htm).

De språkdata-baser som finns i Sverige fyller alltså inte litteraturvetenskapens behov. I Sverige startade Litteraturbanken (litteraturbanken.se) med stöd av Riksbankens Jubileumsfond 2004-2005. Litteraturbanken fungerar nu som en ideell förening, medlemmar är Kungl. biblioteket, Språkbanken vid Göteborgs universitet, Svenska Akademien, Svenska Vitterhetssamfundet och Vitterhetsakademien. Den skall tillgängliggöra digitala versioner av svenska litterära och vetenskapliga verk i original och översättning, samt vissa verk översätta till svenska. Verk skall kunna läsas och bearbetas on-line eller sparas ner till hårddisken som pdf-filer. Vissa verk kan också läsas som faksimil av förstaupplagorna. De första åren har främst ägnats förberedelser och grundarbete, databasen är ännu inte så stor men utvidgas efter hand. Litteraturbanken skall samla den litterära delen av det svenska kulturarvet, dels genom att själv samla in texter, dels genom samarbete med pågående utgivningsprojekt: Almqvist, Strindberg, Lagerlöf. 2006 startade Kungl. biblioteket sitt projekt Öppna samlingar. Litteraturbanken och Öppna samlingar kommer att kunna komplettera varandra, men resurser är en förutsättning. Just nu drivs de i hög grad med hjälp av Kulturrådets tvååriga Access-projekt.

Vid Riksarkivets Svenskt Diplomatarium finns rikliga mängder material varav en del är direkt relevant för vissa litteraturvetenskapliga inriktningar (<http://www.ra.se/ra/diplomat.html>). Det finns textdatabaser som är avsedda främst för lingvistik och språkteknologi, t.ex. i Språkbanken vid Göteborgs universitet (<http://spraakbanken.gu.se/>) där man kan göra sökningar i stora språkkorpora.

Litterära texter tillgängliggörs förvisso i högt tempo på Internet. Sådana databaser och webbplatser riktar sig i hög grad till den intresserade allmänheten, och kan vara till nytta också för grundutbildningen när t.ex. studenter behöver få tag i kurslitteratur (<http://www.runeberg.org>, <http://www.omnibus.se/eBoklagret>). Sammantagna erbjuder

sådana webbplatser erbjuder ett ganska stort antal texter, men de utgör en relativt osovrad mängd. Ofta är texterna moderniserade, ofta är de tagna från en sen utgåva utan att det anges vilken: det är oklart om den är förkortad eller utvidgad, om stavningen är originalets eller inte. Texterna är i regel inte korrekturlästa, vilket gör att forskaren inte kan använda dem utan brasklappar och inte kan citera direkt ur dem – däremot erbjuder Runeberg mycket användbara faksimil av hela böcker. Sökfunktionerna fyller inte forskarens behov, och korrekturfelen gör att resultaten av sökningar blir opålitliga.

5 Databaser som kan utveckla och utvecklas av editionsfilologin.

På flera håll används stora databaser för att inte bara tillgängliggöra utan också, i första hand, åstadkomma vetenskapliga utgåvor. Utgivningen av Henrik Ibsens Samlade skrifter i Oslo (<http://www.hf.uio.no/ibsenenteret/>) och Wittgenstein i Bergen (<http://gandalf.aksis.uib.no/wab/>) är två exempel, Søren Kierkegaard-institutet i Köpenhamn (<http://www.sk.ku.dk/>) ett annat. Gemensamt för dem är att de utvecklat sofistikerade tekniska lösningar för att åstadkomma en arbetsmiljö för editionsfilologer där samtidigt resultaten skall bli lätt tillgängliga för forskare och allmänhet.

I avsaknad av utgivningsprojekt som Norges och Danmarks Ibsen, Wittgenstein och Kierkegaard får den editionsfilologiska frontlinjen i Sverige beskrivas som det nordiska samarbetsprojektet Menota, Medieval Nordic Text Archive (<http://www.menota.org>). Menota är ett nätverk av nordiska forskningsinstitutioner och sällskap för textutgivning. Syftet är att bevara, förmedla och utge medeltidstexter i elektronisk form, både på folkspråken och latin. Hittills har man lagt stor vikt vid att utveckla en standard för uppmärkning av texterna, dels för att det egna arkivet skall bli homogent men också för att åstadkomma en standard som alla kan använda. I deras arbete har ingått nära kontakter med TEI, och man har varit mycket aktiva när det gäller att få in tecken som behövs för nordisk paleografi i Unicode. Fokus ligger på e-text, men man skall också kunna se faksimil av manuskript.

Språkbanken vid Göteborgs universitet, Svenska fornskriftsällskapet och Svenskt Diplomatarium är knutna till Menota. Projektet är fokuserat på medeltidstexter, men det vore av stort värde om man kunde knyta utgivningen av svensk litteratur i allmänhet till Menota och utnyttja dess standarder.

Många projekt med samma ambitionsnivå finns i Europa och USA, men det finns inga i Sverige. Utgivningsprojekt kring enskilda författarskap har hittills inte kunnat åstadkomma sådana miljöer, men den utgivning av Selma Lagerlöfs Samlade verk som just startat har potentialen att utnyttja den digitala tekniken på ett nytt sätt. Planen är att Litteraturbanken och Lagerlöfutgivningen tillsammans skall etablera en utgivningsverkstad som dels drar nytta av befintliga databaser, dels utvecklar helt nya databaser och en ny webbmiljö där resultaten senare också kan tillgängliggöras för forskarsamhället. Lagerlöfutgivningen och Litteraturbanken deltar också i COSTs (European Cooperation in the field of Scientific and Technical Research) Action A32 (http://www.cost.esf.org/index.php?id=233&action_number=a32).

6 Databaser som erbjuder en arbetsmiljö över gränser.

Fler och fler internationella webbplatser strävar att skapa en kreativ forskningsmiljö som möjliggör samverkan och inspiration oberoende av var forskarna befinner sig fysiskt. The Rossetti Archive (<http://www.rossettiarchive.org>) är ett tidigt exempel, andra är tidigare nämnda Electronic Text Center Hyper-Nietzsche och Hyper-Learning (<http://www.hypernietzsche.org>). En viktig ambition är ”dynamic contextualization”, ett sätt att utnyttja mediets alla potentialer för att skapa nya och större sammanhang kring den enskilda texten eller verket. COST ägnar sin Action A32 åt *Open Scholarly Communities on the Web*.

Detta slags databaser utgår ofta från tillgängliggörandet av ett enskilt författarskap, och kan resultera i helt nya vetenskapliga landvinningar. Så har t.ex. Jerome McGann utvecklat den teoretiska diskussionen av inte minst editionsfilologi i samband med sitt arbete med The Rossetti Archive.

Framåtsyftande verksamhet i Sverige

Först av allt måste jag betona att min utsiktspunkt är från platsen som Litteraturbankens föreståndare. Det finns många goda språkteknologiska resurser i Sverige, men här ser jag alltså till litteraturvetenskapens specifika behov. Språkbanken vid Göteborgs universitet har en central position som platsen för stora korpusar avsedda för lingvistisk och språkteknologisk användning, och som samarbetspartner i ett antal internationella projekt. Språkbanken är också huvudman för Textil (Textteknologi i forskning och lärande). Projektet inbegriper nordistik, germanistik, lingvistik, litteraturvetenskap, religionsvetenskap, historia, idéhistoria och vetenskapsteori och syftar till att genom textteknologi skapa dynamisk interaktion med texter. Den litteraturvetenskapliga delen i detta är litterär onomastik, i detta fall studiet i hur olika slags namn används i ett specifikt verk eller författarskap, eller i en större korpus. Textil, ännu i startgroparna, är nog det mest tvärvetenskapligt upplagda projektet hittills, och det accentuerar att textdatabaser är en viktig källa för en lång rad discipliner. Det finns ett stort behov av att samordna både material och tekniker, och att åstadkomma gränssnitt som gör databaserna tillgängliga för så många sorters forskning som möjligt.

Det är också vid Språkbanken som Litteraturbanken utvecklas, och även om Litteraturbanken ännu är i begynnelsen finns där en potential att tillgodose många av litteraturvetenskapens – och andra discipliners – behov. Hittills har fokus legat på att utveckla en stabil plattform för tillgängliggörande av texter i olika former och med olika ingångar – för olika behov. Teknikerna för etablerande av material har utvecklats så att importen av texter kan bli betydligt större framöver. Samtidigt fortsätter den teknologiska utvecklingen dels med bättre funktioner för vanliga användare, dels med en helt ny sorts verksamhet: tillsammans med Lagerlöf-utgivningen förbereds en editionsfilologisk verkstad där nya redskap skall utvecklas och den fortskridande utgivningen skall tillgängliggöras för forskarsamhälle och allmänhet. Litteraturbanken har under uppbyggnadsskedet hållit informell kontakt med Menota, men är nu i en situation där ett verkligt samarbete kan börja. Det som avgör Litteraturbankens utveckling är dess resurser.

Men det är också viktigt att komma med i de europeiska initiativen. Det vore förödande om Sverige halkade efter när COST skapar digitala forskningsgemenskaper baserade på nationella databaser. Litteraturbanken utvecklas i fortlöpande kontakt med COST. När det gäller i2010, Digital Libraries, är KB den naturliga kontakten, och det är lika viktigt att skjuta till resurser där.

Sammanfattningsvis

Jag konstaterar alltså att bevarande och bibliografering är två verksamheter som bäst ges resurser till Kungl. biblioteket, även om bibliografering av litteraturvetenskaplig forskning också kunde föras till litteraturvetenskapens ansvarsområde.

Huvudfrågan: databaser för litteraturvetenskapligt bruk: Allt pekar mot att den bästa lösningen för framtiden är att åstadkomma databaser som kan användas av så många discipliner som möjligt och internationellt. Dagens språkdata-baser är utformade för språkteknologisk och lingvistisk forskning, och kommer mer undantagsvis till användning för annat. Samtidigt är litterära och andra texter en självklar källa för så många humanistiska och samhällsvetenskapliga ämnen, att man rimligen borde sträva mot en korpus som dels är av sådan omfattning och kvalitet att den verkligen blir användbar, dels erbjuder sådan information och sådana sökmöjligheter som lämpar sig för många discipliner. Litteraturvetenskapens behov är den humanistiska forskningens: en nationell databas anpassad för så många användningsområden som möjligt. För att den skall vara användbar krävs att den byggs upp med just litteraturvetenskaplig, editionsfilologisk kompetens. Det bör ske i direkt kontakt med internationella initiativ som drivs nu, inte efteråt.

Databaser och digitalisering inom humaniora: konst- och bildvetenskap.

av Lena Johannesson

Vid den sammankomst som DISC inviterat till den 12.12.06 rekapitulerade jag i korthet, hur osannolikt snabbt utvecklingen i fråga om bilddigitalisering har gått. När Europas största bildarkiv, det av Getty's Foundation stödda Bildarchiv Foto Marburg, 1989 kunde ge svaret, att digitala bilder skulle inte vara aktuella i forskningssammanhang förrän tidigast vid slutet av 1990-talet, erinras man om hur fort utvecklingen har skett. I mer än tjugo år satt en forskargrupp i USA och funderade över de klassifikationshierarkier av dataelement och ämnesord som måste upprättas, innan man skulle kunna samordna bildregistren vid bibliotek, arkiv och museer rörande den antika mytologin (Lexicon Iconographicum Mythologiae Classicae. Detta skulle bli det senare s.k. ICONCLASS-systemet, förebildligt för skilda disciplinrelaterade klassifikationssystem såsom t.ex. Out-line, EAC/EAD, SWETERM, Dublincore). Men på knappt ett år var det digitala bildarkivet ett faktum, när programvaran väl var etablerad 1994-1995. 1993 öppnades den första svenska bildbanken för humanistisk forskning vid Tema Kommunikation vid Linköpings universitet. För att kunna åstadkomma det som varje bildforskare efterlyser, nämligen bildåtergivning tillsammans med textdata om proveniens, upphovsman, tillkomstår, förvaringsort, teknik etc, fanns inga lösningar för samkörning av registerdata inom det då existerande utbudet av s.k. biblioteksprogram. Vår arbetsgrupp i Linköping fann dock paradoxalt nog lösningen i ett terminalprogram för Volvos verkstäder, där reservdelar avbildades och samordnades med tekniska data, uppgifter om lagertillgång m.m. Detta "hemgjorda" program för industriell lagerhållning blev via linköpingsinstitutionen för flera år framåt en attraktion för museer och arkiv som ville bygga upp egna bildbanker. Om detta skede och dess enorma respekt för den utrymmes-, kraft- och tidskrävande digitala bildmaterialet kan man läsa i den av HSFR och Riksarkivet stödda rapporten *Bilddigitalisering vid bibliotek, arkiv och museer. En lägesrapport för Nationella arkivdatarådet (NAD) och Humanistisk-Samhällsvetenskapliga Forskningsrådet (HSFR), utförd av Yvonne Eriksson, Margareta Ståhl och Gary Svensson under ledning av professor Lena Johannesson. Arbetsrapporter från Tema K, 1995:24.*

Idag är de enklaste persondatorer för privatbruk ofattbart kapabla utifrån vad vi trodde skulle vara möjligt för bara fem-sex år sedan. Lika ofattbart snabbt har det gått att insocialisera forskare av alla åldrar i integrerad användning av ordbehandling + bildhantering; det finns ingen aktiv bild- och konstforskare idag som INTE använder sig av digitala medier. Det må

gälla att komma åt bildarkiv, textkällor, kataloger mm eller att framställa föreläsningmaterial i form av Power-Points, att skanna in bildmaterial eller ladda över och behandla bilder från den egna digitala kameran.

* * *

Fortsättningsvis följer jag det frågeformulär som professor Eva Strangert och DISC lagt ut som rättesnöre och bas för denna inventering.

- Hur ser situationen ut idag inom ämnet? Finns digitala databaser? Andra typer av material- och datasamlingar (tal, text, bilder etc)?

Inom de konstvetenskapliga och konst- och bildvetenskapliga institutionerna som sådana existerar inga bildbanker. Däremot lägger varje enskild forskare i sin dator upp egna bildförråd, hämtade från bilddatabaser, webbsajter, egen fotografering, egen inskanning av material ur böcker, kataloger och tidskrifter. Likaså skapar varje konstvetare sin egen bank av länkar till bildbanker vid större museer i världen samt till enskilda konstnärer och gallerier etc. Som en konsekvens av detta har institutionerna i landet på sina hemsidor och som hjälp åt doktorander och studenter lagt ut länkar till större bilddatabaser eller metabaser i form av så kallade resursbaser. Se t.ex. hemsidan för Institutionen för konst- och bildvetenskap vid Göteborgs universitet. Rubrik: Konsthistoriska resurser”[Art History Resources on the Web](#); [The Web Gallery of Art](#); [Chris Witcombe's Art History Resources](#); [Yahoos lista](#); [Jeffrey Howes](#); [Columbia University Media Center for Art History](#); [Internationell Arkitekturdatas](#); [Architecture Web Resources](#); [Digital Art Museum](#); [New York Public Library Digital Gallery](#)

- Hur utnyttjas tillgängliga databasresurser?

De resurser som läggs ut av alla större bibliotek, arkiv och museer världen över används regelbundet och rutinmässigt av den enskilde forskaren för att orientera sig inom ett forskningsfält, söka referenser och överblicka samlingar. Det ständigt återkommande problemet är selektiviteten i vad museerna lägger ut på nätet.

- Finns erfarenhet av samverkan och forskningssamarbete gruppvis/nationellt kring uppbyggnad av nu existerande databaser?

Det har inte varit diskussion om att bygga upp någon databas eller nätverksstruktur som skulle serva alla de konstvetenskapliga institutionerna för forskningsändamål. Däremot har det funnits och finns det via disciplinens regelbundet återkommande Ämneskonferenser en samordnad diskussion om hur man skulle kunna rädda de omfattande diabilidsarkiv som byggts upp vid varje institution och som skulle behövas rensas ut resp. digitaliseras för att

kunna användas i forskning och undervisning. Dessa bildarkiv omfattar i Lund mellan 80 000 – 100 000 diabilder, i Uppsala ca 60 000, Stockholm ca 80 000 och Göteborg ca 50 000. Till detta kommer i bl.a. Uppsala och Göteborg unika planschsamlingar, bestående av svartvita fotokopior (ca 30 x 50 cm) uppklistrade på kartong och använda vid de första konsthistoriska seminarierna i landet. Likaså finns vid dessa universitet samt i Stockholm omfattande samlingar av storbildsdiå från 1920-50-talen, svartvitbilder i oerhört god kvalitet.

- Vilken roll spelar befintliga databaser vid arkiv, bibliotek, museer, mm för forskningen inom ämnet?

Fortfarande är traditionen stark att själv söka sig till samlingarna och med självsyn studera konstverk, teckningar, grafik, fotografier etc. Men de flesta forskare genomför före igångsättandet av en undersökning eller före en forskningsvistelse vid utländska bibliotek, arkiv och samlingar via sökningar en ingående inventering av vad dessa samlingar innehåller. Tillgängligheten är dock mycket skiftande, vilket kan konstateras genom de exempel jag anför nedan. Inom gruppen ”Selected Collections” är urvalet mycket hårt styrt och populariserande och kan nätt och jämt tillgodose grundkursstudentens orientering. Forskaren måste därför vanligen komplettera sin sökning med att gå in i textdatabaser. Många museer håller sig med renodlade textdatabaser för forskarbruk, lyckligtvis baserade på fritextsökning. Alltfler museer och bildarkiv försöker emellertid skapa för forskningen öppnare system genom kombinerade text-och bilddatabaser. National Gallery i London är därvidlag föredömligt med sitt Full Collection Index, Nationalmuseum i Stockholm än så länge mer typiskt med sina 24 000 bildåtergivningar av de 92 000 objekt som registrerats i WebArt och de 600 000 objekt som det totalt handlar om.

Selected Collections, på sgs alla större museihemsidor:

Kunsthistorisches Museum Wien; Alte + Neue Pinakothek, München; National Gallery, Washington; Metropolitan Museum, New York; Eremitaget, S:t Petersburg; Nationalmuseum, Stockholm.

Enbart textbas, katalog för fritextsökning: Statens Museum for Kunst, Köpenhamn:

Warburg Institute, London; Bildarchiv Foto Marburg; Moderna Museet “Sök I samlingen”, 50.000 objekt registrerade (bilder skall så småningom tillgängliggöras via Access-insatser)

Text- och bilddatabas samordnade: Nationalmuseum, Stockholm: WebArt; Kungliga Biblioteket (selektivt: resor, Vecko-Revyn, Djävulsbibeln, affischdatabasen, komplett: Suecia Antiqua et Moderna); British Library – selektivt; National Gallery, London, Full Collection Index; ArtCyclopedia, ca 180 000 konstverk, data om 8500 konstnärer, länkat till resp

museum; Art-Guide, Sammlung Kunsthistorischer Internetquellen; AEIOU – Österreichischer Bilder-Album.

- Finns behov av digitalisering av material som förvaras vid arkiv, bibliotek, museer, mm?

Ja, i högsta grad. Riksarkivets satsning på digitalisering av kartor liksom på folkbokförings- och mantalshandlingar, dödsböcker m.m. visar detta, liksom även de satsningar som gjorts – men som nu stannat av - på att digitalisera de rika fotoarkiv som finns i Sverige, på läns museer, folkrörelsearkiv, universitetsbibliotek, Nordiska museet o.s.v. . Önskvärt vore om t.ex. ritningar vid Arkitekturmuseum kunde digitaliseras, liksom teckningar, ritningar, kartor, brev och excerpter i Kungl.Vetenskapsakademiens samlingar och Carolina Redivivas och KB:s rika bild- och handskriftssamlingar. (Sedan detta skrevs har Ulrika Kjellman offentliggjort sin i detta sammanhang betydelsefulla analys av KB:s bildbestånd och digitaliseringspolicy; se doktorsavhandlingen *Från kungaporträtt till läsketikett. En domänanalytisk studie över Kungl.bibliotekets bildsamling med särskild inriktning mot katalogiserings- och indexeringsfrågor* Skrifter utgivna vid institutionen för ABM vid Uppsala universitet. Volym 2, Stockholm 2006.) Här bör förklaras, att dagens konst- och bildforskare inte bara ägnar sig åt konstbilden utan mycket ofta sysslar med bruksbilder och alla varianter av instrumentella bilder som t.ex. vetenskaplig och ingenjörsteknisk illustration.

- Vilka framtida behov finns vad gäller databasuppbyggnad och digitalisering?

Primärt för konstvetare i landet är en satsning på att rädda de omfattande lokala diabildsarkiven vid institutionerna, se ovan. Detta på grund av deras kulturhistoriska värde men även och inte minst för att de baserats på bilder till vilka vi idag har rättigheterna, då mycket fotograferats av lärargenerationer tillbaka. BUS-reglerna lägger idag hinder i vägen för den omfattande visning av bilder som sker inom ämnets undervisning. (Ett exempel: Cork College University på Irland skickar sina lärare till Italien för att de på ort och ställe skall fotografera den konst och den arkitektur som skall behandlas i renässanskurserna; allt för att undgå problem med bildrättigheter och visningsrättigheter)

Därnäst: Ämneskonferensen är enig om att det är akut angeläget att resurser ställs till förfogande för digitalisering av svenska konstsamlingar (som Nationalmuseum, Göteborgs konstmuseum, Norrköpings konstmuseum., Malmö museum). Det nationella kulturarvet blir mindre och mindre synligt i de internationella översikterna och har inte en chans att nå ut i

forskarsamhället om inte museerna kan exponera dem i sina digitala arkiv. Detta är en överhängande fråga också för skolningen i ämnet och vi önskar oss även ett avtal med de svenska museerna som gör att konstvetare skulle kunna få använda svenskt material i undervisningen och utan hinder av BUS-regler.

Önskvärt vore även att bildmaterialet till doktorsavhandlingar kunde läggas in i databaser *in extenso* i analogi med språkvetenskapernas *corpustradition*. Stora mängder bilder samlas in för en avhandling eller annan vetenskaplig undersökning, men av tryckkostnadsskäl är det endast ett fåtal av dessa bildkällor som kan reproduceras och tryckas.

- Finns vetenskapligt intressanta utvecklingslinjer, som inte skulle kunna förverkligas utan tillgång till nya databasresurser?

Tveklöst. Studier kring performances, site specific art, video, film, dans, multimediala konstrepresentationer digitaliseras idag och kan bara studeras via digitala medier. Även kognitionsforskning rörande seendet och spatialitet studeras idag främst via digitala modellsystem. Studium av rumsperception och av den kreativa skapande processen sker bland konstnärer och arkitekter idag helt via datorerna. Konstvetarna skulle kunna studera och följa med i den processen om databasresurserna ökade. Se även nästa fråga.

- Vilka möjligheter finns att samverka i forskargrupper eller i andra konstellationer nationellt och/eller internationellt för att bygga upp och utveckla nya databaser med bred användning i forskningen?

Stora. Redan idag samarbetar t.ex. konstvetare och kognitionsforskare (Göteborg – Lund). Den samsyn som råder inom ämnet dels i fråga om att utnyttja digitaliseringen i allmänhet för forskning, dels för att befrämja tillgängligheten för svensk konst nationellt och internationellt, garanterar att fastare nätverk och forskargrupperingar kommer att etableras, inte minst mellan museer och universitet. Likaså skulle större dataresurser kunna medföra ett närmare samarbete mellan konstforskare och konstnärer, för vilka arbetet med digitala medier idag är en självklarhet och starkt processinriktat. Inom arkitekturforskningen skulle större databaser vara av vital betydelse, då större cluster av satellitdata, demografiska, geologiska och arkeologiska data etc. exempelvis kunde samköras med långa serier av stadsplanedata, ritningsprojektioner, modeller etc. (Ett liknande projekt utvecklas nu i Rom vid Bibliotheca Herziana, där 1700-talsarkitekten Juvaras samtliga verk lagras i digital form för att kunna underlätta restaureringsarkitekternas och konservatorernas analyser av hans byggnader.)

Bland bildforskare idag finns ett intresse att komma de instrumentella bildpraktikerna inom naturvetenskap närmare in på livet, t ex inom medicin eller meteorologi. Möjlighet till allokering av vissa digitaldata från dessa bildkulturer är av högt intresse vid studiet av bildmediets epistemologi i dagens vetenskap.

o Kan tillgång till och användning av databaser påverka forskningens synlighet och närvaro i ett internationellt perspektiv?

Ja, se ovan särskilt resonemanget om den nationella konstens osynlighet i globalitetens era. De rika svenska arkiven attraherar allt flera utländska forskare och sannolikt skulle ett större synliggörande av dessa även i fråga om bildkällor öka intresset ännu mera. Större dataresurser skulle även möjliggöra större internationella forskningsprojekt, då det nu är lagring och sändning av bilder som vanligen är för kapacitetskrävande för att ingå i den dagliga bytestrafiken forskare och länder emellan.

Vetenskapsrådet

Digitalisering inom humaniora
Behovet av databaser som forskningsinfrastruktur

Medie- och kommunikationsvetenskap

Ester Pollack

Medieforskningens två traditioner

Forskning om journalistik, medier och kommunikation har bedrivits internationellt sedan tidigt 1900-tal. I Sverige har viss forskning om pressen och om mediepåverkan funnits sedan 1900-talets tidiga decennier. Under 1950-talet studerades frågor om medier av svenska forskare inom många olika discipliner, exempelvis inom sociologi, statsvetenskap, litteraturvetenskap och pedagogik. En hel del av forskningen var förlagd vid enskilda forskningsinstitutioner och vid olika forskningscentra, såsom Sveriges Radio och dåvarande Beredskapsnämnden. Sedan detta 1950-tal har medieforskningen, som vid den tidpunkten benämndes masskommunikationsforskning, befunnit sig i ständig förändring. De stora samhällsomvandlingarna med de politiska systemens utveckling, journalistikens allt mer centrala roll i samhällslivet och den moderna medieteknologins framväxt har av självklara skäl gett starka avtryck i medieforskningen.

Forskningen om medier och kommunikation har, som beskrivits, rötter i ett flertal olika såväl samhällsvetenskapliga som humanistiska ämnen. Därför kan man också tala om två huvudtraditioner: den humanistiska och den samhällsvetenskapliga. Humanisterna har förenklat uttryckt intresserat sig mer för meningen i det medierade, emedan samhällsvetarna i större utsträckning studerat olika mekanismer i och effekter av medierade budskap och ett medialiserat samhällsliv. Ser vi bakåt i historien har avståndet mellan dessa två huvudtraditioner varit tämligen stort vad gäller metod och teori, men minskat väsentligt över tid. Närmandet mellan humanister och samhällsvetare blev särskilt tydligt under 1980-talets så kallade "kulturella vändning" då framförallt mediernas roll i de kulturella processerna kom att fokuseras och sådant som exempelvis vardagslivets medieanvändning blev av intresse. Idag kan det många gånger vara svårt att tydligt identifiera en skiljelinje mellan en humanistisk respektive samhällsvetenskaplig forskningsinriktning. På många områden sammanfaller forskningsintressena.

Det var under 1980-talet som de båda tvärvetenskapliga ämnena journalistik och masskommunikation - idag betecknat medie- och kommunikationsvetenskap - institutionaliserades. Vid Stockholms universitet skedde det i och med nedläggningen av dåvarande Journalisthögskolan (som sedan 1977 varit en institution vid SU) och bildandet av Institutionen för journalistik, medier och kommunikation (JMK) 1989. Den nybildade institutionen kom att tillhöra humanistiska fakultetens historisk-filosofiska sektion, den redan startade systerinstitutionen i Göteborg kom däremot att tillhöra samhällsvetenskaplig fakultet vid Göteborgs universitet. Mönstret går igen över hela Sverige: medieforskningen är vad gäller institutionell tillhörighet delad mellan humanistiska och samhällsvetenskapliga fakulteter. Ungefär lika många institutioner tillhör humanistisk fakultet som samhällsvetenskaplig. Flera institutioner härbärgerar, oavsett fakultetstillhörighet, bägge arven inom sig. Det gäller exempelvis för JMK vid Stockholms universitet, där just mötet mellan de humanistiska respektive samhällsvetenskapliga traditionerna är särmärket. Forskarutbildning finns vid universiteten i Göteborg, Karlstad, Lund, Stockholm, Umeå,

Uppsala och Örebro. Forskning bedrivs dock vid betydligt fler universitet och högskolor och grundutbildning i medie- och kommunikationsvetenskap finns idag vid cirka 20 olika läroanstalter.

Forskningsområdet idag

Samhällsutvecklingen i termer av medialisering och globalisering ställer idag medieforskare inför stora utmaningar. Att teoretisk inkludera och att empiriskt undersöka den medievärld som befinner sig i accelererande förändring är ett projekt av mångvetenskaplig karaktär. Framväxten av globala mediekonglomerat, det politiska livets anpassning till medierna, en digitaliserad produktion och distribution, nya typer av medieinnehåll och kraftigt förändrade mönster i mediekonsumtionen kräver forskningsinsatser av mångahanda slag. Exempel på sådant som uppmärksammas är den samtidiga konvergensen av medierna - gamla medier växer samman med nya - och divergensen av medieinnehållet - upplösningen av genrer som journalistik, information, underhållning och reklam. En övergripande fråga är hur en ny mediestruktur påverkar samhällets demokratiska processer i en samtid då mediernas betydelse och makt ökar på flera olika sätt. Så ställer exempelvis den europeiska unionens ambitioner som ett demokratiprojekt många medierelaterade frågor på sin spets, bland annat den om medborgarnas tillgång till en gemensam europeisk offentlighet. Huruvida denna europeiska offentlighet alls existerar är för övrigt omdebatterat inom forskningen, liksom förutsättningarna för att skapa en sådan. För närvarande pågår, oavsett forskarsamhällets diskussioner, ett intensivt arbete inom EU-kommissionen med utgångspunkt i dess vitbok om kommunikationspolitiken och med Margot Wallström som huvudansvarig. Medierna och medietekniken betraktas här, utifrån den övergripande visionen om att minska avståndet till medborgarna och överbrygga klyftan mellan EU och folket, som helt centrala.

Tilläggas bör att även om samhälls- och medieutvecklingen idag har implikationer för framtiden och är ett angeläget forskningsfält, så har medieforskningen samtidigt en utmaning i att studera historien - inte minst för att den kan hjälpa oss att förstå vår samtid. Så är exempelvis mediernas roller i den politiska historien ett ännu till stora delar underutforskat område som kan ge värdefull kunskap och öka förståelsen också av dagsaktuella fenomen i samspelet medier, journalistik och politiska konflikter.

Flera forskningsfronter

Att karakterisera forskningsfronten idag är svårt. Forskningsfältet har expanderat kraftigt och utmärks, som ovan beskrivits, av stor mångfald. Olika små specialområden tillkommer i takt med medie- och teknikutvecklingen. Det vore därför rimligare att tala om flera olika forskningsfronter. Några exempel på aktuella forskningsområden från medie- och kommunikationsvetenskapliga institutioner (organiserade inom humanistiska fakultet) får illustrera mångfalden och bredden: mediehistoria, mediernas roll i den politiska historien, politisk kommunikation, medial offentlighet och privatsfär, globaliseringen och komparativa mediestudier, medieutveckling, visuell kultur, medier och sociala problem, medier och religion, krigsjournalistik, public relations som organisations- och samhällsfenomen, risk- och kriskommunikation, publik- och receptionsforskning, tal och samtal i offentlighetens rum. I perspektiv av denna skisserade bredd avhåller jag mig här från att definiera en forskningsfront.

Befintliga databaser

Lika svårt som det är att definiera EN forskningsfront inom medie- och kommunikationsvetenskapen, lika omöjligt är det att ge en översikt över de databaser som medieforskningen idag brukar. Eftersom medie- och kommunikationsvetenskapen till sin

karaktär är ett tvärvetenskapligt forskningsområde, betyder det att databaser av väldigt många olika slag utnyttjas. När mediernas roll i den politiska historien studeras kommer givetvis flera av de databaser som historiker och sociologer brukar till användning, när tal och samtal i offentligheten undersöks kan lingvistiska databaser vara värdefulla och så vidare. Jag har därför valt att i det här sammanhanget koncentrera mig på några för medieforskningen gemensamt angelägna databaser och arkiv och problem förknippade med tillgång eller utvecklingen av dessa. Till grund ligger svar på en förfrågan utsänd till ett urval kollegor inom medievetenskapliga institutioner sorterade under humanistisk fakultet.

Inom den språkvetenskapliga medieforskningen nämns ett behov av att bygga upp flera databaser/korpus med sådant som transkriberade mediesamtal (exempelvis intervjuer) som kan digitaliseras och göras kollektivt tillgängliga för forskares analyser. Det är idag mycket svårt att få medel för transkribering vilket innebär att många projekt bygger på små enskilda datamaterial. Möjligheten att kunna söka medel för uppbyggnad av sådana datamaterial skulle väsentligen bidra till forskningens utveckling på detta område.

De databaser/arkiv som nämns som väsentliga av samtliga tillfrågade medieforskare är Nordicom, Statens Ljud- och Bildarkiv, Kungliga Bibliotekets tidningsenhet och utbudet av andra press- och bildarkiv. Farhågorna gäller framförallt upprätthållandet och utvecklingen av Nordicom samt SLB. Som problematiskt anges också den bristande tillgången till SVT:s och SR:s interna arkiv samt de höga kostnaderna för tillgång till de textdatabaser som håller tidningarnas arkiv.

Vilken roll spelar befintliga databaser vid arkiv, bibliotek, muséer?

Befintliga databaser spelar naturligtvis en stor roll för medieforskningen. Sådana är exempelvis de ovan nämnda Nordicom och Statens ljud- och bildarkiv (dessa återkommer jag till nedan). Därutöver kan man exempelvis nämna Riksarkivets samlingar, SCB, Kungliga biblioteket med sitt microfilmsarkiv med sparade tidningar, Uppsala universitets och Lunds universitets motsvarande tidningsarkiv, Journal Digital och Grängesbergsarkivets dokumentära material (för filmvetarna).

Kungliga biblioteket har forskare som arbetar med historisk pressforskning stor erfarenhet av. Här finns några problem som gäller forskarservicen att lösa: stora forskningsprojekt som ska samla mycket stora mängder microfilmstexter och där vi själva har behov av att kopiera dessa på papper, arbetar under samma villkor som exempelvis privata släktforskare. Kopieringskort ska betalas kontant, dessa kort får endast laddas med mindre summor åt gången, man måste sätta upp sig i fulltecknade köer till kopieringsapparater, de fungerande maskiner är få och så vidare. Det gör att själva insamlingsprocessen är onödigt komplicerad och långsam. Kan detta lösas med någon typ av avtal?

En rad olika digitaliserade pressarkiv och bildarkiv uppbyggda i olika syften och av olika intressenter används också, såsom Sveriges Pressarkiv, Pressklipp och Mediearkivet. En del av dessa kan tjäna forskningsändamål, men har begränsningar som gör att de bara kan användas i viss omfattning och för vissa ändamål. Så kan till exempel Pressklipp och Mediearkivet vara användbara om man vill undersöka vad som publicerats under en viss period som inte sträcker sig alltför långt tillbaka i historien. Arkiven, som omfattar ett urval av några av de vanligaste dagstidningarna, har olika startdatum och går ofta inte längre tillbaka i tiden än till 1990-talet.

Det finns också en del kommersiella digitala databaser med söktjänster som man kan prenumerera på såsom Magentanews, där resultat av sökord levereras var dag. Tjänsterna är dock dyra, särskilt om man också vill kunna utnyttja till exempel statistiktjänster, och ingen rabatt ges för forskare. Också här är frågan om forskarsamhället skulle kunna utverka någon typ av specialavtal?

Statens Ljud- och Bildarkiv

Statens Ljud- och Bildarkiv, SLB, är sedan 1979 vårt nationalarkiv för ljud och rörliga bilder. Samlingarna som är tillgängliga för forskare (och till viss del för allmänheten) består av pliktsamlingen, levererad enligt lagen om pliktleverans av ljud- och bildupptagningar, samt fria samlingar. Pliktsamlingen omfattar radio- och tv-program, filmer, video, fonogram, interaktiva multimedier, privata lokala medier och taltidningar i urval. I de fria samlingarna kan man bland annat hitta sådant som SF:s journalfilmer, Publicistklubbens debatter och kommersiell musik på CD.

SLB utreddes senaste 2004 (SOU 2004:53), *Bevara ljud och bild*. Utredningen hade bland annat två bra förslag som skulle tjäna forskningen: referensbandningen skulle byggas ut, det vill säga SLBA:s egen insamling av radio- och televisionsprogram som besvärats av en eftersläpning på sex månader. Detta är under utveckling men ännu inte fullbordat. Sveriges Radio och TV4 sänder numera sina program närmast omedelbart till SLBA. Sveriges Television förhandlar man för närvarande med. Att få radio- och tv-material fullt ut digitaliserat och också kunna skapa möjligheter för on-line-baserade sökningar är oerhört viktigt för den moderna radio- och tv-forskningen. Aktuella etermediefenomen inom nyhetsjournalistik, samhälls- och debattprogram blir annars svåra att belysa och exemplifiera. Bristen har lett till att forskare tvingas till privata inspelningar, ofta ofullständiga. Att lösa den frågan är oerhört viktigt för att öka arkivbeståndets tillgänglighet och hanterbarhet och skulle i hög grad gynna den samlade medieforskningen.

När det gäller den delen av arkivet som redan är digitaliserat finns dock ett problem som är av juridisk karaktär och kanske det viktigaste att på sikt lösa: nämligen frågan om digital access. Det är lagligt att per post sända digitalt material, alltså kan du som forskare i till exempel Örebro begära att få vissa program, varvid SLBA överför dem på en CD och sänder dem med det "gamla" postverket. Många av dem finns digitalt tillgängliga och skulle kunna sändas med över med e-post - om det inte vore i lag förbjudet. Frågan gäller upphovsrätten. Alltså hjälper det inte att få allt digitaliserat - såvida inte tillgängligheten samtidigt förbättras. **Som moderna medieinstitutioner har vi idag inte möjlighet att utveckla de stora tv-projekten vi skulle vilja, därför att vi saknar tillgång till aktuella inspelningar av tv-utbudet.**

Det andra utredningsförslaget som gällande SLBA var ett migreringsprojekt som handlar om att rädda det historiska material som arkivet förfogar över. Detta är för närvarande i full gång (i Årjäng).

Nordicom

Den främsta erfarenheten av forskarsamverkan kring nu existerande databaser är Nordicom. Det första som alla medieforskare tillfrågade om behovet av databaser nämner är vikten av Nordicom, nordiskt informationscenter för medie- och kommunikationsforskning. Nordicom är en institution inom Nordiska Ministerrådet, som står fritt från partsintressen. Ledning och huvudsaklig verksamhet är förlagd vid Göteborgs universitet. Nordicom startade för lite drygt 30 år sedan när forskare, pressens företrädare, radio- och tv-företagen tillsammans agerade för att inrätta en nordisk dokumentationscentral för masskommunikationsforskning. Det skedde

mot bakgrund av TV:s stora genomslag på 1960-talet och mediernas, som man uppfattade, allt mer centrala roll i samhällslivet. Idén var att genom nordiskt samarbete kunna åstadkomma ett kunskapscenter med bättre resurser än respektive enskilt land. Det hela bygger dock på självständig nationell verksamhet som samordnas på nordisk basis.

Nordicom har idag tre huvudsakliga arbetsområden: dokumentation av medieforskningen och dess resultat i Sverige och de nordiska länderna, framtagandet och förmedlingen av kunskap om medieutvecklingen och medietrender i Sverige och de nordiska länderna samt forskning om barn, unga och medier i världen.

Medieforskningen i Sverige och Norden: det handlar om dokumentation av den forskning som bedrivs, vilka den bedrivs av och var man kan inhämta resultaten. Genom att ha byggt upp ett gott forskarnätverk, med god förankring i forskarvärlden och genom publiceringsverksamheten fungerar den här systematiska dokumentationen av forskningen väldigt väl.

Publiceringsverksamheten startade ungefär 1980 med två tidskrifter, *Nordicom Information* (fyra nummer per år, 700 prenumeranter), *Nordicom Review* (engelskspråkig refereed tidskrift, 2 nummer per år, 2000 prenumeranter). En poäng med Nordicom har varit att göra medie- och kommunikationsforskningen i Norden internationellt känd samt att bygga upp kontakter. Av de 2000 prenumeranterna på *Nordicom Review* finns 1500 i 125 länder utanför Norden (ett osedvanligt stort antal, många internationella tidskrifter har en prenumerantstock på 400-500). Tidskriften har av ledande internationella medieforskarorganisationer rankats som en av de främsta på sitt område.

Publiceringen hjälper till att föra ut den nordiska medieforskningen, skapar uppmärksamhet och återpublicering och därmed ökar antalet citeringar i Social Science Citation Index.

Nordicom ger också utöver tidskrifter och nyhetsbrev ut *nordiska och internationella antologier* på engelska, så på många håll betraktas det också som ett bokförlag.

Vartannat år hålls nordiska medieforskarkonferenser. Ett urval konferenstexter publiceras i en specialantologi utgiven av Nordicom, något som positivt bidrar till en nordisk forskarsamverkan.

Medieutveckling och medietrender: Detta är kanske det som Nordicom idag är mest känt för, mediestatistik framtaget av en oberoende aktör. Efterfrågan är stark på statistiskt underlag av olika slag och det finns också väldigt många som producerar sådan, utifrån olika typer av intressen. Här gör Nordicom en insats av stor betydelse, för forskarsamhället men också för samhället i stort. Tilläggas kan att Nordicom har många olika brukargrupper, förutom forskare och studenter (som utgör en tredjedel), också medieföretag/mediebranscher (en tredjedel) och politiska beslutsfattare (en tredjedel).

Den svenska statistiken publiceras i *MedieSverige*, *Mediebarometern* och *Den Svenska Mediemarknaden*. Här redovisas statistik om dagspress, tidskrifter, vecko- och månadstidningar, böcker, radio, television, video, film och biografier, fonogram, reklammedier och Internet. Analyserna handlar om ägande, ekonomi, innehåll, distribution och tillgång/konsumtion/användning. Nordicom samarbetar med forskare och arbetet redovisas i olika återkommande publikationer. Uppgiften att samla svensk mediestatistik tillkom 1989 på

efter en statlig utredning och riksdagsbeslut (*MedieSverige*). SCB motsvarade inte behoven. Redan tidigare, sen 1979, fanns en årlig medievaneredovisning (*Mediebarometern*).

Efter många års diskussion tillkom så vid mitten av 1990-talet en samlad nordisk mediestatistik, *Media Trends*, med analyser och kommentarer.

I europeiska och internationella samarbeten spelar statistiken om medier en växande roll. Man kan nog säga att genom Nordicom har de nordiska länderna kommit att betraktas som en enhet.

Kunskap om barn, unga och medier baserad på forskning i hela världen: 1997 fick Nordicom i uppdrag av UNESCO att etablera ett internationellt kunskapscentrum om barn och medie våld, stöds av svenska regeringen och UNESCO. *International Clearinghouse on Children, Youth and Media*. Det handlar inte bara om att forskning om medie våldet, utan om ett större kunskapsområde och om att stimulera till mer forskning om barn och medier. Det finns ett globalt nätverk om ungefär 1000 medlemmar i 125 länder. Clearinghouse samarbetar med flera organisationer, institutioner och nätverk.

Hur ser Nordicoms organisation och finansiering ut? Verksamheten bygger på kontrakt och överenskommelser med Nordiska Ministerrådet, UNESCO, Kulturdepartementet (regleringsbrev) och flera mer tillfälliga uppdragsgivare. Hela 13 gånger har Nordicom utvärderats externt. Institutionen leds av professor Ulla Carlsson, har en styrelse som består av nordiskt ledande forskare och använder sig av vetenskapliga råd och referensgrupper. I Göteborg finns cirka 10 medarbetare, det finns personal i varje nordiskt land knuten till Nordicom och så anlitas forskare vid olika universitet på kontraktbasis.

Finansieringen sker genom olika typer av medel, fakultetsanslag, direkta bidrag från departement och myndigheter, anslag från överstatliga organ, bidrag från den privata sektorn samt genom försäljningsintäkter.

En av lärdomarna vad gäller Nordicom är att en stor andel externa medel och många givare samt ett mångnationellt samarbete kan göra det svårt att upprätthålla en samlad verksamhet (enligt professor Ulla Carlsson). Det finns en frestelse att låta modetrender leda till mer kortsiktiga och tillfälliga beslut, något som snabbt kan tillintetgöra en verksamhet. Det gäller att inte falla för den typen av frestelser och ha förmåga att anpassa sig till mer långsiktiga förändringar i samhället. Nordicom vill vara såväl förvaltare som entreprenör och strävar efter att nå också nya brukargrupper.

Nordicom är ett gott exempel på vilken stor betydelse tillgången till och användningen av databaser kan ha för att skapa synlighet och internationell närvaro för, i det här fallet, medieforskningen.

Bygga upp och utveckla databaser med bred användning i forskningen

Nordicoms databaser används av såväl enskilda forskare som forskargrupper. Det nordiska informationscentret med dess databaser är det främsta exemplet inom medie- och kommunikationsvetenskapen på forskningssamverkan nationellt kring uppbyggnad av databaser. Som framgått ovan har forskare från många olika discipliner glädje av Nordicoms verksamhet, liksom många andra brukargrupper. Utvecklingen på medieområdet går, som tidigare beskrivits, väldigt snabbt och efterfrågan på kunskap är stor. **Den ständiga frågan handlar om finansieringen av Nordicom. Idag finns ett stort behov av att utveckla**

tjänsterna, något som dock inte låter sig göras utan bättre finansiellt stöd. Bland medieforskare ses utvecklingen av Nordicoms tjänster som högt prioriterad för att stödja forskningens infrastruktur.

Etnologi och forskningsdatabaser

Magnus Bergquist

Inledning

Det ska sägas direkt. Datorisering, i form av att skapa databaser för bearbetning av empiriskt material, är ingen självklar väg i ett forskningsprojekt för dagens etnologer. Det är inte detsamma som att informationsteknologin inte spelar någon roll i ämnet. Det gör den i allra högsta grad, men inte på samma sätt som inom vissa andra humanistiska ämnen, som exempelvis lingvistik och språk. Anledningen till varför det blivit så är naturligtvis intressant att fundera över. Denna artikel är ett försök att, dels ge en sammanfattande överblick över vad som finns i form av databaser - analoga och digitala - och hur de används, dels problematisera relationen mellan samlande, samlingar och forskning för att peka ut några för framtiden strategiskt viktiga frågor. Inledningsvis görs en fundering kring varför det blivit så att etnologin inte i så hög grad kommit att arbeta med digitala databaser. Förutsättningar för detta har nämligen funnits.

Databasen som kom av sig

I artikeln *Ett ämne väljer väg* beskriver Orvar Löfgren det intresse för klassificering som fördes in i dåvarande folkminnesforskningen på 1910-1920-talen. Här hämtades inspiration från botaniken och ledde till skapandet av ett genresystem för att kunna sortera folktraditionsmaterial i olika lådor: undersagor för sig och legender för sig, allt prydligt organiserat i realkatalogen (Löfgren 1996:29). Även om intresset inom ämnet förändrades mot folklivets olika uttrycksformer skulle klassifikationstänkandet leva vidare även i senare studier av spridning av traditioner eller innovationer. En "uppteckning" kan också ses som ett avgränsat stycke "data" vilket finns inskrivet på ett papperskort som kan registreras och sorteras. I nästa steg plockas kort ut ur samlingen utifrån vissa kriterier för att därmed sorteras igen och jämföras i hopp om att se nya mönster och teman i materialet. Därmed liknar det också det sätt på vilket information organiseras och hanteras i en databas.

I denna tradition börjar också Jonas Frykman arbeta med sitt avhandlingsarbete om utomäktenskapligt födda barn (Frykman 1977). Frykman inleder arbetet med att bygga ett dataregister över uppteckningsmaterial om utomäktenskapliga barn. Frykman databehandlar statistiken från samtliga församlingar inom landet vid vissa tidsintervall, och kan därmed göra en "totalkartering av den utomäktenskapliga fruktsamhetens geografiska förekomst" (Frykman 1977:9).

På ett sätt är datorisering en helt naturlig utveckling på den forskningsinriktning han i början av sitt avhandlingsarbete arbetar mot, eftersom det handlar om att bearbeta kvantitativt sett stora material i syfte att nå en "helhetsbild" vilken ses som möjlig att uppnå genom att sammanställa många mindre "delar". Detta ska ses mot bakgrund av den långa tradition som tidigare hade präglat ämnet där bra "belägg" var hårdvaluta i den vetenskapliga argumentationen i kombination med intresse för utbredningsmönster, variationer och förändringar (Löfgren 1996). Frykman drar saker till sin spets och

utnyttjar de nya möjligheter som datorerna vid den tiden började skapa och kunde därmed åstadkomma snabbare forskningsresultat än sina föregångare. Åtminstone finns detta som en latent möjlighet som ofta lyfts fram när datoriseringens möjligheter ska beskrivas (jfr Willim 2002). Datorn gör arbetet snabbare, effektivare och med högre precision än den manuella bearbetningen, men bygger samtidigt på en vetenskapssyn där kunskapen ses som något som går att bryta ned i mindre stycken för att därefter sammanfogas på nya sätt.

Detta material och dess bearbetning kommer i den färdiga avhandlingen mer att spela rollen av bakgrundsteckning. Under arbetet är det helt andra frågeställningar som tar över. En dag under en promenad finner han "av en händelse" i en reaback ett exemplar av den amerikanska antropologen Mary Douglas bok *Purity and danger. An analysis of concepts of pollution and taboo* (Douglas 1976/1966). Avhandlingsarbetet tar en ny vändning. Den omfattande databearbetningen finns kvar, men sjunker in i bakgrunden, medan det mer kvalitativt analytiska och teoridrivna hamnar i fokus (Frykman & Gerholm 1993). Frykmans avhandling skulle i denna mening komma att bli stilbildande för stora delar av etnologin under 1980-talet och framåt.

Dagens situation

Så skulle alltså en bakgrundsteckning av forskningsdatabasens korta historia i etnologin kunna se ut. Men historien rymmer många spår och vissa av dessa leder trots allt mot digitala databaser. Inför denna översikt gjordes en undersökning bland etnologer och folklorister verksamma på institutioner, arkiv och museer. Utskick gjordes även på e-postlistan *Etnolist* som når ett 80-tal etnologer i främst Sverige, men också Finland och Norge.¹

Som inledningsvis nämndes: om man går strikt på definitionen att redogöra för hur etnologer i forskningsprojekt skapar och använder egna (stora) databaser återfinns inga större ansatser efter Jonas Frykmans försök i mitten av 1970-talet. Det verkar heller inte finnas några uttalade ambitioner att starta sådana projekt, även om de flesta som svarat i princip tycker att det verkar vettigt. Men, om perspektivet vidgas går det att se två användningsområden. Det första och största handlar om arkivens databaser. Det är äldre pappersbaserade samlingar som nu digitaliseras och görs tillgängliga för både forskare och allmänhet.

Arkiven har historiskt sett alltid varit relaterade till forskningen: etnologer initierar tillsammans med arkiv och museer insamlingsarbeten som blir ett forskningsmaterial. Efter det att forskningen redovisats, förvaras och bevaras materialet i arkivet. Ibland kan ett äldre material bli föremål för forskning på nytt genom att nya teoretiska landvinningar och perspektiv gör materialet intressant på nya sätt. Vissa arkivarier anser att traditionsarkiven faller utanför arkivväsendet just av det skälet. De är i själva verket att

¹ Tack till alla som delat med sig av kunskaper erfarenheter och synpunkter till denna översikt. Här ska särskilt nämnas Gösta Arvastson, Joakim Forsemalm, Kerstin Gunnemark, Charlotte Hagström, Truls Erik Johnsen, Annika Nordström, Christian Richette, Göran Sjögård, Robert Willim och Magnus Åberg.

betrakta som forskningsdatabaser. Att göra en uppdelning i arkiv/samlingar å ena sidan, och forskningsdatabaser å den andra, stämmer alltså inte för hur etnologer har arbetat och arbetar med större samlingar av forskningsdata. Det finns alltså en hel del erfarenhet av samverkan och forskningssamarbete här.

Några av de viktigare datasamlingarna härvidlag är de som finns på de olika dialekt-, ortnamns-, och folkminnesarkiven i Göteborg (DAG), Lund (DAL), Umeå DAUM och Folkminnesavdelningen i Uppsala (FA), som ingår i Institutet för språk- och folkminnen.² Härutöver kan nämnas LUF (Folklivsarkivet i Lund) som utgör en del av etnologiska institutionen i Lund, samt Nordiska museets samlingar av uppteckningar. Samtliga dessa arkiv har olika uppbyggnad och historia, men har det gemensamt att det insamlade minnesmaterialet ofta tillkommit i olika samarbetsformer med forskare och institutioner. Materialen består framför allt av levnadshistoriska intervjuer, nedtecknade minnen, inspelningar och svar på frågelistor.

För att ge ett exempel på hur sådana samarbetsformer växt fram och utvecklats över tid kan DAG i Göteborg tas som exempel. DAG ingår från 1 juli 2006 i myndigheten Institutet för språk och folkminnen, men arkivet har en lång historia. Redan 1919 startade verksamheten att uppteckna folkminnen med bildandet av Västsvenska folkminnesföreningen (VFF). Föreningen medverkade till att upprätta Institutet för folkminnesforskning 1926 vid Göteborgs högskola (IFGH). Samma år erhöles en donation av en rikstäckande folkminnessamling från docent Waldemar Liungman (Liu). Tillsammans kom dessa samlingar att gå under namnet Västsvenska folkminnesarkivet. 1962 förstatligades IFGH och knöts till Göteborgs universitet. Under 1960- och 70-talen skedde flera namnbyten för att 1979 samordnas med dialekt- och ortnamnsarkivet vilket ledde till att Dialekt-, ortnamns- och folkminnesarkivet i Göteborg bildades. Det finns alltså en bakgrund i både föreningsverksamhet och forskning. Arkivet har sedan bildandet haft samarbete med och under perioder varit knutet till Göteborgs universitet. Samarbetet med etnologiska institutionen har funnits i många år. En stor del av det forskningsmaterial som genererats vid institutionen (framför allt inspelade intervjuer, men också annat material) har successivt förts över till arkivet. Detta gäller insamlat material på olika nivåer från B-uppsatser till avhandlings- och forskningsprojekt. Detta har skett mer eller mindre kontinuerligt sedan slutet av 1970-talet. En hel del av detta material - tillsammans med övrigt minnesmaterial i arkivet - har katalogiserats, och metadata kring materialet har också digitaliserats och gjorts sökbara. Arkivet har över en miljon poster och mer än 85.000 av dessa är åtkomliga via ett webb-baserat söksystem.

Likartade beskrivningar, med vissa variationer, gäller också övriga folkminnessamlingar som lyder under Institutet för språk- och folkminnen.

Folklivsarkivet i Lund är det arkiv som har tydligast koppling till forskning genom att vara placerat vid Etnologiska institutionen i Lund. Här finns en tät kontakt mellan insamling, forskning och undervisning. Arkivets upptagningsområde omfattar södra Sverige. Initiativ har tagits för att göra samlingarna tillgängliga för allmänheten.

² Fram till juli 2006 var namnet Språk- och folkminnesinstitutet (SOFI).

Materialet utgörs av frågelistsvar, intervjuutskrifter, brevsamlingar, dagböcker och foton. Sedan mer än 15 år registreras allt material via dator. Äldre manuella register har lagts in i en databas och är sökbara via Internet. Det mesta i arkivets samlingar är sökbart på ämne och ort. Digitalisering av bild och fotosamlingar pågår också, och arbetet med att digitalisera själva textsamlingarna har påbörjats.

Nordiska museets har ett omfattande material med de två samlingarna av folkminnen och frågelistor (Nm). I dessa undersökningar fokuseras människors liv, villkor, materiella och imateriella världar, och utgör ett brett källmaterial i form av bland annat föremål, texter och bilder. Metodiskt utgörs det av uppteckningar, fältundersökningar och frågelistsvar. Flera av frågelistorerna i Nm-serien har tillkommit i samarbete med forskare eller andra aktörer. Några exempel är Nm 234, *Våra kungligheter* (2005), som genomförts i samarbete med Livrustkammaren och som empiriskt material för en avhandling i etnologi; Nm 233: *Nu är det jul igen* (2004) är ett samarbete mellan Nordiska museet, LUF och DAG. Ett resultat är boken *Nu gör vi jul igen* (2006) där ett flertal etnologer och folklorister medverkar; Nm 232: *Broderfolk emellan* (2003) gjordes inför en utställning om unionsupplösningen och är ett samarbete mellan Nordiska museet, Norsk Folkemuseum och Norsk etnologisk granskning.

Att ett material samlats in och analyserats i ett forskningsprojekt måste inte innebära att det är "förbrukat". Flera forskningsprojekt har haft redan befintliga samlingar som fokus i analysen. Det har handlat om att göra kritiska analyser av såväl befintliga samlingar som samlande som institutionell praktik. Bakgrunden till dessa nytolkningar har i flera fall varit nya teoretiska perspektiv som skapat förutsättningar för att arbeta med det befintliga materialet på nya sätt. Tre exempel ska nämnas. Agneta Lilja gör i *Föreställningen om den ideala uppteckningen* (1996) en kritisk granskning av hur idé- och maktpraktiker formade insamlingsarbetet av folkminnen vid arkivet i Uppsala 1914-1945. Bo G. Nilsson (1996) använder diskursanalys för att belysa villkoren för hur Nordiska museet representerat svenska arbetares levnadsvillkor via insamlingen av arbetarminnen. Lennart Zintchenko gör i *Stadens tidsbilder* (2003) en analys av hur samlingar av tidningsklipp strukturerar förandet av kollektiva minnen om stadens historia och dess omvandlingar.

Internet - etnologiskt fältarbete i världens största databas

Det finns ett annat perspektiv på databaser som är relevant för etnologins arbetssätt. Under de senaste drygt 10 åren har Internet etablerats som globalt nätverk, inte bara för att samla information på, utan också som en social plats för interaktion och meningsskapande. Etnologins intresse för människans kulturella uttrycksformer har gjort att Internet blivit en viktig arena där sådana uttryck kan studeras, och användning av material från Internet har blivit mycket vanligt. På så vis kan man säga att etnologin hämtar data från en gigantisk världsomspännande databas. Samtidigt uppfattas den inte så som vi traditionellt tänker på begreppet databas. Det är visserligen information som enligt vissa kriterier söks ut och hämtas in till forskningen, men etnologen uppfattar det snarare som en ny typ av fältarbete. Etnologer vill fånga pågående interaktion och kommunikation mellan människor och Internet är idag en viktig arena där detta sker. Internet fogas in i en etablerad etnologisk praktik - "fältarbetet". Istället för att vara en av

forskaren kontrollerad värld, blir databasen intressant som ett resultat av många aktörers praktiker och uttrycksformer. Detta skapar nya metodologiska och analytiska utmaningar, men också nya krav på exempelvis källkritik och etik (Sveningsson, Lövheim & Bergquist 2003).

Ett par exempel på studier baserade på fältarbete på Internet är egna studier om öppen källkod (open source), om hur individer i olika delar av världen gemensamt utvecklar programvara på Internet. Ett av de mer kända exemplen är operativsystemet Linux. Öppen källkod är inte bara programkod, det handlar lika mycket om de nätbaserade sociala världar, normer, praktiker och kommunikationsformer som dessa programmakare skapat, där själva programkoden är ett viktigt meningsskapande element (Bergquist 2003b). En annan studie behandlar så kallade "hoax", det vill säga e-postmeddelanden och hemsidor som skapas i syfte att luras. Dessa analyseras utifrån genreteori som en intertextuell lek som går ut på att de falska breven och sidorna hela tiden försöker dölja sitt uppsåt genom att framträda som andra genrer (Bergquist 2003a).

IT för kommunikation och interaktion kring forskning

Internet är alltså allt viktigare som en plats för etnologiskt fältarbete och kan sägas vara en utvidgning av den traditionella förståelsen av begreppet forskningsdatabas.

Ett annat allt viktigare område är hur digitala resurser mobiliseras för forskare att kommunicera med varandra, för att informera och för att föra ut forskning till olika avnämargrupper. Också här handlar det om databaser i en vidgad betydelse, som ett medel för samarbete i och kring forskning. Här finns många exempel bland etnologer och en stor entusiasm inför utvecklingen.

Ett exempel som redan nämnts är e-postlistan *Etnolist* som är en Sunet-baserad e-postlista med ett 80-tal medlemmar, administrerad av Magnus Åberg, doktorand vid göteborgsinstitutionen. Listan används i första hand för forskningsinformation, men fungerar ibland också som en kanal för debatt och diskussion eller - som i fallet med denna översikt - för att snabbt samla in kunskap inom ett visst område.

Uppsalainstitutionen har tagit flera initiativ för att använda digitala resurser för att presentera forskning, bland annat i undervisningssammanhang. Umeå samarbetar med HumLab kring olika projekt. Från Göteborg rapporteras om internationella projekt som organiseras via elektroniska kommunikationsplattformar. I Lund finns HEX - humanistisk experimentgrupp (www.hex.lu.se) där humanister, med etnologen Robert Willim som koordinator, utvecklar nya former för att bedriva tvärvetenskapliga studier och experimentera med nya gestaltungsformer för humanistisk forskning. Etnologen Charlotte Hagström deltar i delprojektet *Historia på plats* som handlar om platser - symboliska, virtuella och reella. Ett av de mer uppmärksammade HEX-projekten är *Surreal Scania* (www.surrealscania.se) som är ett samarbete mellan Willim och filmaren Anders Weberg. Digitala videofilmer kombineras med geografisk positioneringsteknik (GPS) och Google Earth för att skapa nya upplevelser av platser, samtidigt som dessa problematiseras med hjälp av tankar om platsmarknadsföring och branding i förening med konstnärlig gestaltning. Projektet har uppmärksammats på en rad internationella

seminarier och festivaler. Charlotte Hagström forskar om namngivningspraktiker i spelet World of Warcraft (WoW). Spelet är också platsen för fältarbetet och till detta finns ett europeiskt nätverk av forskare, The Truants, som tillsammans gör en antologi om WoW.

Framtiden

Etnologer inte sägas vara storanvändare av forskningsdatabaser, åtminstone inte om databas förstås i en traditionell mening. Det innebär att det heller inte finns konkreta krav och önskemål på framtida satsningar. Etnologer har kanske svårt att ha önskemål eftersom få tycks fundera över vad som går att göra med källmaterial lagrat i en digital databas. Vi ser dock flera ansträngningar att göra befintliga samlingar mer tillgängliga vilket kan ha en forskningsinitierande effekt. Här gäller det att inte dra alltför skarpa gränser mellan det som kallas arkivdatabaser och forskningsdatabaser eftersom detta är något som för etnologins del alltid gått hand i hand.

Vi ser också nya möjligheter vad gäller metod- och teoriutveckling runt fältarbetet som etnologisk metod, med Internet som fält. En möjlighet för framtiden är att etnologer i större utsträckning använder programvaror för att analysera material som samlas in på detta sätt. Här finns program som NVivo från QSR International som möjliggör kvalitativa analyser på större mängder data, som exempelvis intervjuer eller diskussionslist- och chatt-material. Också samverkan mellan forskare via nätverk och andra nätgemenskaper tycks vara framväxande och av stor betydelse.

En viktig aspekt av att arbeta med material som samlas in från Internet är de etiska frågor som blir aktuella. Här har etnologer lång erfarenhet av att forska på personnära intervju-, observations- och uppteckningsmaterial, och att hantera etiska frågor kring anonymisering, upphovsrätt etc.

Litteratur

- Bergquist, Magnus 2003a: *Hoax - sociala virus på Internet*. Paper presenterat på Nordisk Etnolog- og Folkloristkongres, Helsingör 6-8 maj 2003.
- Bergquist, Magnus 2003b: "Open source software development as gift culture: work and identity formation in an Internet community." Garsten, Christina & Wolff, Helena (eds.): *New technologies at work. People, screens and social virtuality*. Oxford: Berg Publishers.
- Douglas, Mary 1976/1966: *Purity and danger. An analysis of concepts of pollution and taboo*. London: Routledge & Kegan Paul.
- Frykman, Jonas 1977: *Horan i bondesamhället*. Lund: Liber.
- Frykman, Jonas & Gerholm, Lena 1993: "Samtal om etnologin." Gerholm, Lena, red.: *Etnologiska visioner. Femton forskare reflekterar kring sitt ämne*. Stockholm: Carlssons.
- Lilja, Agneta 1996: *Föreställningen om den ideala uppteckningen. En studie av idé och praktik vid traditionssamlade arkiv. Ett exempel från Uppsala 1914-1945*. Uppsala: Dialekt- och folkminnesarkivet.
- Löfgren, Orvar 1996 "Ett ämne väljer väg." Ehn, Billy & Löfgren, Orvar: *Vardagslivets etnologi. Reflektioner kring en kulturvetenskap*. Stockholm: Natur och Kultur.

- Nilsson, Bo G. 1996: *Folkhemmets arbetarminnen. En undersökning av de historiska och diskursiva villkoren för svenska arbetares levnadsskildringar*. Stockholm: Nordiska museet.
- Sveningsson, Malin, Lövheim, Mia, Bergquist, Magnus 2003: *Att fånga nätet. Kvalitativa metoder för Internetforskning*. Lund: Studentlitteratur.
- Willim, Robert 2002: *Framtid.nu. Flyt och friktion i ett snabbt företag*. Eslöv: Symposion.
- Zintchenko, Lennart 2003: *Stadens tidsbilder. Om minnen, erfarenheter och förväntningar utifrån stadens omvandlingar i Sverige 1880-1990*. Göteborg: Etnologiska föreningen i Västsverige.

Stora databaser inom ämnet historia

Lars Nilsson

Historikerna var relativt tidiga med att börja bygga stora databaser för framtida forskning. De första satsningarna gjordes i början och mitten av 1970-talet. Ämnet historia hade vid den tiden en orientering mot samhällsvetenskap. Demografisk historia, socialhistoria och ekonomisk historia stod till exempel högt i kurs. Denna inriktning ökade säkerligen behovet av och intresset för stora databaser. Parallellt ökade intresset för släktforskning, hembygds- och lokalhistoria, ett uttryck för detta var den så kallade "Gräv där du står rörelsen". Trycket mot kyrkböcker, mantalslängder och liknande materialgrupper blev stort och slitaget hårt. Ett viktigt skäl för byggandet av stora databaser, vid sidan av de vetenskapliga argumenten, blev därför att man på så sätt skulle kunna rädda arkivmaterialet från förslitning och kanske förstörelse.

Kvantitativa data och kvantifierbara uppgifter lämpar sig särskilt väl för databaser, och de första registreringarna avsåg folkbokföringsmaterial. Detta material har fortsatt spelat en viktig roll och har samtidigt kompletterats med andra typer av registreringar. Arkiven var redan från början aktiva och medverkade vid uppbyggandet av de stora databaserna. Senare har också bland andra museer och bibliotek engagerat sig i till exempel tillkomsten av bilddatabaser. I det följande ger jag till att börja med några exempel på befintliga större databaser med en grov indelning i mer renodlade Demografiska databaser och Andra typer av databaser.

Demografiska databaser

Demografiska databasen, DDB (www.ddb.umu.se)

Demografiska databasen (DDB) vid Umeå universitet startade i Riksarkivets regi 1973. Syftet var att datorisera kyrkoboksmaterial från främst 1800-talet och göra det tillgängligt för forskning. Samtidigt var det ett sätt att minska slitaget på det vid den tiden mycket efterfrågade materialet, inte minst för släktforskning. De första registreringarna avsåg sju församlingar spridda över Sverige, där forskning pågick. Därefter inriktades registreringarna på två regioner, städerna Sundsvall och Linköping med grannförsamlingar. Sundsvallsregionen omfattar idag 16 församlingar jämte själva staden. Linköpingsregionen kommer, när den är färdigregistrerad, att inkludera 35 församlingar utanför staden. Det datoriserade materialet består av individuppgifterna i husförhörslängderna, in- och utflyttningslängderna, samt längderna över födda, döda och vigda. Detaljrikedomen är hög och uppgifterna länkas så det blir möjligt att ta fram biografier över enskilda personer. Kronologiskt täcker man vanligen tiden från sent 1700-tal eller tidigt 1800-tal fram till mitten av 1890-talet, med vissa variationer mellan församlingarna.

Senare tillkom Skellefteåregionen, som omfattar Skellefteå stad jämte ett mindre antal, men till ytan stora, församlingar runt staden. I tid sträcker sig det materialet ända från 1721 fram till 1899. En fjärde region där registreringar inletts och i några fall slutförts är Övre Norrlands inland och fjälltrakter. Totalt skall där ingå 18 församlingar inklusive icke-territoriella samiska församlingar. För Gällivare och Karesuando är materialet registrerat, länkat och klart för användning. Med sökprogrammet Indiko kan man söka och hitta personer i det registrerade materialet. Det är framförallt värdefullt för släktforskare och undervisning, men har förstås även relevans för den vetenskapliga forskningen.

Verksamheten vid DDB har med tiden vidgats inte bara genom att kyrkoböcker för fler församlingar registrerats utan också genom att nya materialgrupper inkluderats. Ett exempel är Tabellverkets befolkningsformulär från 1749 till 1859. För varje församling finns uppgifter för i princip vart femte år över folkmängdens storlek och fördelning på kön, ålder, yrken mm. Dessutom finns årliga uppgifter över nativitet, mortalitet och giftermål och för 1800-talet även flyttningar. En annan registrering avser folkmängden i varje stad och socken vart tionde år mellan 1810 och 1990.

Tillgängligheten till det datoriserade kyrkobokföringsmaterialet begränsas av uttagsavgifter. En eventuell användare måste således betala för att få del av uppgifterna. Tabellverkets material är dock fritt åtkomligt över nätet och där man kan söka på församlingar.

Stockholms historiska databas, SHD (www.ssa.stockholm.se)

Stockholms historiska databas (SHD) startade vid ungefär samma tid som DDB, eller i mitten av 1970-talet med Stockholms stadsarkiv som huvudman. Syftet var att datorisera Rotemansarkivet, dvs Stockholms stad speciella form av kyrkobokföring under perioden 1878-1926. Att göra materialet tillgängligt för forskning var också här ett viktigt skäl för att påbörja registreringarna liksom att skydda materialet från slitage, vilket det redan var svårt utsatt för. Rotemansarkivet innehåller precis som kyrkoböckerna individuppgifter över alla personer som bott och bokförts i Stockholm under den aktuella tiden. Färdigregistrerade församlingar är Maria och Katarina på Södermalm, Nikolai, Klara och Kungsholm. Men ännu återstår cirka tio år innan hela Stockholm är täckt. Materialet finns till viss del utgivet på CD-ROM och DVD skivor. Det är också fritt åtkomligt på Stadsarkivet men man kan behöva viss teknisk hjälp för att göra sina sökningar. Vissa sökningar kan göras direkt på Internet.

SHD har likt DDB vidgats och omorganiserats sedan starten för cirka trettio år sedan. Exempel på andra materialgrupper som datorregistrerats utgör Hälsovårdsnämndens dödsbevis samt register till mantalsuppgifter, kyrkoböcker och bouppteckningar. Dessa materialgrupper är fritt åtkomliga på Stadsarkivets hemsida.

Skånes demografiska databas, SDD (www.ddss.nu)

Tillkomsten av Skånes demografiska databas initierades 1983 av Lunds landsarkiv och Ekonomisk-historiska institutionen vid Lunds universitet. Här finner man uppgifter om födda, döda, vigda, inflyttade och utflyttade för 12 skånska församlingar mellan 1646 och 1894. Program finns för länkning av individer och familjerekonstruktion. SDD har senare integrerats med den större demografiska databasen för södra Sverige.

Demografisk databas södra Sverige, DDSS (www.ddss.nu)

Demografisk databas södra Sverige startade 1997 på initiativ av Lunds landsarkiv som också är huvudman. Projektet är en fortsättning på och utvidgning av Skånes demografiska databas. Samma typ av uppgifter registreras som i SDD, dvs. födda, döda, vigda samt in- och utflyttade. Tidsperioden är fortsatt 1646-1894. Det långsiktiga målet är att församlingsvis registrera all befolkning i Skåne, Halland och Blekinge fram till 1900 och hålla databasen tillgänglig online.

Exempel på andra typer av databaser

SVAR, Svenskt arkivregister (www.svar.ra.se)

SVAR är en enhet inom Riksarkivet som bland annat håller på att bygga en databas över 1800-talets folkräkningar. Den innehåller en registrering församlingsvis av alla individer uppförda i folkräkningarna 1860, 1870, 1880, 1890 och 1900. Folkräkningar hölls vid den tiden vart tionde år. För åren 1890 och 1900 är materialet komplett och tillgängligt på SVAR:s hemsida. Sexton län är i skrivande stund klara 1880, två stycken 1870 och ett 1860. På SVAR:s hemsida finns dessutom en rad andra registreringar utlagda mer eller mindre kompletta, t.ex. bouppteckningar, sockenstämmor, by- och gårdsdata, konseljakter, brandförsäkringar och frigivna fångar. Sökningarna är främst anpassade för släktforskare och man kan till exempel inte få upp folkräkningen för en hel socken, utan man söker på namn.

Folkräkningen 1890 finns dock för fem län – Norrbotten, Västerbotten, Västernorrland, Jämtland och Värmland – även på Forskningsarkivets vid Umeå universitet hemsida (www.foark.umu.se). Och där kan man skapa egen statistik per församling ur databasen.

Svensk lokalhistorisk databas, SLHD (www.lokalhistoria.nu)

SLHD startade 1997 och drivs i samarbete mellan Region Halland och Länsarbetsnämnden i Hallands län och med Region Halland som huvudman. Syftet är att digitalisera protokollen från sockenstämmor, kommunalstämmor, kommunalfullmäktige och stadsfullmäktige från äldsta tid fram till storkommunreformen 1952. Materialet hålls tillgängligt över Internet och är sökbart. De första registreringarna avsåg socknarna och kommunerna i Hallands län, och har senare utsträckts till Jönköpings, Oskarshamns och Vänersborgs kommuner. På längre sikt är målet att täcka hela Sverige.

Medicinhistoriska databasen (pub.epu.liu.se)

Denna databas initierades 1997 av länsarbetsnämnden i Säter som ett arbetsmarknadsprojekt. Numera drivs verksamheten av Linköping University Electronic Press i samarbete med ArkivData i Norrköping. Databasen innehåller digitaliserade dokument som på olika sätt belyser den medicinska och sociala situationen i Sverige sedan mitten av 1700-talet, bland annat årsberättelser från provinsialläkare, lasarett och hospital samt protokoll från medicinalstyrelsen och hälsovårdsnämnder. Materialet är fritt tillgängligt för alla och envar.

Kulturmiljöer

I anslutning till arbetet med kommunala kulturmiljöprogram har till exempel Stockholms läns museum (www.lansmuseum.a.se) gjort det möjligt att på hemsidan via rullister klicka sig fram till en beskrivning av de olika miljöerna. Vidare tillhandahåller Riksantikvarieämbetet digitaliserade bilder på kulturmiljöer från olika delar av Sverige (www.kultur.nu). Liknande information finns på vissa kommuners hemsidor. En stor regional kulturhistorisk satsning utgör Kulturarv Östergötland. På hemsidan (www.kulturarvostergotland.se) finns ingångar till en rad olika databaser och register av skilda slag inklusive länkar till demografiska databaser med flera.

Historiska kartor

Lantmäteriet (www.lantmateriet.se) har digitaliserat flera typer av historiska kartor, bland andra ekonomiska kartan, häradskartor, stads- och sockenkartor. I rullister söker man sig fram till en viss karttyp för den ort man är intresserad av. Man får dock inte tillgång till själva kartan utan att betala för den. En CD-ROM med historiska stadskartor utgavs 1998.

Bilddatabaser

Under senare år har på åtskilliga håll ett omfattande arbete lagts ned på att bygga bilddatabaser. Många museer, arkiv, bibliotek och andra institutioner har digitaliserat sina innehav av historiskt bildmaterial och ofta gjort det lätt åtkomligt över Internet. Här finns inte plats att ange alla. Den intresserade kan söka på www.google.se.

CyberCity

Stads- och kommunhistoriska institutet tillhandahåller via sin hemsida ett antal digitala produkter bland andra CyberCity (www.urbanhistory.su.se). CyberCity, eller Svenska städer på nätet, är ett slags uppslagsverk över de 133 orter i Sverige som var städer fram till kommunreformen 1971. Informationen är vanligen bearbetad för att i största möjliga utsträckning vara jämförbar över tid och mellan orter. CyberCity innehåller idag följande variabler:

- Folkmängd från 1570-tal till 1995 i absoluta tal och som andel av rikets stads- och tätortsbefolkning
- Födda, döda, födelse- och flyttningsnetto 1835-1950
- Nativitet, mortalitet samt relativa födelse- och flyttningsnetton 1835-1950
- Yrkesfördelningar, absolut och relativt 1805-1855, 1900-1960
- Näringsprofil
- Kommunala val, mandatfördelningen i stadsfullmäktigevalen 1919-1966 och kommunvalen 1970-2006
- Litteratur
- Korthistorik

Samtliga dataset i CyberCity är fritt tillgängliga och kan hämtas hem i såväl pc- som macformat.

Ytterligare några exempel

Exemplen ovan är just exempel och inte någon fullständig inventering av befintliga databaser av viss storlek. Många forskare har byggt eller bygger egna register i anslutning till sina projekt och försöker göra dem tillgängliga på olika sätt. Ett exempel på ett sådant påbörjat arbete är "Svenska jordbruket 1570-1805" av Lennart Andersson Palm. Där är avsikten att registrera data om befolkning och jordbruksförhållanden. Folkmängdsdata på sockennivå finns redan lagrade från ett tidigare projekt.

Ett annat exempel är Örjan Simonsson som till Upplandsmuseet donerat de databasfiler över Torstuna härad på 1600-talet han tog fram i anslutning till sitt avhandlingsarbete. Det är uppgifter ur domböcker, jordeböcker och mantalslängder. Tanken är att länka detta material till andra databaser vid museet. Källorna till Göran Samuelssons avhandling "I godsets skugga" (2004) finns utgivna på CD-ROM skiva. Det rör sig om kartor, kyrkoböcker med mera från 1700- och 1800-talen.

Användare

Befintliga historiska databaser har en bred avnämning. Där ingår inte bara historiker och ekonomhistoriker utan också till exempel geografer, sociologer, arkitekter, medicinare, tekniker och naturvetare. Det finns dessutom en internationell efterfrågan särskilt på de demografiska databaserna och det unika svenska kyrkoboksmaterialet. En omständighet som begränsar användningen är de avgiftssystem som tillämpas av till exempel DDB och Lantmäteriet.

Framtida behov

För framtiden finns det minst två alternativ att fundera kring. Skall vi investera ytterligare i redan påbörjade registreringar av bland annat demografiskt material och i någon mening slutföra dem innan vi på allvar startar med att bygga nya stora databaser. Eller skall resurser huvudsakligen satsas på att datorisera nya materialgrupper. En tredje variant är förstås att göra litet av både och, såväl fortsatt uppbyggnad av befintliga system som visst nytänkande.

Det finns många goda skäl till att fortsätta och vidareföra de arbeten som så framgångsrikt inletts och redan pågått under lång tid. Stora investeringar har gjorts och det vore oklokt att inte tillvarata dem på bästa sätt. Att få de kommunala protokollen datoriserade för hela Sverige (SLHD) och enkelt tillgängliga är önskvärt för många historiker. Likaså vore det en guldgruva att ha 1800-talets kyrkböcker datoriserade för hela Sverige. Nu finns materialet för några områden i norr och några i söder och den demografiska forskningen styrs dit. Den historiska stadsdemografien har således under senare tid nästan enbart behandlat städerna Sundsvall och Linköping. Hade vi datoriserade dataset för hela Sverige vore säkert bilden en annan och kanske skulle även vår uppfattning av den historiska utvecklingen påverkas.

Byggandet av en stor databas är utsträckt över lång tid, vilket innebär att forskningstrenderna hinner skifta en del. När DDB och SHD/Rotemansarkivet startade i mitten av 1970-talet var intresset för demografisk och social historia stort, vilket bidrog till att motivera satsningarna. När en mer kulturhistorisk inriktning tog över sjönk efterfrågan inom ämnet på demografiskt och kvantifierbart material och ökade på så kallade "kvalitativa" material. Det är också då man börjar investera i kulturarvsdatabaser, bilddatabaser och liknande. Vill man att nya registreringar skall fånga aktuella trender bör man förstås främst satsa resurserna på nya och inte på befintliga databaser anpassade efter andra strömningar. En sådan praktik leder dock till ryckighet, då trender skiftar snabbare än den tid det tar att bygga upp en stor rikstäckande databas. Det verkar därför oklokt att bara satsa på datoriseringar av nya materialgrupper. Jag förordar personligen den gyllene medelvägen, en fortsatt uppbyggnad av befintliga baser i kombination med strategiska nysatsningar.

En kort rundfrågning som Marie Clark Nelson och jag själv gjort bland historiker runt om i landet ger inget klart besked om framtida önskemål. Det tycks inte finnas någon entydig opinion vare sig för enbart fortsatta investeringar i befintliga databaser eller för enbart omfattande datorisering av nya materialgrupper. Snarast är det alternativet "både och" som förordas, men framförallt talar var och en för sin vara.

Vad man kan önska för framtiden är också en bättre samordning av pågående registreringar, till exempel mellan DDB, SHD och DDSS som alla arbetar med demografiska dataset. Men även mellan SVAR och SLHD som båda registrerar sockenstämmoprotokoll. Man bör också undvika dubbelarbete, dvs. att samma material, t.ex. folkmängdsuppgifter, datoriseras på flera håll. Inte för att det är störande att finna samma information i ett antal databaser mer för att

det kan vara ett slöseri med resurser. Många enskilda forskare datoriserar numera sina källmaterial i egna databaser och även här kan man göra betydande vinster om forskaren efter utförd uppgift gör materialet tillgängligt för hela forskarsamhället. Tillgängligheten kan vidare ordentligt förbättras om man tar bort de avgiftssystem som idag hindrar många forskare att använda datoriserat material. Det kan vara enklare och billigare att bygga sin egen databas än att dyrt köpa det redan datoriserade materialet.

Samverkan mellan nationer borde också behöva stärkas. Vissa insatser har förvisso redan gjorts. Ett exempel är HISCO, som är ett internationellt gemensamt system för klassificering av yrken. I Sverige används det bland annat av SHD/Rotemensarkivet. Men mycket mer kan göras. En stor databas med historiska dataset över det nordiska och europeiska stadsväsendet är något jag som urbanhistoriker gärna ser fram emot.

**Till vetenskapsrådets Disc grupp:
(Att: Eva Strangert, Jan Larsson)**

Rapport angående databaser inom humaniora rörande ämnet historia

Databaser inom humaniora är ett stort och amorfiskt ämne, som är svårt att få ett bra grepp över, dels pga. mångfalden och dels pga. den tillväxt på alla möjliga nivåer. Utvecklingen av datorer och mjukvaror möjliggör för var och en att ge sig i området och skapa sin egen liten databas inom ramen för ett forskningsprojekt. Databaser har mycket skiftande karaktär och har definierats olika. Förutom att samla in material har Lars Nilsson och undertecknad skickat ut ett brev till historiker i landet för att försöka samla synpunkter. Tidpunkten (kring jul, nyår och terminens slut och början) innebar få svar, men försöket genomfördes.

Stora databaser

För att försöka bringa ordning till ett något kaotiskt område har den följande presentation delats upp i ett antal områden. Ordet databas var länge synonymt med ett antal *stora befolkningsdatabaser* som skapade tidigt i Sverige: *Demografiska databasen vid Umeå universitet*, *Stockholmsstads databas* och även databaserna vid Lunds universitet, framförallt i anslutning till Institutionen för ekonomiskhistoria.

Den Demografiska databasen i Umeå innehåller registrerat material från kyrkoarkiven för ett flertal områden: Sundsvallsområdet, Linköping med omnejd, samt vissa andra utvalda kommuner. Registrering pågår av vissa kommuner i det norrländska inlandet. Mycket av materialet är länkat, som underlätta sökandet efter individer. Materialet i denna befolkningsdatabas, Popum, är tillgängligt via INDIKO, ett söksystem på nätet. Systemet är gratis för studenter vid landets universitet under en tvåmånaders tidsperiod för t.ex. uppsatsskrivande efter överenskommelse; för övriga användare kan ett abonnemang täckas av såväl individer som grupper. Om forskare har speciella behov som inte kan fyllas genom att använda INDIKO kan en offert erhållas. Tillgängligheten har förbättrats avsevärt under de senaste åren. Det bör även noteras att Linköpingsdelen av databasen är också tillgängligt via Linköpings universitet.

Tillgänglig utan kostnad är "Folkmängdssiffror för varje församling i landet 1810-1990" samt "Tabellverket på nätet" som innehåller mer detaljerade siffror på församlingsnivå. Detta aggregerat material är användbart för forskare, studenter och amatörhistoriker och genealoger. Databasens material är även av stort intresse för utländska forskare som har använt databasens resurser genom beställningar och genom vistelser i Umeå såväl som forskare inom andra vetenskapliga områden. Nya användningsområden är under utveckling (se nedan).

Inte oväsentligt är det faktum att Databasen i Umeå är en bra ingång för att söka andra väsentliga databaser av kvalitet på olika nivåer inom och utanför Sverige.

(<http://www.ddb.umu.se>)

Vid Lundsuniversitet finns det ett antal databaser, men mest känd är Skånes demografiska databas (SDD), som är en forskningsdatabas över nio landsbygds församlingar i Skåne som har utvecklats i samarbete med Landsarkivet i Lund. (Den kan lätt förväxlas med Demografisk Databas Södra Sverige (DDSS) som är framförallt en släktforsknings databas.) Denna databas är longitudinell och innehåller individbaserad information på demografiska förhållanden, men även socioekonomiska variabler såsom yrke, jord, typ av boställe, ägandeförhållande samt hushållsstruktur. Allt material är inte tillgängligt direkt på nätet, men är tillgänglig via institutionen. Demografisk material kan dock sökas via DDSS. (<http://www.ekh.lu.se/ed/databaser/sdd.asp>). Materialet används inte enbart av historiker, men även av forskare inom andra vetenskapliga grenar från Sverige och utlandet.

Lokalhistoriska databaser

Ett annat viktigt område omfattar *de lokalhistoriska databaserna*. Den äldsta, störst och mest känd är *Stockholmsstads databas* som är också en av de gamla med ett mycket stort material om staden. Denna rapport kommer inte att ta upp mer detaljer pga. Lars Nilssons mer ingående kunskap inom området. Det nära samarbetet dock mellan Stockholmsstads databas och det arbete som utförs vid Institute of Urban History, Stockholms universitet (som Lars Nilsson är ansvarig för) är ovärderligt. Här bör "Cybercity" sidan på nätet nämnas som en stor källa till information om svenska städer över tid.

(<http://www.historia.su.se/urbanhistory/cybcity/stad/stockholm/historia.htm>).

Men det finns även databaser på en mindre skala som gör försök att täcka ett landskap eller region. Här kan det inte vara en fråga om en omfattande registrering av material som, t.ex. Stockholmsstads databas har gjort sen sin start. Ett exempel är Svensk lokalhistorisk databas (SLHD), baserad i Halmstad, som rör huvudsakligen sydvästra Sverige. Målet är "att digitalisera de svenska kommunal- och sockenstämmoprotokollen samt göra dessa tillgängliga via Internet." Utöver detta finns framförallt ingångar till information angående källor som har anknytning till lokalhistoria via landsting, museer, arkiv, mm. (<http://www.lokalhistoria.nu>)

Ett annat exempel är KulturarvÖstergötland, en sida som ger ingångar till diverse databaser och nätsidor som tar upp olika aspekter av Östergötlands historia. Här finns flera omfattande bild databaser som berör Linköping och Norrköping samt länkar till databaser som innehåller material relevant till området, t.ex., DDB, Arkivdata i Norrköping, mm. Länkar till alla arkiv och museer som kan tänkas vara av värde finns också med. Särskilt uppmärksamhet har riktats åt att sprida användningen till skolor. (<http://www.kulturarvostergotland.se/>)

Ämnesriktade databaser – exempel medicinhistoria

Utöver databas typerna som nämndas ovan finns databaser inriktade mot särskilda ämnen. Exemplet här har tagits från medicinhistoria som väl kan illustrera vilka diverse databaser kan finnas. *Medorg* är en typ. Denna databas innehåller beskrivningar av drygt 50 000 artiklar, utredningar, eller skrifter med relevans för medicin historia som publicerades mellan 1663

och 1967. Den är i princip en kortkatalog som registrerades digitalt för att göra detta tillgängligt. Skrifterna måste sedan sökas på bibliotek. (<http://www.medorg.scb.se/>)

Den *Medicinhistoriska databasen* är en databas med en brokig historia. I början var denna arbetsmarknadsåtgärd baserad i Säter på Utbildningscentrum i Dalarna inriktad på att registrera historiska källor, i det här fallet provinsialläkare berättelser, i en databas. År 2002 fördes materialet över till DDB i Umeå för att så småningom tas över av Arkivdata i Norrköping och Linköpings universitet. Andra historiska källor, som Medicinalstyrelsens protokoll, Inspektionsberättelser för hospitalen, och årsberättelser för lasarett och hospital har registrerats. (<http://pub.ep.liu.se/medhist/>)

Via olika sjukhus, Karolinska institutet och Sahlgrenska, t.ex. har så kallade databaser utvecklats inom medicinhistoria som är brokiga samlingar av alltifrån bilder av gamla dokument, länkar till forskningsbibliotek och arkiv i utlandet till egenhändiga listor och diverse hemsidor av högst skiftande kvalitet och tillförlitlighet.

Arkivdata i Norrköping som nämndes ovan är en arbetsmarknadsåtgärd som sysslar med registrering av historiska källor. Utöver arbetet med det medicinhistoriska materialet som nämndes ovan, har källor från rättsväsendet för Linköping och Norrköping (exempelvis fångceller, spinnhuset, poliskammaren) gjorts tillgängliga.

(<http://www.norrkopings.se/jobb/arbetsmarknadsatgarder/arkivdata/>)

I allmänhet kan dessa exempel illustrera en del av de stora problemen som finns. Insamlingen och avskriften av material har ibland påbörjats utan någon större eftertanke eller reflexion i frågor kring kvalitetskontroll och utan förståelse för de tekniska kraven som ställs av en sökbar databas. I andra fall har värdefullt material kunnats räddas och göras tillgänglig för framtida forskare. Att hitta rätt och att kunna urskilja tillförlitliga källor och material från dem som man inte kan lita på kommer att ställa stora källkritiska krav på individen.

Små individuella databaser

Det finns ytterligare en nivå av databaser, som här har döpts till *Individuella insatser*. Inom ramen för sina forskningsprojekt har individuella forskare ibland samlat på sig excerpter av olika slag som sparas i personliga databaser. Precis som de gamla pappersexcerpter som hamnade på vind och i källare och även undantagsvis på institutionsarkiv (där de egentligen bör finnas), samlas nu denna forskning på diverse hårddiskar och CD-skivor.

En annan typ som är mer tillgänglig är databaser som finns som bilagor till doktorsavhandlingar som har publicerats elektroniskt. Ett exempel är en Linköping avhandling av Claes Westling, *Småstadens dynamik*, där den elektroniska upplagan innehåller mycket av det källmaterial som samlades.

(<http://www.diva-portal.org/liu/abstract.xsql?dbid=4698>).

Arkiv

Ett stort område är databaser som ger tillgång till arkiv, ett område som ligger utanför denna översikt, men som är mycket viktigt för historiker. NAD är givetvis ett fint hjälpmedel för alla historiker och har lagt grunden till ökad tillgänglighet till arkivförteckningar över hela landet. Men de individuella arkiven på alla nivåer har i varierande grad databaser som ibland ger tillgång även till forskningsmaterial. Utvecklingen går snabbt och är svår att överblicka. Riksantikvarieämbetets hemsida är ett bra exempel på hur en del material har gjorts tillgängligt. (<http://www.raa.se>)

Internationella kontakter och projekt

När det gäller *internationella kontakter* är det de stora befolkningsdatabaser framförallt som har kontakt med sina motsvarigheter i utlandet och som har haft besök av utländska forskare som har utnyttjat materialet. Institute for Urban History (och även databasen) i Stockholm har omfattande kontakter med utländska universitet via internationella nätverk i stadshistoria. När det gäller de stora databaserna kan nämnas, förutom de nordiska, kontakter och utbyte med EHESS (L'École des Hautes Études en Sciences Sociales, Paris); IPUMS (Integrated Public Use Microdata Series, University of Minnesota, Minneapolis, Minnesota, USA); NIDI (Netherlands Interdisciplinary Demographic Institute); och IREP (Institut interuniversitaire de recherches sur le population, Montréal, Canada).

Det finns även ett par exempel på pågående internationella forskningsprojekt som bör nämnas. The EurAsia project är ett internationellt samarbete som inkluderar forskare från USA, Kina, Japan, Belgien och Sverige, där forskare från Institutionen för ekonomisk historia vid Lunds universitet har bidragit. Projektet har publicerat ett stort antal artiklar och nu har monografierna börjat komma från förlagen. Ett projekt som befinner sig på ett tidigare stadium är "Causes and effects of consanguinous marriages", som är ett samarbete mellan Umeå universitet och Perth i Australien. Tillgång till stora longitudinella databaser av hög kvalitet har varit avgörande för sådana projekt.

Framtiden

Hur kan databaser utvecklas? Svaren är många. Men ett par utvecklingslinjer kan dock skönjas. Ett konkret exempel är "äkta" tvärvetenskapliga studier, där forskare med bakgrund i olika discipliner samarbeta för att undersöka problem tillsammans. Ett projekt vid bl. a. forskare från Centrum för befolkningsstudier vid Umeå universitet har beviljats pengar från Krisberedskapsmyndigheten för studien "Influensapandemiers påverkan på individ och samhälle – nödvändig erfarenhetsbakgrund för pandemiplanering". Ett annat exempel är Twin Cities gruppen vid Linköpings universitet där medicinare, sociologer, psykologer, historiker för närvarande utvecklar studier av hälsa under övergången från den industriella till den post-industriella staden.

Ett annat sätt är att arbeta över gränserna för att hitta nya användare av databaserna. Ett område, som delvis kan ses i exemplen ovan, är medicinska studier, där t.ex. epidemiologer och genetiker börja se nya användningar för historikernas data.

Hur ser historiker på framtiden? Idéerna kring vilka typer av databaser behövs i framtiden berör dels på vilka man fråga. Många historiker som koncentrerar sig på äldre tider (medeltid, tidigmodern) vill se mer källor på nätet. Varje specialitet vill att just sitt område ska stå i rampljuset. Många trycker på vikten av att verkligen stödja de gamla stora databaserna: resonemanget är att så mycket har investerats att det är synd att låta dessa förfalla.

Ett tema som genomsyrar nästan alla kommentarer är behovet av att få någon typ av ordning. Hur kan vi hitta bland alla möjliga och omöjliga databaser? En databas över databaser är ett stort önskemål.

Hur kan kvaliteten säkras? En del av exemplen ovan illustrerar dessa problem. Kunde det finnas vissa krav som skulle uppfyllas innan en databas kan inkluderas i basen över databaser?

Hur kan nya användningsområden uppmuntras? Det är en viktig fråga.

Utvecklingen av databaser och den utökade användning av datorer är en stor utmaning och att våga ta itu med att säkra en sund användning är en stor och viktig uppgift inom alla ämnesområden.

Linköping, den 21 januari 2007.

Marie Clark Nelson

Medeltidsforskning: databaser och intern forskningsstruktur

Inom medeltidsforskningen skapar några forskare egna databaser för de speciella syften de för tillfället arbetar efter, och dessa databaser har sällan ett vidare användningsområde: de är tillkomna ad hoc. Vid uppbygganden av den egna databasen är de normalt helt beroende av de databaser som finns på de lärda verken och muséerna. Dessa är m.a.o. oundgängliga för medeltidsforskningen inte bara inom ämnet historia utan också för konstvetare, arkeologer, språkforskare (främst nordiska språk och latin).

Det tillhör beklagligheterna att en ny lärdomsfientlig ideologi åren omkring 1990 slog igenom på de statliga institutionerna och inom några år ödelade förutsättningarna för flera grenar av denna verksamhet så att ambitiösa och angelägna projekt inte har kunnat fullföljas.

För Vetenskapsrådet kan det vara väsentligt att veta att flera av projekten initierades och årligen tillfördes medel från dåvarande Humanistisk-samhällsvetenskapliga forskningsrådet, tills de överfördes till ordinarie stat på mottagande statliga verk, varvid forskningsrådet samtidigt berövades samma medel. Vad som sedan skett (under 1990-talet) är att de tilldelade medlen exproprierats och använts för annat ändamål än som avsett var, då riksdagen fattade beslut om överföringen.

Riksantikvarieämbetet (RaÄ)

Vid RaÄ fanns tre s.k. lärda verk Runverket, Det medeltida Sverige och Sveriges kyrkor, som alla är mer eller mindre nedlagda, liksom Fornminnesinventeringen.

1. Äldst är *Runverket* som skulle vårda men också edera/tolka befintliga runinskrifter genom utgivning i verket *Sveriges Runinskrifter*. Arbetet närmar sig slutet, bara några Norrlandslän återstår med ett tjugotal inskrifter och delar av Gotland. De tidigt utgivna banden behöver dessutom kompletteras och korrigeras. Nyupptäckter och lösfynd publiceras regelbundet i *Fornvännen*. Allt publicerat material är infört i den samnordiska rundatabasen i Uppsala (www.nordiska.uu.se/forsk/samnord.htm) som alltså upptar inskrifter från hela Norden. Ett problem för den vidare användningen är att inskrifterna i databasen inte är tolkade/översatta till modernt språk, vilket ger upphov till problem då de utnyttjas av icke-filologer, t.ex. arkeologer, historiker och konstvetare.

En tjänst finns ännu kvar på Runverket, men vad som händer, när vederbörande går i pension, är nog att tjänsten inte återbesätts. Det betyder att en institutionaliserad kompetens försvinner som kan slutföra arbetet och ta hand om och bearbeta de nyfynd som årligen görs.

2. Sveriges kyrkor var ett byggnadsarkeologiskt inventarium, där enskilda kyrkor fått en monografisk beskrivning. Verket är nedlagt, men det lever i viss mån vidare genom det s.k. sockenkyrkoprojektet (finansierat av Vitterhetsakademien och Riksbankens jubileumsfond), där Sverige alla sockenkyrkor bearbetas efter en viss mall. 12 rapporter har publicerats, och återstoden torde komma att omfatta ytterligare 9 rapporter med en tidsåtgång på ca fyra år, låt vara att alla grunddata är insamlade. Genom sin generella inriktning och sin utvecklade systematik erbjuder projektet ett värdefull grundmaterial till förståelse av kyrkornas byggnadshistoria. Insamlade grunddata finns tillgängliga på projektet genom en axess-databas som hålls tillgänglig för forskning. Bearbetade grunddata matas löpande in i och görs allmänt tillgängliga genom kulturmiljövårdens bebyggelseregister (www.bebyggelseregistret.raa.se), som omfattar grunddata av kulturhistoriskt värdefull bebyggelse.

3. *Fornminnesregistret* (FMR) hade sin grund i en stor inventering av rikets fornlämningar som var sammankopplad med utgivningen av den ekonomiska kartan där de fasta fornlämningarna skulle läggas in. Riksantikvarieämbetets Fornminnesinventering startade mot slutet av 1930-talet och avslutades 2002. Viss fornminnesinventering pågår fortfarande i länsstyrelsernas regi. Uppgifterna fördes in i anteckningsböcker och på papperskartor (vanligen ekonomiska kartan). Uppgifterna i FMR har under senare år digitaliserats och förts in en stor databas, FMIS (Fornminnesinformationssystemet), ett geografiskt informationssystem. FMIS har två "ingångar", ett som är tillgängligt för allmänheten (www.kms.raa.se/cocoon/fmis-public/index.info), och ett som kan nås via särskild inloggning som ges till yrkesverksamma inom samhällsplanering och miljövård, samt till anställda på universitet och högskolor. FMIS uppdateras kontinuerligt med ny information utifrån arkeologiska undersökningar och olika inventeringar.

4. I *Bebyggelseregistret* finns information om kulturhistorisk bebyggelse. Informationen är sökbar via Internet (www.bebyggelseregistret.raa.se) och nya uppgifter läggs in kontinuerligt. Uppgifterna kommer från kulturhistoriska inventeringar vanligtvis utförda av landets läns museer på uppdrag av bl.a. kommuner och Svenska kyrkans stift. Dessa registerförare ansvarar för registrets kulturhistoriska sakinnehåll och värderingar. Rikstäckande information finns för byggnadsminnen och kyrkor. Texter och planer från "Sockenkyrkoprojektet" förs in när material för inskrivning levereras. Även information om olika typer av profan bebyggelse finns att söka ojämnt fördelat över landet.

5. *Det medeltida Sverige* (DMS) är en topografisk-historisk handbok, som i bearbetad form anger grunddata för bebyggelsen i medeltidens Sverige (fram till 1550) på landsbygden och i städerna. Utgångspunkt för bearbetningen var den minsta dåtida kamerala enheten (gården, byn) varpå data systematiseras till flera aggregerade nivåer som socken, härad, städer, landskap/stift. Stora delar av Uppland är utgivet, liksom delar av Södermanland och Småland. Digitalisering av uppgifterna ur de på senare utgivna volymerna pågår och kommer inom kort att inordnas i Riksantikvarieämbetets databas för Sveriges kulturmiljöer.

Statens historiska museer

Två stora databaser finns här allmänt tillgängliga.

6. Det större är ett s.k. föremålsinformationssystem som hanterar museets samlingar (www.historiska.se/data <<https://owa.vinnova.se/exchweb/bin/redirect.asp?URL=http://www.historiska.se/data>>) och som omfattar ca 465 000 föremålsposter och 265 000 dokument och bilder. Merparten av de enskilda uppgifterna rör de arkeologiska samlingarna och endast en mindre del de kyrkohistoriska samlingarna. Av det totala beståndet av föremål är ändå bara drygt 5% införda, men digitalisering sker nu snabbt tack vare statliga Access-pengar, ca 500 poster per dag.

7. Den mindre databasen är direkt förknippad med medeltida kyrkokonst, och den består av en rikt illustrerad nationell översikt över ett urval av de bäst bevarade medeltida föremålen och kalkmålningarna i kyrkor och museer, omfattande totalt 4200 objekt och 19 000 fotografier. Databasen heter Medeltidens bildvärld och nås via medeltidbild.historiska.se

Riksarkivet (RA)

Vid Riksarkivet finns två projekt av intresse: Svenskt Diplomatarium och Nationalutgåvan av äldre geometriska kartor (1630-55).

8. *Diplomatariet* har till uppgift att edera de svenska medeltidsbrev efter moderna textkritiska principer. Hittills har alla brev till 1376 och 1401-1420 givits ut i tryck. Allt detta finns tillgängligt i digital form, dels som råtext, dels via PDF-filer av boktrycket. Därtill kan användarna söka i det övriga brevmaterialet genom de register (kortfattade avskrifter) som finns online i Diplomatiets huvudkartotek, som är den databas som fungerar som nav i onlinepresentationen av medeltidsbrev. Totalt rör det sig om ca 41 000 brev som på så vis

är tillgängliga. Dessutom sker en digitalisering av originalhandlingarna. För närvarande har detta arbete resulterat i digitala bilder av pergamentsbrev för perioden 1420-1480, vilka under våren planeras bli tillgängliga via databasen, medan bearbetningen av det äldre brevmaterialet inte har påbörjats. Sökning sker via www.ra.se/ra/diplomat.html.

9. Från perioden 1630-1655 finns omkring 12 000 storskaliga kartor över svenska gårdar och byar. I projektet "Nationalutgåva av de äldre geometriska kartorna" skannas originalakterna och skrivs ut i färg och skalenligt på arkivpapper samt lagras i en digital kopia. All text på kartorna och Notarum Explicatio transkriberas och läggs kontinuerligt ut på projektets hemsida (www.ra.se/ra/geometriska). I dagsläget är 28 kartsamlingar sökbara (ca. 3500 kartor). Därutöver skapas ett GIS för samtliga bebyggelseenheter och tillhörande statistisk information registreras i en relationsdatabas. Via ett sökformulär kan forskare, utifrån egenvalda frågeställningar, göra tematiska och geografiska urval och ta del av kartorna digitalt. Databasen och pappersarkivet finns tillgängliga på Riksarkivet i Marieberg. Projektet pågår fram till 2009 och finansieras av Vitterhetsakademien och Riksbankens Jubileumsfond.

Stockholms universitet

10. Numismatiska forskningsgruppen vid Arkeologiska institutionen fortsätter det arbete som tidigare låg under **Kungl. Myntkabinettet i Stockholm** och utfördes via två särskilda tjänster som Myntkabinettet självsvåldigt omvandlat till annat arbete. Det aktuella arbetet som sker vid sidan av annan tjänstgöring avser registrering av myntfynd i främst Sverige (lösfynd, hopade fynd, skatter) som är uppdelade på flera databaser som tjänar som forskningsredskap. Databaserna omfattar i nuläget ca 270 000 poster, mycket återstår att göra avseende det svenska materialet som i bearbetat och inmatat skick torde omfatta två gånger så många poster som de nu tillgängliga. Delar av det bearbetade materialet finns tillgängligt på nätet: www.numismatiskaforkningsgruppen.su.se.

Sammanfattning

När det gäller användningsområden kan en schematisk uppdelning på akademiska discipliner ge viss information, varvid siffrorna svarar mot den numrering som följer den löpande redovisningen:

	Arkeologi	Språk	Konstvet	Kulturgeografi	Historiker	Utländsk efterfrågan
1.	x	x	x			x
2.	x		x			
3.	x		x			
4.	x		x	x	x	
5.	x	x	x	x	x	
6.	x					
7.			x		x	x
8.		x			x	x
9.	x			x	x	x
10.	x					x

 Arkeologi innefattar även numismatik, Språk avser nordiska språk (inkl ortnamnsforskning) och latin, Konstvet innefattar även byggnadsarkeologi, Historiker innefattar även ekonomhistoriker.

Som framgått behöver flera av de nämnda projekten stöd för att fullfölja sina åtaganden. Det gäller närmast Runverket (1), Numismatiska forskningsgruppen (9), diplomatarieverket (7) och Det medeltida Sverige (4). För DMS gäller det inte bara stöd för digitalisering av det som redan är bearbetat utan en restaurering av projektet så att nya områden kan bearbetas och framtagna data inordnas i en databas. Vid sidan av det jättelika fornminnesregistret (3) och diplomatarieverket har DMS bearbetningen störst avnämarpotential i forskningshänseende, och det bygger på en bearbetning av ett internationellt sett unikt källmaterial.

Frågor ang. databaser och digitalisering inom arkeologi¹

Generellt: arkiver med tillhörande databaser spiller en central roll för arkæologisk forskning. AL arkæologisk information är baserat på systematisk dokumentation och registrering av fortdisminder i landskabet (inventeringen), och av arkæologiska genstande i museernes magasiner og arkiver med tillhörande beskrivelser. Der findes i Sverige som i de fleste lande flere typer nationale registre, ikke alle datoriserede. Dessutom finns en stor mängd data från miljöarkeologiska undersökningar lagrade i både analog och digital form.

1. Fornminnesinventeringen med fortdisminderegistret og lænker til udgravninger mv.(er digitaliseret). Under denne findes en række specialregistre over f.eks. helleristninger, runesten etc. som idag ikke er digitaliserede (eller i begrænset omfang)
2. Oplysninger om fund fra fortdisminder i museerne. Disse findes i varierende grad dokumenteret og der savnes en national database. Vitterhetskademien har samlet megen information i deres Antikvarisk topografiske arkiv (ATA) som er inddelt efter sogn. Dette register burde digitaliseres, som en lænke mellem inventeringen, museernes samlinger og udgravningsberetningerne.
3. Dokumentation af arkæologiskes udgravninger i form af beretninger. Dokumentationen av arkeologiska undersökningar sker sedan 1990-talet i huvudsak digitalt i form av geografiska databaser och andra digitala dokument. Att skapa en policy för lagring och tillgänglighet för dessa är en angelägenhet av nationellt intresse.
4. Der findes mange kataloger over udvalgte emner, udarbejdet i forbindelse med doktorsafhandlinger eller lignende, som også har stor betydning. Visse af disse fremstilles idag digitalt.
5. Flera naturvetenskapliga laboratorier med anknytning till arkeologi producerar kvantitativa data. Hit hör bl a miljöarkeologiska analyser, C14-dateringar, kemiska analyser, osteologiska bestämningar mm.

Tilsammen udgør dette grundniveauerne i al national arkæologisk dokumentation, men den findes ofte spredt på mange museer og i publikationer. Idag udføres de fleste arkæologiske udgravninger med digital feltdokumentation, og kan således umiddelbart gøres tilgængelig.

¹ Frågorna har besvarats av Kristian Kristiansen och Karl-Göran Sjögren, Institutionen för arkeologi, Göteborgs universitet. Institutionen för arkeologi och samiska studier, Umeå universitet har genom Phil Buckland kompletterat med uppgifter angående miljöarkeologi. Delar av dessa uppgifter har inarbetats i den danska texten och andra delar föreligger på engelska.

Der burde oprettes en national handlingsplan for arkæologiske databaser, da forskningen mere end i de fleste andre discipliner er helt afhængig af sådanne.

Arkæologiske databaser, som inventeringen, anvendes også i forskningen sammen med geologiske og geografiske databaser, i det som kaldes Geografiske Informationssystemer (GIS). Samarbejde med disse andre discipliner er således vigtigt for forskningen fremtid. Også naturvidenskabelige databaser anvendes (f.eks over C14 dateringer eller økologiske data). Det är väsentligt att arkeologiska databaser utformas för att kunna sammanlänkas med dessa.

Kristian Kristiansen

Kortfattede svar på spørgsmålene:

1. Situationen idag.

Databaser finns på en rad olika nivåer, internationella, nationella, regionala och individuella. Som exempel på internationella kan nämnas Archaeology Data Service (<http://ads.ahds.ac.uk/index.html>) eller C14-databaser som Banadora (<http://www.archeometrie.mom.fr/banadora/>). Internationella miljöarkeologiska databaser finns också. Bl a kan nämnas BugsCEP (<http://www.bugscep.com/>) som är en databas över fossila insekter i Europa, och som även tillhandahåller en del analysverktyg. Många sådana resurser är samlade på World Data Center for paleoclimatology (<http://www.ncdc.noaa.gov/paleo/paleo.html>) och The PAGES (Past Global Changes) network (<http://www.pages.unibe.ch/>).

SEAD – the Strategic Environmental Archaeology Database (<http://www.sead.se/>) is planned for development at the Environmental Archaeology Lab, in cooperation with HUMLab, in Umeå. It will provide an international database of environmental archaeology raw data (including that held by MAL), along with a number of GIS and query based tools for interrogating and analyzing these data.

På nationell nivå finns Riksantikvarieämbetets fornlämningsdata, FMIS, eller dess danska motsvarighet DKC, historiska kartor och andra kartdata vid Lantmäteriet och SGU, samt samlingar vid vissa museer (SHMs samlingar är sökbara över nätet liksom flera regionala museers). Här kan också nämnas databaser över analysresultat från arkeologiska laboratorier. Dessa är oftast inte tillgängliga externt.

På regional nivå finns databaser över många museisamlingar, både föremål och bilder. Dessa är av varierande fullständighet och kvalitet. Databasuppbyggnad pågår på många museer inom Accessprojektet.

De flesta databaser torde vara individuella eller knutna till vissa projekt. De är oftast inte tillgängliga utifrån. Några är dock nedladdningsbara, se t ex <http://hum.gu.se/institutioner/arkeologi/publikationer/download/> . Senare års undersökningsrapporter är i många fall nedladdningsbara som pdf eller i form av GIS-databaser, se t ex <http://hildebrand.raa.se/uv/rapporter.htm> .

Som påpekades ovan finns dock den stora mängden data lagrad i icke digital form, i en rad arkiv och musiemagasin på både nationell och regional nivå. Ett arbete med att skapa en översikt över europeiska arkeologiska arkiv har under senare år bedrivits inom AREA-projektet (<http://www.area-archives.org/>).

2. Utnyttjande av tillgängliga databaser.

Svårt att uttala sig om användningen, förmodligen mest på individ- eller projektnivå i samband med uppsatser och avhandlingar. De större databaser som idag är sökbara över internet används regelbundet på alla nivåer av forskning liksom inom den antikvariska verksamheten, men i ökad utsträckning också av allmänheten och i den allmänna samhällsplaneringen. Hit hör väl främst FMIS och SHMs samlingar. Man kan också notera en ökande användning av digitala resurser i samband med undervisningen.

Databasernas användning internationellt begränsas till stor del av språket eftersom de flesta inte är på engelska. Undantag är t ex BugsCEP.

3. Samverkan vid databasuppbyggnad.

Det förefaller som om samarbetet vid databasuppbyggnad i huvudsak varit begränsat till projektgrupper, laboratorier eller institutioner, undantag är dock FMIS och BugsCEP, jfr ovan. Databaser över museisamlingar torde inte vara samordnade i någon större grad. Et nyt projekt med støtte fra RJ om en national helleristningsdatabase er eksempel på samarbejde mellem universitet (Göteborgs Univ.), Riksantikvarämbetet, og Västra Götalands region (landsstyrelse, läns museer etc.). Flere sådanne projekter burde kunne initieres, jfr nedan.

4. Databasernas roll för forskningen.

Databaserna spelar allt större roll allteftersom innehåll, tillgänglighet och kvalitet ökar, bl a genom den ökade digitaliseringen av arkeologisk fältverksamhet och museisamlingar. Dessutom har datorvana och tillgång till programvaror (GIS, statistik osv) ökat, vilket ger en ökad efterfrågan på data av god kvalitet. Samtidigt leder de ökande datamängderna till krav på verktyg för att hantera, integrera och analysera dem. Möjlighet att integrera data från olika datakällor via internet (t ex naturgeografiska, arkeologiska och museala data), skulle kunna ge utvecklingen ett kvalitativt språng.

5. Behov av digitalisering.

Stort behov. En del tillgodoses genom Accessprojektet och pågående digitalisering av historiska kartor, dock återstår mycket. Exempel: Nationell C14-databas, nationell databas över arkeologiska undersökningar, nedladdningsbara rapporter, digitalisering av äldre arkivmaterial, digitalisering av analysresultat från miljöarkeologiska undersökningar osv.

6. Framtida behov av databasuppbyggnad.

Se pkt 5. Att utveckla möjligheter till sammankoppling av data från olika källor, t ex fynddata från museisamlingar med fornlämningsregistret i FMIS är angeläget. Allmänt sett bör man främja samarbete mellan producenter av data och utvecklare av mjukvara med sikte på att främja samkörning och länkning av databaser samt tvärvetenskaplig forskning.

Ett stort behov är också att utveckla tillgängligheten genom användarvänliga gränssnitt (bl a GIS-baserade) och tillgänglighet via internet, kanske också en gemensam portal för arkeologiska data. Ikke mindst udviklingen af brugersoftware bør fremmes for at øge anvendeligheden i forskerverdenen.

7. Vetenskapliga utvecklingslinjer.

Ja, främst storskaliga mönster som kräver stora mängder av högkvalitativa data, t ex möjlighet att samköra fornlämningsdata med geografiska data på regional eller nationell nivå, samt att testa utfallet av olika hypoteser mot stora och komplexa material.

Existing databases increase the efficiency with which data can be retrieved, collated from disparate sources, and interrogated multi-dimensionally. They thus empower the researcher with both more time for analyses and greater scope for detailed interpretations.

The ability to cross-query multiple datasets through a variety of disciplines will stimulate new lines of research that are currently infeasible due to the complexity and enormity of the data.

8. Möjligheter att samverka.

Dette skulle kunne ske gennem tilsætning af en national projektgruppe som udarbejder en handlingsplan. Det nystartede hellristningprojekt er eksempel på en sådan ny type databse som fra beyndelse designes for både forskning og formidling.

- Collaboration between the national archaeological labs, towards a common gateway for the storage and dissemination of scientific archaeological data.
- Collaboration with other university groups with similar research interests.
- Collaboration with HumLab, especially on the development of data mining, visualization and interface components.
- An expansion of the end user scope of archaeological data and

interpretations, into the fields of modern environmental science and conservation, through collaboration with researchers in these fields and state bodies such as Länsstyrelsen, or private contractors.

- Environmental archaeology data is practically useless without modern, comparative reference data and material. Collaboration with experts in modern biology, environmental science and perhaps ethnography should be increased.

9. Forskningens synlighed ökas?

Ja, dock beroende av språk. Databaser demokratiserer forskningen og gør data tilgængelige for alle. Om bare databaserne udvikles også med brugerflader for forskningen. Dette er ikke altid tilfældet idag, hvor hovedformålet er myndighedsarbejde. Dvs at kunne dokumentere arkæologiske interesser i planlægningen. Her skulle meget kunne gøres med allerede eksisterende databaser.

Internet access always leads to increased interest, the search engines not only directing people to what they are looking for, but to less obvious links. The internet is international, and has an indiscriminate audience, whereas scientific journals are targeted and less likely to reach an audience outside of those already interested in the subject.

Making well structured, queryable and linkable data available will undoubtedly increase the potential for cooperation on an international level.

Databaser i religionsvetenskap (jämför i någon mån utomeuropeiska språk och kulturer samt antikforskning)

Jørgen Straarup
Högskolan i Gävle

Introduktion

För snart tio år sedan var vi några stycken som hjälpte en kamrat att flytta. Det var de sedvanliga problemen att få möblerna från villan (somliga arvegods) att passa in i 1960-talslägenheten. Särskilda problem orsakades av ett skrivbord som Rolf ärvt från sin pappa. Det gick bara inte in i det avsedda arbetsrummet. Det var stort, förstås, men också fint, plus att det ju hade affektionsvärde. Efter en stund höll Rolf på att ge upp och hotade med att ta en såg, dela bordskivan mitt itu och ta in delarna var för sig.

Då vaknade flytthjälparna till liv: ”icke på villkor fick han göra det!” På baksidan av skrivbordet var det nämligen ett speciellt hyllsystem avpassat efter kyrkoåret (pappan hade varit präst) och en hylla för varje sön- och helgdag i kyrkoåret. Hyllsystemets innehåll, predikningarna var också bevarade. Pappan hade inför stundande helg kunnat plocka gamla predikningar, se efter vilken inspiration han haft förra gången denna årgångstext varit aktuell, och bestämma sig för att ta om den, eller skriva en ny predikan med ny inspiration.

Det löste sig med Rolfs pappas skrivbord. Något ljushuvud kom på att man kunde montera av skivan och sedan lirka in resten, och väl framme montera tillbaka skivan. Hyllsystemet – och predikningarna – överlevde även denna flyttning.

Vad händer när Rolfs barn ärver skrivbordet – och predikningarna? Troligtvis kommer de att behålla möblerna men slänga predikningarna. Är det bra eller dåligt? Med tanke på att arkivering, databaser och digitalisering tvingar fram ett urvals- eller gallringstänkande – vi kan inte rimligen behålla allt – är det kanske bra att predikningarna slängs. Men jag skulle inte ha svårigheter att hitta en religionsvetare, t ex en kyrkohistoriker eller praktisk teolog, som pläderar för att det är av största vikt att predikningarna bevaras.

Det är inte säkert att Rolfs pappa har varit en särskild *stor* teolog eller en särskild *betydelsefull* präst i Svenska kyrkan, så ett biografiskt intresse för att koppla de skrivna predikningarna till ett viktigt livsverk kanske inte finns. Men som just exempel på troget och träget prästarbete under mitten av 1900-talet, med den givna input som evangeliebokens texter levererar och den påverkan som politiska och kulturella händelser utgör, kan predikningarna vara en tidsspegel, en bild av livet på 1950-talet i just den församlingen.

Religionsvetenskapliga ”ämnen”, ”delämnena”, ”discipliner”, ”kunskapsområdena”

Religionsvetenskap kan liknas vid groggvirket *gin*, så till vida att den blandas på ett fint sätt med en rad andra drycker, alias vetenskapliga discipliner och diskurser. En uppgräddning av religionsvetenskapliga deldiscipliner är som en promenad på ett väl utbyggt universitetsområdes humanistiska och samhällsvetenskapliga domäner. Överallt ser man discipliner som religionsvetenskapen bildat bindestrecksvetenskap med. En fulltäckande lista skulle bli lång, varför här presenteras en kort variant: religionsantropologi, -fenomenologi, -didaktik, -filosofi, -historia, -pedagogik, -psykologi, -sociologi, och kyrkohistoria.

Därtill kommer ett antal discipliner där bindestreckskaraktären inte är lika utpräglad, men där det likafullt är uppenbart att det finns samröre med andra discipliner: dogmatik (med filosofi), etik (med praktisk filosofi), homiletik (med retorik), islamologi, kateketik (med

utvecklingspsykologi och –pedagogik), kyrkorätt (med juridik), liturgik (med ritual studies), nya och gamla testamentets exegetik (med litteraturvetenskap och språk).

Slutligen finns det en rad områden som av språkliga skäl inte har kunnat uttryckas som bindestrecksvetenskaper, men där det finns ambitioner att bli egna ämnen eller delämnena, som t ex:

religion och medier (med MKV), religion och politik (med statsvetenskap).

Material och metoder i religionsvetenskap

Det ger sig själv att en sådan mångfald på det teoretiska planet leder till användande av en lång rad material och ett något mindre antal vetenskapliga metoder. I det typiska fallet är det materialet som styr metodvalet, men givetvis är det ibland så att en forskare som övat upp en färdighet på en specifik metod kan leta efter material att fortsätta på.

Religionsvetenskapens bruk av databaser och digitalisering

En ofullständig förteckning över sådana databaser som forskare i religionsvetenskap (jämte i någon mån utomeuropeiska språk och kulturer samt antikforskning¹) använder återfinns i tabell 1.

¹ Det upptäcktes att en icke obetydlig del av det humanistiska forskningsfältet saknades bland förlagorna till bidragen i denna volym, varför jag på eget bevåg riktade förfrågan till representanter för forskningsområdena utomeuropeiska språk och kulturer samt antikforskning, kolleger i VRs beredningsgrupp HS-K.

Tabell 1 Databaser som används i religionsvetenskap (jämfte i någon mån utomeuropeiska språk och kulturer samt antikforskning)

Beskrivning @ URL
<i>Bibliska studier</i>
Documentation for biblical literature Innsbruck @ http://bildi.uibk.ac.at/index-en.html
<i>Kyrkohistoria (mm)</i>
Demografiska databasen vid Umeå universitet @ www.ddb.umu.se
LUKA: Lunds universitets kyrkohistoriska arkiv
<i>Religionssociologi</i>
Svensk Samhällsvetenskaplig Datatjänst @ www.ssd.gu.se
Norsk samfunnsvitenskapelig datatjeneste: World Value Study, European Value Study Group @ www.nsd.uib.no
Svenska kyrkans statistikdatabas @ www.svenskakyrkan.se/forskning/statistikdatabasen.htm
<i>Antikforskning och utomeuropeiska språk</i>
Klassiska texter, Perseus Digital Library, Tufts university @ www.perseus.tufts.edu
PHI, Searchable Greek Inscriptions, The Packard Humanities Institute @ http://epigraphy.packhum.org/inscriptions
Patrologia Latina, the works of the Church Fathers from Tertullian in 200 AD to the death of Pope Innocent III in 1216 @ http://pld.chadwyck.co.uk
Biblioteca Teubneriana
Claros - Konkordans för olika publikationer av grekiska inskrifter @ www.dge.filol.csic.es/claros/cnc/2cnc.htm
LGPN - Grekiska personnamn @ www.lgpn.ox.ac.uk
CSAD - "Imaging Projects", fotografier av avtryck av ett urval av grekiska inskrifter av paleografiskt intresse @ www.csad.ox.ac.uk/CSAD/Images.html
Demos; athensk demokrati @ www.stoa.org/projects/demos/hom
Läkekonst (antika och moderna texter) @ www.medicinaantiqua.org.uk/index.html
Thesaurus Linguae Graecae: Grekiska texter från Homeros till det Byzantiska rikets fall 1453 @ www.tlg.uci.edu
Hellenic Culture: bilder på föremål, platser samt antikvarisk organisation i Grekland @ www.culture.gr
The Beazley archive @ www.beazley.ox.ac.uk/BeazleyAdmin/Script2/default.htm

Andra typer av material- och datasamlingar förekommer också, men uppgift om systematisering av sådana individuella kompillerade samlingar har inte gått att få. Inom många av religionsvetenskapens forskningsområden hör det till vedertaget arbetssätt att själv aktivt kompilera sitt material, vare detta är intervjuer, tidningsurklipp eller bilder. Ett exempel – som inte nödvändigtvis är typiskt – är utvecklandet vid Lunds universitet av en databas för gravstenar.²

Befintliga databasresurser används relativt lite. Sannolikt gäller detta inte för de databaser som liknar uppslagsverk, men arkiven för analyserade data präglas inte av en hög grad av utnyttjande. Enskilda forskare, såväl som forskargrupper, kan vara användare.

En del materialsamlingar har sin grund i större mer eller mindre tvärvetenskapliga forskningsprojekt, varför bruket av dem ibland har tvärvetenskaplig karaktär.

De (del)discipliner som arbetar med historiska perspektiv och material nyttjar i stor utsträckning databaser vid arkiv o likn. Från kyrko- och religionshistoriker, arkeologer och antikvetare finns det önskemål om fullständiga museikataloger/föremålsförteckningar on-line.

Forskare berättar om önskemål om digitalisering av material som förvaras vid arkiv, bibliotek, och museer. I vissa avseenden kan det räcka med skannade versioner av sådant material, så att åtkomsten underlättas. I andra avseenden handlar det om en bearbetning av det slag som

² Brev till författaren januari 2007 från CD.

Demografiska databasen i Umeå har gjort med kyrkoarkivalier. Utöver skanning handlar det om överförande av materialet till länkade databaser, vilket möjliggör att följa personers levnadsöden genom deras upptagande i olika arkivhandlingar. I samband med 200-årsminnet av Lars Levi Læstadius födelse (år 2000) väcktes tanken om ett arkiv (beläget i Pajala) med bland annat skannade dokument rörande Læstadius. Förverkligandet av denna idé pågår.

Uppbyggandet av databaser sker i (minst) två sammanhang. Forskarvärlden är given. Med forskningsfinansiering av olika slag (fakultetsmedel och riktade forskningsmedel) sker dagligen sådant som bidrar till uppbyggande av databaser. Det finns emellertid en värld utanför forskarvärlden i vilken också uppbyggande av databaser sker. Organisationer, föreningar, stiftelser, företag har uppmärksammat behovet av att i databasform systematisera och tillgängliggöra olika slag av uppgifter. I merparten fall har dessa databasbyggare den kompetens som behövs för att databasen ska bli tillgänglig och användbar för forskningen, men det kan finnas sammanhang i vilka sådan kompetens saknas. Man kan överväga om det vore väl använda forskningsmedel att för den typen av databasuppbyggande inrätta en rådgivningstjänst, dit föreningar, församlingar, klubbar, småföretag m fl kunde vända sig för stöd och råd.

En rad religionsvetenskapliga forskningsfrågor kan endast besvaras med tillgång till stora, högkvalitativa och internationella databaser. Som exempel kan nämnas att frågor om värderingars utveckling nationellt och internationellt inte kan besvaras utan tillgång till ett högkvalitativt surveymaterial. Också för andra frågor vidkommande gäller att de endast kan besvaras om ett brett empirisk material tas in och bearbetas, vilket mer eller mindre automatiskt betyder att stora databaser skapas.

I hägn av de stora internationella organisationerna (EU och underorganisationer, FN och underorganisationer, internationella lärda sällskap o s v) pågår överläggningar och utvecklingar som leder till uppbyggande av databaser. Mycket av det arbetet (t ex textredigeringsarbetet) syftar i första hand till att öka tillgängligheten, medan det är mer tveksamt om bruket av forskningsmaterialet ifråga kommer att öka på ett påtagligt sätt. Man kan bakom detta byggande se spår av ett kulturarvstänkande. Det upplevs vara ett ansvar för den nuvarande forskargenerationen att möjliggöra att andra forskare än de geografiskt närmaste får tillgång till material som utgör en del av detta kulturarv, plus att en digitalisering innebär ett minskat tryck på önskemålet att egenhändigt hantera originalmaterialet och därmed undvika slitage.

På samma sätt som en ökad tillgänglighet till databasmaterial leder till att de frågor som databasen handlar om blir med synliga kanske man kan säga att samma process leder till ökad synlighet för forskningen. Möjligheten att analysera databasmaterial kan eventuellt leda till att värdet med (humanistisk) forskning tydliggörs, nationellt och internationellt.

Uppfyllda behov

För en rad religionsvetenskapliga discipliner och ämnen är behovet i stort sett uppfyllt.

Konstruktiv teologi/religionsvetenskap

En hel del religionsvetenskaplig forskning äger rum inom något slags praxisfält bestämt av å ena sidan akademins krav om vetenskaplighet och å den andra livet i den religiösa gemenskapen.³ I den forskningen kan behovet av databaser sägas vara uppfyllt i och med att sådan forskning har paradigmatiser karaktär och i huvudsak använder sig av andra vetenskapliga texter som material, texter som säkrats inom ordinarie bibliotekssystem.

Avgränsade, äldre textkorpusar

Ett annat exempel är att det behov av databaser som kan ha visat sig i ett tidigare forskningsskede, och som lett till att vissa forskare varit trogna besökare på universitetsbiblioteken, har försvunnit i och med att textutgåvor gjorts tillgängliga på datamedia (hieroglyftexter, klassiska medeltida latintexter, etc). Det rör sig om sådana discipliner där materialet av historiska skäl är avslutat och fixerat. De bibliska vetenskaperna är ett sådant område. De olika aktuella handskrifter har identifierats och dokumenterats, textkritiska utgåvor har publicerats, konkordanser och bibliografiska hjälpmedel har utvecklats.

³ Se Thomas Girmalm, *Teologi och det moderna universitetet. Perspektiv på polariseringen mellan teologi och religionsvetenskap under svenskt 1900-tal*. Umeå universitet, 2006.

För sådana discipliner finns förvisso en möjlighet till omformulering av teoretiska perspektiv, men i detta sker det oftast genom inspiration från intilliggande discipliner, inte på grund av att materialläget (inkl databasmöjligheterna) ändras. Bibelvetenskaperna kan t ex inspireras av arkeologiska fynd och rön, av semiotik eller litteraturvetenskap. Religionsvetenskapen står alltså inte ensam om ett önskemål om databaser och digitalisering, utan samma önskemål drivs även från annat vetenskapligt håll.

Framprovocerat material (enkäter etc)

För religionsvetenskapliga discipliner med anknytning till samhällsvetenskap finns det en redan upparbetad kontakt med *Svensk Samhällsvetenskaplig Datatjänst* i Göteborg och motsvarande organ utomlands. Tillgänglighet och lagringssäkerhet för den World Value Survey som insamlats i ett stort antal länder sedan början av 1980-talet och som inneburit en tydlig forskarinsats av ett antal religionsvetare har t ex säkrats av *Norsk samfunnsvitenskapelig datatjeneste* i Bergen.

Historiska urkunder

För vissa discipliner har materialåtkomlighet och lagringsansvar löst på det sätt som exemplifieras av *Demografiska databasen* i Umeå. Databasen, som bland annat innehåller datoriserade uppgifter från kyrkböcker, har uppkommit som följd av ett antal lyckliga och samtidiga omständigheter, nämligen (a) att det (sent 1960-tal och tidigt 1970-tal) började finnas datorer för universitetsbruk, (b) att ett nyetablerat, hungrigt universitet behövde sätta sig på kartan, (c) att den region som universitetet skulle betjäna hade en hög arbetslöshet, i synnerhet bland lågutbildade kvinnor, och (d) att en inspiratör av ovanlig lyskraft (Egil Johansson) dök upp på detta universitet.

Databasen ifråga är en resurs för demografer, historiker, religionsvetare, läskunnighetsforskare, epidemiologer, m fl.

Icke-uppfyllda behov

Databaser (arkiv) för redan analyserade data

Intervjudata

Sedan 1980-talets mitt har en ökande mängd av undersökningar inom religionsvetenskapen inneburit att det genomförts intervjuer. En typisk avhandling har byggt på 25 intervjuer som sedan analyserats innehållsligt. Detta intervjumaterial hittar inte vägen in i de etablerade forskningsdatabaserna. Många orsaker kan ligga bakom detta: oklarhet om de etiska problemställningarna bakom materialet, ovana vid det samhällsvetenskapliga hantverket, ovilja att släppa ifrån sig "sitt" material, med flera orsaker.

På detta område behövs klarare regelverk och kanske en skraddarsydd databas. SSD fungerar bra för hårt strukturerade enkätdata, medan data från intervjuer ofta bygger på ett slags tematisk struktur, som är svårare att systematisera.

Databaser för icke-analyserade data

Internetdata

Fler och fler forskningsinsatser bygger på sådant som tankats hem från Internet. Hemsidor av mycket varierande slag och kvalitet lämnar viktigt material till religionsvetenskapligt teoretiserande.

Som exempel kan nämnas att undersökningar av religiöst inspirerad våldsanvändning lutar sig hårt mot olika organisationers hemsidor i sina försök att spåra inspirationen till politiskt våld. Därmed försvåras kvalitetskontrollen och -granskningen, i och med att det ingenstans finns möjlighet att rekonstruera en hemsida i det skick den hade på den dag som en forskare tankat hem material från den. Att detta har sammanhang med egenskaper hos de organisationer som använder sig av hemsidor för informationsspridning är uppenbart. Information är i vissa lägen färskvara med bäst-före-datum.

Att ha tillgång till en databas med dag-för-dag-variationer över innehållet på Internet skulle kunna underlätta dels kontroll, dels diakrona studier av information på Internet.

Radio- och TV-data

Också för radio- och TV-innehåll gäller att det får allt mer framskjuten position bland forskningsmaterial i religionsvetenskap. Speciella sändningar riktade till konsumentminoriteter blir betydelsefulla och effektiva kommunikationskanaler till sådana minoriteter. För många av de till Sverige invandrade blir parabolantennen och TV-apparaten en viktig länk till en avlägsen, eller bara oåtkomlig kulturell och religiös verklighet.

Att kompilera en databas för mediainnehåll världen över är en orealistisk tanke. Man kan däremot fundera på om en databas med sådant innehåll på stickprovsbasis skulle kunna bli av.

Arkivmaterial

En rad arkivmaterial är tillgängliga endast vid det bibliotek där materialet deponerats. Predikosamlingar av det slag som artikeln inleddes med kan få utgöra exempel. Att digitalisera sådant material är inte realistiskt, i och med att tolkningen av handstilar kan vara svår. Att ett sådant material fotograferas och görs tillgängligt som bilder är emellertid inte lika arbetskrävande. Ett av underlagsmaterialen för Demografiska databasen, nämligen mormonernas avfotografering av bl a svenska kyrkoböcker kan nämnas som exempel.

Från antikforskningens håll förs särskilt fram att ”ökad tillgänglighet via nätet, inte minst av kataloger med goda beskrivningar, är självklart ett desideratum. Virtuella presentationer av bibliotek, forskningsinstitutioner, museer mm kan spara mycket tid i förberedelser inför forskningsresor... Mängden svårtillgängligt material är närmast oändligt... Behovet av enhetlighet och standardisering är genomgående för att de olika projekten skall bli mest möjligt användbara”.⁴

Från arkeologins område anförs att ”däremot är systematisk digital dokumentation av arkeologiska föremål och platser, särskilt i ett internationellt perspektiv, eftersatt inom vårt ämne. En stor del av vår kunskapsproduktion bygger på relationer i utseende, rum och tid hos de materiella lämningar vi studerar. I traditionella publikationer tvingas man välja ett presentationssätt där vissa relationer bedöms vara viktigare än andra. På grund av ämnets internationella karaktär är det dock mycket svårt att skapa plattformar för övergripande och internationell digital dokumentation, som t.ex. fornminnesregister, keramikdatabaser, myntdatabaser, etc.”⁵

Sammanfattning

Eftersom religionsvetenskap tenderar att hålla sig ganska nära urkunder, har en rad av religionsvetenskapens problem vad gäller databaser och digitalisering funnit en tillfredsställande lösning inom befintliga dokumentationssystem: bibliotek, databaser, arkiv. Vad som inte funnit en tillfredsställande lösning är sådana analyser som gäller dagsaktuella eller gårdagsaktuella studier.

Religionsvetenskapernas icke-tillgodosedda behov ligger inom området intervjudata, där ett system för arkivering och lagring av intervjudata saknas.

⁴ Brev januari 2007 till författaren från GV.

⁵ Brev januari 2007 till författaren från GN.

Inventering av språkliga korpusar och datoriserade lexikon vid Språkvetenskapliga fakulteten¹

Nedanstående är en lista på korpora som etablerats av forskare vid Språkvetenskapliga fakulteten. Listan är inte fullständig, men ger en överskådlig bild av den korpusbaserade forskningen som bedrivs.

Asiatiska språk

Kinnauri Corpus

Etablerad av: Anju Saxena

Språk: Kinnauri-Engelska

Omfattning: c:a 30 000 ord

Språktyp: muntlig sakprosa, berättelser, sagor, sånger, konversationer

Period: 1989 och framåt

Egenskaper: Det är en muntlig korpus, fonetiskt transkriberad med åtföljande annotering: (1) morfemsegmentering; (2) morfem-för-morfem-översättning av hela korpusen; (3) en del är också syntaktiskt annoterad; (4) fri översättning till engelska.

Användning: forskning, forskarutbildning, undervisning inom kurserna: Funktionell grammatik/Grammatik 2, Lingvistik 2, Lingvistik 1, Syntax, Grammatik

Tinan corpus

Etablerad av: Anju Saxena, delvis i samarbete med CIIL, Indien inom projektet ”Digital documentation of Indian minority languages”

Språk: Tinan-Engelska

Omfattning: c:a 20 000 ord

Språktyp: muntlig sakprosa, berättelser, sagor, sånger

Period: 1989-90, 2003-2005

Egenskaper: Det är en muntlig korpus, fonetiskt transkriberad med åtföljande annotering: (1) morfemsegmentering; (2) morfem-för-morfem-översättning av hela korpusen; (3) fri översättning till engelska.

Användning: forskning, forskarutbildning och för dokumentation av hotade språk. Korpusen kan också användas i undervisningen på samma sätt som kinnaurikorpusen.

Engelska

CENE - The Corpus of English Newspaper Editorials (1900–1993)

Etablerad av: Ingrid Westin

Språk: Brittisk engelska

Omfattning: 502,834 ord (864 tidningsledare)

Språktyper: tidningsledare från *Daily Telegraph*, *Guardian* och *The Times*

Period: 1900–1993 (10 perioder: 1900, 1910, 1920, 1930 osv.)

Användning: avhandling

¹ Inventeringen gjordes hösten 2006. Komplettering av korpusar i engelska (i slutet av dokumentet) har gjorts av Merja Kytö 2007-06-07 med tillstånd av Anna Sägval-Hein.

UPC - The Uppsala Press Corpus

Etablerad av: Margareta Westergren Axelsson

Språk: Brittisk engelska

Omfattning: 354,084 ord (88 texter)

Språktyper: tidningstexter som motsvarar tidningstexter i LOB (Lancaster/Oslo-Bergen Corpus med texter från 1961): reportage, tidningsledare och recensioner

Period: 1994

Användning: forskarutbildning

UCTT - The Uppsala Corpus of Travel Texts

Etablerad av: Tore Nilsson

Språk: Brittisk engelska

Omfattning: 497,091 ord (139 texter)

Språktyper: turistbrochyrer, researtiklar, resehandböcker

Period: 1994–1999

Användning: forskarutbildning

The Victorian Corpus and the Contemporary Corpus of British Children's Literature

Etablerad av: Hanna Andersdotter Sveen

Språk: Modern brittisk engelska

Omfattning: 2 x 95,000 ord (2 x 10 texter); Contemporary Corpus är baserad på ett urval av texter i BNC

Språktyper: barnlitteratur för pojkar och flickor

Period: 1851–1858 och 1981–1993

Användning: forskarutbildning

Finsk-ugriska språk

Estnisk morfologi

Etablerad av: Virve Raag

Språk: Estniska

Omfattning: pluralformer: 30 000, illativ sg 9000, superlativ 2700

Texttyp: tidningstext

Period: år 1887, 1907, 1923, 1938, 1961, 1991

Användning: forskarutbildning, forskning, uppsatskrivande

Estniska telefonsamtal

Etablerad av: Leelo Keevallik

Språk: Estniska

Omfattning: 103 000 ord

Språktyp: samtal

Period: 1998/99

Användning: forskning, forskarutbildning, undervisning

Sverigeestniskt talspråk

Etablerad av: Raimo Raag, Virve Raag, Leelo Keevallik

Språk: Estniska

Omfattning: 253 timmar

Språktyp: radioprogram, forskningsintervjuer, sociala tillställningar
Period: 1979-2006
Användning: forskning, undervisning

Vrajavali ordlistor

Språk: Vrajavali & English
Omfattning 2-300 sidor
texttyp: drama, lyrik (skönlitteratur)
Period: 15-1600 talet
användning: forskning

Klassiska språk

Bysantinsk hagiografi

Etablerad av: Jan-Olov Rosenqvist
Språk: bysantinsk grekiska
Omfattning: 18,3 miljoner tecken
Texttyp: hagiografi
Period: 300-1300-talen
Användning: forskning

CLEDBase

Etablerad av: Christer Henriksén, Henrik Vitalis, Patrik Granholm
Språk: latin, viss grekiska
Omfattning: f.n. ca 4500 ord
Texttyp: romerska, icke-kristna gravinskrifter på daktylisk vers
Period: ca 50 f.Kr – 350 e.Kr.
Användning: forskning

Romanska språk

CCUCFI - Le Corpus Contrastif d'Uppsala du Conditionnel français et italien

Etablerad av: Hans Kronning
Språk: franska och italienska
Omfattning: Motsvarande 8 böcker, 1000 konditionalisförekomster
Språktyp: romaner
Period: 1900-tal
Användning: forskning

Corpus de journaux francophones

Etablerat under ledning av Mats Forsgren
Språk: franska från Schweiz och Belgien
Omfattning: 49 nr
Språktyp: dagstidningar
Period: 80-90-talet
Användning: forskarutbildning, forskning, uppsatsskrivande

Corpus des periodicas hispanos

Etablerat under ledning av Mats Forsgren

Språk: spanska

Omfattning: 116 nr

Språktyp: dagstidningar

Period: 80-90-talet

Användning: forskarutbildning, forskning, uppsatsskrivande

CUCF - Le Corpus d'Uppsala du Conditionnel français

Etablerad av: Hans Kronning och Coco Norén

Språk: franska

Omfattning: Motsvarande 30 böcker, 5000 konditionalisförekomster

Språktyp: romaner, teaterpjäser, dagstidningar, vetenskaplig prosa

Period: 1700-tal, 1800-tal, 1900-tal, 2000-tal

Dialogues entres jeunes Français

Etablerat av: Coco Norén

Språk: franska

Omfattning: 5½ timmar, transkriptioner

Period: 1990-tal

Språktyp: informella tvåpartssamtal

Användning: forskarutbildning, forskning, undervisning

Fransk-svenskt Parallellkorpus

Etablerat av: Carina Andersson

Språk: franska-svenska

Omfattning: 2 miljoner ord

Period: 1900-tal

Språktyp: romaner

Användning: forskarutbildning, forskning

ISU - Interaktionsstudier Stockholm - Uppsala

Etablerad av: Cilla Häggkvist, Johan Gille

Språk: spanska

Omfattning: 7 samtal à 40 minuter

Språktyp: vardagliga samtal mellan universitetsstuderande

Period: 2000-tal

Användning: forskning, C- och D-uppsatser, inspelat under kurserna Samtalsanalys och Diskursanalys

The Uppsala French TAP-Corpus ("Think-Aloud-Protocols")

Etablerad av: Kerstin Jonasson

Språk: Franska-Svenska

Omfattning: ca 50 sidor

Texttyp: "tänka-högt-protokoll"

Ålder: etablerat 1993-1995

Användning: forskning

Semitiska och turkiska språk

Muntliga berättelser från Östra Turkiet, Norra Syrien och Jemen

Etablerad av: Bo Isaksson, Ablahad Lahdo

Språk: arabisk dialekt och arameisk dialekt

Omfattning: 20 timmars ljudfiler varav ca 10 % transkriberat

Språktyp: muntliga berättelser

Period: 2000-2005

Användning: forskning, avhandling, uppsatskrivande på Orientalistikprogrammet och Arabiska C-D (kandidat- och magisteruppsatser)

Turkisk-Svensk Korpus / Turkish-Swedish Corpus

Etablerad av: Prof. Anna Sågvall Hein, Prof. Éva Á Csató Johanson, dr. Beáta Bandmann Megyesi, dr. Bengt Dahlqvist

Språk: Turkiska-svenska / Turkish-Swedish

Omfattning: 118 000 token på turkiska och 144 000 token på svenska.

Språktyp: skönlitterär text och facktext

Period: 1900-tal och 2000-tal

Användning: forskning, uppsatskrivande, kurser i ämnet turkiska språk samt orientalistikprogrammet med turkisk inriktning

Hebreiska texten i UUB O Heb 32

Etablerad av: Mats Eskhult

Språk: hebreiska

Omfattning: ca 100 trycksidor från handskriftens 200 foliantsidor.

Språktyp: bibeltext

Period: 1700-talet

Användning: forskning

Vrajavali ordlistor

Etablerad av: William Smith

Språk: Vrajavali & English

Omfattning 200-300 sidor

Språktyp: drama, lyrik (skönlitteratur)

Period: 1500-1600 talet

Användning: forskning

Slaviska språk

The Uppsala Russian Corpus -- Upsal'skij korpus russkich tekstov

Etablerad under ledning av prof. Lennart Lönngren

Språk: ryska

Omfattning: 1 miljon löpord

Språktyp: skönlitteratur sakprosa, stort urval

Period: 1900-talet, 1960-1988, 1985-1989

Användning: forskning, grundutbildning

Svenska

The Swedish Map Task Corpus

Etablerad av: Pétur Helgason

Språk: svenska

Omfattning: 70 minuter, ca 8 000 ord

Språktyp: samtal

Period: samtida

Användning: 3 forskarutbildningar, C- och D-uppstater

SCARRIE SvD9596

Etablerad av: Anna Sågvall Hein

Språk: Svenska

Omfattning: 47.4 miljoner ord

Texttyp: dagstidningstext

Period: 1995-96

Användning: forskning, forskarutbildning, uppsatskrivande, inom kursen/programmet: STP

SCARRIE UNT9596

Etablerad av: Anna Sågvall Hein

Språk: Svenska

Omfattning: 22.8 miljoner ord

Texttyp: dagstidningstext

Period: 1995-96

Användning: forskning, forskarutbildning, uppsatskrivande, inom kursen/programmet: STP

UNT92

Etablerad av: Anna Sågvall Hein

Omfattning: 6.5 miljoner ord

Texttyp: dagstidningstext

Period: 1992

Användning: forskning, forskarutbildning, uppsatskrivande, inom kursen/programmet: STP

ECD - Error Corpus Database (<http://www.lingfil.uu.se/ling/ecd/>)

Etablerad av: Anna Sågvall Hein

Språk: Svenska

Omfattning: ca 9 000 meningsfragment

Texttyp: dagstidningstext

Period: 1995-96

Användning: forskning, forskarutbildning, uppsatskrivande, inom kursen/programmet: STP

SCANIA (<http://perdix.lingfil.uu.se/scania.html>)

Etablerad av: Anna Sågvall Hein

Språk: Svenska

Omfattning: ca 200 000 ordformer

Texttyp: facktext

Period: 1995-2005

Användning: forskning, forskarutbildning, uppsatskrivande, inom kursen/programmet: STP

Lärobokstexter från projektet "Elevers möte med skolans textvärldar"

Etablerad av: Caroline Liberg, Agnes Edling, Jenny Folkeryd, Åsa af Geijerstam

Språk: Svenska

Omfattning: 27 953 ord

Språktyp: skönlitterär text, diskursiv lärobokstext

Period: insamlat 1999-2003 (tryckår ca 1980-2001)

Användning: forskning, forskarutbildning

Elevskrivna texter från projektet "Elevers möte med skolans textvärldar"

Etablerad av: Caroline Liberg, Åsa af Geijerstam, Agnes Edling, Jenny Folkeryd

Språk: Svenska

Omfattning: ca 400 kortare texter

Språktyp: skönlitterär text, sakprosa

Period: insamlat 1999-2003

Användning: forskning, forskarutbildning

Samtal, åldrande och identitet 1: identitetsskapande strategier i äldresamtal

Etablerad under ledning av Bengt Nordberg

Språk: svenska

Omfattning: dialoger 27 timmar, gruppsamtal 14 timmar, delvis transkriberat

Språktyp: arrangerade dialoger med medelålders och pensionerade kvinnor, dels av informella gruppsamtal (kafferep, kortspel o.l.) mellan äldre kvinnor.

Egenskaper: I dialogerna sammanlagt 40 kvinnor som inte är bekanta med varandra i syfte att lära känna varandra, dels över generationerna, dels inom dem. Deltagarna i gruppsamtalen känner i de flesta fall varandra sedan tidigare.

Period: 1997–2000

Användning: forskning

ÅbEsk - Kontinuitet och förändring i nutida talspråk. Återbesök i Eskilstuna

Etablerad under ledning av Bengt Nordberg

Språk: svenska

Omfattning: 78 timmar, delvis transkriberat

Språktyp:

Egenskaper: Korpusen är insamlad för ett projekt som följer upp Eskilstunaundersökningen från 1967. De infödda informanterna har samma sociala fördelning och inspelningarna samma stilistiska kvaliteter som i den tidigare undersökningen (se Svenska stadsmål 1). 13 talare ingick också i 60-talsundersökningen, medan övriga 72 är nya.

Användning: forskning

GIC - Samtal i nödsituation. Telekommunikationen vid en giftinformationscentral

Etablerad under ledning av Bengt Nordberg

Språk: svenska

Omfattning: 17 timmar, transkriptioner

Språktyp: Samlingen består av 377 autentiska telefonsamtal från privatpersoner över hela landet till giftinformationscentralen gällande akuta eller befarade förgiftningar.

Egenskaper: Den rådgivande personalen består av 13 apotekare/informatörer.

Texter i europeiska skrivsamhällen

Etablerad under ledning av Britt-Louise Gunnarsson

Språk: svenska

Omfattning: 70 intervjuer

Språktyp: 70 intervjuer med personer ansvariga för transkriptioner skrivande av olika texter inom en bank, en ingenjörbyrå, en historisk institution samt en yrkesmedicinsk institution i tre länder: Sverige, Tyskland och Storbritannien.

Period: 1994–96

Interaktionen vid seminarier

Etablerad under ledning av Britt-Louise Gunnarsson

Språk: svenska

Omfattning: 20 doktorandseminarier, transkriptioner

Språktyp: doktorandseminarier från tre institutioner vid Uppsala universitet: en humanistisk, en samhällsvetenskaplig, en naturvetenskaplig.

Period: 1992–95

FORTIS - Sverigefinnars två språk – språkbruk och attityder hos två generationer

Etablerad under ledning av Bengt Nordberg

Språk: finska-svenska

Omfattning: 54 timmar, transkriptioner

Språktyp: formella och informella intervjuer, gruppsamtal med självrekryterade deltagare

Egenskaper: Situationellt varierade inspelningar med två generationer Sverigefinnar (16–19 och 35–55 år, barn och föräldrar, båda könen) boende i en Stockholmsförort. Ämnesvalet i gruppsamtalen är fritt, de informella intervjuerna tar upp vardagsliv, informanternas migrationshistoria och situation som sverigefinne, de formella intervjuerna är autenticitetstroga anställningsintervjuer och radiointervjuer.

Talutveckling

Etablerad av: Birgitta Garne

Språk: svenska

Omfattning: Ca 170 samtal

Språktyp: samtal med elever i olika åldrar, från skolår 2 till gymnasiet, som arbetar med olika typer av samtalsuppgifter. Oftast samtalas eleverna i små grupper utan vuxen ledare, men korpusen innehåller även lärarledda diskussioner.

Period: 1988–92

Kommunikationen vid en larmcentral

Etablerad av: Bengt Nordberg

Språk: svenska

Omfattning: 8 timmar, transkriptioner

Språktyp: Autentiska inspelningar av telefon- och radiokommunikationen vid en länsalarmeringscentral mellan uppringare (hjälpökande), operatör och utförare. Samtalen rör akuta sjukdomsfall, olyckor, bränder, ambulansbeställningar, driftstopp etc.

Period: 1986

Läkarens och lekmannens begreppsvärldar

Etablerad av: Ulla Melander Marttala

Språk: svenska

Omfattning: 95 intervjuer, 15 läkar-patientsamtal

Språktyp: En semantisk och samtalsanalytisk undersökning av föreställningar och begrepp kring reumatiska sjukdomar. Två delar: 95 semantiska djupintervjuer, audioinspelade; 15 läkare-patientsamtal med intervjuer och videovisningar, audio- och videoinspelningar.

Period: 1986-89

Ungdomars samtalsstil

Etablerad av: Bengt Nordberg

Språk: svenska

Omfattning: 2 timmar, transkriptioner

Språktyp: Fria samtal mellan nära vänner

Egenskaper: Flickor och pojkar i åldern 12–16 år i enkönade grupper. Samtalen kretsar kring personliga intressen, vardagliga händelser, kamratskvaller. Deltagarna uppväxta i Uppsala.

Period: 1984–88

Språkanvändning och språkmiljöer i staden och på landet

Etablerad av: Bengt Nordberg

Språk: svenska

Omfattning: 29 timmar

Språktyp: Informella intervjuer

Egenskaper: 20 informanter i åldern 30–50 år, hälften boende i urban, hälften i rural miljö (Mellansverige, södra Norrland), hälften män, hälften kvinnor, hälften arbetare, hälften tjänstemän. Intervjuerna rör deltagarnas kommunikationsmiljö, kontaktnät och språkaktiviteter, upplevda förändringar i dessa och attityder till dessa.

Stad och omland: Urbaniseringen speglad i språket

Etablerad av: Mats Thelander

Språk: svenska

Omfattning: 52 timmar

Språktyp: Halvstrukturerade telefonintervjuer

Egenskaper: 120 informanter från norra Västerbotten resp. Eskilstuna kommun. De intervjuade personerna (i åldern 20–80 år) är slumpmässigt valda ur sju geografiskt definierade populationer: födda och boende på landet i Västerbotten (12 informanter), födda och boende i Skellefteå stad (12), födda på landet i Västerbotten men inflyttade till Skellefteå stad (24), födda och boende på landet i Eskilstuna kommun (12), födda och boende i Eskilstuna stad (12), födda på landet utanför Eskilstuna men inflyttade till staden (24) och födda på landet i Västerbotten men vid inspelningen bosatta i Eskilstuna stad (24). Varje intervju är ca 30 min. lång och i mer eller mindre ledig samtalston. Mest diskuteras flyttning och språk. Intervjuaren är en riksspråkstalande man i 35-årsåldern.

Period: 1978-79

Barnets språkliga identifikation (BSI)

Etablerad av: Bengt Nordberg

Språk: svenska

Omfattning: 78 timmar, delvis transkriberat

Språktyp: Inspelade situationer: högläsning, kommunikationsspel, kortspel och diskussion.

Egenskaper: Situationellt varierade inspelningar av 85 infödda grundskoleelever i Eskilstuna under tre på varandra följande år, årskurs 1–3, 4–6, 7–9. Eleverna är jämnt könsfördelade men socialt differentierade.

Period: 1977–79

Språk, roll och sociala relationer (Burträskundersökningen)

Etablerad av: Mats Thelander

Språk: svenska

Omfattning: 29 timmar gruppsamtal + 3,5 timmar intervjuer.

Språktyp: gruppsamtal

Egenskaper: 14 situationellt varierade gruppsamtal med fyra deltagare i varje grupp – alla 56 talarna (i åldern 14–64 år) från dåvarande Burträsk kommun. Varje inspelning varar ca 2 timmar och samtalsstilen är för det mesta otvungen. Under andra hälften av samtalet medverkar också en utomstående forskare. Ämnet för samtalen är Bygd i förvandling (tankar om Burträsk inför den då aktuella kommunsammanslagningen). Nio av deltagarna i gruppsamtalen har senare spelats in i mer formella intervjuer.

Period: 1973-75

Svenska stadsmål 1. Eskilstuna

Etablerad av: Bengt Nordberg

Språk: svenska

Omfattning: 53 timmar, transkriptioner

Språktyp: Samtalsliknande intervjuer

Egenskaper: 83 infödda, socialt, åldersmässigt och könsmässigt differentierade Eskilstunabor om vardagliga ämnen, personliga minnen och lokalhistoriska notiser. Några dialoger utan intervjuare ingår.

Period: 1967–68

Facktexter under 1900-talet

Etablerad av: Britt-Louise Gunnarsson

Språk: svenska

Omfattning: Totalt omfattar den inlästa 1900-talskorpusen 1340 normalsidor om 3000 tecken; de vetenskapliga texterna 650 normalsidor och de populärvetenskapliga 690.

Språktyp: 180 inskannade vetenskapliga och populärvetenskapliga texter (90 inom varje genre) inom områdena (national)ekonomi, (lung)medicin och (elektro)teknik (60 inom varje område) från perioderna 1895–1905, 1935–1945 och 1975–1985.

Egenskaper: Från varje period är 60 texter inskannade knutna till två ämnen: bank- och kreditväsen samt skatter inom ekonomi, lungsjukdomar samt hud- och könssjukdomar inom medicin, elteknik samt telekommunikationer inom teknik. Varje område är således representerat med 60 texter, fem från vardera ämne, från respektive genre och period (se nedan).

Fackspråkens framväxt

Etablerad av: Britt-Louise Gunnarsson

Språk: svenska

Omfattning: Totalt omfattar 1700- och 1800-talskorpusen 1590 normalsidor om 3000 tecken; de vetenskapliga 940 normalsidor och de populärvetenskapliga 650 normalsidor.

Språktyp: 180 inskannade vetenskapliga och populärvetenskapliga texter inom områdena (national)ekonomi, medicin och teknik från 1700-talet samt perioderna 1800 1849 och 1850 1880. Från varje period är 60 texter inskannade knutna till två ämnen: bank- och kreditväsen samt handel inom ekonomi, lungsjukdomar samt hud- och könssjukdomar inom medicin, elteknik samt telekommunikation och mekanik inom teknik. Varje ämne är således

representerat med 60 texter, fem från vardera ämne, från respektive genre och period (se nedan).

Texter i europeiska skrivsamhällen

Etablerad av: Britt-Louise Gunnarsson

Språk: svenska

Omfattning:

Språktyp: Insamlade och systematiserade svenska, engelska och tyska texter tillkomna inom miljöerna bank, ingenjörbyrå, yrkesmedicinsk institution och historisk institution i Sverige, Storbritannien och Tyskland.

Egenskaper: Texterna representerar fyra olika genrer: 1. riktade till personalen (personaltidning); 2. riktade till ägare, kunder och allmänhet (pressmeddelande, årsberättelse); 3. riktade till (presumtiva) kunder (anbud, brev, broschyr, expertutlåtande, forskningsansökan, läromedel, populärvetenskaplig artikel, projektbeskrivning, protokoll, vetenskaplig artikel); 4. avsedda för imagearbete (allmän presentation, annonser). Antal genrer från varje närmiljö och område varierar (se nedan).

Samnordisk runtextdatabas

Etablerad under ledning av: Lennart Elmevik och Lena Peterson

Språk: runskrift

Omfattning: ca 6 000 inskrifter.

Språktyp: nordiska runtexter, inklusive sådana funna utanför Norden,

Egenskaper: Texterna återges i translittererad och normaliserad form, dessutom i översättning till engelska. Till texterna är knutna uppgifter om tidsperiod, fyndområde, excerperad källa, typ av föremål m.m. Genom speciellt utformade program kan sökningar av olika slag göras i textfilerna.

Användning: forskning inom ett flertal discipliner

Komplettering av korpusar i engelska

(av Merja Kytö, 2007-06-07)

The Uppsala Student English Corpus (USE)

Etablerad av: Ylva Berglund och Margareta Westergren Axelsson

Språk: Engelska

Omfattning: 1 221 265 ord (eller 1 489 uppsatser av 440 studenter)

Språktyper: A-, B- och C-uppsatser

Period: 1999–2001

Användning: forskning, avhandlingar, uppsatser

A Corpus of English Dialogues 1560–1670 (CED)

Etablerad av: Merja Kytö och Jonathan Culpeper (Lancaster University)

Språk: Äldre nyengelska (Early Modern English), brittisk engelska

Omfattning: 1 183 690 ord (av vilka 894 100 ord i kategorin 'direct speech')

Språktyper: rättegångsprotokoll, vittnesmål, drama, didaktiska manualer, fiktionsprosa (och blandade skrifter)

Period: 1560–1760 (fem perioder: 1560–1599, 1600–1639, 1640–1679, 1690–1719, 1720–1760)

Användning: forskning, avhandlingar, uppsatser

A Corpus of Nineteenth-century English (CONCE)

Etablerad av: Merja Kytö och Juhani Rudanko (Tammerfors universitet)

Språk: Modern engelska (Late Modern English), brittisk engelska

Omfattning: 986 814 ord (bortsett ord inom metatext)

Språktyper: riksdagsdebatter, rättegångsprotokoll, drama, historieböcker, fiktionsprosa, personliga brev, vetenskapliga skrifter

Period: 1800–1900 (tre perioder: 1800–1830, 1850–1870, 1870–1900)

Användning: forskning, avhandlingar, uppsatser