



Ann Catrin Lagerkvist

Utredare, Tekn dr

Områdeskansliet för humaniora, juridik
och samhällsvetenskap

2015-01-12

Yttrande över remissen *Unik kunskap genom registerforskning*

Områdesnämnden för humaniora, juridik och samhällsvetenskap har inbjudits att yttra sig över rubricerad internremiss.

Övergripande kommentarer

Utredaren har genomfört ett gediget och konstruktivt arbete, som bör leda till ökade möjligheter och en ökad klarhet och rättssäkerhet vad gäller forskningens möjligheter att för angelägna forskningsändamål återanvända och kombinera data som ursprungligen insamlats för andra ändamål, t.ex. data insamlade av myndigheter för administrativa ändamål.

Utredaren analyserar forskningens behov på ett detaljerat och klagörande sätt och visar på att det finns goda möjligheter att förtydliga och komplettera lagstiftningen på sådana sätt att man ökar forskningens möjligheter att för forskningsändamål återanvända och kombinera data som insamlats av olika aktörer, främst myndigheter, för olika administrativa ändamål – utan att andra angelägna samhällsbehov, främst behovet av personlig integritet, behöver eftersättas. De förändringar och kompletteringar av lagstiftningen, som föreslås, bör också göra förutsättningarna för tilltänkta forskningsprojekt mera förutsägbara och mindre utsatta för beslut från fall till fall av t.ex. Datainspektionen, som under senare år i flera uppmärksammade fall skapat stora problem både för befintliga och planerade forskningsprojekt och därmed sammanhängande register, databaser, dataarkiv m.m. Som beskrivs i utredningsrapporten är registerbaserad forskning förknippad med stora möjligheter, men det finns även en hel del problem, bl.a. av metodologisk natur.

I yttrandet ges exempel på vilka kunskaper och kompetenser som Stockholms universitet och dess institutioner skulle kunna bidra med i detta sammanhang.



Nämnden vill även lyfta fram att det finns vissa oklarheter i utredningens och utredningsdirektivens användning av begreppet ”register”.

Betänkandet innebär ingen åtstramning, utan snarare det motsatta, vilket ur ett forskningsperspektiv är att välkomna. De bedömningar och förslag som utredaren presenterar gällande etikprövning, den nya sekretessbestämmelsen, kodnyckelförfarande, forskningsdatabaser och nationellt biobankregister är väl underbyggda.

Den forskningssekretess som av utredningen föreslås komplettera den gamla statistiksekretessen i offentlighets- och sekretesslagen blir viktig för att bevara allmänhetens förtroende. Nämnden delar uppfattningen att integritetsskyddet generellt inte hotas av registerforskning. Problematiken kring befintliga register som skall återanvändas för andra syften än de som de hade skapats för hanteras väldigt bra av etikkommittéerna och nämndens bedömning är att hanteringen av registerforskningen enligt de befintliga etiska principerna kan anses räcka för att tillgodose kraven som ställs av PUL och allmänna integritetsaspekter.

Medan den rättsliga dimensionen utreds ingående är dock finansieringen av forskningsdatabaserna något styvmoderligt behandlad. Regelverket bör tydligare än nu garantera databasernas långsiktiga fortbestånd än endast en omprövning vart åttonde år. En långsiktig vetenskaplig kunskapsuppbyggnad måste vara fast institutionellt säkrad.

En annan aspekt av registerforskningen som utredningen borde tagit mer fasta på är att vissa databaser idag är interaktiva, d.v.s. informationen fortsätter editeras efter den första insamlingsfasen. Editeringen är en del av användningen som görs av användare med särskild licens. Att ett forskningsregister kan vara föränderligt bör beaktas av nya regelverk.

Förslaget om ett nytt slags infrastruktur i form av utökad användning av redan befintliga forskningsdatabaser ökar den vetenskapliga nyttan av redan insamlat material. Äldre dataset, som ur samhällsplaneringssyfte kanske innehåller föråldrad information, kan med utredningens förslag om forskningsdatabaser få en ny betydelse som historiska databaser. Inom humaniora och samhällsvetenskap är det också vanligt att man digitaliserar äldre analoga kortregister.

Slutligen kan lyftas fram att forskningens syfte inte enbart handlar om att finna mönster i en population, utan att det också kan röra sig om att finna mönster i hur en population blivit bemött, hanterad, omhändertagen eller det motsatta: diskriminerad, missbrukad av olika institutioner (offentliga eller privata) etc. Den kunskap som registerbaserad forskning kan bidra med berör med andra ord inte enbart befolkningen i sig utan också hur samhällets organisering påverkar befolkningen.

Forskningens behov av att återanvända och kombinera data

Möjligheterna att för forskningsändamål återanvända och kombinera data som insamlats av olika aktörer (t.ex. myndigheter) för andra ändamål har många fördelar, t.ex.:

- Longitudinella studier, som annars vore omöjliga att genomföra, kan genomföras.
- Data som samlas in av ett specifikt forskningsprojekt, t.ex. genom intervjuer eller enkäter, kan berikas med data från andra undersökningar och register. Dyrbara specialinsamlingar av data kan därigenom ibland undvikas helt eller åtminstone begränsas till sin omfattning, vilket leder till lägre kostnader och mindre belastning av uppgiftslämnare.
- Återanvändning av registerdata och andra redan insamlade data kan i många fall bidra till högre kvalitet i de data som samlas in av ett forskningsprojekt, dels genom att registerdata ibland kan vara en säkrare och mera fullständig datakälla än data som samlas in genom särskilda undersökningar, dels genom att registerdata kan användas för att kompensera för bortfall och skevheter (bias) i specialinsamlade data med hjälp av vetenskapligt underbyggda och väl beprövade statistiska metoder.
- Stora samhällsekonomiska vinster kan uppnås genom att återanvändning av registerdata och andra redan insamlade data är förknippade med kostnader som bara utgör en bråkdel av kostnaderna för att samla in motsvarande data med hjälp av speciella undersökningar.

Jämförelse med statistikproduktionens återanvändning av redan insamlade data

När det gäller återanvändning för forskningsändamål av registerdata och andra redan insamlade data kan många paralleller dras med återanvändning av registerdata och andra administrativa data för statistikändamål. EU-kommissionen var länge motståndare till denna form av statistikproduktion och menade att officiell statistik måste baseras på specialdesignade statistiska undersökningar: totalundersökningar (t.ex. folkräkningar) eller urvalsundersökningar. EU-kommissionen har emellertid numera helt ändrat sin inställning, vilket tydligt återspeglas i följande dokument:

Commission of the European Communities (2009): The production method of EU statistics: a vision for the next decade, COM(2009) 404:

<http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=COM:2009:0404:FIN:EN:PDF>

Man anbefaller numera medlemsländerna att utveckla sina statistiksystem i riktning mot de principer och metoder som sedan länge tillämpas i bl.a. de nordiska länderna. Skälen som anförs är i mångt och mycket desamma som vi nyss angav för forskningens användning av registerdata och andra redan insamlade data, t.ex.:

- Möjligheterna att genomföra longitudinella studier och studier som ger ett innehållsmässigt integrerat och konsistent underlag över bredare samhällsområden, t.ex. den typ av integrerad ekonomisk statistik som man försöker uppnå i Nationalräkenskaperna.
- Data som samlas in genom en speciell statistisk undersökning, t.ex. genom intervjuer eller enkäter, kan berikas med data från andra undersökningar och register. Tidsödande specialinsamlingar av data kan därigenom ibland undvikas helt eller åtminstone begränsas till sin omfattning, vilket leder till lägre kostnader för samhället och mindre belastning av uppgiftslämnare.
- Återanvändning av registerdata och andra redan insamlade data, främst administrativa data insamlade av myndigheter för andra ändamål än statistiska, men även data



från andra statistiska undersökningar, inom ramen för ett enhetligt, väl genomtänkt och integrerat statistiskt system, har en rad viktiga fördelar ur kvalitets- och effektivitetssynpunkt utöver de som redan nämnts:

- registerdata är ofta en säkrare och mer fullständig datakälla än data som samlats in genom särskilda undersökningar
- registerdata kan användas för att kompensera för bortfall (som ofta är systematiska snarare än slumpmässiga) och andra skevheter (bias) i specialinsamlade data, och detta kan och bör ske med hjälp av vetenskapligt väl underbyggda och väl beprövade statistiska metoder
- registerdata och administrativa data kan användas som underlag för ett mera kontinuerligt statistikflöde än traditionella statistiska undersökningar, som bara kan genomföras vid vissa tidpunkter, t.ex. årsvis eller (mera sällan) kvartalsvis eller månadsvis eller som engångsundersökningar; administrativa register och databaser uppdateras normalt förlöpande och med förhållandevis korta fördröjningar mellan det att en händelse inträffar i verkligheten och att data om händelsen finns registrerade i det administrativa systemet
- ett enhetligt, väl genomtänkt och integrerat system, innehållande data från många olika datakällor (traditionella statistiska undersökningar, register, databaser och administrativa system av olika slag etc.), och baserat på standardiserade definitioner och förfaranden, kan säkerställa att statistiken blir mer jämförbar i olika dimensioner: i tiden, i rummet och mellan samhällssektorer och ämnesområden

Även i USA, där tillgången på registerdata och administrativa data, som går att använda för statistiska ändamål, av tradition är betydligt sämre och inte alls lika accepterad som t.ex. i de nordiska länderna, har man numera ambitionen att vid nästa folkräkningstillfälle år 2020 ersätta eller komplettera den traditionella datainsamlingen via intervjuer och enkäter med data från administrativa källor. Flera arbetsgrupper vid United States Bureau of the Census arbetar intensivt med dessa frågor. Rapporter kring detta arbete finns på censusbyråns webbplats: <http://www.census.gov/>. (Använd t.ex. sökorden administrative data 2020 census.)

Det finns givetvis även problem och nackdelar med att återanvända register och administrativa data för statistiska ändamål (eller forskningsändamål). Bl.a. har statistikproducenten (respektive forskaren) normalt sett inga eller mycket små möjligheter att påverka definitioner av observerade objekt, populationer och variabler utan måste acceptera de definitioner som gjorts för att tillgodose de administrativa behoven. Även här kan dock (liksom vid systematiska bortfall och andra skevheter och fel i data) bristerna i viss mån kompenseras med hjälp av statistiska metoder och med hjälp av begränsade statistiska undersökningar som specialutformas just för att underlätta vetenskapligt välgrundad kompensation av de nämnda bristerna.

Avvägningar mellan forskningens behov och personlig integritet

När forskare vill återanvända och kombinera data som insamlats för andra ändamål kommer liksom hittills att behövas en avvägning av forskningsändamålen gentemot andra angelägna mål, främst personlig integritet. Dessa viktiga avvägningar har hittills alltför ofta gjorts från fall till fall, främst av Datainspektionen, som i ett antal uppmärksammade fall har tagit beslut med mycket långtgående och negativa konsekvenser för forskningen. Utredarens förslag beträffande förändrad och ny lagstiftning bör innebära ett stort och mycket önskvärt steg i riktning mot mera positiva, tydliga och konsekventa förutsättningar för angelägen forskning

baserad på återanvändning och kombination av data som insamlats av andra aktörer för andra ändamål.

Utredaren pekar på att bl.a. ett antal regeringsbeslut samt även praxis utvecklad i etikprövningsnämnderna tyder på att registerbaserad forskning, forskningsregister och forskningsdatabaser inte kan anses innebära betydande intrång i enskildas personliga integritet. Som utredaren också påpekar är det dock angeläget att det finns ett tydligt lagligt stöd för denna verksamhet. Utredaren föreslår sålunda ett antal lagändringar och nya lagar:

1. Ändringar i lagen om den officiella statistiken
2. Ändringar i lagen om etikprövning av forskning som avser människor
3. Ändringar i offentlighets- och sekretesslagen
4. Ändringar i patientsäkerhetslagen
5. Ny lag om forskningsdatabaser
6. Ny lag om Nationella biobanksregistret

Vissa frågor kräver fortsatt utredning. Som utredaren pekar på, behöver förutsättningarna för Svensk Nationell Datatjänst (SND) utredas ytterligare. SND är tänkt att fungera som ett arkiv för forskningsdata, så att data insamlade av andra forskningsprojekt kan återanvändas av nya projekt. Ett sådant arkiv har också en viktig forskningsetisk roll för att i praktiken möjliggöra att forskare kan granska och ifrågasätta andra forskares resultat. Efter ett ingripande från Datainspektionen hotas SND:s existens. Erfarenheter från Danmark och Norge bör tas till vara vid utformningen av ny lagstiftning för SND.

Synpunkter på det registerbegrepp som används i utredningen

Inom statistiken definierar man traditionellt ett register som ”en fullständig och aktuell förteckning av objekt tillhörande en viss population”. Objekten kan vara t.ex. personer, företag och andra organisationer, fastigheter, fordon. Förteckningen skall helst vara auktoriserad av någon behörig instans, t.ex. en myndighet, och objekten skall helst vara identifierade med hjälp av någon auktoriserad identifierare, t.ex. personnummer eller organisationsnummer. För att ett register skall vara användbart för statistiska ändamål måste det utöver objektidentifierare också innehålla vissa variabler, som gör det möjligt att lokalisera och kontakta de listade objekten, t.ex. namn, adress och telefonnummer. Därutöver är det vanligt att register som används för statistiska ändamål också innehåller vissa basdata om objekten, som t.ex. kan användas vid urvalsdragningar och stratifieringar och vid framställning av viss grundläggande statistik avseende objektens och objektpopulationernas demografi och struktur.

Utredaren utgår i princip från en förenklad variant av ovanstående definition av begreppet ”register”. I utredningen och utredningsdirektiven förekommer dock många andra definitioner av registerbegreppet, i vissa avseenden vidare och i andra avseenden snävare än det registerbegrepp som definierats ovan.

I utredningens sammanfattning ges följande beskrivning av registerbegreppet under rubriken ”Vad menas med register?”:

”Med register avses i detta sammanhang i första hand strukturerade samlingar av personuppgifter som har upprättats inom myndigheter för primärt andra ändamål än forskning. Som exempel kan nämnas register hos statliga myndigheter som har tillkommit för administrativa eller statistiska ändamål och register hos regioner och landsting som har skapats för utvärdering och kvalitetsssäkring av hälso- och sjukvården. Även biobanker, som är

samlingar av biologiska prover från individer tagna huvudsakligen i samband med vård och behandling, är i sammanhanget av stort intresse.

Med register avses i utredningsdirektiven dock även vissa uppgiftssamlingar som primärt är skapade för forskningsändamål. Uppgifterna i dessa kan hämtas direkt från individer eller från de nämnda registren. Oftast skapas sådana forskningsregister för ett specifikt forskningsprojekt, men det förekommer också att register byggs för att kunna ge underlag för flera framtida, från början inte definierade, forskningsprojekt. I den här utredningen kallar jag sådana forskningsregister för forskningsdatabaser. De utgör en sådan infrastruktur som omnämns i utredningsdirektiven.”

Den just citerade definitionen eller snarare beskrivningen av registerbegreppet är uppenbarligen i flera avseenden betydligt vidare än statistikens traditionella registerbegrepp. Dessa vidgningar av begreppet kan dock ses som begripliga och försvarbara med hänsyn till utredningens syften och det språkbruk som faktiskt förekommer bland forskare och även bland statistiker. Utvidgningarna kan alltså ses som pragmatiska och verkar inte ha föranlett några större problem i utredningsarbetet, snarare tvärtom.

Mera tveksam och besvärande är dock inskränkningen till personuppgifter i den nyss citerade definitionen, och som också är genomgående i hela utredningen. Det må vara att många av de problem som utredningen haft att tackla väsentligen varit kopplade till motsättningar mellan forskningens behov och behovet av att skydda den personliga integriteten. Men många forskare, inte minst inom de ekonomiska vetenskaperna, är också intresserade av register och data avseende företag och organisationer. Även inom nationalekonomin, där man traditionellt har fokuserat på makrodata, finns ett ökande intresse för mikrodata, och dessa data kan på sitt sätt vara känsliga för enskilda företag på liknande sätt som mikrodata om enskilda personer kan vara känsliga ur integritetssynpunkt. För företagen är dock den skada som ett företag kan åsamkas genom utlämnande av data om företaget snarare av ekonomisk natur, t.ex. konkurrensskäl, och handlar sällan om integritetsintrång.

Det är inte riktigt möjligt att överblicka om de lösningar som utredaren föreslår även de facto skulle täcka forskningens behov vad gäller möjligheterna att återanvända och kombinera mer eller mindre känsliga data om andra objekt än personer, t.ex. företag och organisationer, men det är viktigt att kommande lagstiftning tar hänsyn även till dessa behov.

Vad kan Stockholms universitet och dess institutioner bistå med?

Som utredaren påpekar, rymmer forskningens användning av register, administrativa data, forskningsdatabaser och andra besläktade typer av datakällor stora möjligheter för forskningen och därmed för samhället – men också en hel del utmaningar och problem. Både möjligheterna och problemen är förknippade med mer eller mindre avancerade metodfrågor, bl.a. av statistisk natur, men även när det gäller frågor kopplade till ämne och innehåll. Här kan framhållas att Stockholms universitet (SU) och dess institutioner besitter kunskaper och kompetenser som kan och bör utnyttjas i samband med det fortsatta arbetet med att genomföra utredningens förslag, och när det gäller Vetenskapsrådets och andra myndigheters och organs arbete med att på olika sätt stödja forskningens behov som de beskrivs i utredningen. Det kan t.ex. gälla teknisk och innehållsmässig utformning av forskningsdatabaser, frågor kring stora och delvis heterogena databaser, dataanalys och data mining, s.k. ”big data”, frågor kring dokumentation och metadata, kvalitetssäkring, att kombinera data från olika datakällor, att kombinera registerdata med data från traditionella statistiska undersökningar.



Inom Samhällsvetenskapliga fakulteten har på initiativ av Statistiska institutionen bildats ett s.k. Survey Education and Research Centre (SuREC) som bl.a. har till uppgift att arbeta med dessa frågor, både vad gäller problem och möjligheter. Inom SuREC samarbetar metodstatistiker från Statistiska institutionen med ämnes- och tillämpningsexperter från andra samhällsvetenskapliga institutioner. SuREC fokuserar också på internationellt samarbete. Det är förhållandevis lätt för Stockholms universitet och SuREC att locka gästforskare inom de nämnda områdena till Sverige både för undervisning och forskningssamarbeten, eftersom Sveriges unika förutsättningar inom registerforskningsområdet är väl kända internationellt, i kraft av våra sedan lång tid tillbaka väl utvecklade register och administrativa system. Sedan länge finns också ett väl utvecklat samarbete mellan Statistiska institutionen vid SU och Statistiska centralbyrån (SCB) i dessa frågor. Bl.a. har det i många år funnits en särskild professur i officiell eller allmän statistikproduktion med kopplingar till både SU och SCB.

Karin Helander

Ann Catrin Lagerkvist