

Datajournalistik – ett växande område

Gunnar Nygren, Ester Appelgren & Helge Hüttenrauch

Det finns många namn på den växande genre inom undersökande journalistik som använder statistisk analys, datorstödd textanalys och olika typer av grafisk visualisering. Detta område har varit föremål för ett samarbete mellan forskare vid Södertörns högskola och Pejlrredaktionen vid Sveriges Television (SVT) under 2011/12. Syftet med samarbetet har varit att utveckla kunskap om arbetsmetoder inom datajournalistik och om publikens användning av den. Projektet har haft stöd från den statliga forskningsstiftelsen Vinnova, och går från hösten 2012 över i en ny fas där ytterligare sju medieföretag ska vara med under en tvåårsperiod.

Datajournalistik har sina rötter i USA. En av pionjärerna var Philip Meyer som redan i slutet av 1960-talet använde datorer för att analysera orsaker till upploppen i Detroit 1967. Genom att jämföra utbildningsnivå kunde han slå hål på myterna att det var utbildade ungdomar som stod bakom upploppen, något han fick Pulitzerpriset för (Vallance-Jones/McKie 2009). Begreppet ”computer assisted reporting” (CAR) växte i takt med att allt fler journalister lärde sig att använda kalkylprogram (spreadsheets) och statistik. Vid ett besök på Columbia School of Journalism i New York 1992 visade en av lärarna, Steven Ross, hur hela kurser lades upp utifrån inhämtning och analys av data (eget besök i mars 1992/GN).

Philip Meyer gav ut sin klassiker ”Precision Journalism – a Reporters Introduction to Social Science Methods” redan 1973 (senaste omarbetade utgåvan 2002) Där förklarar han hur CAR utmanar två av journalistikens starka traditioner – att passivt referera vad källorna säger och att en reporter ska vara naivt okunnig för att kunna se verkligheten med friska ögon. Istället ville Meyer anknyta till vetenskapliga metoder i journalistiken, och därmed kom han också att ligga nära den amerikanske journalistikens ideolog Walter Lippman (Meyer 2002). Meyer betonar insamling och analys när han beskriver arbetsprocessen i CAR, från hur data kan samlas in från myndigheter och andra källor till hur data kan göras åtkomligt och analyseras för att hitta mönster och orsaker.

I USA och Kanada har datajournalistiken utvecklats med betoning på inhämtning och analys. En av de senaste handböckerna definierar området så här:

At heart, it is simply reporting. Reporters use the raw data-chrunching power of the computer to manipulate information in order to uncover trends, patterns and facts that otherwise would remain hidden on some government hard drive (Vallance-Jones/McKie 2009: 1).

En organisation knuten till University of Missouri och organisationen Investigative Reporters and Editors (IRE) bildades 1989 (NICAR), och vid årliga konferenser samlas journalister från hela USA för att lära av varandra om arbetsmetoder (<http://www.ire.org/nicar/>).

Även om NICAR är knutet till ett universitet, så finns det inte särskilt mycket forskning om datajournalistik. Cox beskriver hur datajournalistiken växt fram i USA och vad begreppet CAR står för (2000). Det finns också några studier som beskriver exempel på datajournalistik på t ex New York Times (Royal 2010, Cohen 2011). De senaste tio åren har datajournalistiken också tagit steget över Atlanten, och arbetsmetoderna har etablerats i stora europeiska medier som t ex brittiska Guardian som har en hel avdelning på sin sajt som de kallar ”Datastore” (<http://www.guardian.co.uk/data>). Inom det fristående European Journalism Center (EJC) utarbetades under 2011/12 en fritt tillgänglig handbok i datajournalistik (<http://datajournalismhandbook.org/>), och intresset växer nu även bland europeiska journalister. En finländsk översikt 2011 ger många exempel på hur ledande europeiska medier arbetar med datajournalistik (Sirkkunen et al 2011). I Europa används oftast termen datajournalism istället för CAR. Det återspeglar ett intresse inte bara för analysdelen av datoranvändningen, utan också för presentation och visualisering av både resultat och databaser på nyhetssajterna. Den finska studien betonar att datajournalistik i allt högre grad handlar om att ge publiken möjlighet att interagera med de data som presenteras, att publiken själv kan gå in och söka och skapa sin egen bild av vad databaserna kan berätta (Sirkkunen et al 2011). Därmed flyttar en del av den urvalsmakt som tidigare varit förbehållen journalisten över till användarna som kan göra egna urval och analyser utifrån sina egna intressen. De journalistiska tjänster som skapas kan vara både underlag för ett traditionellt journalistiskt berättande och för sökfunktioner som utgår från publikens intresse.

PSI-direktivet och journalistikens värde

Det finns flera faktorer bakom det ökande intresset för datajournalistik. En avgörande faktor är naturligtvis utvecklingen av ett nätbaserat mediesystem där publiken kan vara aktiv och söka den information man är intresserad av. Med nätets interaktivitet kan nya tjänster utvecklas som går utöver journalistikens traditionella enkelriktade modell. Nya programvaror för visualisering kan göra siffror och fakta tillgängliga för breda grupper.

En annan faktor är EUs direktiv om att myndigheters offentliga data ska göras allmänt tillgängliga (det sk PSI-direktivet). Bakom detta finns både kommersiella och demokratiska skäl, EU-kommissionens uppskattning är att marknaden för utveckling av nya tjänster kring offentlig data i Europa är värd 27 miljarder i form av nya affärsområden, jobb och tjänster som underlättar medborgarnas vardag. Men tillämpningen av detta direktiv är mycket ojämn, och bara i vissa länder pågår en långsam process av att göra offentliga data tillgängliga (Sirkkunen 2011).

En tredje faktor är medieföretagens sökande efter nya typer av innehåll som kan locka publiken till medierna, oavsett vilken typ av plattformar det handlar om (papper, online, mobil, TV, radio). I USA har publikerna till stor del flyttat från papperstidningarna till nätet. Men intäktskällorna för kvalificerad journalistik på nätet har varit begränsade vilket i sin tur inneburit stora neddragningar för den resurskrävande granskande journalistiken. I detta tomrum har flera journalistiska ”non profit”-organisationer uppstått, privatfinansierade stiftelser som värnar om samhällsutvecklingen och demokratin. Organisationerna liknar en form av privatfinansierad public service, där vinstintresse inte är ett mål utan istället samhällsnyttan för medborgarna. Det mest framgångsrika exemplet i USA är ProPublica (propublica.org) som vunnit flera Pulitzer-priser de senaste åren för

sina avslöjanden. Andra exempel är Center for Public Integrity (iwatchnews.org), Sunlight foundation (sunlightfoundation.com) och California Watch (californiawatch.org).

För att nå maximal genomslagskraft arbetar dessa organisationer med kommersiella partners för att publicera sina avslöjanden. Den dyrbara, tunga researchen och det journalistiska grovjobbet görs alltså av de icke vinstdrivande organisationerna, medan själva publiceringen görs i samarbete med de traditionella medieorganisationerna för att nå maximalt antal tittare och läsare. Efter publiceringen har det sedan varit möjligt för alla andra som vill återpublicera materialet i sina tidningar eller webbeditioner att göra detta, utan kostnad. Detta har inneburit att lokal samhällsnyttig journalistik som annars inte gjorts har blivit möjlig att publicera.

Även om situationen i Europa inte är lika krisartad som i amerikanska medieföretag, finns samma sökande efter nya typer av innehåll som kan dra publik till medierna. Det finns också en stark public service, där nya typer av innehåll kan utvecklas i en ”non-profit”-miljö. I Sverige är det public service-tv som gått i spetsen för datajournalistiken, med en särskild redaktion sedan 2010 inom SVT. Den första tjänsten som SVT i samarbete med Sveriges Radio utvecklade var Valpejl som var en väljartjänst inför valet 2010 där alla drygt 50 000 kandidater som ställde upp i riksdags-, landstings- och kommunval fanns med. Besökaren kunde förutom uppgifter som fanns på valsedlarna ta del av deras inkomster och vilka bolagskopplingar de hade. Dessutom fanns det enkäter publicerade som hade skickades ut till samtliga riksdagskandidater och en valkompass där användarna själva kunde jämföra sina åsikter med såväl partiernas som de enskilda kandidaternas svar. SVT och Sveriges Radio fick Stora Journalistpriset som ”årets förnyare” 2010 för Valpejl. Sedan dess har Pejlrredaktionen utvecklat nya tjänster samtidigt som man arbetar tillsammans med både nyhetsredaktionen och andra program när det gäller datajournalistik (<http://www.svt.se/pejl>).

Arbetsprocessen i datajournalistik

Under 2011/12 har forskare på Södertörns högskola arbetat tillsammans med journalister från SVT-Pejl i en förstudie till ett större utvecklingsprojekt i datajournalistik. Forskarna kom från olika ämnen; journalistik, medieteknik och företagsekonomi. Förstudien finansierades med anslag från den statliga forskningsstiftelsen Vinnova.

Under förstudien har det genomförts en serie workshops med inbjudna experter. Forskarna och journalister från SVT gjorde en resa till den årliga NICAR-konferensen och IRE (Investigative Reporters and Editors) samt besökte redaktioner på medieföretag som New York Times och på stiftelser som ProPublica. Dessutom följde forskarna arbetet på SVT-Pejl under perioden, gjorde löpande intervjuer och även en användarstudie för en av de databastjänster som lades ut (se nedan). Förstudien dokumenterades löpande, och resulterade i en ny ansökan för ett större projekt från Vinnova.

Arbetsprocessen i datajournalistik har samma grundelement som i all annan journalistik, från idé till publicering och återkoppling från publiken. Processen kan förenklat beskrivas linjärt, men är lika lite linjär som all annan journalistik. De olika stegen i processen glider samman och arbetet förändras under hela processen. Den stora skillnaden jämfört med traditionell journalistik är samma som när det gäller andra former av undersökande och fördjupande journalistik – i datajournalistik handlar det inte i första hand om att referera vad olika källor påstår, utan att istället själv undersöka och ta fram

fakta som gör att journalisten kan påstå något. Med analys av olika typer av data kan journalisten själv berätta hur något förhåller sig – och sedan ta kommentarer från både de som berörs ("drabbade") och de som är ansvariga; myndigheter, politiker, företagsledare och andra maktthavare (Hanson 2009).

Så här ser grundelementen ut i arbetsprocessen för datajournalistik:

Idé – ett datajournalistiskt projekt kan starta med en idé om något som borde undersökas, och sedan söker man fram data som gör detta möjligt. Men det kan också börja med ett dataset som kommer i journalistens hand, och man börjar ställa frågor till data för att se om det finns något av allmänintresse i detta. Båda sätten att börja projektet är möjliga, även om det enligt intervjuerna är vanligast att starta med en journalistisk idé.

Datainsamling – det vanligaste i CAR var från början att utgå från myndighetsdata. Med stöd av offentlighetsprincipen kan journalisten begära ut data, till exempel alla kandidater i det svenska valet 2010 till Valpejl. Valmyndigheten lämnade ut detta, men var inte skyldig att göra det digitalt. Därför fick redaktionen för Pejls scanna in namn och personnummer för alla 50 000 kandidater. Andra myndigheter som lämnat ut data digitalt är t ex Skolverket, Brottsförebyggande rådet och Statistiska Centralbyrån (SCB).

Även om myndighetsdata är den viktigaste källan, finns det också andra. Publiken/läsarna kan själva bidra med data på ett strukturerat sätt genom "crowdsourcing". Ett exempel är Svenska Dagbladets räntekarta: när bankerna inte ville lämna ut uppgifter om vilka bolån som deras kunder betalar, vände sig tidningen till läsarna som fick fylla in ett begränsat formulär med de viktigaste uppgifterna om deras bolån. Sedan sammanställdes allt i en databas som hösten 2012 kom att innehålla uppgifter om närmare 14 000 bolån. Dessa lades också in på en karta, och det blev möjligt att på nyhetssajten gå in och jämföra sitt eget bolån med vad grannarna betalade. Dessutom genererade "räntekartan" en stor mängd artiklar där bankernas bolån granskades. Det är också möjligt att kombinera myndighetsdata och läsardata, något som SVT Pejls gjorde i projektet Brottspejl (se nedan).

En tredje typ av data är kringdata som kan användas för att illustrera och analysera, t ex geografiska data och demografiska data. Data kan också tas hem från nätet, t ex om man vill beskriva nätverk som växer fram i sociala medier som Twitter eller Facebook.

Bearbetning och analys – oftast måste data bearbetas innan det går att analysera dem, data måste rensas och läggas in i någon programvara för analys. Många journalister använder vanliga kalkylprogram som Excel eller Google spreadsheets för analys, andra använder mer avancerade statistikprogram som SPSS eller SAS. Det finns också olika programvaror för "text mining", för att kunna hitta mönster i stora textmängder t ex program som Wordle som skapar ordmoln ur texter.

I analysen ställer journalisten frågor till sitt dataset, istället för att fråga en intervjuperson. Med statistiska metoder försöker journalisten se samband och mönster i materialet. Ett annat syfte med analysen är att hitta "case" som är typiska och som kan illustrera det man vill berätta.

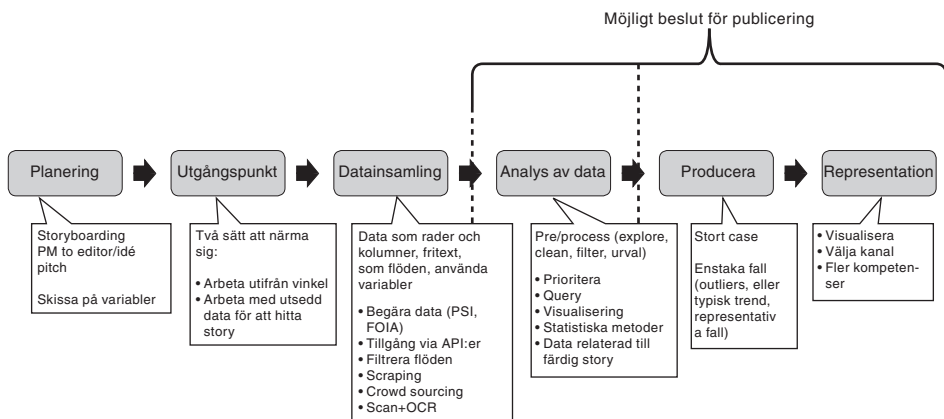
Journalistisk presentation – här går datajournalistiken över i välkända former där det gäller att berätta en historia så att den når fram till publiken. Det gäller att visa fall som är typiska för det som analysen säger, att hitta de ansvariga och fråga vad de gör åt detta. De nyheter som går att utvinna ur datamaterialet går att berätta i alla typer av kanaler, både som faktatexter om resultaten och som berättelser om människor. Dessutom tillkommer en ny dimension i nätpubliceringen, där kan hela datamaterialet läggas ut och göras sökbart för publiken. Det innebär att läsaren kan skapa sina egna nyheter och genom att hela grundmaterialet blir synligt skapas också en transparens i journalistiken. Läsaren som har skolbarn behöver inte nöja sig med Rapports inslag om de stora skillnaderna i betygsgenomsnitt – han/hon kan själv i Skolpejl gå in och ta fram betygsgenomsnitt i skolorna i sin egen kommun, jämföra lärartäthet och andelen behöriga lärare.

Med nya typer av programvaror kan statistik visualiseras på nyhetssajterna, siffrorna måste göras begripliga även för en publik som inte kan statistik. Adresser kan översättas till en karta med hjälp av Google Fusion, figurer kan göras rörliga för att beskriva en förändring som t ex Pejls siffror över hur inkomstfördelningen i Sverige spridits ut och skillnaderna ökat de senaste 20 åren. Dessa typer av visualiseringar har utvecklats av t ex Hans Rosling, professor i internationell hälsa, och hans organisation Gapminder, och kompetens därifrån har också knutits till SVT-Pejl.

Återkoppling till publiken – denna finns med under hela processen, särskilt om redaktionen arbetar med crowdsourcing. Det kan t ex handla om att efterlysa publikens berättelser om det ämne som är aktuellt, att ge utrymme för diskussioner och interaktiva inslag som t ex valkompassen där publiken testade sina åsikter i relation till partiernas politik. Det är också möjligt för redaktioner att lägga ut källmaterial och be publiken om hjälp för att analysera, att lägga ut rådata är både ett sätt att skapa transparens i processen och att inbjuda publiken att göra egna analyser och bidra med ny information (Sirkkunen 2012).

Beskrivningen av arbetsprocessen visar vad som hänt sedan Philip Meyer utvecklade CAR under 1960-talet. Då handlade det enbart om reportrar som använde datorer för att analysera myndighetsdata – nu är datajournalistik i hög grad en interaktiv process där publiken är viktig både som källa/producent och som aktiv användare av de datamängder som publiceras.

Figur 1.



Arbetsprocessen i datajournalistik har en större tyngd på datainsamling och analys än i annan vardaglig journalistik. Figuren visar en linjär process, men i praktiken glider de olika stegen ihop. Det gäller till exempel de projekt som arbetar med crowdsourcing där publikens medverkan oftast kommer efter de första publiceringarna. Det sista stadiet i processen saknas också på bilden – publikens användning av de datajournalistiska tjänsterna och den återkoppling som sker till redaktionen.

Vardagsbrott i fokus

Ett exempel på datajournalistik är Brottspejl som producerades av Sveriges Television hösten 2011. Brottspejl lanserades 15 december under nyhetssändningar i Rapport och Aktuellt. Fokus var vardagsbrotten, och alla SVTs regionala redaktioner hade under en och samma dag den 6 december följt de brott som anmäldes, intervjuat drabbade människor och följt upp vad polisen gjorde. Samtidigt publicerades en databastjänst som byggde på data från den myndighet som ansvarar från statistik om brottslighet, BRÅ. I denna databas kunde användarna se vilken typ av brott som anmäldes lokalt och även se i hur stor grad dessa anmälningar klarades upp. De kunde också jämföra siffrorna med andra kommuner och med ett riksgenomsnitt, det fanns siffror på hur många poliser som arbetar i länet och hur allt detta relaterar till befolkningmängden. På sajten fanns också en intervju med en känd kriminolog (professor Jerzy Sarnecki) samt namn och telefonnr till den lokala polischeferna.

Sajten presenterades i samband med nyhetsinslagen, och första dagen kom det 22 300 besökare. Sedan sjönk antalet, och hela första veckan kom det ca 46 000 unika besökare till sajten. Att jämföra med SVTs sajt som helhet som hade ca 2,6 miljoner unika besökare under veckan. Men även om det inte var så många som hittade till Brottspejls databas, så nådde resultatet ut i form av nyhetsinslag i de stora nyhetsprogrammen. En liten engagerad grupp gick sedan vidare till databasen.

Någon månad senare gjordes ett försök med crowdsourcing i samband med Brottspejl. Genom ett samarbete med programmet Veckans brott efterlystes publikens egna berättelser om hur de drabbats av vardagsbrott, och deras erfarenheter av hur polisen tagit hand om anmälningarna. 3000 tittare gick in på sajten och lämnade sina tips om att de varit utsatta för vardagsbrott och hur polisen agerat i samband med detta. Det fanns även en möjlighet att skicka in sin egen berättelse för publicering på SVT Pejls sajt, vilket mer än 300 personer gjorde. Alla dessa berättelser gjordes sökbara på Brottspejls sajt (utan att avslöja uppgiftslämnarnas namn), och dessutom kunde de regionala nyhetsprogrammen få bra ”case” till sina uppföljningar på vardagsbrotten. Veckans brott kunde följa upp satsningen med nya berättelser som visade verkligheten bakom statistiken. Rikspolischefen fick i studion svara på den kritik mot polisen som kom fram i många berättelser.

Inom forskningsprojektet gjordes en begränsad användarstudie av Brottspejl. Genom en webbenkät på sajten samlades svar in från 57 användare, en svarsfrekvens på 19 procent av dem som nåddes av enkäten. Av de som svarade hade 75 procent besökt sajten tidigare, de flesta var män (70 procent) och snittåldern var 48 år. Det vanligaste motivet för att besöka sajten var att man ville fördjupa sig för egen del (45 procent). Vanliga skäl var också att man ville fördjupa något man sett i ett tv-program samt för att få fakta till diskussioner med familj och vänner. I textkommentarer uttryckte många svars personer

positiva kommentarer – tacksamhet mot att SVT ger möjlighet till fördjupning på ett område där några av användarna ansåg att myndigheter försöker dölja viktiga fakta.

Brottspejl var en kombination av myndighetsdata och publikdata, men också en kombination av breda och smala kanaler. Redaktionen hade inga problem att få ut data från BRÅ i digital form, och gensvaret var stort när Veckans brott frågade efter tittarnas egna berättelser. Publiken fanns på två plan – den breda publiken såg nyhetsinslagen och programmet Veckans brott som byggde på både analysen av brottsdata och publikberättelserna, medan en ganska smal publik själv kunde fördjupa sig i brottsstatistiken i databasen på Brottspejl. Genom cross promotion kunde de breda formerna dra publik även till den smalare databasen, men bara under en begränsad tid. Databasen och de berättelser som kom från tittarna var dock förutsättningen för de nyheter som kunde produceras till de breda publikerna.

Många frågor för nya studier

Under hösten 2012 startar nästa etapp i projektet Datajournalistik. Södertörns högskola har fått fortsatt anslag från Vinnova för att under två år utveckla arbetsmetoder och en gemensam databank tillsammans med SVT och sex andra medieföretag samt SAS Institute som utvecklar mjukvara för statistisk analys. Försättningen har en miljöprofil, och den gemensamma databanken ska göra miljödata tillgängliga för journalister. Dessutom ska metoder utvecklas och kompetensen höjas på de deltagande redaktionerna. Metoderna är delvis desamma: en serie workshops, en nationell konferens om datajournalistik under 2014 och metodutveckling i samband med uppbyggnad av databanken.

Det finns många frågor runt arbetsprocessen som behöver fördjupas och undersökas:

- Hur är det med tillgängligheten på offentliga data? Hur fungerar offentlighet och ”open government” i praktiken?
- Många myndigheter kommersialiserar sina data genom att sälja dem genom olika typer av tjänster. Hur påverkar det tillgängligheten?
- Hur kan publikdata samlas in och struktureras på användbara sätt? Hur kan de verifieras och bearbetas?
- Hur kan datajournalistiken göras transparent, vad kan läggas ut och göras sökbart?
- Vilka etiska problem finns det i datajournalistik när det gäller t ex personlig integritet?
- Hur kan metoder i datajournalistik tillämpas på redaktioner? Vilka behov av ny kompetens finns det bland journalister?
- Hur kan data göras tillgängliga för publiken utan krav på förkunskaper när det gäller statistik? Vilken grafik fungerar och hur kan man bygga olika lager för användare med olika förkunskaper?
- Hur ser användarna på datajournalistik? Vilka krav har användare på datajournalistiska tjänster?
- Hur förhåller sig datajournalistik till medieföretagens traditionella affärsmodeller? Läsarbetaling och/eller annonsfinansiering?

För forskarna i projektet är rollerna dubbla; dels att delta i projektet för att utveckla arbetsmetoder och redskap och dels att dokumentera utvecklingsarbetet genom intervjuer, enkäter och ”fältanteckningar”. Projektet skiljer sig från traditionell forskning genom att vara en del av den process som forskarna samtidigt ska studera. Därmed ligger projektet nära det som tidigare kallats aktionsforskning, när forskaren inte som brukligt står vid sidan av som en ren observatör utan istället finns mitt ibland det som ska beforskas och deltar i en process (Rosengren/Arvidsson 2002). Metoden är inte särskilt vanlig i medieforskningen, men kanske just därför desto mer spännande.

Resultat kommer att publiceras löpande under projektet. Både i form av journalistik i de medier som deltar och i form av vetenskapliga artiklar från forskarna.

Referenser

- Cohen, S., et al (2011) *Computational Journalism*. Communications of the ACM, October 2011, vol. 54, no. 10
- Cox, M. (2000) *The Development of Computer Assisted Reporting*. A paper presented to the Newspaper Division, Association for Education in Journalism and Mass Communication, Southeast Colloquium, March 17-18, 2000, University of North Carolina, Chapel Hill.
- Hanson, Nils (2009) *Grävande journalistik*. Stockholm: Ordfront förlag.
- Meyer, Philip (2002) *Precision Journalism. A Reporters Introduction to Social Science Methods*, forth edition. Lanham: Rowman & Littlefield Publishers.
- Rosengren, Karl Erik& Arvidsson, Peter (2002) *Sociologisk metodik*. Stockholm: Liber förlag.
- Royal, C. (2010) *The Journalist as Programmer, A case study of The New York Times Interactive News Technology Department*, Sala de Prensa, WEB PARA PROFESIONALES DE LA COMUNICACION IBEROAMERICANOS, Junio 2010, Año XI, Vol. 6.
- Sirkunen, Esa; Aitamurto, Tanja; Lehtonen, Paulina (2011) *Trends In Data Journalism*, Helsinki: Next Media.
- Vallance-Jones, Fred & McKie, David (2009) *Computer Assisted Reporting, a Comprehensive Primer*. Oxford University Press.

GUNNAR NYGREN, Ph.D., Professor, Institutionen för kommunikation, medier och IT, Södertörns Högskola, Huddinge gunnar.nygren@sh.se

ESTER APPELGREN, Ph.D., Högskolelektor, Institutionen för kommunikation, medier och IT, Södertörns Högskola, Huddinge ester.appelgren@sh.se

HELGE HÜTTENRAUCH, Ph.D., Högskolelektor, Institutionen för kommunikation, medier och IT, Södertörns Högskola, Huddinge helge.huttenrauch@sh.se