Mediepubliken och stora datamängder

Ester Appelgren, lektor, Södertörns högskola
Sara Leckner, lektor, Malmö högskola

Introduktion

2015 introducerade Svenska Dagbladet en personaliserad förstasida baserad på en algoritm. Förstasidans nyhetsvärdering hade tidigare utförts uteslutande av journalister men kunde nu genomföras på ett mer automatiserat vis. Vid första anblicken låter automatisering kanske som att människors kunnande och expertis har rationaliseras bort, men en algoritm kan snarare utgöra ett exempel på motsatsen. Algoritmer består av olika sorters villkor som någon på förhand måste ställa upp, i detta fall någon med stort kunnande om nyhetskonsumtion och journalistik. En algoritm för en personaliserad förstasida på en dagstidning skulle exempelvis kunna positionera och välja ut nyheter efter närhetsprincipen, det politiska läget, förekomsten av personer med stort allmänintresse, men också efter den enskilde läsarens personliga preferenser och tidigare beteendemönster.

Listan av villkor för hur nyheterna ska presenteras kan alltså fyllas på och uppdateras efter hand med kriterier som motsvarar en del av det löpande journalistiska kunnandet. Men ytterligare en finess med algoritmer är att de också kan agera på egen hand med någon form av intelligens, det vill säga de kan lära av en situation, minnas detta och därefter förfinas till att göra smarta val åt medieanvändarna. Hur detta lärande sker är beroende av vilka användardata som finns tillgängliga för analys och agerande i kombination med vilka kriterier som ställts upp av de som programmerat algoritmen. Användarens preferenser gör dessutom att algoritmerna med tiden tränas upp till att förutspå en specifik användares val och därefter underlätta för denne, genom att på förhand göra de innehållsmässiga valen. För exempelvis Svenska Dagbladet torde de valda kriterierna främja affären men också det journalistiska uppdraget, eftersom Svenska Dagbladets algoritm är anpassad för just nyhetsförmedling och nyhetskonsumtion. Men i en tid då mycket av det nyhetsinnehåll vi konsumerar på nätet når läsaren via aktörer som primärt inte sysslar med nyheter, såsom Facebooks flöde av olika sorters innehåll, kan de algoritmer som väljer nyheter vara programmerade för helt andra syften än nyhetsrapportering. Detta kan få såväl positiva som negativa konsekvenser för vilket innehåll medieanvändarna får presenterade för sig, vad de får se och inte får se, men också hur nyheterna som väljs ut är vinklade. Algoritmer som används för att utvinna kunskap ur beteenderelaterad data brukar på engelska kallas “machine-learning technology”. När maskiner, så

som i fallet med Facebooks algoritmer, väljer åt människor är det ett uttryck för så kallad teknologisk paternalism. Enligt Hilty (2014) handlar paternalism om att någon tror sig veta en lösning på en annan individs problem och sedan implementerar denna lösning utan att den person det egentligen berör gett sitt medgivande. När datorer är inblandade kan sådana antaganden och beslutsprocesser byggas in i system. Det krävs dock att systemen har tillgång till data om personer. I dag baseras sådana data om individen ofta på olika former av digitala spår som aktivitet och interaktion på internet, men också positioner som loggats via exempelvis mobiltelefoner. Beteendedata kan innebära stora mängder data, och i dag talar man ofta om så kallade ”stora datamängder” (big data); data som skapas med hög volym, hastighet och mängd och vars storlek överstiger det som vanligen hanteras av mjukvaror (t.ex. Manovich, 2012). För mediebolagen utgör beteendedata bland annat klickstatistik, det vill säga, vilket innehåll som klickas på, den tid som spenderas på innehållet, liksom graden av engagemang hos användarna. Denna typ av data är viktiga mått för att motivera annonsförsäljning, men också för att utveckla digitala tjänster, automatisera produktion av text- och video-berättelser, samt för försäljning till tredje part (t.ex. Stone, 2014. Se Bolins kapitel i denna bok för en diskussion kring medieproduktionsmodeller).

I viss mån kan sådana data också användas för att justera och uppdatera journalistiskt innehåll. Analys av stora datamängder kan vidare också användas som redaktionella verktyg för journalistiskt berättande och research, också kallat datorstödd rapportering och datajournalistik, liksom för så kallad robotjournalistik där algoritmer skapar innehållet i en artikel (t.ex. Coddington, 2014; Lewis och Westlund, 2015).

Efter en tids hype kring stora datamängder och beteendedata argumenteras det för att det finns en övertro på att stora datamängder skall komma att lösa medieföretagens alla problem. Dessa data är förstås användbara om man analyserar dem utifrån de förutsättningar de har (se också Andersson Schwarz et al., 2014), men det finns risk för ”apophenia” att man ser korrelationer och samband som egentligen inte är relevanta (t.ex. Boyd och Crawford, 2012). Det finns också en risk för att man samlar in ”all” data utan att veta vad de skall användas till, vilket bland annat kan skapa problem med lagring och hantering. Insamling och samkörning av stora databaser kan också vara problematiska ur etisk synpunkt för varje enskild individ som

ingår i en större datamängd. Beroende på hur man väljer att se det kan alltså användning av beteendedata utgöra såväl ett hot mot individens integritet, som ett erbjudande i form av en personaliserad tjänst riktad till den individuelle användaren. Gränsen för vad som kan klassas som ren övervakning och vad som utgör företagens behov av data för att tillgodose affärsmodeller, kan därför vara hårfin. Och då antaganden om människors behov och avsikter i dag alltmer delegeras till datorer, hävdar många att teknologisk paternalism, som beskrivs tidigare i detta kapitel, kraftigt kan inkräkta på den personliga integriteten.

På en vanlig nyhetssajt kan i dag upp mot 50 olika aktörer spåra en persons beteende (dataskydd.NET, 2015), ofta genom cookies (se t.ex. Carlsson och Jacobsson, 2012; Fuchs, 2014; Wong et al., 2014). Det finns en oro hos allmänheten kring datainsamling, men den är fortfarande inte alarmerande. Troligtvis eftersom gemene man inte är fullt medveten om, eller intresserad av, hur mycket, vad och när denne faktiskt delar data med ett företag eller organisation, eller i nästa steg, med en tredje part. Statistik tyder dock på att människor gradvis håller på att bli mer medvetna om att de har ett val när det gäller insyn i deras internet-aktivitet, som att sluta dela med sig av digitala spår eller ta kontroll över hur innehåll de inte valt, visas för dem. I senaste mätningen av Orvesto Konsument uppgår exempelvis andelen av den svenska befolkningen som använder annonsblockerare, till 21,5 procent (TNS Sifo, 2015; se också Findahl, 2014). En femtedel av befolkningen har numera alltså valt att stänga av annonser. Steget är inte långt till att också stänga av möjligheten för företag att spåra den enskilde individens aktivitet.

Området kring lagstiftning, teknisk utveckling, men inte minst medieanvändarnas attityd till att ge sitt medgivande till att dela med sig av digital beteendedata, kan få stora konsekvenser för mediebolagens affärsmodeller på kort såväl som lång sikt. I detta kapitel kommer vi därför fokusera på hur teknisk innovation, hårdare lagstiftning och personlig integritet, blir en utmaning för medieindustrins framtida affärsmodeller, i relation till insamling, bearbetning och samkörning av medieanvändarnas beteendedata.

Forskningsöversikt

Den internationella forskningen kring insamlingen och användningen av beteenderelaterad trafikdata i stora datamängder är på framväxt. I dessa sammanhang är det inte ovanligt att använda ordet "dela" (vilket vi också valt att göra i detta kapitel) – som är ett positivt värdeladdat ord – för att beskriva individens potentiella möjlighet, upplevelse och inställning till distributionen av personlig data till andra och tredje part. Begreppet "dela" utnyttjas exempelvis av Facebook som inte en enda gång nämner ordet "sälja" i sin datapolicy kring personlig data som användargenererat innehåll och demografi, utan beskriver att de data som samlas in (och säljs vidare) just är information, som "delas" (se Andersson Schwarz kapitel i denna bok för en fördjupad analys av delningslogiker).

Integritetsparadoxen och svårigheten att sluta dela med sig

Enligt EU-kommissionens Special Eurobarometer 359 från 2011, ansåg 74 procent av Europas befolkning att delandet av personlig information blir en allt större del av det dagliga livet, där den främsta anledningen till att dela med sig var att få tillgång till olika digitala tjänster såsom sociala nätverk eller vid e-handel (Eurobarometer, 2011). Forskning visar också att användare internationellt är oroade över att de inte har någon kontroll över sina internetgenererade data, men också att deras data kan användas i andra kontexter än de ursprungligen delades i (Eurobarometer, 2015; Lilley et al., 2012; Pew, 2014). Individer är också bekymrade för att tredje part, såsom annonsörer eller andra kommersiella aktörer, ska få tillgång till deras personliga information (t.ex. Findahl, 2014; Kshetri, 2014; Narayanaswamy och McGrath, 2014; Pew, 2014). I en studie utförd av Pew Research Centre anser så många som 91 procent av de amerikaner som tillfrågats att människor har förlorat kontrollen över hur personlig information samlas in och används av företag (Pew, 2014). Samtidigt finns forskning som tyder på att många individer inte har gjort några större förändringar i sin datadelning eller sitt integritetsskydd under senare år (Christensen och Jansson, 2015; Findahl, 2014; Light och McGrath, 2010; Martin et al., 2015). Anledningar kan vara att människor inte har den tekniska kunskapen och inte vet hur de skall göra, eller att de inte anser sig ha

något att dölja (t.ex. Findahl, 2014; Light och McGrath, 2010; Martin et al., 2015). Detta kan beskrivas som en så kallad integritetsparadox (privacy paradox); medieanvändare använder frekvent tjänster som potentiellt kan vara integritetskränkande, samtidigt som de säger sig vara mycket oroade för att data samlas in när de använder produkter och tjänster på nätet (se Bechmann, 2014).

Det är dock inte primärt lättja som styr om användare begränsar den data som delas. Att hoppa av eller avstå insamling av digital data, så kallad opt-out är i dag rent praktiskt krångligt för den enskilde individen eftersom en stor del av det sociala livet kräver närvaro på internet och de tjänster som finns där. Forskning visar vidare att människor inte läser den information som står i det digitala medgivandet utan accepterar villkoren ändå (se t.ex. Alverén, 2012; Bechmann, 2014; Fuchs, 2014; Pitkänen och Tuunainen, 2012). När medgivandet har getts har individen ofta gått med på att dess data delas med, eller säljs till, tredje part (som t.ex. i Facebooks användaravtal), eller att data kan komma att korsköras med en rad tjänster som ingår i företagets produktportfölj (som t.ex. Googles insamlade användardata). Troligtvis bedömer därför användaren att det är tillräckligt mycket värt att använda tjänsten trots de risker hen potentiellt utsätter sig för när beteendedata delas. Studier har visat att en sådan bedömning är beroende på flera saker. En viktig aspekt är de vänner och bekanta som redan använder tjänsten, som ses som en garant för tjänsten eller företaget (Bechmann, 2014). Individen har dessutom ofta förtroende för företag som de har någon form av relation till (t.ex. genom tidigare köp, prenumeration eller renommé), vilket kan öka benägenheten att dela (t.ex. Jai och King, 2015; Leon et al., 2013). Förtroendet ökar också om företaget beskriver hur de beteendedata som samlas in kan komma att användas (Jai och King, 2015; Leon et al., 2013; Sas, 2015). Men viljan att dela har också visat sig bero på typ av data. Aktiviteter som mediepreferenser och e-handelsvanor anses mindre känsliga att dela (se t.ex. Eurobarometer, 2011; Sas, 2015), än uppenbart personlig information såsom telefonnummer, adress, personnummer, kreditkortsnummer eller hälsorelaterade uppgifter (Jai och King, 2015; Leon et al., 2013; Pwc, 2012; Sas, 2015). Samtidigt argumenterar forskare för att enkelheten i att underhålla relationer via internet, där det som delas ofta är mycket personligt, också har visat sig vara en viktig faktor (Ellison et al., 2007). Därmed handlar viljan att dela mycket om sammanhanget där information delas (se också

van Gool et al., 2015). Sammanhanget kan dock också innebära en risk för att användare struntar i, mer eller mindre medvetet, huruvida de delar känslig data eller ej. Det kanske beror på bristande teknisk kunskap kring hur individdata kan samlas in och medieanvändarna kanske inte ens inser att information om dem kan lagras (se Carlssons kapitel i denna bok för en diskussion kring betydelsen av medie- och informationskunnighet). Forskare argumenterar för att okunskap är vanligt vid yngre personers användande av sociala nätverk (t.ex. Pitkänen och Tuunainen, 2012; Vanderhoven et al., 2014), och yngre personer förefaller mindre oroade över ofrivilligt delande med andra människor än äldre personer är (t.ex. Findahl, 2014). Däremot tycks inte attityden till att dela data vara en generationsfråga vad gäller inställningen till huruvida myndigheter och företag övervakar och inkräktar på den personliga integriteten, och inte heller vara relaterat till kön (Findahl, 2014).

Frågan kring benägenhet att dela data är därmed komplex; spåren av individernas internetaktiviteter kan ju innebära nackdelar, men också nyttjas i form av fördelar. I en studie utförd 2014 såg vi exempelvis att individer som delar mer data gagnas av det genom att personaliserade tjänster blir mer träffsäkra, och att värdet av dem då kan öka för den enskilde individen (Appelgren et al., 2014; se också Fuchs, 2014). Tidigare studier har också visat att användare är skeptiska till vissa av de fördelar som delning sägs ge, men är villiga att dela mer personlig information under vissa förutsättningar, exempelvis om det medför gratis eller effektivare tjänster (Pew, 2014).

Personlig integritet

När personer delar något explicit, har de möjlighet att bedöma situationen och sammanhanget och välja om och vad de vill dela. När användare i stället producerar information implicit befinner de sig i en situation där de inte nödvändigtvis är medvetna om att de delar. Detta, påpekar boyd (2010), formar de data som produceras och människors förväntningar kring dem, liksom om det öppet redogörs för att det delas och det som ges eller tas outsagt. Medieanalys, vilket för mediebolagens del ofta handlar om processer där datamängder av klickdata lagrats och sedan analyseras, skulle kunna ses som en form av offentlig övervakning. Nissenbaum (2004:116) beskriver hur sådan

övervakning involverar ny teknik och applikationer som har utvecklats för att möjliggöra observation av människor, liksom insamling av information om människor, behandling av insamlad data, samt analys och spridning av den information som kan utvinnas. Begreppet personlig integritet är högst närvarande i detta sammanhang och omfattar olika sorters mänskliga behov, såväl inom egendom, som rättigheter (Carlsson och Jacobsson, 2012). Begreppet kan också delas upp i personlig och medborgerlig integritet (Olsson, 2008). Personlig integritet är närvarande i mänskliga relationer, medan medborgerlig integritet handlar om var gränsen går för vad samhället behöver veta om en individ. Dessa typer av integritet behöver inte nödvändigtvis sammanfalla, även om de sannolikt oftast gör det. Liksom Nissenbaum (t.ex. 2004), menar boyd (2010) att integritet är kontextuell och huruvida information är känslig, beror på kontextens normer. Tar man innehåll som är producerat explicit eller implicit ur sin kontext riskerar man att kränka sociala normer; tänjer man på gränserna för vad användarna medgivit eller distribuerar utan deras medgivande, riskerar man att kränka deras integritet.

Normer är också centrala i forskning kring integritet, där informationsteknik också kan ses som ett verktyg, som individer kan använda för att förebygga oönskade beteenden och effekter (Light och McGarth, 2010). För närvarande finns ingen lagstadgad eller teknisk lösning som ger människor total kontroll över sin personliga information på nätet. Carlsson och Jacobsson (2012) nämner flera orsakerna till detta: 1) Begreppet personlig integritet är, som nämnts ovan, kontextberoende och saknar en enhetlig definition. 2) Den personliga integriteten är på många sätt en paradox; för att vissa uppgifter skall kunna skyddas måste annan information exponeras. För att exempelvis få tillgång till Facebooks och Googles lösenordskyddade tjänster krävs att individen uppger personliga uppgifter så som exempelvis namn, ålder och lösenord och medgivande till att viss typ av data delas till tredje part. 3) Syftet med totalt skydd kan motsäga andra rättigheter såsom pressfrihet eller insamling av uppgifter för forskningssyften. Således, i praktisk mening, är det nästan omöjligt för enskilda individer att kontrollera all information som de själva genererar.

Rättsliga ramverk

Samhället har i allmänhet närmat sig oron för hot mot den personliga integriteten genom regleringar. Den nuvarande EU-förordningen dataskyddsdirektivet – skydd av personlig integriteten vid behandling av personuppgifter i de europeiska länderna – antogs 1995 (Datainspektionen, 2015). Mot bakgrund av den ökade användning av ny teknik för att bearbeta stora datamängder personuppgifter, har direktivet kritiserats för att vara oflexibelt och dåligt anpassat till den rådande tekniska utvecklingen. Även om direktivet skyddar den grundläggande rätten till individers skydd av personuppgifter, kan det appliceras på olika sätt i EU:s medlemsländer, vilket har lett till varierande skyddsnivåer för personuppgifter beroende på var en person bor eller köper en vara eller tjänst (EU-kommissionen, 2012a). Å andra sidan har det gett varje land rätt att säkerställa att landets konstitution faktiskt gäller, som till exempel pressfrihetsförordningen i Sverige (Dahllöf et al., 2013).

År 2014 antog Europaparlamentet ett förslag till en ny dataskyddsförordning, som ska ersätta det nuvarande direktivet. Förordningen förväntas träda i kraft tidigast 2018 (Datainspektionen, 2015), och kommer att innehålla en del förändringar. Till exempel kommer förordningen att gälla för hela den europeiska unionen, vilket ger ett enhetligt dataskydd med samma regler i alla medlemsstater och kommer att gälla för alla företag (inklusive icke-europeiska företag) som agerar inom EU. Förordningen kommer att ge användarna kontroll över hur deras personuppgifter kan användas; när privat information samlas in, vem som utför insamling, vem som kommer att använda informationen, och om den kommer att säljas eller på annat sätt nyttjas av tredje part. Huvudpunkterna omfattar kravet på aktivt, informerat och frivillig samtycke (opt-in) från användaren för behandling av personuppgifter. (Datainspektionen, 2015; EU-kommissionen, 2012a). För både företag och annonsörer ställer detta stora krav på att kunna förklara när, var och hur sådant samtycke sker, och vad det gäller för. Förordningen omfattar även rätten för personer att ”glömmas bort” i alla kanaler där den personliga informationen har spridits (Datainspektionen, 2015; EU-kommissionen, 2012a). Det föreslås också ett krav på aktivt tillstånd från användaren för att (som t.ex. företag) få genomföra urval, särskilt vad gäller bedömning av kreditvärdighet, pålitlighet eller beteende.

I princip kan kraven begränsa möjligheten till selektion för företag och annonsörer, vilket kan minska noggrannheten och relevansen av den segmentering som utförs vid analys. Det innebär också ett högre ansvar för uppgifter som säljs till tredje part, exempelvis i annonssyfte. Källor såsom bloggar, hemsidor, Twitter-meddelanden och öppna Facebook-konton kommer att bli föremål för samtycke för publicering av personuppgifter, vilket innebär att personer som nämns i gärningar, åsikter eller namn måste godkänna publicering (Dahllöf et al., 2013).

Ur ett journalistiskt perspektiv har oro framförts för att detta i slutänden kommer att ge myndigheter och domstolar rätten att avgöra vad som definierar journalistisk verksamhet, eftersom journalistiska arbeten är undantagna från kravet på samtycke. Journalister har också protesterat mot att personligt samtycke och rätten att bli bortglömd kan göra det svårare att undersöka personer i journalistiska syften (Brandel, 2013). Vidare kommer nya regler att gälla för ”inbyggd integritet” (privacy by design) och ”förvald inställning för integritetsskydd” (privacy by default), som syftar till att skydda personuppgifter genom hela livscykeln av produkter och tjänster, och föreskriver att integritet byggs in direkt i designen av system och processer som hanterar personlig data. Företag som bryter mot förordningen riskerar bötesbelopp om 2–5 procent av dess globala omsättning eller minst en miljon Euro (EU-kommissionen, 2012a). För bolag som Facebook och Google kan detta innebära mångmiljardbelopp. Av förekommen anledning har därför inflytelserika amerikanska it-bolag reagerat med intensiva lobbykampanjer mot förändringarna, med stöd av den amerikanska regeringen (Brandel, 2013).

Analys av utvecklingen i Sverige

För mediebolagen är data en nödvändig resurs av affärsskäl, men gentemot individen hävdas ofta att insamlingen av beteenderelaterad data används för att förbättra användarupplevelsen. Individen uppfattar kanske vid första anblicken inte att insamlad beteendedata från mediebolagens sajter är så känslig, men det är trots allt möjligt att till synes okänslig data kan användas på ett integritetskränkande sätt om det hamnar i fel händer (se Kshetri, 2014).

Oro och integritet – ett svenskt forskningsfokus

Människors inställning till insamling av beteenderelaterad data är ännu inte särskilt väl undersökt i Sverige. Framför allt är det oro för spridning av personlig information som har undersökts, och i synnerhet på sociala medier. Bland annat visar Bergström (2013) att 34 procent av de som använder sociala medier känner oro för att personliga uppgifter ska komma att missbrukas (bara användning av betalkort på nätet oroade fler personer, 51 procent). I 2014 års SOM-undersökning visar Ghersetti att uppfattningar om att människor bör vara försiktiga med vad de lägger upp i sociala medier är vanligt förekommande, liksom uppfattningen att det är nästan omöjligt att radera uppgifter. I Svenskarna och Internet (Findahl, 2014) visas att oron bland användare för vad stora it-företag gör med personlig information har ökat bland den svenska befolkningen, men samtidigt uppger majoriteten att de inte är oroade (73 procent) och att de inte har något att dölja (67 procent). De flesta användare anser att den personliga integriteten är viktig (84 procent) och att ägna sig åt oro över hot inte är överdrivet (77 procent).

Pågående studier

Vi har under våren 2015 undersökt människors inställning till om delning av personlig data är känslig, och studerat loggfiler över 22 svenska hushålls internetkonsumtion under en vecka i januari 2015 (Leckner et al., manuskript inskickat). Loggfilerna baseras på alla data som dessa hushåll genererar vid användning av internet, och kan registreras till exempel i fem-minutersintervall. För att visa hur till synes avidentifierad data över internetkonsumtion, genom tolkning av data eller vid korskörning med andra datakällor, kan uppfattas som känslig, har vi från dessa data tagit ut tre exempel (av många) där loggfilerna visat på frekvent användning av en specifik tjänst eller datasuffix från hushållens internetanvändning under en avgränsad tid. Exemplen vi har valt rör följande ämnen: 1) Personlig blogg. 2) Politiska preferenser. 3) Porr.

1. Det första exemplet handlar om beteenden som kan avslöja identiteten hos den person som genererat en loggfil. Ett hushåll gjorde flera dagliga besök på en viss blogg, där de registrerade

länkarnas namn tydde på inmatning av text. Sekvensen av besöken följdes av besök till Google Analytics. Sådant beteende tyder på att bloggen tillhörde någon i hushållet. Om dessa data såldes eller delades med tredje part är det alltså möjligt att identifiera någon i hushållet genom att spåra den personliga bloggen.

2. Det andra exemplet behandlar politiska åsikter, som kan identifieras genom en person i hushållets val av webbplatser. Hushållet hade under undersökningsperioden enbart besökt nyhetskällor som tydligt stod till höger på den politiska skalan. Om sådana data hamnade i fel händer skulle det vara möjligt att katalogisera hushållens politiska preferenser utan deras vetskap.

3. Det tredje exemplet tar upp känsliga aktiviteter som porrsurfning. Detta hushåll bestod av en tonårsfamilj där loggfilerna visade att besök på porrsajter hade gjorts tidigt på eftermiddagen under vardagarna. Det är sannolikt att dessa besök gjorts av en tonåring i hushållet som kommit hem från skolan. Om sådana data delades med resten av familjen, eller om den blev öppet tillgänglig, skulle denna information kunna göra stor inverkan på individers personliga integritet och potentiellt också negativt påverka familjemedlemmarna.

För att kunna säkerhetsställa information i exemplen ovan skulle analys av loggfiler behöva ske under en längre tidsperiod. Det visar dock vad som är potentiellt möjligt att utröna från loggfiler, d.v.s. sådana användardata som operatörer har tillgång till, men som delvis kan registreras av företag via cookies. Exemplen visar också på risken att tolka data fel. Det är inte säkert att den trafik som registreras används för de syften som först ter sig uppenbara. De tre hushållen som här fått statuera exempel är alla medvetna om att de delar med sig av sina internetaktiviteter för forskningssyften, och har skrivit under ett medgivande som tillåter det. Teoretiskt är de alltså medvetna om vad de gått med på att dela, men frågan är om de i praktiken förstått vad det innebär, och om de skulle kunna tänka sig att dela den här typen av data med andra aktörer? Detta är en problematik som delvis är realitet för gemene man i dagens internetsamhälle.

Hushållen som deltagit i studien tillfrågades också kring viljan att dela med sig av beteenderelaterad data, med hjälp av enkäter. Urvalet kompletterades med ett urval av personer som inte hade skrivit på

något medgivande (kunder i ett shoppingcenter på samma ort), som fick svara på samma frågor. De icke-representativa urvalen indikerade att de tillfrågade inte ville dela data med kommersiella aktörer. Detta gällde i synnerhet för de hushåll som faktiskt gett sitt medgivande att dela med sig av sina data till sin operatör för forskningssyfte. Kanske hade medgivandet till att bli loggad också resulterat i ett mer försiktigt förhållningssätt?

Våra resultat tyder också på att viljan att dela ökade något om de tillfrågade personerna erbjöds någon form av ersättning, speciellt för aktiviteter där delandet resulterade i en bättre användarupplevelse. Detta kan kopplas till begreppet reciprocitet (se även SOU, 2015:94) vilket innebär ömsesidigt utbyte mellan två eller fler aktörer. Detta är centralt i det samtida medielandskapet då olika sorters sociala utbyten är vanligt förekommande och viktiga för att medskapande ska ske (jämför t.ex. online communities och reciprocal journalism) (se t.ex. Lewis et al., 2014; Sankaranarayanan och Vassileva, 2009). För våra resultat var det särskilt tydligt bland de svarande som hade undertecknat medgivande, och något mer för kvinnor än män. Resultaten visade annars ingen signifikant skillnad mellan mäns och kvinnors svar. Vad gäller viljan att dela specifikt innehåll med olika aktörer, visar våra resultat att detta beror på interaktionsnivå; besök eller klick på innehåll (t.ex. Wikipedia och nyhetssajter) upplevs som mindre känsligt att dela med sig av, än när det krävs större interaktion och engagemang (t.ex. besök på Facebook och Google), och framför allt vad gäller delning av rent personlig information (t.ex. e-post) (Leckner och Appelgren, 2015).

Våra resultat överensstämmer med tidigare forskning när det gäller viljan att dela beroende på sammanhang och typ av data, där användarna blir mer ovilliga att dela ju mer personlig data blir, liksom att de är mer villiga att dela om de har en personlig vinning i delandet. En annan slutsats är att det är skillnad på vad personer tycker när de i enkäter utfrågas om delning och vad de faktiskt gör i realiteten när de använder internet. Det skall dock tilläggas att våra studier är av explorativ karaktär med ett litet urval. Fler studier behövs för att kunna uttala sig med statistisk säkerhet.

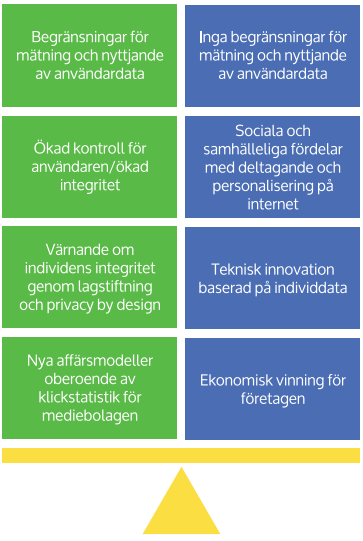
Slutsatser och implikationer

Enligt EU-kommissionen är nyckeln till framgång på nätet att ett förtroende finns mellan medieanvändare och företag. Om förtroendet brister kan det leda till att användare avvaktar med köp eller låter bli att testa nya tjänster. Kort sagt kan ett bristande förtroende motverka innovation (EU-kommissionen, 2012b). I nationella digitala agendor runt om i Europa syns ett fokus på individers rätt till digital integritet. Den osäkerhet som föreligger kring hur kommersiella företag och myndigheter bör hantera beteenderelaterad data, i relation till de ekonomiska vinster som kan göras vid användning av insamlad data, ökar trycket på att nya lagar behöver utformas för att skydda individers integritet. Som våra studier visar är det relativt lätt att avidentifiera insamlad beteendedata, och därmed inverka på individens personliga integritet. Figur 1 visar på balansförhållanden som i dag råder mellan individens integritet och företagens nyttjande av användardata.

Konsekvenser av begränsningar

Det finns en risk att affärsmöjligheterna baserade på personlig data har ett bäst-före-datum. I och med att en växande andel svenska medborgare redan valt att blockera annonser i digitala mediekanaler är steget troligtvis inte långt till att de också väljer att stänga av möjligheten för medieföretag att spåra aktivitet vid mediekonsumtion. När människor inte längre går med på att dela med sig av sina data i den utsträckning de i dag gör får mediebolagen stora problem – de rådande affärsmodellerna upphör att fungera.

Figur 1 **Balansförhållandet mellan individens integritet och företagens nyttjande av beteenderelaterad data.**



För mediebolagen är användarna inte bara en publik utan också kunder. Försvinner användarnas data går företagen miste om kundinsikt, vilket påverkar deras digitala affärsmodeller negativt. Detta påverkar i sin tur medieföretagens överlevnad, då det inte längre går att följa beteenden för specifika kundsegment, rikta annonser baserat på positioneringsdata eller i realtid mäta klick. Om lagstiftning dessutom begränsar insamling och därmed tillskriver mediebolagens användning av användarstatistik negativa epitet av ”övervakning” eller ”missbruk”, tror vi att en åtgärd som primärt skyddar individen i slutänden blir kontraproduktiv. Detta kan jämföras med hur man 2009, genom den så kallade Ipred-lagen, förändrade skyddet för immateriella rättigheter och därmed begränsade individers nedladdning av medieinnehåll genom att göra beteendet illegalt. Resultatet blev en förändrad syn på nedladdning. Människor skrämdes till att upphöra med sitt beteende och företagens kommersiella vinstintresse lyftes fram på ett negativt sätt.

Utveckling som rör exempelvis inbyggd integritet kan leda till att möjligheter tekniskt byggs bort så att företag inte längre kan samla in så kallad rik data om sina kunder. Från ett integritetsperspektiv är det lockande att begränsa tekniska möjligheter till analys av beteendedata, men en sådan åtgärd innebär också att teknisk innovation

utsätts för censur och bromsar den digitala affärsutvecklingen för de svenska medieföretagen. Att på detta sätt hämma innovation inom informationsteknikområdet för att värna den personliga integriteten är ett enkelt sätt att slippa sätta press på de stora it-företagen såsom Google och Facebook, samt ett ignorant sätt att slippa utbilda och informera människor i att ta eget ansvar för sin personliga integritet på nätet. Det är inte heller givet att förbud mot att samla och bearbeta beteenderelaterad data har avsedd verkan.

För att komma runt problematiken med insamling och användning av beteendedata, kräver många företag i dag att deras användare signerar användaravtal och använder inloggningsfunktioner, vilket gör att individer fråntas rätten att kräva kontextuell integritet och ger företagen kontroll över specifik individdata. Exempelvis boyd (2010) påpekar att detta må vara lagligt, men frågan är om det är etiskt? Som tidigare nämnts är det svårt för individer att avstå från att använda internet och de sociala och samhälleliga funktioner som blir alltmer närvarande där.

Valet att inte ge sitt digitala medgivande till företag, innebär ofta att individen helt måste avstå från att använda företagets tjänst. Vidare är den text som presenteras för användaren i steget då digitalt medgivande efterfrågas ofta lång och komplicerad. Här finns alltså mycket att göra för att förenkla för medieanvändaren och öka individens insyn vad gäller delning av personlig data vid mediekonsumtion. Ett sätt att stärka den digitala kompetensen och öka den informationskunnighet som krävs för att medborgarna fullt ut ska kunna delta i demokratin och den demokratiska utvecklingen, är därför att medieföretagen blir än tydligare gentemot användarna med hur och när användargenererad data samlas in och används.

Åtgärder för hållbara digitala affärsmodeller

Även om det i dagsläget är potentiellt få privatpersoner som medvetet ägnar sig åt självcensur vid mediekonsumtion eller aktivt stänger av möjligheterna till att mediekonsumtion spåras – som att tacka nej till att cookies lagras – kan en ökad medvetenhet hos medieanvändarna ses som både positivt och negativt. För mediebolagen är det riskfyllt att bygga affärsmodeller på användardata som användarna kan bestämma sig för att inte vilja dela med sig av. Minskar användarnas

vilja att låta sig spåras med en ökad medvetenhet om digital integritet, som våra resultat pekar på, måste mediebolagen i stället hitta medel som gör att användarna börjar se fördelar med att dela beteendedata.

Lagstiftning skulle inom en snar framtid kunna komma att förbjuda insamling av beteenderelaterad data som inte användarna gett sitt medgivande till. Det är, som tidigare påpekats, vid en första anblick positivt för den enskilde individens integritet. Men då stora internationella aktörer, som Facebook och Google, ändå kan komma att kringgå nationell lagstiftning i denna fråga och fortsätta samla data om medieanvändarna, kan en följd bli att mindre och nationella aktörer, såsom svenska mediebolag, försvagas. De kan då komma att skapa en beroenderelation till de multinationella bolag som fortfarande kan samla in data av integritetskänslig natur. Om så sker, skulle ett företag som Facebook kunna förvandlas till Sveriges största nyhetsförmedlare, där svenska medieföretag endast blir en av många innehållsleverantörer. Detta är redan en realitet i viss mån. Utvecklingen pekar mot att mediebolag i och med detta får färre egna möjligheter att påverka nyhetsvärderingen gentemot användaren. Detta skulle kunna få effekter på demokratin och samhället. Tillåts stora bolag som Facebook att ta över den nationella marknaden utan begränsningar, sätts nyhetsvärderingens logik ur spel och den kunskap som nationella och lokala mediebolag besitter kommer rekapitulera inför den så kallade filterbubblan och algoritmers betydelse för nyhetsvärdering.

Men utvecklingen av en ökad medvetenhet hos befolkningen kring den personliga integriteten på nätet behöver inte få ett fullt så nattsvart utfall. Om svenska mediebolag, liksom deras internationella motsvarigheter, drivs mot att insamling av användardata blir en mer transparent process, som styrs av etiska värderingar för den personliga integriteten, gynnas företagen också långsiktigt, då ökad lojalitet från medieanvändarna kan göra affärsmodellerna mer hållbara. Av detta skäl behöver mediebolagen motiveras, genom politiska åtgärder, att ytterligare fokusera på hur de kan stärka sina band med sina användare. Banden stärks genom öppnare processer, primärt synlig i design, där tydlighet kring insamling av individdata förstärks. När data samlas in och för vilka syften detta görs, måste motiveras av att användarna också känner att de vinner på att dela sina data.

Det varje medieföretag nu borde inspireras till att ägna sig åt, är därför att på olika sätt hitta möjligheter att motivera medieanvändarna

till att vilja dela med sig, utan att kompromissa med det journalistiska uppdraget, så att den ur integritetssynpunkt försvarbara data som kan samlas in, blir så rik som möjligt. Datainsamling ska inte kännas som att tv-pejlaren hotfullt knackar på dörren för att en tv-apparat detekterats i hushållet, medieanvändaren ska i stället tycka att det är lika relevant att i dagens samhälle dela beteendedata, som det för tidigare generationer var att betala för en tryckt papperstidningsprenumeration. Att skrämma upp medieanvändarna kommer inte att lösa problemet med integritet på internet, men att vinna deras förtroende blir i stället nyckeln till framgång i morgondagens medielandskap.

Avslutande ord

Vi har i detta kapitel argumenterat för att relationen mellan användarnas och företagens för- och nackdelar med användningen av insamlad beteendedata, liksom de åtgärder som behöver göras för att generera en integritetsbevarande men affärsmässig process, är komplexa frågor utan helt självklara lösningar. Skulle lagstiftare hårdare reglera datainsamling, exempelvis genom inbyggd integritet eller genom att förbjuda att data används för andra syften än de samlats in för, skyddas individen. Men den potential, såväl ekonomisk som samhällelig, som stora datamängder erbjuder går samtidigt om intet. En av de branscher som då påverkas negativt blir mediebranschen, där domstolar och myndigheter skulle ges rätten att kunna begränsa tillgången till källmaterial och på detta sätt också påverka journalistiskt innehåll. Då allt som sker på internet mer eller mindre går att följa får, i ljuset av vad som diskuterats ovan, frågan om den personliga integriteten därför en allt större betydelse för samhällsutvecklingen. Skyddas individen är baksidan på myntet att medieföretagens affärsmöjligheter kring personlig data begränsas, vilket får demokratiska konsekvenser då utbudet av kvalitetsjournalistik minskar. När journalistik inte längre kan finansieras med digitala affärsmodeller, begränsas i sin tur människors kunskaper om aktuell samhällsutveckling.

Vi vill dock hävda att balansen mellan insamling och förtroende kan få ett positivt utfall om mediebolagen uppmuntras till att på ett mer transparent sätt samla data om sina användare. Känner sig medieanvändaren trygg med att insamling sker, liksom orsakerna bakom att den sker, kan detta till och med bidra till ökad lojalitet hos användarna. Om lagstiftning påskyndar och underlättar synliggörandet

av mediebolagens metoder och syften med insamling av data, finns en möjlighet att människors förtroende för den insamling som ändå lagligen sker, är av godo, och görs för att svensk journalistik ska kunna finansieras med hållbara affärsmodeller.

Referenser

Alverén F (2012) *Såld på nätet: priset du betalar för gratis*. Stockholm: Ordfront.

Andersson Schwarz J, Hammarlund J, Kjellberg M och di Grado S (2014) Åsikter på sociala medier är inte den allmänna opinionen. *DN.se*, 25 december.

Appelgren E, Leckner S och Mejtoft T (2014) Mediekonsumentens medvetna och omedvetna val: En nyckel till morgondagens mediekonsumtion. *Medie-Sverige 2014*, Göteborg: Nordicom, Göteborgs Universitet, 29–37.

Bechmann A (2014) Non-informed consent cultures: Privacy policies and app contracts on Facebook. *Journal of Media Business Studies,* 11(1): 21–38.

Bergström A (2013) Internetanvändningens kontexter. I: Weibull L, Oscarsson H och Bergström A (red.) *Vägskäl*, Göteborg: SOM-institutet vid Göteborgs universitet, 493–505.

boyd M D (2010) *Privacy and Publicity in the Context of Big Data*. (WWW to print). Raleigh, North Carolina, USA, 29 april.

boyd M D and Crawford K (2012) Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon. *Information, Communication & Society*, 15(5): 662–679.

Brandel T (2013) Vem skall äga makten över dig på nätet? *Svenska Dagbladet*, 27 maj.

Carlsson B och Jacobsson A (2012) *Om säkerhet i digitala ekosystem*. Lund: Studentlitteratur.

Christensen M och Jansson A (2015) Complicit surveillance, interveillance, and the question of cosmopolitanism: Toward a phenomenological understanding of mediatization. *New Media & Society*, 17(9): 1473–1491.

Coddington, M. (2014) Clarifying Journalism's Quantitative Turn: A typology for evaluating data journalism, computational journalism, and computer-assisted reporting. *Digital Journalism*, 3(3): 331–348.

Dahllöf S, Funcke N och Dahlin F (2013) EU:s nya datalag oförenlig med svenska grundlagar. *DN.se,* 12 april.

Datainspektionen (2015) *EU:s dataskyddsreform: Nya lagar om personuppgiftsbehandling från 2018*. Tillgänlig på: http://www.datainspektionen.se/lagar-och-regler/eus-dataskyddsreform/ (hämtad 30 maj 2015).

Dataskydd.NET (2015) *Integritet och privatliv är grundläggande mänskliga rättigheter*. Tillgänglig på: dataskydd.net/ (hämtad 30 maj 2015).

Ellison B N, Steinfield C och Lampe C (2007) The benefits of Facebook ”Friends:” Social capital and college students use of online social network sites. *Journal of Computer-Mediated communication*, 12(4): 1143–1168.

Eurobarometer (2011) *Attitudes on Data Protection and Electronic Identity in the European Union* (Special Eurobarometer 359). European Commission. Tillgänglig på: http://ec.europa.eu/public_opinion/archives/ebs/ebs_359_en.pdf. (hämtad 30 maj 2015).

Eurobarometer (2015) *Data Protection* (Special Eurobarometer 431). European Commission. Tillgänglig på: http://ec.europa.eu/justice/newsroom/data-protection/news/240615_en.htm (hämtad 8 juli 2015).

EU-kommissionen (2012a) Why do we need a data protection reform? *Commission proposes a comprehensive reform of the data protection rules.* Tillgänglig på: http://ec.europa.eu/justice/newsroom/data-protection/news/120125_en.htm (hämtad 30 maj 2015).

EU-kommissionen (2012b) *Proposal for a regulation on the European parliament and of the council on the protection of individuals with regard to the processing of personal data and on the free movement of such data* (COM (2012) 11 final). Brussels: European Commision. Tillgänglig på: http://ec.europa.eu/justice/data-protection/document/review2012/com_2012_11_en.pdf (hämtad 30 maj 2015).

Findahl O (2014) *Svenskarna och Internet 2014*. Göteborg: .SE.

Fuchs C (2014) *Social media: a critical introduction*. London: SAGE Publications.

Ghersetti M (2015) Sociala medier till nöje och nytta. I: Bergström A, Johansson B, Oscarsson H och Oskarson M (red.) *Fragment*, Göteborg: SOM-institutet vid Göteborgs universitet, 511-522.

Hilty L M (2014). Ethical Issues in Ubiquitous Computing: Three Technology Assessment Studies Revisited. I: Kinder-Kurlanda K och Ehrwein C (red) *Ubiquitous Computing in the Workplace: What Ethical Issues? Advances in Intelligent Systems and Computing*. Heidelberg: Springer.

Jai T-M och King N J (2015) Privacy versus reward: Do loyalty program increase consumers willingness to share personal information with third-party advertisers and data brokers? *Journal of Retailing and Consumer Services*. Online.

Kshetri N (2014) Big data's impact on privacy, security and consumer welfare. *Telecommunications Policy*, 38(11).

Leckner S och Appelgren E (2015) The audience's willingness to share internet traffic data: an emerging ethical challenge for the media industry Paper presented at *NordMedia 2015*, Copenhagen University 13–15 August.

Leon P G, Ur B, Wangz Y, Sleeper M, Balebako R, Shay R, Bauer L, Christodorescuy M och Faith Cranor L (2013) What Matters to Users? Factors that Affect Users' Willingness to Share Information with Online Advertisers. *Symposium on Usable Privacy and Security* (SOUPS) 2013, July 24–26, Newcastle, UK.

Lewis SC, Holton A E och Coddington M (2014) Reciprocal Journalism. *Journalism Practice*, 8(2): 229–241.

Lewis S C och Westlund O (2015) Big Data and Journalism: Epistemology, expertise, economics, and ethics. *Digital Journalism*, 3(3): 447–466.

Light B och McGrath K (2010) Ethics and social networking sites: a disclosive analysis of Facebook. *Information, technology and people*, 23(4): 290–311.

Lilley S, Grodzinsky F S och Gumbus A (2012) Revealing the commercialized and compliant Facebook user. J*ournal of information, communication and ethics in society*, 10(2): 82–92.

Manovich L (2012) Trending: The promises and the challenges of big social data. I: Gold M K (red) *Debates in the Digital*

Humanities (pp. 460-475) Minneapolis, MN: University of Minnesota Press.

Martin S, Rainie L, och Madden M (2015) *Americans Privacy Strategies Post-Snowden*. Pew Research Center. Tillgänglig på: http://www.pewinternet.org/2015/03/16/Americans-Privacy-Strategies-Post-Snowden/ (hämtad 30 maj 2015).

Nissenbaum H (2004) Privacy as Contextual Integrity. *Washington Law Review*, 79(1): 119–157.

Olsson A (2008) *Att stänga det öppna samhället*. Kristianstad: Tusculum förlag.

TNS Sifo (2015) Orvesto Konsument 2015:1. Tillgänglig på: www.tns-sifo.se/rapporter-undersokningar/rackviddsrapporter-orvesto/orvesto%C2%AE-konsument (hämtad 24 juni 2015).

Pew (2014) *Public Perceptions of Privacy and Security in the Post-Snowden Era*. Pew Research Center. Tillgänglig på: http://www.pewinternet.org/2014/11/12/public-privacy-perceptions/ (hämtad 30 maj 2015).

Pitkänen O och Tuunainen V K (2012) Disclosing personal data socially: An empirical study on Facebook users privacy awareness. *Journal of information privacy & security*, 8 (1): 3–29.

Pwc (2012) *Consumer privacy: What are consumers willing to share. Consumer Intelligence series*. PriceWaterCoopers. Tillgänglig på: http://www.pwc.com/us/en/industry/entertainment-media/publications/consumer-intelligence-series/consumer-privacy.jhtml (hämtad 30 maj 2015).

Sankaranarayanan K och Vassileva J (2009) Visualizing Reciprocal and Non-Reciprocal Relationships in an Online Community. Workshop on *Adaptation and Personalization for Web 2.0*, Trento, Italien. 22–26 juni.

Sas (2015) *Finding the Right Balance Between Personalization and Privacy*. Research Paper. SAS Institute, USA.

SOU (2015: 94) *Medieborgarna och medierna: En digital värld av rättigheter, skyldigheter – möjligheter och ansvar*. Tillgänglig på: http://www.sou.gov.se/wp-content/uploads/2015/11/SOU-2015_94_Hela_Webb.pdf (hämtad 10 december 2015).

Stone M L (2014) *Big Data for Media*. Report (Nov). Reuters Institute for the Study of Journalism. Tillgänglig på: http://

reutersinstitute.politics.ox.ac.uk/publication/big-data-media (hämtad 23 juni 2015).

Vanderhoven E, Schellens T, Valcke M och Raes A (2014) How Safe Do Teenagers Behave on Facebook? An Observational Study. PLoS ONE 9 (8): 1–9.

van Gool E, Ouytsel J V, Ponnet K och Walrave M (2015) To share or not to share? Adolescents' self-disclosure about peer relationships on Facebook: An application of the Prototype Willingness Model. *Computers in Human Behavior*, 44: 230–239.

Wong K, Wong A, Yeung A, Fan W och Tang S K (2014) Trust and privacy exploitation in online social networks. *IT professional*, 16(5): 28–33.