

## **Internet profiling: The economy of data intraoperability on Facebook and Google**

**Anja Bechmann**

MedieKultur 2013, 55, 72-91

Published by SMID | Society of Media researchers In Denmark | [www.smid.dk](http://www.smid.dk)

The online version of this text can be found open access at [www.mediekultur.dk](http://www.mediekultur.dk)

*This article investigates online profiling and data strategies by identifying and comparing data strategies of the two most visited internet companies, Google and Facebook. The aim of the article is to use media economics and management perspectives to enrich the discussion on profiling from a political economy perspective. The article maps differences in the data strategies of the services and the potential data collected through a data point analysis, and suggests conceptual distinctions between vertical and horizontal data strategies, touch point and social network, integrated and diversified application programming interface (API) structures, and relevance and reputation data strategy perspectives. Furthermore, the findings in the article suggest distinguishing among profiling for advertisers, developers, and government agencies. Addressing these stakeholders through the identified data strategic differences, the findings point to different implications for privacy, digital divides, algorithmic adoption, and societal segregation and intolerance.*

### **Introduction**

One of the current most lucrative economic models of large internet services is targeted commercial information through profiling of the target group or individual. In 2012, 83% of Facebook's income came from advertising and an estimated 96% of Google's income

came from advertising in 2011 (Ebersman, 2012; Kelly, 2012). As addressed by other scholars (Castells & Arsenault, 2009; Van Couvering, 2003; Mansell, 2004), the internet has an oligopolistic market structure in which ownership and power (of advertising) is in the hands of a few companies. In 2010, Google had 71.27% of the search market and Facebook had 60.7% of the total social media market (Anderson, 2012). In Denmark for instance, 18.3% of total internet time is spent on Facebook and 15.2% on Google searches (Thunø, 2013). These internet companies compete for advertising money by being able to create a lucrative advertising environment through precise targeting.

In order for advertising agencies to target the relevant groups and individuals, internet companies need both a palette of products that generates relevant data and data mining techniques that are able to scrutinize large data sets and single out relevant profiles through more-or-less advanced pattern recognition. This type of dossier is a digital footprint or “signature” that can be found in the ocean of transactional data created in the course of everyday life” (Dempsey & Flint, 2004, p. 1464). On the basis of these data, internet services can track and document activities and develop comprehensive profiles based on certain demographics, purchases, social interactions, interests, locations, and clickstream patterns (Rubinstein, Lee, & Schwartz, 2008).

To obtain these valuable data, users often receive messages from Google and Facebook that allow the companies to collect data on their whereabouts or asking them to allow the companies to cross-reference data from different devices and different services in the name of user convenience. These end-users are users of cloud computing; seamlessly switching between digital devices such as smartphones, laptops, tablets, and services such as games, social networks, searches, video streaming, and clothing and grocery stores.

Giving away data online has especially been seen as an act of convenience at the cost of security in what has been called the “the dancing pigs” phenomena (McGraw & Felton, 1999) or the “privacy paradox” (Barnes, 2006); users do worry about security and privacy, but not enough to compromise convenience. With social media, the increasing amount of data that is freely shared has been analyzed as an aspect of user impression management (eg. Boyd, 2007), where the data shared play a central role in the image of a person, for the self portraits of who we are. In this interpretation, we are therefore freely participating in a “surveillance” (Albrechtslund, 2008) that we can benefit from socially. Functionally, the registration of, for instance, location can give access to information on what or who is near the user when moving around, along with tips and ideas from users on what to read, watch, buy, or like.

From a commercial point of view, freely accessible data about users are a valuable commodity because they make (predictive) behavioral targeting (Dwyer, 2009) more precise in terms of interest in a specific product category, proximity to a physical store, or other indicators of whether a certain product would be interesting and therefore more likely to sell to a certain user. Furthermore, in cloud services the integration of data from different user modes, such as, searches, social networking, or mail, makes it possible to mine complex profiles of users that are attractive to advertisers and third party companies.

The strategy of personalized targeting through data mining can be treated as a schism between providing relevant information to users and the usage of data for commercial purpose; or as phrased by Millar, the internet companies have an interest in the best match between “the predictive profiling and the ‘individual’s underlying psychological properties’” (2009, p. 114). Hence arises the growing awareness of privacy that has characterized both the public debate and internet research recently (Gross & Acquisti, 2005; Capurro, 2005; Fernback & Papacharissi, 2007; Albrechtslund, 2008; Ess, 2009; Boyd & Hargittai, 2010).

The academic literature on internet profiling and privacy has strong roots in surveillance research (e.g., Foucault, 1995; Lyon, 2001) and political economy (e.g., Marx in Fuchs, 2011; Fernback & Papacharissi, 2007). Both fields of research address the uneven power relation between the superior commercial company or state agency and the repressed user. On the other hand, we have a growing literature that tries to analyze empirically how users behave in this economy and not least how researchers methodologically use application programming interfaces (API) to document such online behavior (e.g., Neuhaus & Webmoor, 2012). The aim of this article is not to conduct a user study on profiling, but rather to detail and enrich the discussion on profiling in the political economy field (Wasko, Murdock, & Sousa, 2011) through a media economics and management perspective, by analyzing the *differences* in data strategies and the implications of profiling for *different* types of stakeholders. By doing so, the article hopefully will contribute to comparative studies of data strategies as an emerging and necessary research topic. The key contributions of the article are the development of preliminary, grounded, and comparative data strategy concepts and the attempt to broaden the political economy discussions through media economics.

### **Theory and concepts: Internet profiling in communication studies**

Political economy is preoccupied by the “structural” as well as “processual” power relations in society (Mansell, 2004, p. 97) and is often considered to be critical, normative, and macro-level based in contrast to the more micro-level descriptive media economics. Instead of accepting the status quo, political economy scholars often challenge the unjust and inequalities (Wasko, Murdock, & Sousa, 2011, p. 3). However, this article will use the media economics’ descriptive and grounded micro-level approach as a way to inform further work on profiling within a political economy perspective. More precisely, the article is inspired by a theoretical understanding of economy based on the mapping of ownership structures in media industries as executed by, for instance, communication scholars Castells and Arsenault (Castells & Arsenault, 2008; Castells, 2009) and will use this approach to identify possible data gathered through acquiring and developing different features and services to profile users. Furthermore, the article will use Mintzberg’s five focus areas in strategic analysis (pattern, position, perspective, plan, and ploy) to sum up the differences in the data strategies of the companies (Mintzberg, 1987).

In this article, internet profiling as a concept is defined as the ability of an organization or company to create an understanding of a person on the basis of personal information or data points (Neuhaus & Weboor, 2012) gathered by data companies through specific data strategies (e.g., behavior, attitudes, and motives, and potential behavior and interests). Profiling and personalization are not synonyms in this article. Typically, profiling is used as personalization in third party apps or advertisements, but can also be used to single out individuals in risk assessments by government agencies or the financial industry.

The concept of *intraoperability* (Sutor, 2011) has been adapted here to characterize the uneven power relation in the data economy. Interoperability is defined as the way in which services and databases are able to “talk” to one another and share data across domains and platforms through the programming interface. Sutor criticizes the use of the term “interoperability” for not taking into consideration the intent of opening up the code. He argues that we need to distinguish between actual *interoperability*, where parties connect in symmetrical power relations, and *intraoperability*, where software providers are dominant in terms of market share, attitude, or acquiescence, and want to “suck all-important data and processing into the central software ecosystem” (Sutor 2011, p. 214). In the case of Facebook and Google, we see *intraoperability*. Developers agree upon an asymmetrical power relationship where they connect to Facebook and Google, thereby enhancing the importance of their standards, making them more powerful as data hubs and passage points (Castells, 2009; Latour, 1987; Bechmann, 2009). Following this argument, it is the description of the *differences* in these data *intraoperability* strategies that is the focal point of this article, and who the stakeholders are in this *intraoperable* data economy.

Profiling, data exchange, and predictive behavioral targeting are hot topics in contemporary internet research, and the literature addresses different stakeholders dependent on the research field and perspective. Often, profiling is studied from a privacy policy perspective (e.g., Stutzman, Gross, & Acquisti, 2012; Nissenbaum, 2010; Bodle, 2011a), exemplifying certain privacy issues with selected cases of extreme profiling; not as a data infrastructure and data strategy analysis as this article is aiming to conduct. However, this body of literature addresses the relationships among data companies, users, advertisers, and government agencies (e.g., Dempsey & Flint, 2004; Solove, 2008). Other profiling studies are either qualitative user studies of personal data usages (e.g., Taddicken, 2012; Marwick & boyd, 2011; Bechmann, 2013b) or predictive behavior studies of, for instance, social media users (e.g., Kosinski, Stillwell, & Graepel, 2013; Jernigan & Mistree, 2009). These studies do not focus on data flow structure and profiling as economy, but on optimization and evaluation of profiling or on profiling as sense-making (Bechmann & Lomborg, 2013). This literature places an emphasis on the relationships among the data company in question, the users, and in some cases advertisements as instances of communication. However, a growing literature on application programming interfaces (APIs) as tools for data retrieval introduces a new stakeholder of user profiling: developers and third party companies (e.g., Neuhaus & Webmoor, 2012; Bechmann & Lomborg, in review; Bodle, 2011b).

Thus, three stakeholders can be identified in the economy of data intraoperability apart from the data companies and the users themselves:

1. Advertisers
2. Third party companies (developers)
3. Government agencies

The article will consider these three stakeholders in order to discuss profiling and the economy of data intraoperability.

Even though existing studies have not compared data strategies of Google and Facebook, this article will build on knowledge gained via several studies on different *aspects* of Google's and Facebook's data strategies. These strategies are thoroughly discussed in non-academic literature about Facebook and Google, but with a focus on the profile of the companies (e.g., Kirkpatrick, 2010; Vise & Malseed, 2005; Battelle, 2005). Roosendaal (2012) and Gerlitz and Helmond (2013) focus on the economy of Facebook's "like" and the data flows affiliated with these plugins used by external websites. These plugins are developed to increase the amount of data on usage patterns for Facebook on third party websites. Gerlitz and Helmond show how Google Analytics is by far the most widely used tracker based on the top 100 global websites (according to Alexa.com) and that 18% of the websites have installed Facebook social plugins or Facebook connect. Both plugins are tracking tools for developers in order to integrate their websites with Google and Facebook (often as an authentication procedure or Facebook like-button). This article will build on the findings of Gerlitz and Helmond and expand and update their findings in order to generalize to the data strategy of Google and Facebook. Supplementing Gerlitz and Helmond, Bodle (2011b), Stutzman, Gross, and Acquisti (2012) and Raynes-Goldie (2012) outline the history of data collection, tracking, and interoperability, with a focus on privacy issues. This article will draw on these mappings as part of the comparison of data strategies of Facebook and Google. Similar works on Google are primarily found in the critical literature on Google as a monopolistic filter bubble (e.g., Elgesam, 2008; Halavais, 2010; Fuchs, 2011). This literature will provide insight into the structure of Google Search, but when it comes to data economy, it is important to take into consideration that Google owns other services apart from Search that enable them to collect data across domains.

### **Methodology: Data point analysis**

This article focuses on analyzing the data strategies of the top two internet services in the world, Facebook and Google (alexa.com, April 2013), in order to outline and compare their data economies. Turning to Neuhaus and Webmoor's concept of data points as instances of "personal information used in a digital context" (2012, p. 46), the differences in data strategies will be identified through a data point analysis that seeks to answer three ques-

tions: How do the companies gather different data points about users? What does that say about their overall strategies? And how do the strategies affect profiling?

Methodologically, the article will use a “follow the medium” (Rogers, 2013) inspired approach in which the different possible data points of the companies are mapped through a descriptive outline of their product portfolio and how they add to profiling. In order to qualify the case descriptions, the article will build on triangulation of analytical sources: the above outlined existing literature, supplemented with API structure analysis, and online information from and about the companies. As the article will show, there are important differences in their data strategies. Besides the limitations to outlining data flows through APIs (e.g., Bechmann & Lomborg, forthcoming), the knowledge of the API structure is more profound in Facebook than Google in the following sections. Facebook has one shared API in contrast to Google’s many different APIs for each service. Also, the size of the Google product portfolio has led to a more selective and exemplifying description than in the case of Facebook. Due to the economic value of data algorithms and the earlier mentioned schismatic relationship between convenience and exploitation, both Google and Facebook are black boxes to researchers, as the companies have no interest in revealing what kind of data they have on users and exactly how they retrieved it. This often gives the research contributions on this topic a somewhat speculative nature, because we are not able to make in-depth tests of the actual algorithms and data strategies (e.g., Bucher, 2012); we can only move towards an understanding of the potential profiling principles. However, strategic pattern analysis is always produced from the outside and in retrospect (Mintzberg, 1987). The result is that the data strategic pattern analysis and the mapping of potential profiling are made from a company-centered *public* perspective.

In the next two sections, the cases of Google and Facebook will be presented in terms of product portfolio, advertising interface, and API structure, and how they potentially add to profiling. Due to methodological limitations, the outline cannot be exhaustive, but it will highlight the most important services and products developed, acquired, or achieved through collaborations. Afterwards, the article will summarize the differences in data strategies and discuss the implication for advertisers, third party companies, and government agencies. The data strategy concepts developed are grounded in the cases, but steered by the theoretical framing and identified stakeholders.

## Google

Google was founded in 1998 as a search engine that would rank search results according to estimated relevance and importance with the algorithm PageRank (Brin & Page, 1998; Vise & Malseed, 2005; Battelle, 2005). In 2000, Google started to use advertisements on their search engine, connected to search keywords through Adwords. The data points that advertisers are able to choose among are metrics, such as, gender, age, interests, remarket-

## Article: Internet profiling

ing, topics, placements on, for instance, specific external websites, search or display ads, keywords, device, and connection.

The screenshot shows the Google AdWords interface. At the top, there is a navigation bar with tabs for Home, Campaigns, Opportunities, Tools and Analysis, Billing, and My account. Below this is a search bar and a breadcrumb trail: Select campaign settings > Create ad group > Create ads. The main content area is titled 'Type: Display Network only - All features'. It includes a description: 'An ad group contains one or more ads and targeting methods. For best results, try to focus all the ads and targeting in this ad group on one product or service. Learn how to structure your account.' Below this is a form for 'Ad Group name' with the value 'digifeet'. The 'Choose how to target your ads' section has three radio buttons: 'Display keywords' (unselected), 'Interests & remarketing' (unselected), and 'Use a different targeting method' (selected). A dropdown menu for 'Select a targeting method' is open, showing options: 'Topics' (Show ads on pages about specific subjects), 'Placements' (Show ads on websites that you choose), 'Age', and 'Gender'. On the left side, there is a sidebar with 'All online campaigns' (No matching campaigns, New campaign), 'Shared library', 'Bulk operations', and 'Reports and uploads'.

*Illustration 1: Examples of metrics from Google Adwords.*

Google registers the clickstream pattern and search terms of users when using Google Search. Furthermore, the personal ad profile becomes finer grained with Google's contextual targeting services. Among others, AdSense registers advertising categories (released in 2003) when clicking on Google advertisements on external webpages, DoubleClick (acquired in 2008) registers ad interactions and page views on primarily display ads, and Google Analytics (released 2005) registers on-site behavior across digital media platforms. The acquisition of Teracent (in 2009) makes Google able to tailor ads instantly according to the tracking results and the acquisition of Admobs (2010) strengthens Google's ad market on mobile platforms (Skou, 2010).

These advertising strategies mirror their product strategies. In 2000, Google began a new distributed strategy of search. Instead of forcing users to visit their site, they make a plug-in so that users can search from other websites. The buying and releasing of new kinds of services in 2002 and onwards under the Google brand follow this diversification strategy. In 2004, Google acquired Keyhole and Where 2 Technologies, digital mapping companies, and later in 2005 introduced the services Google Earth and Google Maps on the basis of

this technology. Today it is possible to integrate with, for instance, Google Maps through API so that other web services can make different mash-ups. Google Maps has become the most used mapping service for location based social networks and other smartphone geo-services.

In 2004, Google released a beta version of their ad-based online email client Gmail that allows the user to keep all their mails in the cloud of Google, and in 2005, Gmail became available for mobile platforms along with many other Google services. IGoogle (released in 2007, terminated in 2013) was a personal tailored web portal that helped users keep an overview of the (potentially) many Google services through single sign-on. This integration later became a standard setting of the Google services, tying them closer together both visually and legally.

In 2005, Google also released YouTube and acquired the location-based social networking service Dodgeball (founded 2000) and in 2009 opened a similar service called Google Latitude. Google Latitude locates the user on different platforms through IP address (PC), GPS, or Cellular Positioning (mobile platforms). It is possible through layers to integrate the geographical positions of your network with, e.g., geo-tagged status updates from Google+ (replacing Buzz released in 2010). In 2005, Google also bought the Linux-based operating system Android. From a data strategic perspective, this was an important move, as Google then could collect data from the smartphone that was formerly unavailable to the company. This data strategy is a continuation of the diversification strategy of Google that started with Google Search and other web application, spread to websites (through Google Analytics, DoubleClick, and AdWords), browser (Chrome), and desktop, and then to operating system and last but not least hardware. With the hardware data strategy, Google is able to collect data directly from a device like Apple that gives renewed data points to the pool. This strategy is executed in collaboration with hardware companies such as Sony (Google TV released in 2010), Asus, LG, Samsung (different Nexus smartphone versions), Acer (Chromebook), and Toyota (Google driverless car). As of March 2012, Google introduced one privacy policy to all their services, which allowed them to profile from data points across services, treating the user as "single users" (Reitman, 2012), and in 2013, Google released beta versions of Google Glass to selected developers. Google Glass again adds new data points as wearable computing, not only as a mobile device, but as an accessory that has the potential to generate data on eye-tracking technology. This adds new dimensions to profiling in terms of potential data points, such as unconscious user eye-movements.

As of 2013, Google has over 100 different services ranging from web and mobile stand-alone application to desktop applications, operating systems, and hardware. As a developer, this horizontal data strategy can be integrated through different APIs that make fine-grained profiling for them more difficult than if they were to just access one API with all the data points.

Google Analytics Data Export API	Google Apps APIs
Google Base Data API	Blogger Data API
Google Booksearch Data API	Google Calendar Data API
Google Code Search Data API	Google Contacts Data API
Google Documents List Data API	Google Finance Portfolio Data API
Google Health Data API	Google Maps Data API
Picasa Web Albums Data API	Google Project Hosting Issue Tracker API
Google Sidewiki Data API	Google Sites Data API
Google Spreadsheets Data API	Google Translator Toolkit Data API
Google Webmaster Tools Data API	YouTube Data API

*Illustration 2: As of 2013, Google has 20 different APIs for developers to integrate with, to obtain different data points in user profiling (<https://developers.google.com/gdata/docs/directory>).*

As an Internet user it is difficult not to come across some of Google's services because of their expansive strategy and the heavy impact of especially Google Search and Youtube. As this short and non-exhaustive overview of Google's services shows, the functions available are innumerable, but the functions run fairly single standing with little advanced data cross-reference despite the integrated privacy policy.

## Facebook

At first glance, the products of Facebook are much less complex than Google's, because Facebook "only" offers social networking. Google has a diversification strategy with acquisitions of different services on the internet, for instance, video streaming service Youtube, online work facilities (Drive), Gmail, and Android. In comparison to Google's strategy, Facebook integrates services into the same framework when collecting data on its users. Even so Facebook like Google also builds on the strategy of acquiring companies in order to obtain the relevant user functions to generate detailed usage data, and relies heavily on collaboration with strong partners, which will be the focus in this section.

Facebook was released in 2004 as a social media that encouraged university students to network with existing circles of friends through a profile page called "the wall" (like the university facebook) and user created groups and events (for its early development see, for instance, Kirkpatrick, 2010; Raynes-Goldie, 2012). In 2005 Facebook added the popular photo album feature that allowed users to tag friends on photos and thereby connect a new data point to another user's profile (Raynes-Goldie, 2012). In 2006 the public version of Facebook became available to everyone, including companies, and Facebook expanded its data collecting strategy by introducing new functionalities such as status updates, Newsfeed (stories sorted by the EdgeRank algorithm), and the share button (Gerlitz & Helmond,

2013; Bucher, 2012). In 2007 Facebook made an advertising alliance with Microsoft, which became “the exclusive third-party advertising partner for Facebook” and after the alliance owned 1.3% of the shares (Raynes-Goldie, 2012; Microsoft, 2007). The same year Facebook made an aggressive move in its data strategy with the introduction of especially Facebook for iPhone, Pages, Facebook Platform, and later Facebook Ads and Pages (Raynes-Goldie, 2012; Bodle, 2011b).

From a data strategic perspective, Facebook Platform was of particular importance, as this initiative expanded the Facebook data network into thousands of external sites collecting usage patterns outside the domain of Facebook (Kharif, 2007; Bodle, 2011b). Building on a beta version from 2006, Facebook Platform consisted of integration with ten data points: Canvas, profile box, profile action, News feed, Mini-feed, left navigation, Request, Notifications, emails, messages, and share (Facebook, 2007). Today the ten integrative data points have grown to an incredible number of data points that can be integrated in different solutions through the Facebook Graph API; as Facebook Apps, such as FarmVille, Pet Society, Facebook Quizzes, and myBirthdayCalendar; as Facebook Connect (2008), which lets external websites use Facebook login data as authentication and at the same time transport activity data back to Facebook; as plugins (Roosendaal, 2012) on external websites, such as like-buttons or instant personalization (2010); or as external mobile apps on Android or IOS, using live data from Facebook in return for sending activity data back to Facebook (Bodle, 2011b).

Achievement (Instance)	Album	Application
Check in	Comment	Domain
Errors	Event	Friend List
Group	Insights	Link
Message	Note	Offer
Order	Page	Pagination
Payment	Photo	Pictures
Post	Privacy Parameter	Publishing
Question	Question Option	Real-time Updates
Review	Search	Selecting Results
Status message	Thread	User
Video		

*Illustration 3: Facebook has 34 objects available for integration as of April 2013. Every object has subcategories of data that can be used to integrate with. For instance, the object User has 39 sub fields that in total give developers an incredible amount of data points (Developers.facebook.com).*

Using the valuable extensive data collection, Facebook began to generate positive “cash-flow” in 2009 (Carlson, 2009). Since then Facebook has acquired important companies, such as, the social networking aggregator FriendFeed (2009), the friend finder service Octazen (2010), Friendster patents (2010), the check-in service Hot Potato (2010), the location based service Gowalla (2011), the photo sharing service Instagram (2012), and the face recognition platform Face.com (2012). Apart from Instagram, all services and teams have been used to build and strengthen features on Facebook, such as Facebook Places (2010).

In 2012 Facebook was registered on Nasdaq and recently the company has added features that go in direct competition with, for instance, Google. In 2012 Facebook released App Center, an equivalent to App Store (Apple), with apps that connect to Facebook (Scott, 2012). In 2013 Facebook released the Graph Search feature in collaboration with Microsoft Bing, which was an attempt to create a social search engine that allows users to search on specific queries involving the activities and likes of friends. This is an important feature from a data strategic perspective as it adds behavior that before was primarily reserved to the Google domain. Another aggressive data strategy in 2013 was the introduction of Facebook Home as an integrated feature to HTC smartphones (Svensson, 2013). Facebook Home is integrated in the operating system and thus has access to collect smartphone data that has otherwise been off-limits for Facebook (e.g., geo-location data).

Even though all these data points, which have voluntarily been provided by users or collected “behind the scenes” by Facebook, give an extremely accurate profile of the end-user (e.g., Kosinski, Stillwell, & Graepel, 2013), the profiling for advertising purposes is still limited. At the time of writing, it is only possible to profile on specific metrics: Zip code, gender, age, interests, connection, users (or friends of users) of specific pages, apps or events, relationship status, interested in, education, graduation year, and workplace. Furthermore, the advertiser can choose to show the activity of the user’s friend related to the product (e.g., likes or shares).

Facebook has the potential to make fine-grained profiling with the increased amount of data points collected, but it is not at the time of writing reflected in the advertising choices. Compared to for instance Gmail or Facebook Apps, it is not possible to make context (semantic) sensitive advertisements yet despite the before mentioned dual intention to make profiling ads more relevant to users who in turn are more likely to engage with the advertised content.

### **Findings: Differences in data strategies**

The analysis of data points collected through the features and services of Google and Facebook reveals both similarities and differences. Both Facebook and Google have the ambition of becoming data companies that generate different data points across a variety of features and not least external websites. Through tracking services such as Google analytics, AdSense, DoubleClick, Facebook Connect, and Facebook Social Graph, both Facebook

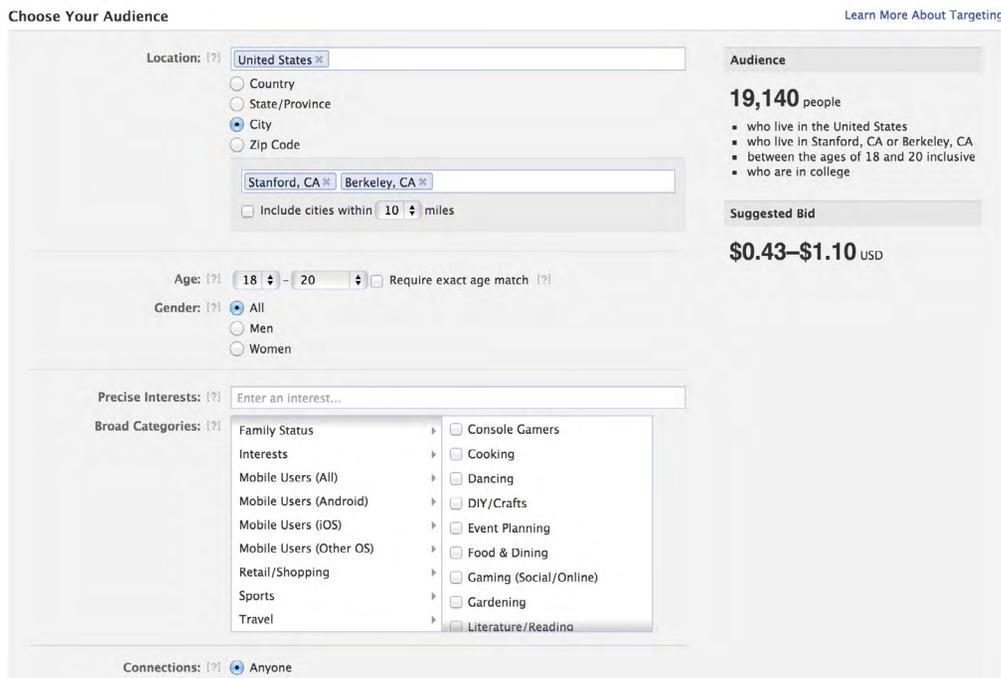


Illustration 4: Examples of the metrics that advertisers can use to target their products on Facebook ad service (<https://www.facebook.com/advertising>).

and Google collect data in an intraoperable way. Also both companies have expanded their data strategies to the operating system level to increase their territory for data collection and they have acquired and collaborated with companies in order to optimize their data collecting features and services. However, inspired by Mintzberg’s general strategic categories (1987) and based on the case descriptions, significant differences in data strategies can be mapped and outlined as well:

Data Strategy	Google	Facebook
Strategic pattern	Horizontal data strategy	Vertical data strategy
Position (inter- vs. intraoperability)	Market leader in Search (intraoperability)	Market leader in social media (intraoperability)
Perspective	Relevance	Reputation
Plan & Ploy	Acquiring mostly <i>services</i> and patents, <i>integrated</i> API, focusing on <i>touch points</i>	Acquiring mostly <i>features</i> and patents, <i>diversified</i> API, focusing on <i>social networks</i>

Table 1. Differences in data strategies

To summarize the main difference, the terminology *horizontal* (Google) and *vertical* data strategies (Facebook) are used. Google has a much more aggressive approach when it comes to diversified data points. Google focuses on different usage situations such as work (Drive,

Calendar), socializing (Google+), and video watching (YouTube & Google TV). In contrast to Facebook, this makes Google capable of collecting data points across a user's everyday doings thereby offering a more complex set of digital footprints, if Google manages to keep the user locked into its portfolio and touch points. Facebook on the other hand collects data points related to socializing as the overarching topic of the data points and collaborates with especially Microsoft to deliver data from other products (e.g., integrating with Bing Search and Bing Map). One could argue that this is a weak data point strategy, but the strategy can be strong if they are able to expand the use of the Facebook platform for different purposes and thereby generate data from a variety of everyday usage situations (e.g., through the use of Facebook Connect and Social Graph integration). Secondly, Google also includes hardware solutions in its product portfolio and possible data point collections such as Google driverless car, Google Glass, Chromebook, and different versions of Nexus. In comparison, Facebook only integrates with the operating system of some smartphones through Facebook Home. This again adds to Google's diversified data points, but it also makes the company weaker in terms of the anti-trust claims that affected particularly Microsoft and Apple. It has a decisive effect on the data points collected and the type of profiling available to advertisers, developers, and government agencies. A horizontal data strategy allows for user profiling and targeting according to user needs throughout a daily routine and daily touch points, whereas a vertical data strategy allows for profiling and targeting according to specific needs and routines. At the same time, Facebook breaks this strategy as an ideal-type by using intraoperable solutions and incorporates user behavior from external touch points as well (Facebook connect, social graph).

Secondly, another important difference is in terms of the strategic positions of Google and Facebook. Whereas Google is market leader in Search, Facebook is market leader in social media. This provides an important legal difference in Facebook's closed and Google's semi-closed product portfolio. Facebook demands user authentication in order to give access to Facebook features. On the other hand, Google mixes public non-authenticated services such as Search and Map with services that demand authentication (Gmail, Hangout, and Google+). Despite the shared privacy policy on all services, this means that it is difficult to use the data points across services in order to make and sell the valuable profiling; the data points simply become less valuable because they cannot be cross-correlated. Given the fact that the public Search still drives Google's revenue, the traffic and data gathered here is still isolated legally for mining with closed services such as Gmail, because informed consent has not been collected.

Thirdly, Google and Facebook have different strategic perspectives. Whereas Google has made "Relevancy" the corporate brand and overall strategic benchmark, Facebook focuses on social connections and reputation. In terms of data points, Facebook thereby has a social bias compared to Google's functional one.

Fourthly, the difference in API structures sketched out in the cases, with Google having a diversified developer approach and Facebook with an integrative approach, adds to the

strategic plan of collecting data points in product silos for Google and as a data pool for Facebook. Facebook creates features to monitor user actions in networks whereas Google creates services to monitor user actions in functional touch points.

In the last section, the data strategic analysis will be used as a stepping stone to discuss the societal implications for the identified stakeholders in a political economy perspective.

## **Discussion: Profiling for different stakeholders**

In a political economy perspective intraoperability as data phenomena pose a threat and risk to our society, because personal data are stored in few services that are owned and controlled by commercial interests. This means that regardless of motives everyone who is able to pay for it can access a large amount of the total end-user online footprints through few services (APIs). This uneven power relationship among subordinate users and third party companies and superior data companies calls for detailed discussion. As the analysis shows, differences occur in the data strategy of different intraoperable services and may lead to different societal implications in terms of profiling.

### ***Profiling for advertisers***

In terms of advertising, profiling equals personalized ads based on a predictive behavioral targeting as a result of fine-grained segmentation (Dwyer, 2009). This fine-grained segmentation is not only based on demographics, but also as shown regarding behavior and participation patterns, and differs in the Facebook and Google cases. Google validates “relevance” of the ad to users through Quality Score, Facebook seems more concerned with the reputation of the account (Carswell, 2011). The horizontal data strategy provides advertisers with easier access to the everyday rhythms and doings of an individual if the person uses the entire Google product portfolio and provides advertisers the opportunity to track users very precisely in the desirable touch point. On the other hand, Facebook’s vertical data strategy provides advertisers with a strong and unique knowledge of the social interaction and network of friends that potentially can enhance brand value and reputation and furthermore strengthen the buying incentives due to social referral.

However, from a political economy perspective, the horizontal data strategy poses challenges to transparency, because users deliver more data points than intended; either because they forget the vast amount of services used with one company or because they are unfamiliar with the complex network of tracking and monitoring facilities, spanning from sensors to digital footprints left behind, when interacting with the services and usage of these data (Tavani, 1999). In different ways Facebook provides the same “creep factor” (Turow, 2013). Facebook users have collected a massive amount of personal likes, shares, messages, and photos, and tagged the data into an identifiable personal network of friends and acquaintances. Even though limited to one service, the participation is so excessive that users equally may lose track of the data points and the exposure of their friends, and

the connections they make when interacting. It is therefore impossible to control the targeting once the user starts using Facebook as a service, because the user is exposed through friends. In both cases, the predictive behavioral (and semantic) targeting may pose implications to society in terms of both social cohesion and privacy. Social segregation is made through price targeting and segmented product delivery according to those who are willing and able to pay and those who are not, or those who are interesting (brand-wise) customers or not. In the case of Facebook, privacy is disturbed when a user's name is used in ads because a friend liked a product on Facebook. In the non-personalized settings such as Google Search, privacy can be invaded when personalized ads are shared with the wrong individual. This was the case when a father found out that his daughter was pregnant through ads from Target (Duhigg, 2012). This case suggests that future studies on advertising and profiling not only should focus on the data companies, but on the available data points and possible data mining across intraoperable data companies.

### ***Profiling for developers and companies***

In the discussion on profiling for third party companies (developers), personalization is often the goal. For instance when companies use Google or Facebook's data APIs to customize apps (Bodle, 2011b). Whereas the horizontal service strategy of Google provides developers with silo access to service integration in which developers still mainly engage with products (e.g., Maps), Facebook's vertical strategy and integrated API structure provide developers with a more profound possibility for personal customization. This personalization cannot take place without the informed consent of users, but statistics and qualitative studies show that users do not read the agreements and therefore do not actually consent (e.g., Bechmann, 2013a). Informed consent as click-wrap agreements therefore becomes a legal paradox that needs to be handled for instance by removing informed consent possibilities from the regulation on privacy altogether.

However, profiling for third party companies is even more important when it occurs as selection or de-selection in the name of risk assessment and management. Again it is difficult to retrieve behavioral patterns from Google's diversified API infrastructure, but straightforward with Facebook's integrated API. Social media intraoperable data hubs such as Facebook provide companies with background data that can be utilized to predict risk, for instance, in hiring, insurance pricing, and bank loans (Quittner, 2012). Predicting risk from our networks of friends and our everyday participation may again threaten the cohesion of society and lead to unwanted social segregation. Furthermore, risk assessment creates an ethical paradox in which users are unaware of the assessment and therefore do not have the ability to act accordingly or adapt to the rules and values of the algorithms, and if this knowledge is obtained, the data value in risk assessment profiling will be decreased. If only some users obtain knowledge of this kind of predictive data mining, there will be imminent risks of creating a very profound digital divide in our society.

### ***Profiling for government agencies***

Despite no immediate financial gain for data companies, government agencies can legally gain access to data from the intraoperable data companies in the name of national security and, at the time of writing, Prism by NSA is an example of an initiative that tries to make predictive profiling to fight terrorism and other national threats. As government agencies access raw log-data across data companies, differences in product portfolios become more important than differences in strategy. Potentially, in the case of Facebook and Google, government agencies circumvent APIs and they will get access to both vertical depth and the horizontal breadth. In other words, through clickstreams and user networking, participation, and data sharing, they are in theory able to analyze the whole spectrum of interactions. Profiling in this case is not a question of personalization, but again a matter of risk assessment and selection of potential dangerous or criminal individuals (Dempsey & Flint, 2004). Every user of the internet becomes a suspect compared to earlier methods, where only prior convicted individuals were screened for crimes that had already taken place (Rubenstein, Lee, & Schwartz, 2008). From a political economy perspective, such methods build on the idea that terrorist planning creates data patterns and crime can be predicted on the basis of standardized data behavior and, for instance, semantic values. As Solove (2008) writes: "data mining programs are often not visible to the public to quell much fear. Instead, their benefits come primarily from their actual effectiveness in reducing terrorist threats, which remains highly speculative." (p. 352) If the Prism algorithm had a maximum value of three mentions of bombing on our Facebook, Youtube, and Gmail accounts before we would be registered as a suspect and denied entry in the U.S., we would adjust our actions accordingly, without knowing if the algorithm proved effective in actually catching terrorists. Apart from compromising freedom of expression, algorithms as government methods highlight deviant behavior as something noticeable and bad. Minorities stand out immediately (Jernigan & Mistree, 2009) and algorithms leave little room for contextual sensitivity (Nissenbaum, 2010), which would take into account the individual in the system. The filter bubble has far bigger societal reach than in the critique of Google Search (Elgesam, 2008; Halavais, 2010). Political economy scholars could argue that we are creating an intolerant society at the expense of the individual.

### **Conclusion**

This article has provided an example of how media and communication studies can use media economics and management to identify differences in data strategies that, apart from the analytical value in itself, can enrich a detailed political economic discussion on profiling. The data point analysis shows significant differences despite the obvious similarities between Google and Facebook as data companies. The key concepts that summarized these differences were the distinction between vertical and horizontal data strategies,

social network versus touch points, integrated versus diversified API structures, and reputation versus relevance as guiding data strategic perspective.

Furthermore, the article suggests distinguishing between profiling for advertisers, third party companies or developers, and government agencies in order to encompass the societal implications of user profiling from a political economy perspective. The article argues that the data strategic differences pose different societal implications in the profiling for different stakeholders. Whereas, for instance, Facebook threatens individual privacy in exposure through friends, Google threatens privacy by individualizing non-individual services. However, when it comes to government agencies, differences in data strategies play a minor role because in theory they can run algorithms on raw log-data across companies. The societal implications therefore become larger than the role of the specific data company.

When we look at the conclusion of this article from a broader perspective, the differences in profiling and data strategies outlined here between the two most visited internet companies stress the importance of not generalizing on “big data” and profiling on the internet. Instead, we have to take into consideration the different versions and aspects of data strategies and profiling, for instance, in privacy discussions in order to address the complexity of profiling. Further work has to be done, among others, within media and communication studies in order to improve our understanding of this complexity grounded in detailed analysis and in this connection we need to constantly improve our methods to provide a more solid empirical basis.

## Acknowledgements

The research was conducted with funding from Digital Humanities Lab in Denmark, and the author would like to thank anonymous reviewers, as well as former Google employee Mia Jung, and my colleague in the Digital Footprints Research Group, Peter Vahlstrup, for helpful discussions and comments on earlier versions of this article.

## References

- Albrechtslund, A. (2008). Online social networking as participatory surveillance, *First Monday*, 13(3).
- Anderson, S.P. (2012). Advertising on the Internet. In Peitz, M. & Waldfoegel, J. (Eds.) *The Oxford Handbook of the Digital Economy* (pp. 355-396). New York: Oxford University Press
- Barnes, S. (2006). A privacy paradox: Social networking in the United States. *First Monday*, 11(9)
- Battelle, J. (2005). *The Search: How Google and its rivals rewrote the rules of business and transformed our culture*. Boston, Mass., London: Nicholas Brealey.
- Bechmann, A. (2013a). Non-informed consent cultures: Privacy policies and app contracts on facebook, *Journal of Media Business Studies*, 10(4),1-20.

## Article: Internet profiling

- Bechmann, A. (2013b). Managing the interoperable digital self, Nordmedia13, Oslo, Norway, 8 August – 11 August 2013.
- Bechmann, A. & Lomborg, S. (2012). Mapping actor roles in social media: different perspectives on value creation in theories on user participation, *New Media & Society*, 15(5), 765-781.
- Bechmann, A. (2009). *Crossmedia: Innovation Networks for Traditional Media Organizations* (in Danish), PhD Thesis, Department of Information and Media Studies, Aarhus University.
- Bechmann, A. & Lomborg, S. (in review). Open APIs as a method for data collection on social media.
- Bodle, R. (2011a). Privacy and participation in the cloud: ethical implications of google's privacy practices and public communications. In K. German, & B.E. Drushel (Ed.), *The Ethics of Emerging Media: Information, Social Norms, and New Media Technology* (pp. 155-174). New York: The Continuum International Publishing Group.
- Bodle, R. (2011b). Regimes of sharing, *Information, Communication & Society*, 14(3), 320-337.
- Boyd, D. & Hargittai, E. (2010). Facebook privacy settings: Who cares? *First Monday*, 15(8).
- Boyd, D. (2007). Why youth (heart) social network sites: The role of networked publics in teenage social life. In Buckingham, D. (Ed.) *Youth, Identity, and Digital Media*. Cambridge, MA: MIT Press.
- Brin, S. & Page, L. (1998). The PageRank citation ranking: Bringing order to the web. *Technical Report, Stanford InfoLab*, <http://ilpubs.stanford.edu:8090/422/1/1999-66.pdf> (retrieved April 30, 2013).
- Bucher, T. (2012). Want to be on the top? Algorithmic power and the threat of invisibility on Facebook. *New Media & Society*, 14(7), 1164-1180
- Capurro, R. (2005). Privacy. An intercultural perspective. *Ethics and Information Technology*, 7, 37-47.
- Carlson, N. (2009). Facebook: Cash-Flow Positive With 300 Million Users, Retrieved 30 April, 2013, from <http://www.businessinsider.com/facebook-cash-flow-positive-with-300-million-users-2009-9>.
- Carswell, G. (2011). Is there a Quality Score for Facebook Ads? Retrieved 3 May, 2013, from <http://ppcblog.com/facebook-quality-score/>.
- Castells, M. & Arsenault, A. (2008). The Structure and Dynamics of Global Multi-media Business Networks, *International Journal of Communication*, 2, 707-748.
- Castells, M. (2009). *Communication Power*, Oxford: Oxford University Press.
- van Couvering, E. (2003) Media Power on the Internet: Towards a Theoretical Framework. Paper presented at the Research seminar for media communication and culture, London School of Economics, London, UK, 25 April 2003.
- Dempsey, J.X. & Flint, L.M. (2004). Commercial data and national security, *The George Washington Law Review*, 72 (6), 1459-1502.
- Duhigg, C. (2012). How companies learn your secrets. Retrieved 16 February, 2013, from <http://www.nytimes.com/2012/02/19/magazine/shopping-habits.html?ref=magazine>.
- Dwyer, C. (2009). Behavioral targeting: a case study of consumer tracking on levis.com. Paper presented at the Fifteenth Americas Conference on Information Systems, San Francisco, California 6 August – 9 August, 2009.
- Ebersman, D. A. (2012). Registration statement under the Securities Act of 1933 Facebook Inc., *United States Securities and Exchange Commission*, Washington, Retrieved 17 April, 2013, from <http://www.sec.gov/Archives/edgar/data/1326801/000119312512034517/d287954ds1.htm>.
- Elgesam, D. (2008). Search engines and the public use of reason, *Ethics and Information Technology*, 10, 233-242.
- Ess, C. (2009). *Digital Media Ethics*, Cambridge: Polity Press.
- Facebook (2007). Facebook Platform launches, Retrieved 30 April, 2013, from <http://web.archive.org/web/20110522075406/http://developers.facebook.com/blog/post/21>.

- Fernback, J. & Papacharissi, Z. (2007). Online privacy as legal safeguard: The relationship among consumer, online portal, and privacy policies. *New Media & Society*, 9(5), 715-734.
- Foucault, M. (1995). *Discipline and Punish*, New York: Vintage Books.
- Fuchs, C. (2011). A contribution to the critique of the political economy of Google, *Fast Capitalism*, 8(1). Retrieved 17 April, 2013, from [http://www.uta.edu/huma/agger/fastcapitalism/8\\_1/fuchs8\\_1](http://www.uta.edu/huma/agger/fastcapitalism/8_1/fuchs8_1).
- Gerlitz, C. & Helmond, A. (2013). The Like Economy: Social Buttons and the Data-intensive Web, *New Media & Society*, online first: February 4, doi:10.1177/1461444812472322.
- Gross, R. & Acquisti, A. (2005). "Information revelation and privacy in online social networks," In: *Proceedings of the 2005 ACM Workshop on Privacy in the Electronic Society* (pp. 71-80). New York: ACM.
- Halavais, A. (2009). *Search Engine Society*, Polity Press, Cambridge.
- Jarvis, J. (2009). *What would google do?* New York: Harper Collins.
- Jenkins, H. (2006). *Convergence Culture: Where old and new media collide*. New York: New York University Press.
- Jernigan C. & Mistree B.F. (2009) Gaydar: Facebook friendships expose sexual orientation. *First Monday*, 14(10).
- Kelly, M. (2012). 96 percent of Google's revenue is advertising, who buys it? Retrieved 29 January, 2013, from <http://venturebeat.com/2012/01/29/google-advertising/>.
- Kharif, O. (2007). Social-networking sites open up. Retrieved 30 April, 2013 from <http://www.businessweek.com/stories/2007-02-13/social-networking-sites-open-upbusinessweek-business-news-stock-market-and-financial-advice>.
- Kirkpatrick, D. (2010). *The Facebook effect: the inside story of the company that is connecting the world*. New York: Simon & Schuster.
- Kosinski, M., Stillwell, D. & Graepel, T. (2013). Private traits and attributes are predictable from digital records of human behavior, *PNAS*, online first. Retrieved 11 March, 2013, from <http://doi:10.1073/pnas.1218772110>.
- Lyon, D. (2001). *Surveillance society: monitoring everyday life*. Buckingham: Open University Press.
- Mansell, R. (2004). Political economy, power and new media, *New Media & Society*, 6(1), 96-105.
- Marwick, A. and Boyd, D. (2011). Social privacy in networked publics: teens' attitudes, practices, and strategies. Paper presented at Oxford Internet Institute's A decade in internet time, 22 September, 2011.
- McGraw, G. & Felton, E.W. (1999). *Securing Java: Getting Down to Business with Mobile Code*. New York: John Wiley & Sons.
- Microsoft, (2007). Facebook and Microsoft expand strategic alliance. Retrieved 30 April, 2013, from <http://www.microsoft.com/en-us/news/press/2007/oct07/10-24FacebookPR.aspx>.
- Millar, J. (2009). Core privacy: a problem for predictive data mining. In Kerr, I., Steeves, V. & Lucock, C. (Eds.) *Lessons from the identity trail: Anonymity, Privacy and Identity in the Networked Society*. New York: Oxford University Press.
- Miller, D. (2011). *Tales from Facebook*. Cambridge: Polity Press.
- Mintzberg, H. (1987). The strategy concept I: Five ps for strategy. *California Management Review*, 30(1), 11-24.
- Neuhaus, F. & Webmoor, T. (2012). Agile ethics for massified research and visualization, *Information, Communication & Society*, 15(1), 43-65.
- Nissenbaum, H. (2010). *Privacy in Context*, Stanford: Stanford University Press.
- Quittner, J. (2012). Banks to use social media data for loans and pricing. Retrieved 26 January, 2013, from [http://www.americanbanker.com/issues/177\\_18/movenbank-social-media-lending-decisions-bretting-1046083-1.html](http://www.americanbanker.com/issues/177_18/movenbank-social-media-lending-decisions-bretting-1046083-1.html).
- Raynes-Goldie, K. S. (2012). *Privacy in the Age of Facebook: Discourse, Architecture, Consequences*, PhD Dissertation, Curtin University.

- Reitman, R. (2012). What Actually Changed in Google's Privacy Policy. Retrieved 10 November, 2013, from <https://www.eff.org/deeplinks/2012/02/what-actually-changed-google-s-privacy-policy>.
- Rogers, R. (2013). *Digital Methods*, Cambridge, MA: MIT Press.
- Roosendaal, A.P.C (2012). We are all connected to Facebook...by Facebook!, In Gutwirth, S. et al. (Eds.) *European Data Protection: In Good Health?* (pp. 3-19). Heidelberg: Springer
- Rubinstein, I.S., Lee, R.D. & Schwartz, P. M. (2008). Data Mining and internet profiling: Emerging regulatory and technological approaches, *University of Chicago Law Review*, 75, 261-285.
- Scott, C. (2012). Facebook App Center targets mobile users. Retrieved (nd.) from [http://www.computer-world.com/s/article/9227039/Facebook\\_App\\_Center\\_targets\\_mobile\\_users](http://www.computer-world.com/s/article/9227039/Facebook_App_Center_targets_mobile_users).
- Skou, K. (2010). Google from a competitor perspective. Paper presented at IT-City, Aarhus Univeristy, Aarhus, Denmark, 15 October, 2010.
- Solove, D. (2008). Data Mining and the security-liberty debate, *The University of Chicago Law Review*, 75(1), 343-362.
- Stutzman, Gross & Acquisti (2012). Silent listeners: The evolution of privacy and disclosure on facebook. *Journal of Privacy and Confidentiality*, 4(2), 7-41.
- Sutor, R.S. (2011). Software standards, openness, and interoperability. In DeNardis, L. (Ed.) *Opening Standards: The Global Politics of Interoperability* (pp. 209-218). Cambridge, MA: MIT Press
- Taddicken, M. (2012). 'Privacy, surveillance, and self-disclosure in the social web: exploring the user's perspective via focus groups. In Fuchs, C. et al (Eds.) *Internet and Surveillance: The Challenges of Web 2.0 and Social Media* ( pp. 137-145). New York: Routledge.
- Tavani, H.T. (1999). Informational privacy, data mining, and the Internet, *Ethics and Information Technology*, 1, 137-145.
- Thunø, L. (2013). *Medieudviklingen 2012*. Copenhagen: DR.
- Turow, J. (2013). *The Daily You: How the new advertising industry is defining your identity and your worth*, New Haven: Yale University Press.
- Vise, D. & Malseed, M. (2005). *The Google Story*. London: MacMillan.
- Wasko, J., Murdock, G. & Sousa, H. (2011). Introduction: The political economy of communications: Core concerns and issues. In Wasko, J., Murdock, G. & Sousa, H. (Eds.) *The Handbook of Political Economy and Communications* (pp. 1-10). West Sussex, Blackwell

Anja Bechmann  
Associate Professor, PhD  
Department of Aesthetics and Communication  
Aarhus University, Denmark  
anjabechmann@dac.au.dk