Anja Bechmann, Kristine Bilgrav-Nielsen
& Anne-Louise Korsgaard Jensen

Data as a Revenue Model

Sharewall as a Payment Method and Editorial Algorithm in the News Industry

Does data solve the crisis in legacy news companies? This article discusses data as a revenue model and the use of editorial algorithms to curate content and still meet public values. Furthermore, the article criticizes the news companies for using data in traditional advertisement revenue models, which have proved difficult to uphold. Instead we need to focus on public values along with micro segment data in what are here termed *social responsible algorithms*. We also need to continue the discussion on the very concept of news and to experiment with news 'packaging' that are not derived from legacy news companies, but from born-digital ones.

Companies such as Amazon, Apple, Google and Facebook all collect user data and turn it into a commodity that can be sold either as data-enriched advertisement spaces or as data units that can enrich product developments and customer relations for third-party companies (also called data broker industry). Following this trend, Mastercard registered US$ 341 million as 'other profit' in the first quarter of 2014. The label 'other profit' includes data sales and user transactions in all forms, creating large amounts of data that are extremely valuable (Thomasson 2014; Mayer-Schönberger & Cukier 2013). This systematic and automated data collection for user-created news companies such as Facebook and Twitter is spreading in the legacy news media industry as well. *The Guardian* and *Financial Times* prompt users to register before reading news, but how exactly does data translate into rev-

enue for the news industry and what is the value of data?

The aim of this article is to analyze and discuss data as a revenue model for the news industry, understood broadly as companies that produce or facilitate content that relates to new information for a community (be it niche, state or society at large). In doing so, the article will focus on legacy news companies by taking a closer look at a widely used data software called Sharewall that facilitates data collection for content providers. The article exemplifies the use of this software by the first legacy news company in Denmark to adapt this software in a 'permission wall' (instead of paywall). The article thereby seeks to shed light upon and discuss the mechanisms and implications of data as a revenue model in greater detail.

Automated and systematic data collection and enrichment

The social media news industry and the legacy news media industry both have strong traditions for using classic dual products (Picard

Anja Bechmann, Associate Professor at Department of Media and Journalism Studies, Aarhus University, **Kristine Bilgrav-Nielsen** and **Anne-Louise Korsgaard Jensen** are master students.

1989) or three-way-market transaction models (Anderson 2008) in which the companies deliver content to the user (company–user), the companies sell advertising space (company–advertiser) and the advertiser exposes the user for products (advertiser–user). The difference is that data collected earlier was executed as 'prompts' that segmented readers, viewers and listeners on the basis of persona (Cooper 2004; Dahl 1997), surveys or focus group interviews in comparison to the systematic and automated collection of log data and self-reported data that is collected now. This systematic approach to data collection provides a more detailed understanding of the individual user and the context in which the user consumes and creates news. The detailed knowledge allows for *micro segmentation*. Especially as data enrichment is conducted as 'pooling' data about the user from different data sources (e.g., customer databases, log file data and social media accounts).

Sharewall as a payment method and 'editorial' algorithm

Sharewall functions as a permission wall instead of a paywall where users agree to share the news (e.g., from legacy media) through their social network of friends in return for consuming the news in what they call "social sharing as currency" (www.sharewall.co.uk). Furthermore, the user permits the company to collect different data about the user both through social media or emails and by using their own tracking technology. The software flow can be illustrated in three steps:

Step 1: Informed consent

Step 2: Data collection, processing and learning, curating

Step 3: Revenue from data-enriched and targeted advertisement.

The permission takes place as an informed consent 'contract' (see also Bechmann 2014) with the user (step 1) and is followed by data collection (e.g., through a Facebook app and tracking scripts). Based on the data collected,

the Sharewall algorithm 'packages' a news product adjusted to the specific user profile in question (step 2) and serves this to the user (Step 3) along with targeted advertisements on different digital platforms (e.g., websites, apps and newsletters).

According to the CEO of Sharewall, Anders Ibsen (interview conducted in the autumn of 2014), the machine learning algorithm makes calculations based on stochastic variables: random calculations where the outcome is unknown. Thus, it is possible to establish standard deviation in the calculations, expressed by how much the stochastic variable is distributed around the mean or, in other words, the average calculation based on data from Facebook and the script. Standard deviations will always be present, and that is why the user will not always receive 100 per cent relevant content. Another reason is that 25 per cent serendipity is built into the algorithm as a 'randomness generator' – which means that not all content is personalized. Serendipity is 'the art of making an unsought finding' (Corneli, Pease, Colton, Jordanous and Guckelsberger 2014, p. 4) and, therefore, the user will be introduced to content that is not seen as 100 per cent relevant according to the calculations of Sharewall's machine learning engine.

There are more reasons that 25 per cent serendipity is included. First, there is a possibility that content presented to the user could capture his or her interest and thereby keep the user at the content provider (be it Netflix or news apps) for a longer time. Second, filter bubbles or echo chambers must be avoided in the sense that the user is only introduced to the content that his or her behavior has defined. Therefore, serendipity is included to 'challenge your field of interest', as Ibsen describes in the interview.

Sharewall's machine learning engine increasingly becomes smarter as more data about the user and the user behavior has been absorbed and analyzed. Sharewall's machine learning engine does not give personal recommendations but divides users into *micro segments*. A micro segment can consist of

three to four users. The fact that about 40,000 micro segments are defined means that the individual user is introduced to content that is very close to being personal, according to Ibsen. Many different parameters decide which segment the individual user is placed in. Everything is based on calculations that decide how a user can be categorized and what the different micro segments are interested in. This allows for tailored advertisement space that can be sold at a higher price than more generic advertisement space because the probability of sale is higher, as we see in the case of Google and Facebook. In the next section we will exemplify the data as a revenue model through the use of Sharewall in a legacy news provider in Denmark.

The case of Mx and Sharewall

As a part of the Nordic Media Model (Syvertsen 2015, Hallin & Mancini 2004), historically, Denmark has a strong news industry with four leading newspapers and one free national weekly newspaper (Northmedia 2016, Drotner 2011, p. 83). Mx was introduced as a free daily newspaper on the Danish market in 2001 as part of the Swedish media group Metro International and was from the beginning a strong competitor to the existing free newspaper. During the first decade of the 21st century other free daily newspapers were launched, but in 2016 Mx was the only free daily newspaper left. In 2012 it became public that the Swiss media group Tamedia AG now owned all the shares. One year after the change of ownership, Mx underwent a major transformation where both the format and target group were changed. The target group is now 19–39-year-olds (Danish Media 2013). Mx is the most widely read daily printed newspaper in Denmark. It is published on all weekdays and has a circulation of about 330,000 papers.

In 2014, as one of the first to experiment with permission walls in Denmark, Mx used Sharewall to target content and advertisements online at mx.dk and in newsletters emailed to the users. The user needed to permit such targeted advertisements or permission market-

ing (Kumar, Zhang & Luo 2014, Godin 2007) through informed consent. In connection with the informed consent request at mx.dk the data sharing was provided as email addresses only, and Twitter or Facebook accounts. The Facebook account provides Mx with the most valuable data. Through a Facebook app, Mx gains access to email, name, profile picture, age, gender, language, country, network of friends and birthday. When consenting to the newsletter, the user further consents to the use of "tracking technology" in both newsletters and other digital communications.

According to Simon Stilling (editor in chief at Mx, interview conducted in autumn 2014), in contrast to GemiusAudience and Google Analytics cookie data, Facebook data "provides deeper and more accurate information about our users". A cookie registers user behavior patterns online. This includes which functions or links have been visited, what has been read, and the duration of time a user spends on one page or article. Often, a cookie is placed both on the homepage and on each subpage that the user visits (Danish Competition and Consumer Authority 2011, p. 15). Apart from first-party cookies (Mx's own cookies) and the third-party cookies of GemiusAudience and Google Analytics, the user also meets cookies from YouTube, Facebook, Twitter, four advertisement networks and Sharewall (Mx, Cookie and Privacy Policy). According to Anders Ibsen, Sharewall "collects through a script on the publisher's website", which submits behavioral data from the user sessions and, in comparison to other cookies, also tracks a user across devices such as smartphones, laptops and tablets. By pairing these data with other data sources such as Facebook, Mx (through Sharewall) is able to understand when users use different devices and what they read on these devices. These different data sources make Mx able to curate email newsletters and online content tailored to user preferences and past behaviors through the machine learning algorithm of Sharewall. However, according to Thomas Raun, it takes a couple of weeks for the algorithm to process and learn from the user data

in order to generate a tailored pattern based on the 'best guess'-principle. According to Thomas Raun (interview conducted in the autumn of 2014), Mx has chosen not to use the function of social sharing offered by Sharewall in which the user can choose to 'pay' for news by letting Mx share news articles on the user's Facebook wall. This in turn means that Mx does not make full use of Facebook as a social referral channel.

How does tailored content and advertisements convert into profit for Mx then? The revenue model is still the classic advertisement model, but the probability of the user reading or interacting with the content and the advertisements is higher in data-enriched communication (Penn 2012, p. 601). Instead of users typing in their interest areas, the content is curated through the different data inputs from mx.dk and Facebook processed by the Sharewall algorithm. In comparison with earlier models, Mx freely disposes of the advertisement space and does not use advertisement networks on mx.dk. The advertisement network percentage of the revenue stays with Mx and thereby increases revenue.

To illustrate the potential revenue of data we can detail a scenario of estimated revenue. According to Thomas Raun in 2014, the Mx newsletter *potentially* "realizes a CPM on 80 DKK". CPM stands for Cost Per Mille and describes the price for the advertiser of exposing an advertisement 1000 times. In estimates, one exposure costs 0.08 DKK and the newsletter contains three advertisements (0.24 DKK per newsletter). The letter is sent out two times daily and the revenue per year per user is 175.2 DKK. Thomas Raun estimates the number of permissions between 50,000 and 100,000. If the number is 50,000, we can estimate a revenue per year of 8.76 mio. DKK. If the number is 100,000, the estimated revenue per year on the newsletter is 17.52 mio. DKK.

CPM potentially increases with the number of permissions because more segmented advertisements are possible. The price for advertisements in the newsletter is higher than on mx.dk because the user click rate and con-

centration is higher in the newsletter, according to the chief of sales at Mx, Kenneth Madsen. Even though the income for the newsletter potentially is between 8–17 mio DKK, the value per permission is between 10–15 DKK. At 100,000 permissions, the profit per year will potentially be 1.25 mio. DKK in this estimated scenario. Additionally, the newsletter creates referrals to mx.dk when users click on content and this action increases advertisement sales on mx.dk as well. In contrast to the newsletter, advertisement sales on mx.dk happen through advertisement networks (Doubleclick, specific Media, Google Ads and Xaxis) that specialize in behavioral targeting and facilitate contact between advertiser and mx.dk. Still, Sharewall functions in the same way on all content distribution channels, but the revenue that Mx generates on mx.dk comes from the exposure of the advertisement network's targeted advertisements measured in CPM or CPC (Cost Per Click).

However, data-enriched curating of content also has potential pitfalls. Jolley, Lee, Mizerski and Sadeque (2012) emphasize, "personalized salutations tend to increase opt-out (unsubscribe) rates" (p. 2). Also, studies show that unsubscribing will increase if the content immediately seems irrelevant to the user (Kumar, Zhang & Luo 2014, p. 405). Therefore, the week-long duration from data collection to tailored output in the Sharewall algorithm can be a challenge to content providers such as Mx, because users, according to Kumar et al., will tend to unsubscribe if they experience irrelevant and untailored content.

Mx chose to close the data experiment and the collaboration with Sharewall due to lack of sales on the advertisements and commercial content in the newsletter (according to email correspondence with Thomas Raun and Anders Ibsen, spring 2016). Another reason, according to Anders Ibsen, was that Mx did not believe that the value of the newsletter could offset the decline in traffic they had as a result of the wall that blocked content until the users had signed up.

Therefore, the estimated revenue and the data model were not proved, even though

other publishers have succeeded with Sharewall and the data revenue model. Nevertheless, the case study generically illustrates the principles of data as a revenue model, and the next section will discuss the implications of such a model.

Discussion: Does data solve the crisis in legacy news companies?

A data revenue model such as the use of Sharewall by Mx is still a three-way or dual market model that relies on advertising sales. However, in the data model the user pays not only through exposure to advertisements (eyeball economy) or with the data collected at the specific news producer, but she also pays with external data from, for instance, Facebook or Twitter (see table below). We still have a dual market model, but we also have a dual user payment model that does not rely on advertisement and subscriptions, but on advertisement and data (data being converted to advertisements or being sold to a third party).

Does data solve the crisis of the legacy news industry then? Since the analyzed data model still relies on advertising as revenue, the model is vulnerable to the crisis in online advertisement. The 'crisis' is present internationally due to ad-blocking technologies and the smaller interfaces of the mobile platforms. The crisis is also present on the national and regional markets due to strong international companies such as Google and Facebook that succeed in drawing traffic away from national, regional and local players. If advertisements do not travel with content in a way that is un-

breakable to ad-blocking, such a data revenue model is poor.

This calls for an even more important discussion that the news industry fails to engage in. Do users need news from legacy media in the 21st century? What is news, what is it worth and how can technology support this need? From an outsider's point of view, the legacy news industry can look like advanced online supermarkets with high quality products without a good revenue model. These products lure users in just to be exposed to good revenue model products of often low relevance to the users. Can the revenue model of such low relevance products be even better and can the relevance of the products for the user be improved at the same time? Relevance can be heightened through data-enriched curating of news as well as commercial content and products.

At the moment, most legacy news companies are trimming the organization towards digital news production. The problem is that they use data in the same packaging that they have used for a very long time now and that has proved difficult to generate satisfactory revenue. Packaging is here understood as serving news funded by advertising in organization bound formats (in contrast to Facebook's cross-organizational deep linking newsfeed and Spotify's cross-label assortment). The time is ripe for a discussion about whether, in fact, the packaging is right. And to do so, the industry needs to dare to face the discussion about the definition of news, focusing both internally towards journalists and externally towards politicians (as the legacy

	Pre-Sharewall Business Model	**Sharewall Business Model**
Content	Editorial content on mx.dk	Targeted and tailored content on mx.dk and in newsletter
Transactions	Consumer attention on content in indirect transactions	Consumer attention and consumer data on content in direct and indirect transaction
Revenue	Advertisement revenue from traffic on mx.dk	Advertisement revenue from traffic on mx.dk and newsletter
Data type ..	Data from cookies	Data from Facebook and script

media receives substantial state funding in the nordic countries).

Data is not a revenue model in itself (unless the media want to join the data broker industry solely), just as data is not an organizational trimming model in itself. With Netflix as the classic content case, many news companies are looking at solutions such as Sharewall to explore how data can enrich and guide the production of news and news curating for the user through what has been termed micro segments in this article. A higher degree of personalization potentially creates more users/readers/viewers/listeners and leads to the production of news that matches micro segmented needs. Algorithmic curating can potentially lead to organizational changes in the job profiles of the editors. Nevertheless, leaving the editorial algorithm to present news with 75 per cent personalized content based on prior behavior and interest, and 25 per cent on randomness, seems counter-intuitive for news production.

Journalists have a long and strong tradition in serving society and democracy through public values such as a free and independent press; free speech; responsible, informative and reliable promotion of social cohesion, and active citizenship (Danish Ministry of Culture 2011). The 25 per cent serendipity could easily be replaced with content that secures such ideals and serves *controlled serendipity* and *social responsible algorithms*.

Concluding remarks

The article points to the need for a (re)-definition of news from the users' point of view and not only the journalists'. The scope should not be to survive as a media company; the scope should be to concur markets and users in the name of users' need for a free and independent press, free speech, social cohesion, and active citizenship. In doing so, most media companies are born nationally and users transgress national borders in digital solutions such as Facebook. Facebook strives to show the same public values, but are challenged on precisely that in their cross-national algorithmic curating (e.g., free speech and censorship cases). On the other hand, media organizations are storytelling and curating experts that might be able to transgress old ideas of news formats and news packaging and move towards the design of social responsible storytelling and algorithms.

Digital Society Forum is a research-based network between students, companies, organizations and researchers worldwide interested in Grand Challenges in the digital society. A special emphasis in 2016 is placed on digital revenue models in media organizations, internet companies and startups. The network runs from 2016-2018 and companies, organizations, researchers and students interested in joining please contact network administrator: simoneskaarup@gmail.com. The network is co-funded by the innovation network BrandBase.

References

Anderson, Chris (2008) Free! Why $0.00 is the future of business. *Wired Magazine*. Retrieved from http://www.wired.com/techbiz/it/magazine/16-03/ff_free?currentPage=all.

Bechmann, Anja (2014) Non-informed consent cultures: Privacy policies and app contracts on Facebook. *Journal of Media Business Studies*, *11*(1), 1-18.

Cooper, Alan (2004) *The inmates are running the asylum – Why high-tech products drive us crazy and how to restore the sanity*. New Jersey: Pearson Education.

Corneli, Joseph; Jordanous, Anna; Guckelsberger, Christian; Pease, Alison &Colton, Simon (2014) Modelling serendipity in a computational context. *ArXiv 1*, 1-33.

Dahl, Henrik (1997) *Hvis din nabo var en bil.* København K: Akademisk forlag. [in Danish]

Danish Competition and Consumer Authority (2011). *Adfærdbaseret reklame* på internettet. København. Retrieved from http://dokumenter.forbrug.dk/adfaerdsbaseret_reklame_2011/adfaerdsbaseret_reklame_paa_internettet.pdf [in Danish].

Danish Media (2013) *Metroxpress relanceres – og 24timer lukker.* Retrieved from http://danskemedier.dk/nyhed/metroxpress-relanceres-og-24-timer-lukker/ [in Danish].

Danish Ministry of Culture (2011) *Notat om publicismebegrebet samt afgrænsning af relevante digitale medier* (Bilag til mødet den 8. februar 2011). Retrieved from http://slks.dk/fileadmin/user_upload/dokumenter/medier/Mediestoetteudvalget/Notat_om_publicismebegrebet_samt_afgraensning_af_relevante_digitale_medierx.pdf [in Danish].

Drotner, Kirsten (2011) *Mediehistorier.* Frederiksberg C: Samfundslitteratur [in Danish].

Godin, Seth (2007) *Permission marketing.* London: Simon & Schuster.

Hallin, Daniel C. & Mancini, Paolo (2004) *Comparing media systems: Three models of media and politics.* Cambridge: Cambridge University Press.

Jolley, William; Lee, Alvin; Mizerski, Richard & Sadeque, Saalem (2012) Permission email messages significantly increase gambler retention. *Journal of Business Research*, 66, 1617-1622.

Kumar, V., Zhang, Alan, & Lou, Anita (2014) Modeling customer opt-in and opt-out in a permission-based marketing context. *Journal of Marketing Research 51*(4), 403-419.

Mayer-Schönberger, Viktor & Cukier, Kenneth (2013) *Big data: A revolution that will transform how we live, work and think.* London: John Murray Publishers.

Northmedia (2016) *History.* Retrieved from http://www.northmedia.dk/karriere/sondagsavisen/historie/. [in Danish]

Penn, Joanna (2012) Behavioral advertising: The cryptic hunter and gatherer of the internet. *Federal Communications Law Journal 64*(3), 599-616.

Picard, Robert G. (1989) *Media economics: Concepts and issues.* Newbury Park, CA: Sage.

Syvertsen, Trine; Enli, Gunn; Mjøs, Ole J. & Moe, Hallvard (2015) The media welfare state – nordic media in the digital age. *Norsk Medietidsskrift 21*(2), 1-3.

Thomasson, Emma (2011) Mastercard: Real-time consumer trend data is a huge growth area for us. *Business Insider*. 14-07-11. Retrieved from www.businessinsider.com/r-mastercard-expects-big-growth-from-big-datainsights.