Article

The Infrastructure of News: Negotiating Infrastructural Capture and Autonomy in Data-Driven News Distribution

Lisa Merete Kristensen * and Jannie Møller Hartley

Department of Communication and Arts, Roskilde University, Denmark

* Corresponding author (lisak@ruc.dk)

Abstract
The platformisation of news has triggered public and scholarly concern regarding the impact of platforms on the news industry and, more importantly, platforms' potential threat to ideals of autonomy and economic independence. Despite ongoing debate and the increasing investment in technologies for automated distribution and artificial intelligence, the material infrastructures of the news media sustaining this artificial intelligence-driven news distribution remain understudied. Approaching the infrastructural relationship as spaces of negotiation this article investigates how the news media is negotiating their own autonomy vis-à-vis infrastructure capture by platforms. The analysis is grounded in a mapping of technologies sustaining the production, distribution, and commercial viability of the media. This is further combined with ethnographic observations from two large Danish news organisations and 19 in-depth interviews with news organisations and digital intermediaries from Scandinavia, the US, and the UK. The research shows how infrastructure capture is manifested and negotiated through three overall logics in the infrastructure of news: logics of classification, standardisation, and datafication.

Issue
This article is part of the issue "A Datafied Society: Data Power, Infrastructures, and Regulations" edited by Raul Ferrer-Conill (University of Stavanger / Karlstad University), Helle Sjøvaag (University of Stavanger), and Ragnhild Kr. Olsen (Oslo Metropolitan University).

1. Introduction

At a meeting in the spring of 2022, directors of several departments and two of the key developers at the Danish tabloid *Ekstra Bladet* discussed various taxonomies for describing and categorising news content, for example by topic. They discussed the pros and cons of the industry-standard taxonomy developed for digital marketing by the Interactive Advertising Bureau (IAB) and the taxonomy of the International Press Telecommunications Council (IPTC). Among the more than 700 members of the IAB are Microsoft, Amazon, Nielsen, Spotify, Yahoo, and Twitter, while the IPTC standard is developed for media companies. The editors agreed that some 80% of the categories and sub-

categories were usable but that new categories had to be added, as the taxonomy seemed "overly commercial." Hence, they embarked on the task of adjusting this taxonomy to their own context. They coded thousands of articles, removed categories, added their own, and eventually negotiated an adjusted taxonomy, which was a combination of categories from the IAB, IPTC, and their own categories adjusted to the needs of *Ekstra Bladet* and the specific context of news.

This scene from our fieldwork took place almost 50 years since Tuchman (1973) convincingly showed how journalists categorise the news in order to "routinise the unexpected" in everyday news production. In her seminal sociological study of newsrooms, she showed how news stories were categorised and hierarchised and how

a given news flow (and thus reality, she argued) is socially constructed in journalistic practice. Today, news are still categorised according to internal journalistic criteria and economic news values, but they are also categorised to allow the application of performance metrics and to ensure distribution to increasingly personalised digital news sites, search engines, and social media platforms. Thus, the opening example from our fieldwork illustrates how the news media link themselves to larger infrastructures and thereby adapt to certain logics of platformisation, thus negotiating their own autonomy, norms, and values in the process.

It is no understatement that in the past 15 years, we have seen an intensified datafication of news industries. Most significantly, the distribution of news has experienced a radical change, as the communicative system and the infrastructural conditions of distribution have moved from being operated by media companies themselves, as was the case with the printing press, or by states, such as with much of telecom and postal infrastructures in the Western context (Flensburg, 2020). Due to its complexity and ability to transfer data on a global scale, much of the material infrastructure is owned by large technology companies, resulting in what van Dijck et al. (2018) termed "platform societies." Research has begun to examine how this development affects other spheres of society, for example, by showing how the news media adapt to these logics of datafication by increasingly basing decision-making practices on the algorithmic processing of audience and user data (Christin, 2020; Kristensen, 2021; Petre, 2021).

In this article, our focus is on how these technologies become deeply ingrained into the organisational structure of the news organisation. We argue that it is important to examine the material basis of news production, scrutinising the interdependencies between news media and infrastructures to understand how new logics are entering the processes of media production and distribution (Simon, 2022, p. 1833). As such, they are not simply value-free plug-and-play packets of software but actants with purposes and values built-in (Friedman & Nissenbaum, 1996; Thurman, 2011). They exist in what Poell et al. (2022) call "spaces of negotiation," which means that news media to a varying degree adapt to the inherent logics and audience constructions and make them fit their own values and norms, for example how they perceive the audiences. These "fittings" are important because they also make the values and norms durable, as they become part of the technical systems.

The article first positions our research question in the existing literature, arguing that we need to look closer at how the materialities and technologies of news distribution are implemented, but also negotiated along the way. Next, we present the conceptual-theoretical framework of infrastructure capture (Nechushtai, 2018) and media logics (Altheide & Snow, 1979), which leads to the formulation of our research question. This is followed by a methods section. The first part of the analysis maps the infrastructural elements of news production, news distribution, and commercial viability of news. The second explores how media organisations negotiate power over dominant logics by designing their tech stacks. Building on this, we argue that infrastructure capture is negotiated and manifested through three overall logics: logic of datafication, standardisation, and classification.

2. Literature on the Infrastructures of News Distribution

In recent years, scholars have theorized and examined the increasing dependency between news organisations and the infrastructures supplied by commercial platforms, a trend which has resulted in the "platformisation of the news" (van Dijck et al., 2018, p. 49). Drawing on software studies, political economy, and business studies, Poell et al. (2022, p. 5) argued that platforms can be understood as "data infrastructures that facilitate, aggregate, monetize, and govern interactions between end-users and content and service providers." This definition illustrates that platforms simultaneously operate as multi-sided markets, data infrastructures, and governance frameworks.

Several scholars have addressed how various systems sustaining news production and distribution influence the work of journalists. For example, studies of how audience measurement data impact editorial choices (Anderson, 2011). Furthermore, an increasing number of studies are examining the use of recommender systems by news organisations. This research emphasises that news organisations vary regarding the use of recommenders (Møller, 2022) and contradict public service values, such as universalism (Sørensen, 2022). In addition, how they impact diversity (Neyland & Möllers, 2017) and can be designed to support democratic values (Helberger, 2019). We argue that the literature has to some degree overlooked the material aspects of both metrification and personalisation, in that it involves building complect tech stacks and tech systems inside media organisations. By taking an infrastructure approach, we contribute with new knowledge on how decisions concerning the implementation of these systems are negotiated in a news organisation domain, questioning more broadly how the autonomy of news organisations is negotiated in the implementation of systems for production, distribution, and monetisation. Poell et al. (2022) theorised the relationship between news media and the providers of tech solutions as a space of negotiation and argued that the relationship between platforms and news media is not one-sided, as news organisations also adopt the platformisation that they encounter. We find this valuable for examining how news organisations approach the development of tech systems and AI-driven distribution differently, depending on their size and type of news organisation.

We take a material rather than a relational approach to the study of infrastructure, in line with what has been

called for by Flensburg (2020) and Flyverbom and Murray (2018). This entails focusing on the interplay between the technologies and organisational cultures to understand how the technologies shape the institutions and practices they sustain. This approach also helps us understand that the infrastructures are "stacked" via a large number of smaller systems or tech stacks, as they are referred to in the industry. This means including everything from the deepest levels of hardware (e.g., data storage) to the more dynamic layers of software development. Hence, the development of AI within news organisations requires us to analytically go beyond observing relationships between publishers and "traditional" social media platforms to include emerging data, code, and model-sharing platforms such as Github, PyTorch, and HuggingFace. Of interest to this study is work that examines the emergence and implications of cloud infrastructures as preconditions for platformisation. Narayan (2022), for example, provided an analysis of the platformisation of computing assets in which she examined how platform infrastructures expand through cloud infrastructures. She referred to this tendency as "radical outsourcing" and pointed out that very little is still known about cloud providers and their practices of expansion through these outsourcing processes (Narayan, 2022, p. 916). From a social perspective, these infrastructures also give rise to new practices. Such studies show how the development of AI analytics is financed through creative practices of reusing data, codes, and models from one context and fitting them into a different context, as well as how these creative practices involve new conditions of infrastructural dependency and vulnerability because of the risk of infrastructural lock-in, infrastructural decay and new licence models (Thylstrup et al., 2022). The present article contributes to these studies by expanding knowledge about how infrastructural development and platformisation processes unfold in the field of news and the social practices they engender.

3. Theoretical Framework: Media Logics and Infrastructure Capture

Infrastructure, in crude terms, refers to an "underlying foundation or basic framework" (Infrastructure, n.d.). In our research, we zoom in on the media backend as an infrastructure of the individual media organisation and its relation to the larger infrastructure of the internet and platforms (Plantin et al., 2018). Although we fully recognise the importance of tangible large-scale infrastructures, such as undersea cables, this study limits its empirical scope to focus on the media backend, an infrastructural micro-perspective one might say. Thus, we position ourselves in previous research that exemplifies infrastructures as "software, data, and technologies from outside newsrooms" (Ananny & Finn, 2020, p. 1600), "search engines and related systems" (Feuz et al., 2011, para. 13), or "protocols (human and computer), standards, and memory" (Bowker et al., 2009,

p. 97). Infrastructures are often defined in terms of their affordances and characteristics (Flanagan et al., 2008; Star & Bowker, 2002). They are built on top of previously installed infrastructures; thus, it can seem that we are dealing with a patched system with infinite versions (Star & Ruhleder, 1996). A functioning infrastructure requires standardisation across systems and former versions of systems. This also means that elements of infrastructure are embedded in—and therefore cannot be viewed as separated from—the values of former and current structures. Lastly, following Star and Ruhleder (1996), we view infrastructures as shaped by conventions of a community of practice, but simultaneously, they shape practice.

To connect the infrastructural focus to our interest in media, a conceptual lens is provided by the concept of "media capture" and, more specifically, Nechushtai's (2018) concept of "infrastructure capture." This notion refers to "circumstances in which a scrutinising body is incapable of operating sustainably without the physical or digital resources and services provided by the businesses it oversees and is therefore dependent on them" (Nechushtai, 2018, p. 1043). The capture can be both material and non-material, with the first referring to instances in which a regulator is benefitting financially from the industry it is overseeing (Nechushtai, 2018, p. 1046). Non-material forms are cultural and cognitive capture, which refer to capture through formal channels, for example, public relations efforts, and capture through informal relations, for example, personal relationships (Nechushtai, 2018, pp. 1046–1047). Simon (2022) demonstrated the usefulness of Nechushtai's concept of infrastructure capture in his analysis of how AI technologies are increasingly permeating the phases of journalistic gatekeeping. Simon argued that the capture and potential loss of control and media autonomy occur at different paces in the news industry and in news production (Simon, 2022, p. 1843). Eventually, news organisations risk adopting the logics of the external platforms and actors that are sustaining news production and distribution while simultaneously being competitors in seeking the attention of users. Autonomy is one of the key dimensions upholding the journalistic profession and refers in this article to the ability of the media and journalists to carry out professional routines without being influenced or having obligations to external actors, here the providers of infrastructure (Singer, 2007). In line with Simon (2022, p. 1833), we consider infrastructure capture and autonomy to be on opposite sides of a theoretical continuum which delineates the level of dependence between media and platforms. We acknowledge that our mapping does not give us an exact answer as to whether media are "captured" by infrastructure. Instead, our mapping and subsequent analysis make use of the concepts to illustrate the negotiations happening within news organisations regarding infrastructure.

To operationalise infrastructure capture, we apply an adaptation of the theory of media logic first introduced by Altheide and Snow (1979). Media logic

theory concerns itself with the "assumptions and processes for constructing messages within a particular medium" (Altheide, 2016, p. 1). Altheide and Snow (1979) employed the term "logic" in the singular, but as Thimm et al. (2018, p. 3) noted, today's networked media landscape is far more complex than in the mass media tradition from which Altheide and Snow departed. As such, several logics have been proposed in later years to account for the changes in media technologies and conditions (Couldry, 2008; Klinger & Svensson, 2018; van Dijck et al., 2018). In this study, we align with Klinger and Svensson (2018, p. 1244), who argued that "media logics as specific norms, rules and processes both influence and are influenced by the involved actor." Journalistic logics thus influence and are influenced by multiple competing logics. Extending this thinking, our goal is to investigate which logics are at play in infrastructuring as news media implement and develop systems in their tech stacks. Following this, the study aims to answer the following research question: How are the autonomy and infrastructure captured vis-à-vis external tech providers negotiated in the process of implementing and developing tech systems as infrastructure for the production and distribution of news?

4. Methodology

This research rests on a combination of interviews, fieldwork, and desk research. The first analytical part presents a mapping of the backend systems of several media organisations. This is based on a policy and document analysis, combined with our interviews, fieldwork, the StackShare website (https://stackshare.io), and searching the web for software solutions marketing themselves for the media industry. We also attended industry conferences, WebSummit in 2021 and TechSummit in 2021, and participated in industry networks such as the Nordic AI Network, where news media collaborate and exchange ideas on how to implement various tech systems and put together their "tech stacks," meaning the composition of systems on which news sites, news work, and news distribution are built. Methodologically, the aim of this study is not to provide a full picture of the extent to which these systems are used by different organisations, but the methods allow for an overview of the vast types of systems in all infrastructural corners of the news organisations, including production, distribution, and the commercial part of the news industries.

The mapping and the subsequent analysis are also based on in-depth interviews with 13 European and US-based publishers and intermediaries, an analysis of press releases and software documentation from system providers, and ethnographic observations in the development departments of two large Danish news organisations. We selected these two news organisations based on their publicly announced aim of developing independent data infrastructure platforms and personalised rec-

ommender systems. For the interviews, we included US- and UK-based media to assess potential similarities and links in the deployed backend infrastructures in our mapping, and they provided us with a backend understanding helpful for choosing cases for the focus points in the ethnographic observations. The initial fieldwork took place at Jysk Fynske Medier (JFM) from May 2019 to May 2021. JFM has around 1,850 employees and covers parts of North Zealand, all of Fynen, and most of Jutland. It has 15 regional subscription newspapers and 63 local free weeklies. The second fieldwork phase took place at *Ekstra Bladet* in JP/Politikens Hus from February to November 2022. *Ekstra Bladet* is a national newspaper in tabloid format and is one of the most read in its online version. It has around 300 employees, but a total of 2,100 people are employed at JP/Politikens Hus. The observations focused on the development departments and analytics departments and the managerial level to understand which and how systems were chosen, developed, and implemented. For this reason, we focused less on the newsrooms of the two organisations, although the journalists were implicitly present in both the observations and interviews as the "users" of many of the systems implemented during this period. We participated in meetings on project management as well as on everyday work two to three times per week during the observation period.

The interviews were conducted from May 2020 to November 2021 (see Table 1). The news organisations were selected following desk research on which news organisations were and are experimenting with AI in various forms and designs of, for example, recommender systems or in-house metrics and data analytics tools. The system providers were chosen because of their services being aimed at and employed by media companies. The interviews lasted from 40 to 60 minutes and were transcribed shortly after and analysed thematically using the NVivo software package (Braun & Clarke, 2006).

5. Analysis

5.1. Mapping the Backend of News Organisations

In Table 2, we present our mapping of the infrastructural systems sustaining the media. We categorised the systems and services into three levels: (a) production and publishing technologies, (b) distribution technologies, and (c) technologies that sustain the commercial viability of media (monetisation). Publishing technologies refer to the systems that form the basis of the workings of the news site. Production technologies refer to the systems used by journalists in their production processes. A characteristic of these is that they have a user interface, for example, typing in article text in a content management system, choosing photos from a photo library, and using audience measurement systems. There are multiple ways of reaching the audience and we know from previous research that users access content on social

Table 1. Informants.

Identifier	Organisation	Country	Position/Field
1	TV2	Denmark	Journalist and developer
2	Chartbeat	US	Account executive
3	Amedia	Norway	Head of digital development
4	Amedia	Norway	Head of engineering
5	*Ekstra Bladet*	Denmark	Head of research and innovation
6	JFM	Denmark	Editor of data and analysis
7	Mitt Media	Sweden	Data scientist and project manager
8	*The Guardian*	UK	Director of engineering
9	Midtjyske Media	Denmark	Head of digital development
10	*The New York Times*	US	Data science lead
11	*Information*	Denmark	Head of digital development
12	*Altinget*	Denmark	Editor of IT and development
13	*Ekstra Bladet*	Denmark	Data science developer
14	*Ekstra Bladet*	Denmark	Data science developer
15	JFM	Denmark	Digital editor
16	JFM	Denmark	Developer
17	JFM	Denmark	Developer
18	Google	Spain	Partner manager
19	Infomedia	Denmark	Head of data science

media, via search engines, newsletters, and, of course, via the media website itself. These are categorised as distribution technologies. In our mapping, we included technologies that enable monetisation. They may not be directly involved in journalistic practice, but they are key points of exchange of data and standards between systems used in production, distribution, and advertising.

As is the case when dealing with infrastructures, individual technologies sometimes cross categories. For example, content management systems often "solve" several tasks, and other systems are embedded into them. In addition, some technologies, such as cloud services, are foundational for all other systems to run. These deeply rooted infrastructural interdependencies are categorised here as "production and publishing technologies" for simplification purposes.

First, we find it striking that there are so many systems involved on various levels of the news media, which indicates a high level of infrastructuring via tech systems overall. While the printed newspaper also had to be printed and delivered, the infrastructural systems today are increasingly complex and involve many more actors and providers of such services and systems. Interestingly, the mapping further highlights that platform companies are present in all three categories of infrastructural technologies sustaining the media. The representation of Google products is especially striking, suggesting that infrastructure capture can take place on multiple levels. Zooming in on a case from each of the levels of tech systems in the following second part of our analysis allows us to show how different infrastructural logics are at play

in the process of implementing, highlighting logics of classification, standardisation, and datafication.

5.1.1. Infrastructure Capture Through Classification Logics

In Section 1, we presented the Danish tabloid *Ekstra Bladet*. Here, the team of developers experimented with large language models, which are machine-learning algorithms that can recognise, predict, and generate human languages on the basis of very large text-based data sets. Automated textual analysis is particularly useful for the implementation of recommender systems, as well as for pairing certain forms of content with advertisers or for coupling articles with supplementary relevant information from the internet. The development usually involves several steps, the end goal being to automatically analyse articles, content, or pictures and to categorise them so that they can be paired with the interests of the specific user and this user's history and profile. By automatically creating ways for content to flow through the systems, the news media link themselves to larger infrastructures of data. Hence, the purpose of the standardised categories is that they allow for integration with, for example, search engines, whose web crawlers require standardised data categories to "understand," store, and subsequently make news content visible:

> Our ranking systems for news content across Google and YouTube News use the same web crawling and indexing technology as Google Search to continually

Table 2. Technologies sustaining news media.

Categories of technologies	Facilitation in practice	Technologies	Commonly used service providers
Monetisation technologies	Advertising; handling of subscriptions, log-ins, customer profiles, and customer engagement; audience insights and business intelligence; strategy/resource allocation	Ad exchange server, customer data platform, audience measurement (for marketing purposes), customer relationship management, data warehouse	Google ads, Oracle responses, Twilio, Tealium, Adform, Google Analytics, BigQuery (Google), ElasticSearch
Production and publishing technologies	Content storage and content delivery	Server/database and content delivery network	Amazon DynamoDB, Fastly, AWS, CloudFront (Amazon), Cloudflare, Akamai, Microsoft Azure, Firebase
	Producing and publishing news to the news site and/or news app	Content management system/publishing platform	Stibo CUE, Wordpress VIP, Sitecore, Drupal
	Adding storytelling elements and data to news articles	Storytelling and visualisation tools	Infogram.com, Google Fusion Tables, Tableau Public
	Testing different versions of headlines on news articles	A/B testing	Chartbeat, Optimizely, Google Analytics
	News selection and prioritisation of day-to-day editorial resources	Audience measurement system real time and aggregated over time (for editorial purposes)	Google Analytics 360 (Realtime Content Insights), Chartbeat, Parse.ly, ComScore, Gemius, Moat, Facebook Insights
	Content organising, analysis for automation, and tagging	Transformer models (NLP) and topic/language modeling	Think Analytics, Cxense, Contentwise, GDP 1, 2, 3 (OpenAi/Microsoft), Google (BERT), Facebook (XLM Roberta), Huggingface (Huggingface), Google Tag Manager
	Automatic curation of news on the website	Recommender systems	Think Analytics, Cxense, Contentwise
Distribution technologies	News links shared to external platforms (by users and journalists)	Social media/debate fora and message apps	Facebook, Twitter, SnapChat, WeChat, Instagram, TikTok, Reddit
	Newsletters	Email newsletter services	MailChimp, HubSpot, SubStack
	Making news articles available on external platforms	Aggregation services (including podcast apps), voice assistants, pre-load article solutions	Apple News, Google News, Nachtrichten.de, Apple podcasts, Alexa, Google Assistant, Google AMP, Facebook Instant Articles, News APIs
	Search	Search engines	Google Search (Alphabet), Bing (Microsoft), Yahoo

Source: Authors' work based on fieldwork and interviews, provider websites, StackShare, and Newman et al. (2021, 2022).

identify and organize news articles from across the web, taking note of key factors—from keywords to website freshness—and keeping track of it all in the Search index. (Google News Initiative, n.d.)

Our interview with the Danish broadcaster TV2 showed how the organisation was conscious about being able to adapt to outside standards to ensure compatibility across time, platforms and devices: "When we build our

model, we try to look at the open models on the internet, like Google's, work. We try to apply those standards instead of our own to match models and connect content more easily" (journalist and developer at TV2).

In both media organisations in our fieldwork, *Ekstra Bladet* and JFM, the development and implementation of automated text analysis followed a similar pattern. This involved finding a suitable categorisation vocabulary, a taxonomy of content, and suitable tags for content. On a simple level, such tags could be "sports," "finance," or "entertainment," but on a much more finely grained level, text recognition (automated or manual) also involves finding places, names of specific sources, or categories in stories that are linked to a previously covered story. As we indicated in Section 1, the final taxonomy created at *Ekstra Bladet* (originally for different purposes than building transformer models), was a combination of content categories from the IAB, IPTC, and the paper's own categories, a negotiation between outside and inside values.

These negotiations of media autonomy in relation to the taxonomies offered by global marketing organisations, often developed for social media platforms in particular, mainly surfaced as a clash in topics that the audiences were interested in and the interests of the specific audiences of *Ekstra Bladet*, which were somehow not part of the more commercially built taxonomy. As the taxonomy is put into production and used as a cornerstone to train the large language models, it is included and embedded into a larger infrastructure, for example, linked up to databases and existing models provided by other actors available via a site like Huggingface or Github. If a category of content is left out in the first phase, no users will receive the content on this topic, neither as recommended nor as part of a personalised front page, and the developers are acutely aware of this. At both media organisations, the editors often discussed how they would solve the problem of new emerging content, which would then not be recognised by the large language models. For example, it would take a new tag to categorise content on Covid-19, which they argued was of as much democratic importance to the users as other news content, though it did not fit well with the advertising categories in the commercial models. "The IAB taxonomy tends to focus on cars and washing machines, which is far from the content we publish here at *Ekstra Bladet*," a developer said one afternoon, as we discussed how much work had gone into building the adapted taxonomy.

The trained models of text recognition and automated classification are interwoven into complex structures of data, both data on content and audiences, moving users in certain directions through the available content. As touched upon above, industry organs such as the IAB and IPTC, along with platforms such as Google, Facebook, and Yandex, are involved in streamlining categories of content on news websites. This partly pertains to the need to deliver accurate reports of audience data

to advertisers. For media and their potential advertisers to make comparisons on the market, the method of measuring and reporting audience data cannot be entirely up to each media organisation. The interdependency here is driven by the industry level by these classification logics, as seen above, but it is also at the level of both commercial and non-profit actors that agree on standardisations for metadata and structured data markup. As previous research has shown, this, however, means that a news organisation might miss out on being distributed via search engines, for example, if it does not follow the mark-up standards provided, for example, by Google and Schema.org (Kristensen & Sørensen, in press). In the following section, we look more into how these standards work and manifest themselves in negotiations around infrastructure capture.

5.1.2. Infrastructure Capture Through Standardisation Logics

As discussed in the previous section, media organisations are faced with outside technical standards required to produce, distribute and monetise news. An interesting case to examine is the audience measurement systems used in the newsroom and for making editorial decisions. These are not, per se, required to fit outside the standards of measuring methodology and taxonomy. However, in our empirical data, we observed that the systems used often originate or migrate from the marketing departments, which adhere to formal standards, such as the IAB, to allow for comparing measurements across media outlets. Furthermore, most systems used for real-time editorial insights, for example, Chartbeat and Parse.ly, are developed with newsrooms and marketers in mind as end users. Whereas the former (e.g., IAB) involves formal standards, here we see a case of informal standards applied through the use of the same system in multiple settings—for example, online marketers focusing on making "content" rather than "news."

The media organisations in our data were trying to different degrees to negotiate and deal with this. The second-largest media organisation in Norway, Amedia, was working to both eliminate external systems that provide access to their own data (e.g., Google Analytics) and to better tailor the measurements to a media organisation with public service ideals:

> What we saw was that the questions we wanted to ask about our data, we couldn't answer in those kinds of systems and, also, we wanted the ability to customise both the data collection and the observations and how data flowed through our systems. And we really didn't want to be sort of sitting there, just as customers of a third-party product and be limited by the solutions that they offered. But, of course, it's still like every cost, or I'm not sure how many man hours or employees are dedicated to working on this, that wouldn't have been working or that we wouldn't

need it if we have another system. (Head of digital development, Amedia)

As such, Amedia was trying to reclaim its autonomy in setting the standards for operationalising what news is and what could be considered empirical evidence of the "success" of a news story and the media organisation at large. At JFM, the audience measurement system was designed in-house in terms of the user interface, but the data came from Google Analytics and Facebook. Although this meant that the organisation was relying on outside standards of measurement methodology, and content categories, it allowed them to present the data in ways that helped them "qualify how to evaluate the journalism" (head of analytics, JFM). This entailed developing a point system that pooled together relevant metrics and developing custom dimensions in Google Analytics, for example, whether users had spent at least one minute and 30 seconds on the article page.

Standardisation logics come from both universally applied standards, such as taxonomies of Schema.org and Dublin Core, as well as from the systems applied in newsroom analytics. These are grounded in measurement methodologies and fixed metrics, along with visual representations of data with an interface designed by external system providers. Our empirical data reveal that a degree of infrastructure capture is in place. However, media organisations are aware of the potentially contradictory logics between them and the system providers.

5.1.3. Infrastructure Capture Through Datafication Logics

As *Ekstra Bladet* embarked on the project of personalisation and NLP (PIN project) in 2021, they realised that this also meant building their own data platform, eventually named Longboat. The purpose was also to share the data between the different publishers in the same media organisation, in this case, *Ekstra Bladet*, *Politiken*, and *JP*, who are all part of JP/Politikens Hus. Further, it was an attempt to gain autonomy vis-à-vis Google, as the platform was built with its own data analytics system. As the Head of Strategy at *Ekstra Bladet* Kasper Worm-Petersen explained in a press release:

> It's no secret that data is a very central element in the realization of our strategy for the coming years. It is therefore important to us that we have control over and ownership of our data throughout the value chain from collection to processing to activation. Relevance ensures us ownership of the activation. With the PIN project, we are investing heavily in the processing, and with Longboat, we are now also taking ownership of the collection itself. This gives us some completely unique opportunities in the media reality that Ekstra Bladet is moving into.

Across media organisations, we observed an awareness of how developing and maintaining one's own infras-

tructure is expensive and even risky in the case of software and hardware breakdown. As was the case with Amedia, at *The Guardian*, and JFM, developing proprietary systems was at the forefront to avoid technology "giants" profiting from the media's own data and to be able to define measurement categories themselves to a higher degree. However, the audience data infrastructure "pipeline" remained the same, but the visual and statistical presentation of data and the organisational discourse changed following a push to incorporate audience behaviours and preferences through web measurement reports on email and newsroom dashboards. As another example, the Danish daily *Information* experienced a shutdown of its email automation platform, which also interfered with its ability to send purchase receipts to subscribers. *Information* abandoned their major US-based provider following the incident:

> We cannot have a system that is so critical to our business where the provider does not have a phone number…so, first of all, I want a provider I can call, preferably in Denmark, but Germany, Norway or Sweden would also be okay. (Head of digital development at *Information*)

Eventually, they chose a Danish/Swedish email platform that had more functions but was also more expensive. This suggests that media organisations are aware of interdependencies, constantly negotiating their autonomy vis-à-vis these infrastructural systems and providers. Outsourcing infrastructural tasks to external software providers was a way to minimise spending, as the three people employed in the technical department did not have the resources to develop and maintain systems for subscriber login and payment, but functionality and control were still a priority.

Building transformer models and personalisation algorithms at *Ekstra Bladet* also meant negotiating infrastructure capture by not using an external dataset for machine learning and the training of models. Thus, a great deal of work goes into developing various datasets, including those of content, articles, and different kinds of users. This was not only due to the fact that datasets on open source platforms were often not in Danish but also because they did not feel adapted enough to the specific media organisation. For example, *Ekstra Bladet* has more users who are men and of a certain age and thus the baseline dataset used to recommend content needed to reflect this.

When building the transformer models at *Ekstra Bladet*, it was often discussed how much they could rely on open-source models and data provided via the free platform Hugginface, as it created an infrastructure dependence, which it was hard to foresee the consequences of. When a manager questioned this in a meeting, asking whether Huggingface would, at some point, capitalise on the models and code available, the answer from the developer was: "If Hugginface dies, we

die." Interestingly, although the primary aim of building transformer models is to retain autonomy, they make themselves dependent on other providers for the code, datasets, cloud services, etc.

6. Discussion

In the following, we discuss consequences, based on our findings, for journalistic production, monetisation, and the distribution of news.

Autonomy is a key ideal in journalism and through our analysis, we illustrated the extent to which external systems are sustaining parts of the journalistic production process, creating a form of infrastructural interdependence of these systems. The systems also bring with them certain logics. In the pre-digital age, categorisations of news (e.g., in terms of genre and subject matter) and audience members were to a certain degree standardised across the media industry itself. Today, these standards also occur across fields that operate with different logics than is traditionally the case in journalism. The consequence of this could be that news categorised through the same or similar taxonomies as "content" in general might be assimilated. In the case of audience measurement systems, we observed how metrics and visual representations could be standardised, but the large media organisations in our study acknowledged and negotiated the degree of infrastructure capture by implementing their own systems and tweaking the ones they bought externally.

We observe that audience measurement systems and other tech systems used in media organisations are inseparable from other infrastructures. These interdependencies are expressed, for the most part, through logics of standardisation, that is, an alignment of and path dependencies pertaining to practices around the use of systems, methods and data flow, data reporting, and discourses around the practice in and around the media. As illustrated by our cases, infrastructure capture of the news through the tagging of news content using (often open-source) algorithms, audience measurement metrics, and statistical representations becomes a negotiation between media organisations and the providers, that is, the composition of technologies behind media production and distribution.

In the literature on platformisation, we often see the loss of autonomy over distribution on external platforms. This includes Facebook's changes to algorithms in 2016 to focus more on friend relationships and the 2023 revealing of a function within TikTok that allows employees to override the factors that normally determine the position of posts in the feed. Distributing content on these external platforms thus means conforming to external logics of what content is popular and what the platform owner or employees prefer, resulting in a potentially high degree of infrastructure capture.

We found that the picture is somewhat more complex and that media organisations are aware of infras-

tructure capture through potentially competing logics. It is worth noting that the organisations in our sample are relatively well-resourced and that smaller or digital native media likely do not have the same opportunities to negotiate the degree of capture. Contrarily, this means that they potentially lag behind legacy media in their visibility on external platforms for not adhering to standards, leading to a lesser degree of infrastructure capture, perhaps at the expense of monetisation of news.

Dependence on advertising platforms is a key indicator of infrastructure capture (Nechushtai, 2018). In our empirical mapping, we observed how platforms are deeply enthralled in sustaining monetisation through advertising on media websites and externally through ad exchanges. Email services, login, and customer platforms are also deeply intertwined with the media and across systems. This would be considered material infrastructure capture in the sense that the media has the role of a scrutinising body (Nechushtai, 2018) and, at the same time, depends financially on platforms for both advertising and distribution. A future avenue for research is thus the potentially impaired ability to scrutinise the very companies that sustain news distribution and operation.

Although infrastructure capture is indicated to a certain degree in our empirical data, our analysis similarly points to the potential benefits of media organisations' backends being related to and embedded into existing systems. For instance, "outsourcing" technology allows media organisations to abandon maintaining servers for hosting and to use programming resources for tasks other than keeping a user database, for example. In the sense that journalism is important for democracy, opportunities to save money and ensure system stability by buying cloud services and embedding externally maintained and developed software from outside system providers can be a positive shift. Based on our empirical data, large media corporations often have the resources to decide whether to "outsource" their infrastructure, which might not be as accessible for digital native outlets and other smaller publishers. In addition, if media organisations did not structure data on news content to be recognisable to search engine web crawlers, they would not be visible in the search results, which would affect their access to users. The representation of Google products in our mapping is especially striking, suggesting that the company is as involved in sustaining the news media industry as it is in sustaining other parts of society (van Dijck et al., 2018). As such, we might not only see a case of infrastructure capture, but a case of what Plantin et al. (2018, p. 4) described as the "platformization of infrastructure" and "infrastructuralization of platforms." Thus, if we consider journalism in the form of legacy media news as a pillar of democracy, Google and the tech providers that are sustaining the backend of media could be approaching the status of the infrastructure of democracy. We find media logics to be a valuable framework to understand these developments while, for the same reasons, a concept that needs to be expanded

on. We believe that we have contributed to the theory by suggesting classification, standardisation, and datafication as entry points for this.

7. Conclusion

By mapping the elements of the digital infrastructure of media organisations from systems that handle the sales and distribution of advertising to systems that classify news, we have illustrated (Table 2) how data flows through systems originating from both within and outside media organisations. Our findings suggest that the latter is most often the case, illustrating that news and news production are increasingly and inevitably part of the larger infrastructure of the internet, provided by big tech companies, which have been theorised elsewhere as infrastructure capture (Simon, 2022; Nechushtai, 2018). Through case studies centering on the development of different parts of the backend tech stack of news distribution, we have shown that these tasks that were previously performed within the news organisation are now "outsourced" to external systems and providers. The analysis of the interviews and fieldwork illustrated how news organisations deal with this reality and how they are negotiated.

Finally, we discussed the consequences of these interdependencies on the autonomy of news media. To summarise, the dominant logics of media organisations' interdependencies with larger infrastructures are (a) standardisation and (b) embeddedness in former and parallel systems and infrastructures—materially and in terms of values and practices. The interdependencies are evident in our mapping and are of concern to media organisations across our fieldwork and interviews. This, in turn, highlights that infrastructure capture should not be seen as a one-way information highway, but as spaces of negotiation in which the infrastructural power manifests itself through logics of standardisation, classification, and datafication.

With the increasing use of AI in news organisations, the network of backend infrastructures is likely to be even bigger, and, from a research perspective, we need to analyse how this infrastructuring unfolds in different media settings. After all, if news media and journalism were ever the backbone of democracy, the infrastructure supporting them should not be overlooked.

Conflict of Interests

The authors declare no conflict of interests.

References

Altheide, D. L. (2016). Media logic. In G. Mazzoleni (Ed.), *The international encyclopedia of political communication*. https://doi.org/10.1002/9781118541555.wbiepc088

Altheide, D. L., & Snow, R. (1979). *Media logic*. SAGE.

Ananny, M., & Finn, M. (2020). Anticipatory news infrastructures: Seeing journalism's expectations of future publics in its sociotechnical systems. *New Media & Society*, *22*(9), 1600–1618. https://doi.org/10.1177/1461444820914873

Anderson, C. (2011). Between creative and quantified audiences: Web metrics and changing patterns of newswork in local US newsrooms. *Journalism*, *12*(5), 550–566. https://doi.org/10.1177/1464884911402451

Bowker, G. C., Baker, K., Millerand, F., & Ribes, D. (2009). Toward information infrastructure studies: Ways of knowing in a networked environment. In J. Hunsinger, L. Klastrup, & M. Allen (Eds.), *International handbook of internet research* (pp. 97–117). Springer.

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, *3*(2), 77–101. https://doi.org/10.1191/1478088706qp063oa

Christin, A. (2020). *Metrics at work: Journalism and the contested meaning of algorithms*. Princeton University Press.

Couldry, N. (2008). Mediatization or mediation? Alternative understandings of the emergent space of digital storytelling. *New Media & Society*, *10*(3), 373–391. https://doi.org/10.1177/1461444808089414

Feuz, M., Fuller, M., & Stalder, F. (2011). Personal web searching in the age of semantic capitalism: Diagnosing the mechanisms of personalisation. *First Monday*, *16*(2). https://doi.org/10.5210/fm.v16i2.3344

Flanagan, M., Howe, D. C., & Nissenbaum, H. (2008). Embodying values in technology: Theory and practice. In J. van den Hoven & J. Weckert (Eds.), *Information technology and moral philosophy* (pp. 322–353). Cambridge University Press.

Flensburg, S. (2020). *Det Digitale Systemskifte: En historisk analyse af digitaliseringen af det danske Kommunikationssystem* [The digital system change: A historical analysis of the digitalization of the Danish communication system] [Doctoral dissertation, University of Copenhagen]. Københavns Universitet Forskning. https://forskning.ku.dk/soeg/result/?pure=da/publications/det-digitale-systemskifte(39af4a6b-6dfc-46bd-8ce8-2365dde25191).html

Flyverbom, M., & Murray, J. (2018). Datastructuring—Organizing and curating digital traces into action. *Big Data & Society*, *5*(2). https://doi.org/10.1177/2053951718799114

Friedman, B., & Nissenbaum, H. (1996). Bias in computer systems. *ACM Transactions on Information*

Systems, *14*(3), 330–347. https://doi.org/10.1145/230538.230561

Google News Initiative. (n.d.). *Organizing news from around the web*. https://newsinitiative.withgoogle.com/hownewsworks/approach/organizing-news-from-around-the-web

Helberger, N. (2019). On the democratic role of news recommenders. *Digital Journalism*, *7*(8), 993–1012. https://doi.org/10.1080/21670811.2019.1623700

Infrastructure. (n.d.). In *Merriam-Webster's online dictionary*. https://www.merriamwebster.com/dictionary/infrastructure

Klinger, U., & Svensson, J. (2018). The end of media logics? On algorithms and agency. *New Media & Society*, *20*(12), 4653–4670. https://doi.org/10.1177/1461444818779750

Kristensen, L. M. (2021). Audience metrics: Operationalizing news value for the digital newsroom. *Journalism Practice*. Advance online publication. https://doi.org/10.1080/17512786.2021.1954058

Kristensen, L. M., & Sørensen, J. K. (in press). Classifying the news: Metadata as structures of visibility and compliance with tech standards. In J. M. Hartley, J. K. Sørensen, & D. Mathieu (Eds.), *DATAPUBLICS—Constructing publics in datafied democracies*. Bristol University Press.

Møller, L. A. (2022). Between personal and public interest: How algorithmic news recommendation reconciles with journalism as an ideology. *Digital Journalism*, *10*(10), 1794–1812. https://doi.org/10.1080/21670811.2022.2032782

Narayan, D. (2022). Platform capitalism and cloud infrastructure: Theorizing a hyper-scalable computing regime. *Environment and Planning A: Economy and Space*, *54*(5), 911–929. https://doi.org/10.1177/0308518X221094028

Nechushtai, E. (2018). Could digital platforms capture the media through infrastructure? *Journalism*, *19*(8), 1043–1058. https://doi.org/10.1177/1464884917725163

Newman, N., Fletcher, R., Robertson, C. T., Eddy, K. & Nielsen, R. K. (2022). *Reuters Institute digital news report 2022*. Reuters Institute. https://reutersinstitute.politics.ox.ac.uk/digital-news-report/2022

Newman, N., Fletcher, R., Schulz, A., Andı, S., Robertson, C. T., & Nielsen, R. K. (2021). *Reuters Institute digital news report 2021, 10th Edition*. Reuters Institute. https://reutersinstitute.politics.ox.ac.uk/digital-news-report/2021

Neyland, D., & Möllers, N. (2017). Algorithmic IF…THEN rules and the conditions and consequences of power. *Information, Communication & Society*, *20*(1), 45–62. https://doi.org/10.1080/1369118X.2016.1156141

Petre, C. (2021). *All the news that's fit to click: How metrics are transforming the work of journalists*. Prince-ton University Press.

Plantin, J.-C., Lagoze, C., Edwards, P. N., & Sandvig, C. (2018). Infrastructure studies meet platform studies in the age of Google and Facebook. *New Media & Society*, *20*(1), 293–310. https://doi.org/10.1177/1461444816661553

Poell, T., Nieborg, D. B., & Duffy, B. E. (2022). Spaces of negotiation: Analyzing platform power in the news industry. *Digital Journalism*. Advance online publication. https://doi.org/10.1080/21670811.2022.2103011

Simon, F. M. (2022). Uneasy bedfellows: AI in the news, platform companies and the issue of journalistic autonomy. *Digital Journalism*, *10*(10), 1832–1854. https://doi.org/10.1080/21670811.2022.2063150

Singer, J. B. (2007). Contested autonomy: Professional and popular claims on journalistic norms. *Journalism Studies*, *8*(1), 79–95. https://doi.org/10.1080/14616700601056866

Sørensen, J. K. (2022). The shortcomings of the diversity diet: Public service media, algorithms and the multiple dimensions of diversity. In J. Meese & S. Bannerman (Eds.), *The algorithmic distribution of the news: Policy responses* (pp. 289–307). Palgrave Macmillan.

Star, S. L., & Bowker, G. C. (2002). How to infrastructure. In L. A. Lievrouw & S. Livingstone (Eds.), *Handbook of new media: Social shaping and consequences of ICTs* (pp. 151–162). SAGE. https://doi.org/10.4135/9781446206904.n12

Star, S. L., & Ruhleder, K. (1996). Steps toward an ecology of infrastructure: Design and access for large information spaces. *Information Systems Research*, *7*(1), 111–134. https://doi.org/10.1287/isre.7.1.111

Thimm, C., Anastasiadis, M., & Einspänner-Pflock, J. (2018). Media logic or media logics? An introduction to the field. In C. Thimm, M. Anastasiadis, & J. Einspänner-Pflock (Eds.), *Media logic(s) revisited: Modelling the interplay between media institutions, media technology and societal change* (pp. 1–8). Palgrave Macmillan.

Thurman, N. (2011). Making "The Daily Me": Technology, economics and habit in the mainstream assimilation of personalized news. *Journalism*, *12*(4), 395--415. https://doi.org/10.1177/1464884910388228

Thylstrup, N. B., Hansen, K. B., Flyverbom, M., & Amoore, L. (2022). Politics of data reuse in machine learning systems: Theorizing reuse entanglements. *Big Data & Society*, *9*(2). https://doi.org/10.1177/20539517221139785

Tuchman, G. (1973). Making news by doing work: Routinizing the unexpected. *American Journal of Sociology*, *79*(1), 110–131.

van Dijck, J., Poell, T., & de Waal, M. (2018). *The platform society: Public values in a connective world*. Oxford University Press.

About the Authors

Lisa Merete Kristensen is a postdoctoral researcher at the Department of Communication and Arts at Roskilde University, Denmark. Her work is centred around metrics, tech stacks, and technical infrastructures in journalism and the impact of these on media organisations and journalism. She specialises in theories around the journalistic profession and is working with both digital methods and more qualitative research methodologies.

Jannie Møller Hartley is an associate professor at the Department of Communication and Arts at Roskilde University, Denmark. She is currently co-leading the Centre for Big Data at Roskilde University and is the PI of the Velux Funded research project DataPublics (2020–2023). Her research is situated in between the fields of journalism research, audience studies, and data science, and ranges across subjects such as media ethics, #MeToo, and AI in media.