This is an Accepted Manuscript of an article to be published by Taylor & Francis in *Journal of Management Information Systems* as DOI 10.1080/07421222.2021.1912928

The Making of Data Commodities: Data Analytics as an Embedded Process

Aleksi Aaltonen
aleksi@temple.edu
Temple University

Cristina Alaimo
calaimo@luiss.it
LUISS University

Jannis Kallinikos
jkallinikos@luiss.it
LUISS University
and
The London School of Economics and Political Science

Abstract

This paper studies the process by which data are generated, managed, and assembled into tradable objects we call *data commodities*. We link the making of such objects to the open and editable nature of digital data and to the emerging big data industry in which they are diffused items of exchange, repurposing, and aggregation. We empirically investigate the making of data commodities in the context of an innovative telecommunications operator, analyzing its efforts to produce advertising audiences by repurposing data from the network infrastructure. The analysis unpacks the processes by which data are repurposed and aggregated into novel data-based objects that acquire organizational and industry relevance through carefully maintained metrics and practices of data management and interpretation. Building from our findings, we develop a process theory that explains the transformations data undergo on their way to becoming commodities and shows how these transformations are related to organizational practices and to the editable, portable, and recontextualizable attributes of data. The theory complements the standard picture of data encountered in data science and analytics and renews and extends the promise of a constructivist IS research into the age of datafication. The results provide practitioners, regulators included, vital insights concerning data management practices that produce commodities from data.

Keywords

advertising audience, analytics, big data, case study, data commodities, data-based objects, social practices

INTRODUCTION

This paper examines the process through which data are used as the primary stuff for producing goods that mainly emerge from relations between data. We refer to such entities as *data commodities* to indicate the role digital data play in their development, commercialization, and maintenance. We argue that the processes by which data commodities such as popularity indexes, ratings and rankings of services, indicators from alternative data in quantitative trading and investing[1], usability metrics, or advertising audiences are compiled are often much more uncertain and circuitous (Bechmann and Bowker 2019; Ekbia 2009; Kallinikos et al. 2013) than what is usually seen through the lenses of big data analytics and data science (Dourish and Cruz 2018; Jones 2019).

To study the making of data commodities, we focus on transformations that digital data undergo from the generation or acquisition of individual *data tokens* through several stages of processing and management to the delivery of a commodity to customers. Our starting point is that data are not only tokens of higher or lower representational value (cf. Burton-Jones et al. 2017; Wand and Weber 1995) but also a medium for sensemaking and knowledge creation (Alaimo and Kallinikos 2020; Kallinikos 1999; Kallinikos et al. 2013; Lycett 2013; Tuomi 1999). Data tokens are produced by conventions (and, of course, technologies) that capture facts as socially validated and standardized records. Data practices such as aggregation further diffuse the representational value of data, compiling them into larger *data-based objects* that are often marked by considerable ambiguity as to what they represent. Our approach foregrounds the type of practices and interpretive work required to aggregate data

[1] See, for instance, Lee, J. 2020. "Quants Sound Alarm as Everyone Chases Same Alternative Data," *Bloomberg*, available at: https://www.bloomberg.com/news/articles/2020-06-25/quants-sound-warning-as-everyone-chases-same-alternative-data

into larger objects and turn them into data commodities. Yet, the status of data as a medium of staging and interpreting organizational realities has received relatively little attention in the literature apart from a few exceptions (e.g., Alaimo and Kallinikos 2017, 2020; Alaimo et al. 2020; Bailey et al. 2012; Monteiro and Parmiggiani 2019).

We frame the production of data commodities as an organization- and industry-embedded process of data management, analytics, and interpretation. We examine how such a process is associated with the open-endedness and ambiguity of data as a medium for sensemaking and knowledge creation (Bailey et al. 2012; Hirschheim et al. 1995; Kallinikos 1999; Monteiro and Parmiggiani 2019), particularly under conditions in which data are repurposed, decontextualized, aggregated, and recontextualized. The proliferation of data produced and exchanged by social media, data brokers, data management platforms, and other big data industry actors is progressively weakening the links of data to particular contexts and to their original meaning, stretching instead the practices of data reuse, repurposing, packaging, and further commodification (Martin 2015). While sharing empirical interest in data with data science and analytics (Abbasi et al. 2016; Baesens et al. 2016; Brynjolfsson and McAffee 2014; Grover et al. 2018; Müller 2016; Power et al. 2019), we thus view data as a part of broader organizational arrangements that make them matter in a particular business and industry (Alaimo and Kallinikos 2020; Jones 2019).

To explore how a data commodity is created and maintained, we analyze how a newly founded telecommunications operator, Click (pseudonym), turns mobile network subscribers into advertising audiences. Click sells advertising slots to advertisers whose advertisements are relayed to subscribers' phones, and, subsequently, Click aggregates data from the network infrastructure to measure the reception of the messages. As we show later, the data do not represent an audience but are merely a starting point for a process by which the data commodity is produced for the advertisers. We conduct an intensive case study to investigate the steps and practices that underpin, first, the generation and acquisition of masses of data tokens through various sorts of data management practices and, second, the production of data-based objects that are eventually recognized as valuable audiences by the advertisers. To do this, we construct a theoretically motivated narrative that allows us to examine the links between the characteristics of digital data, the big data industry conditions in which the data are exchanged and used, and various analytical operations and organizational practices that account for transformations data undergo (Eisenhardt 1989; Langley 1999; Yin 2003).

Our findings reveal that work with data includes grappling with the ambiguity of data as standardized inscriptions of events and their open-endedness as they continue to be related with other data, repurposed, and packaged into larger objects (Kallinikos 1999). First, we identify three types of datawork practices that deal with the ambiguity of digital data and show how they entail a range of techniques and skillful operations by which the data are formatted, standardized, tidied, assembled, interpreted, and, ultimately, assessed and calibrated against the demands of the market environment. These practices that we label as *metrics production*, *data-based improvisation*, and *data analytics projects* are essential for understanding data analytics as an embedded process through which the meaning and use of data are established rather than as the enactment of statistical procedures alone. Second, theorizing from our findings, we put forward a processual framework that depicts the making of data commodities. The framework explains how data are transformed along the road to becoming data commodities and how these transformations shape the signifying and knowledge status of data and, ultimately, the market value of the data-based objects they help assemble.

The study aligns with a long-standing research tradition that sees the use of technologies and the development of technology-based solutions as part of a broader system of organizational and human practices (e.g., Burton-Jones 2014; Leonardi and Barley 2008; Leonardi et al.

2012; Orlikowski 2000; Sarker et al. 2019; Swanson 2020; Zuboff 1988). We contribute to this tradition by synthesizing our findings with a growing body of empirical studies that investigate data analytics as a circuitous process of data management that pieces means and ends together with varying and often uncertain outcomes (e.g., Aaltonen and Tempini 2014; Alaimo and Kallinikos 2017, 2020; Bechmann and Bowker 2019; Jarvenpaa and Markus 2018; Jones, 2019; Monteiro and Parmiggiani 2019; Passi and Jackson 2018; Ransbotham et al. 2016). Our theoretical framework explains how the editable, portable, and recontextualizable nature of digital data is managed by different types of practices that transform the data along their journey to becoming commodities.

The paper is structured as follows. In the next section, we review literature that has recently emerged to complement research into big data and analytics, define key concepts used in the analysis, and formulate two sets of research questions. This is followed by the report of our research design and empirical findings and their analysis. We then move on to constructing a theoretical framework that lays out the process that data undergo from their original procurement and repurposing to the making of data commodities. We end by summarizing the contribution our study makes to the literature, discussing its limitations and relevance for the fields of IS and management, and indicating areas for further research.

LITERATURE REVIEW

Much of the important literature on big data and analytics falls into three broad categories. There are papers advancing analytical methods and techniques with a view to improving operations (Baesens et al. 2016; Brynjolfsson and McAffee 2014; Chen et al. 2012; Goes 2014; Power et al. 2019), papers on the adoption and value of big data analytics and related transformations (Abbasi et al. 2016; Baesens et al. 2016; Chen et al. 2012; Goes 2014; Grover et al. 2018; Günther et al. 2016; Kitchens et al. 2018; Lehrer et al. 2018), and papers on data quality (Hazen et al. 2014; Lee et al. 2002; Rieh and Danielson 2007; Wang et al.

1995; Wang and Strong 1996) and techniques through which data are acquired, stored, and integrated (Ali and Wrembel 2017; Bansal 2014; Sivarajah et al. 2017). The types of studies these papers represent typically assume that data are more or less faithful representations of external or 'real-world' facts (Burton-Jones et al. 2017; Jones 2019; Wand and Weber 1995), and that work with data occurs against a typified image of organizations as a pool of resources, capabilities, or functional operations. Such premises tend to steer attention away from the actual conditions and practices through which data are made to matter in organizational and industry settings.

At the same time, current literature attests to the growing awareness of the ambiguity of data as a resource and a medium of sensemaking and to a need to study how such ambiguity is linked to local practices and structural conditions through which organizations interpret and insert data into their operations (Burton-Jones 2014; Jones 2019; Kallinikos 1999; Ransbotham et al. 2016) and, ultimately, to the characteristics of digital data themselves (Aaltonen and Penttinen 2021; Alaimo and Kallinikos 2020; Clarke 2016). There is currently a small yet growing body of literature that reflects these research preoccupations (e.g., Aaltonen and Tempini 2014; Alaimo and Kallinikos 2017, 2020; Bechmann and Bowker 2019; Jarvenpaa and Markus 2018; Monteiro and Parmiggiani 2019; Østerlie and Monteiro 2020; Passi and Jackson 2018) and the insight that "data is more than knowledge," as Tuomi (1999) put it in this journal already 20 years ago. Each in its own way and with reference to diverse empirical evidence, these studies point out that work with data does not and cannot function on automatic pilot, that is, by applying generic techniques and templates alone. The process of transforming data to useful indicators of facts and thus building up data-based objects of organizational or market relevance are shown to be replete with ambiguities in these studies. The making of data is marked by various practices that manage the inflow of data from the environment to the organization while requiring the extensive editing and aggregation of data and their interpretation and eventual calibration against specific organizational objectives. We thus ask the following empirical questions:

RQ1: *What type of data management practices are associated with creating and maintaining a data commodity? How do such practices link to the wider data environment in which many organizations currently operate?*

Monteiro and Parmiggiani (2019) described the politics and practices of data collection and interpretation in the context of scanning the marine arctic environment by an oil and gas company in a way that bears particular relevance to our study. Drawing on and extending prior scholarship on these issues (e.g., Kallinikos 2006; Lycett 2013; Orlikowski 2000; Zuboff 1988), the paper subsumes the defining qualities of data under the concept of digital materiality that confers two major attributes to data. First, digital data *liquefy* reality; they are "vehicles for liquefaction" (Monteiro and Parmiggiani 2019, p. 168) that are divested from the material forms and situations to which they refer and transformed into digital tokens ready for further manipulation and use. As such, liquefaction is closely associated with and often results in significantly decontextualized data. Second, data are *open-ended*, a condition that enables them to be expanded, deleted or amended, modified or reprogrammed (see also Aaltonen and Penttinen 2021; Aaltonen and Tempini 2014; Kallinikos et al. 2013). The open-endedness makes it possible to repurpose data to new uses but, at the same time, often makes the meaning of data ambiguous and laborious to fix as the uses of data become increasingly distant from their external referent and the original conditions under which they have been produced.

The implications of liquefaction and open-endedness are, as briefly noted in the introduction, reinforced by the expanding big data industry in which the use and exchange of data are currently embedded (Martin 2015). Synthesizing these views with ideas from digital objects literature (Faulkner and Runde 2019; Kallinikos et al. 2013), we construe data as editable, portable, and recontextualizable entities. Data are *editable* to the degree that data tokens can be updated, amended, combined, deleted, or rearranged at almost no cost. Editability is a pervasive quality of digital data that bespeaks their propensity to be modified and expanded. In the current context of the big data industry and the web, these data attributes are closely connected to the making of more complex data-based objects via aggregation, clustering, rearrangement, or recombination. At the same time, the editable nature of data confers to them both instability and epistemic (knowledge) uncertainty, conditions that must be managed as they move along the road from raw data through data-based objects to data commodities. Furthermore, decontextualized data become *portable* across settings, platforms, and organizations and, thus by implication, *recontextualizable* in the sense of being possible to use to tell stories other than those linked to their origin and initial use (Bechmann and Bowker 2019; Ekbia 2009; Monteiro and Parmiggiani 2019). We thus ask the following theoretical questions:

RQ2: *What types of transformations do data undergo during the production of data commodities? How are these transformations linked to the attributes of editability, portability, and re-contextualizability of digital data?*

RESEARCH DESIGN AND EMPIRICAL EVIDENCE

The lack of extant theorizing on the interpretive practices of data management motivates a qualitative, intensive field study (Edmondson and McManus 2007; Sayer 1992) that supports our theory-building efforts by allowing us to cycle through empirical evidence, emerging theory, and literature as we seek to make sense of the making of a data commodity (Eisenhardt and Graebner 2007). We focus on innovation at the intersection of two data-intensive industries: media and telecommunications. Available data on people’s media consumption and the capacity to turn these data into information about audiences have historically had a significant impact on industrial structures, product development, and the economic viability of different firms in the media industry (Barnes and Thomson 1994; Webster, Phalen, and Lichty 2006), whereas telecommunications business rests on a massive data collection machinery that operates as a part of the network infrastructure. We study how a newly founded mobile telecommunications operator, Click, struggled to establish an advertising-funded business based on turning network subscribers into advertising audiences. The setting offered direct access to a variety of practices and operations through which data are processed into an audience product. Although the bulk of empirical evidence dates ten years back, the questions we ask have not been addressed, nor have the developments in analytical techniques and methods made the issues less relevant today.

Research Site

Click was founded in 2006 in a European country by two former advertising and telecommunications executives who raised venture funding to develop and launch an advertising-funded telecommunications operator. At the time of the fieldwork, the company sought to create value by bringing advertisers and young consumers together in a new channel; Click had “the soul of media but the body and muscles of a telecoms operator,” as one of the employees put it. Consumers subscribed to the service by answering a few profiling questions and by granting the company the right to relay marketing messages from advertisers to their mobile phones, in exchange for a SIM card with a monthly quota of free communications. Advertisers would then pay for having their messages delivered to highly targeted consumers using Click’s service. If a subscriber ran out of the free quota, he or she would need to top up a prepaid account to keep using the service until the next monthly “refill” day. At the time, the company employed 28 people at its head office divided into six teams, described in Table 1, and had a local sales office in another country where the service was first launched in 2007. The intention was to expand (and the company eventually did) to several countries.

Table 1. Click head office organization	
Brand Office team	Develops the company brand and manages marketing and public relations operations
Commercial team	Develops advertising formats and products for advertisers and is responsible for working with the sales office to generate revenues for the company
Members team	Manages the relationship with consumers, including, for instance, a partner organization who provides the customer service operations
Operations team	Maintains and optimizes the operation of the telecommunications service and other Click systems
Technology team	Develops and procures information systems that support different organizational functions
Administration	Takes care of human resources, financial, legal, and other administrative matters

The six teams were responsible for day-to-day operations and the smooth functioning of the advertising and telecommunications services. Heavily outsourced operations and several partner organizations were held together by numerous IT systems that allowed the company to orchestrate an innovative but also fairly complex service for such a small organization. The systems, for instance, allowed for managing various partners, delivering advertisements, and, importantly, sourcing the data required to produce advertising audiences. Among numerous applications and services, there were 12 systems with analytics features of varying levels of sophistication that supported a host of operations related to the production of advertising audiences. These systems are listed in Appendix 1.

Data Collection

The empirical evidence covers mainly internal operations, but we also collected public statements made by the company contributing to the production of audiences that are recognized by advertisers and other market participants. In early 2009, one of the authors spent 62 consecutive days working as an unpaid intern with Members and Brand Office teams assisting in various day-to-day tasks, while Commercial and Operations teams sat a few desks away in an open-plan office. The physical setting for participatory observation provided excellent access to various company operations, resulting in 270 pages of fieldwork notes. We conducted 34 semi-structured, recorded interviews covering every head office staff member except one, while seven staff members were interviewed twice to follow up on key research themes. All of the interviews were supported by an interview guide that we tailored for each interviewee's role and situation to ensure we were able to probe informants consistently while letting them decide the way in which they wanted to answer the questions (Orlikowski and Baroudi 1991). The interviewees and their organizational roles are listed in Appendix 2. We also stored 26 press releases, 10 marketing case studies, and 60 blog posts from the company website, and the participant observer took 147 photographs, several screenshots from corporate systems and gathered 340 documents related to the company operations and its business.

A list of themes helped focus the participatory observation on tools, practices, and events that are central to establishing a new kind of advertising business. The observations made immediately evident the central role of data in Click's business, but they also laid bare considerable difficulties in producing the data commodity. Voluminous and extremely detailed data on the delivery of messages to subscribers' phones did not as such lead to the recognition of a valid audience among advertisers. Finally, we conducted follow-up interviews in 2017 to capture the evolution of Click and to validate our main findings. The empirical evidence is summarized in Table 2.

Table 2. Empirical evidence	
Participant observation	62 consecutive days from 13 February 2009 to 15 May 2009, resulting in 270 pages of observation notes
Interviews	34 semi-structured, recorded interviews with the entire head-office staff
Press releases	26 press releases
Marketing cases	10 marketing case studies
Blog posts	60 blog posts published on Click's website
Photographs	147 photographs from various situations at the office
Documents	340 documents
Analytical memos	14 memos summarizing each week of observation
Follow-up interviews	3 follow-up interviews with key informants in late 2017

Data Analysis

We focus our analysis on data-related operations at Click, from which we have extensive evidence, while we use external and externally oriented materials to validate our observations about the company's interactions with advertisers and other actors in the industry. Following a common practice in qualitative case research, the analysis began during the fieldwork in the form of weekly memos written at the end of each week of observation (totaling 14 memos). The memos captured emerging insights and reflections from the research site, helping to further focus our data collection efforts (Strauss and Corbin 1998). After the fieldwork, we first divided the material into episodes and coded the episodes using a scheme informed by our evolving ideas about data commodities and the nature of Click's business. By *episode*, we simply refer to an uninterrupted sequence of interactions that revolve around a topic or issue that is of interest from the perspective of data commodity production. For instance, the following excerpt from observation notes was coded as a "reporting systems and practices" episode:

> Head of Operations walks over to tell Member Care Manager that the subscriber count for February is 238,000 [as reported by the telecommunications partner]. Member Care Manager says that he sees slightly over 240,000. Head of Operations is ambivalent and says he needs to talk with the telecommunications partner [what counts as a subscriber].
>
> (Observation log, 4 March 2009)

This process resulted in 689 coded observational episodes and interview excerpts that we used as the basis for constructing a theoretically informed narrative on the establishment and market embedment of a new kind of advertising audience. We adopted a narrative strategy to make sense of a process (Langley 1999), documenting events and their connections from the perspective of a specific outcome that gives direction and meaning to the flow of organizational life (Abbott 2001). Coding served this process by providing a preliminary order and easy access to the empirical evidence. However, narrative analysis does not result from aggregating or otherwise relating codes to each other as the primary analytical operation. It rather links the observed practices, conditions, and events represented by the coded material together in a meaningful and causally plausible plot that is motivated by a specific outcome (Abbott 2001; Llewellyn 1999; Polkinghorne 1988). To help construct the audience-making narrative, we used temporal bracketing (Langley 1999) to chunk key operations in the process into meaningful phases so that critical issues in the production of data commodity become tractable.

INDUSTRY CONTEXT

An advertising-funded business is a classic example of what economists call a *two-sided market*. Companies (advertisers) buy time-limited slots from media companies to deliver marketing messages (advertisements) to specific groups of consumers (segments) who are expected to pay attention to the slots on the opposite side of the media platform. The value of each advertising slot is defined by the audience it attracts, that is, the quantity and quality of people who have demonstrably seen or heard the advertisements. The advertising messages are usually produced by advertising agencies, whereas the measurement and reporting of advertising success is provided by dedicated ratings companies or, as in our case, by the media company itself.

A key problem for a company aspiring to sell advertising slots based on a new technological channel is that the existence of an advertising audience is difficult to verify (Napoli 2003, 2011, 2012). Media companies cannot force people to watch advertisements, nor is there a simple way of knowing whether people are paying attention to advertisements relayed through the channel. Consequently, professional advertisers are hesitant to pay for having advertisements delivered to audiences whose size and relevant qualities cannot be reliably quantified (McGuigan 2015; Turow 2012). Such hypothetical audiences have "no reality for advertisers and, consequently, no value" on the market without measurements along accepted metrics that can validate and make the reception of advertising commensurable across media (Barnes and Thompson 1994, pp. 91–92). It follows from these premises that building a new advertising channel involves establishing the means and dimensions of its measurement as an all but necessary part of the effort.

In traditional offline media such as television, radio, and print, selling and buying advertising slots is mainly based on audience measurement panels that have well-known shortcomings related to calculating ratings points for programs (Schweidel and Kent 2010; Taneja 2013; Taneja and Mamoria 2012). The data collection is expensive and requires discipline from participating consumers who must actively document their media consumption. Internet display and later search and programmatic online advertising started to evolve in the mid-1990s with a promise to solve many of the measurement problems that plague offline advertising (Alaimo and Kallinikos 2018; Turow 2012; Goldfarb and Tremblay 2014). In contrast to traditional audience measurement panels, the data are extracted from server logs with the help of cookies placed in the web browser when a user opens a web page. This allows significantly more detailed targeting of advertisements and largely removes the need for separate data collection arrangements typical of offline channels (Goldfarb 2014).

Click's business model is made possible by the fact that telecommunications infrastructure promises even better opportunities for measuring media consumption than internet servers. A telecommunications network can largely automate the painstaking data capture that underpins audience measurement but also links every action to an individual consumer. The mobile network infrastructure consists of hundreds of components that store log files known as Call Detail Records (CDRs) of every call, click, and message relayed through the network, constantly producing large amounts of detailed data on individual subscriber behavior. While similar to server access logs used to capture people's web browsing habits, CDRs also identify the individual consumer by an MSISDN[2] that is unique to a SIM card. CDR data tokens are thus a seemingly potent resource that can provide, in principle, the full picture of what each individual subscriber does in the channel—even beyond that which is possible on the internet. A concise overview of offline, internet, and mobile advertising measurement approaches can be found in Appendix 3.

THE MAKING OF A NEW ADVERTISING AUDIENCE

Click launched its service in September 2007. The company initially offered 43 free voice call minutes and 217 text messages for eligible 16–24-year-old subscribers to build up the consumer side of the platform and an attractive audience for advertisers. On the advertiser side, there had been limited trials before the commercial launch, but a definitive validation of

---

[2] Mobile Station International Subscriber Directory Number (MSISDN).

the effectiveness of the new advertising channel required data from real consumers interacting with advertisements. In this sense, the company was initially selling a product that was hardly more than a promise that advertising messages delivered to mobile phones would carve out a receptive audience from its network subscribers. The lack of data from real advertising operations made it difficult to articulate the audience product to advertisers and convince them to try Click—despite all the potential advantages of CDRs and other data available to the company. Given the lack of an established measurement regime for the type of advertising slots Click was trying to sell, the company had to repeatedly relay advertisements to its subscribers, collect data from their reception, and try to articulate these data in ways that advertisers would recognize as a valuable audience. This is illustrated by the following excerpt from a press release:

> Click, the new mobile network for 16–24 years old funded by advertising, has signed up over 100,000 members since its launch in [country] at the end of September, 2007. . . . The ad campaigns that fund the service have generated industry leading average response rates of 29%, at a time when trust in other forms of mass advertising is falling and brands are finding it increasingly difficult to engage with young people.
>
> (Click press release, April 2008)

One and half years into the operations, Click was able to start arguing that mobile network subscribers can make a viable advertising audience, based on data from initially heavily subsidized campaigns and, for instance, by publishing marketing case studies to make new advertising opportunities salient. The size of the potential audience that was reachable through the new channel was still relatively small for the respective media market, but the rate at which audience members responded to advertising can be considered high compared to any other marketing channel. Importantly, the previously mentioned press release uses metrics such as “100,000 members” and “average response rates of 29%” to describe the audience. These measures are critical for advertisers to be able to perceive Click's subscribers as a commodified audience. By the time we entered the setting in February 2009, Click had doubled its subscriber base, and advertising sales were steadily growing.

Click's most popular advertising formats are one-way push messages to subscribers and a three-step dialogue that starts with a message requesting a reply and then provides a tailored feedback message to those who respond to the initial message. It is notable that the three-step dialogue is only possible due to the extremely granular, census-like CDR data that reveal the reception and responses for each message sent to every individual subscriber, making the mobile channel "the only media that enable direct, immediate dialogue with consumers," as a Brand Office employee argued in a workshop on the company's market positioning. Once an advertising campaign finishes, a sales manager uses an advertising reporting system to create a post-campaign report to communicate the campaign metrics to the client. The report, which consists of a detailed breakdown of advertising operations with various metrics for each message sent to and received from target subscribers, is a critical artifact in the process of verifying the commodity that was delivered to the client by the operations.

In the following three subsections, we temporally bracket recurrent operations (e.g., the production of metrics) and idiosyncratic events (e.g., a transition to a new consumer offer) to illustrate how the data commodity is produced.

Reporting Advertising Campaign Performance

A capacity to demonstrate that consumers remain responsive to advertisements is contingent on managing hundreds of mobile network elements that produce nearly one million Call Detail Records (CDRs) every day. Individual network elements store CDRs in text files that are collected daily in a large relational database where the records can be centrally accessed, aggregated, and filtered in a number of ways. The records store the time and type of the network operation, its endpoints (MSISDNs), and a few other details about an exchange between two subscribers or between a subscriber and Click's advertising engine. Whatever information an individual CDR token encodes, it is both highly decontextualized and extremely granular:

> 097369D2D7372762D31080000000000000001;1;33668741168;3322208;6;20081101004923;20081101004923;20081101004923
>
> (Call Detail Record[3])

The content of the CDR shows that the 'raw' data obtained from the network infrastructure are no more than strings of alphanumeric characters with very little relevant context. The implicit data fields are generic in nature (ID, type, MSISDN, time, etc.), which makes the data tokens individually rather meaningless to the business. Importantly, the data do not represent an audience but do represent microscopic network exchanges that lack any perceivable information about advertising campaign performance. The records can nevertheless be aggregated along dimensions represented by the data fields, allowing to observe patterns and, subsequently, to contextualize the patterns using unique IDs that the tokens carry. The ID allows for mapping the data token to an advertisement, while the MSISDN code identifies a SIM card that belongs to a specific subscriber.

The central database feeds numerous analytical tools and systems that are used by the Commercial, Members, and Operations teams. The main systems for analyzing the reception and responses to advertising messages are an in-house-built advertising engine that manages message delivery and member interactions, and two different reporting tools that create campaign reports for clients. A key feature of these systems is that they implement the

[3] The example is taken from Advenage SMS Gateway Router 1.0 documentation retrieved from http://www.advenage.com/documents/SmsGatewayRouterB2BAccounting.pdf (26 February 2011). Due to matters of confidentiality, we are not allowed to reproduce an actual CDR from the research site.

metrics for advertising reach and consumer responsiveness based on algorithms that aggregate CDRs so that contextually relevant questions can be answered from the data: *How many people received an advertisement? How large a portion of the recipients responded to an advertisement? How large is the total audience available to receive advertisements through Click?* The ability to construct convincing answers to such questions allows Click to argue that it indeed 'has' an audience. To this end, the metrics in the post-campaign report may seem technically trivial to calculate, but at the same time they entail laborious negotiations and work to counter inherent ambiguities that emerge as the data are aggregated.

Reach of advertising. The size of the audience for an individual advertisement, what is commonly defined as the reach of advertising, is seemingly simple to calculate from the data. However, on a closer inspection, it turns out that the metric can refer to at least two different things: the number of advertising messages delivered and the number of different consumers contacted. Furthermore, the two measures are different from the number of messages sent. A message may be delivered by the network infrastructure immediately, with a delay, or it may not be delivered at all to the subscriber's mobile phone. The ambiguities accumulate even further in the case of advertising formats that are based on several exchanges with each recipient. A three-step dialogue-type advertisement entails sending multiple messages to the same consumer—some of the messages are never delivered, and some that are delivered could be counted either once or more than once, depending on the interpretation of the metric. For instance, did a subscriber 'see' the advertisement if he or she received only the first part of a three-step dialogue? Due to these and other differences in underlying technologies and the data, the metrics cannot be copied from other media channels. The underlying ambiguity about defining an audience is reflected in the shifting labeling of the metric that is alternatively known as "delivered," "delivered contacts," or "contacts" in documents produced by different employees at different times.

Response rate. The response rate is the rate at which subscribers respond to advertising messages sent to their phones. Uncertainties in constructing the metric involve issues like those of establishing reach, as well as the fact that subscribers sometimes reply to advertisements that do not require a response, or they respond several times, which needs to be taken into consideration when calculating response rates. In contrast to reach, response rates can be averaged to describe Click's audience in general and are regularly included in public relations and marketing activities. The average response rate can be found in 10 out of 26 press releases, and the responsiveness of the audience is highlighted in all ten marketing case studies Click used to promote its advertising offering. Yet, it is not always clear whether the average response rate refers to the average of campaign averages or to the average of response rates for all advertisements or messages sent to subscribers asking for a reply. There is a degree of strategic ambiguity that can be used to boost published response rates.

Managing Consumer Engagement

Metrics such as reach and response rate allow post hoc articulation of the commodity delivered to an individual advertiser during an advertising campaign. At the same time, Click needs to cultivate its subscribers' willingness to participate in the audience-making process and act as audience members for future campaigns. Indicative of the relationship the company aspires to build with subscribers, subscribers are referred to as members, as if the subscribers were a part of a community or a club. "Member" is the preferred way to identify the subscribers in public relations and other external communications; 18 out of 26 of Click's press releases talk about members instead of more common industry concepts such as "subscriber," "consumer," or "customer." Click interacts with the members using so-called house messages that are technically identical to advertising messages. The house messages are used, for instance, to onboard new subscribers to the service, to announce service updates, to build communal features around exclusive content, and to prod inactive members, often based on ongoing analysis of subscriber behavior. For these purposes, the central database combines other data sources that allow to study member behavior beyond metrics describing the audience. The most important tool for understanding members is a customer insight tool that is mainly used by Members team employees. The last 30 days of CDR data are replicated daily from the central database to the tool for faster access, analysis, and visualization, whereby a Members team employee further transfers the data to a local database for more custom analytics in a spreadsheet application or a statistical software.

We observed firsthand how the data production and analytics arrangement was put to the test when Click launched a new consumer offering during our fieldwork, on 17 February 2009. The aim of the operation was to cut costs in a difficult economic environment created by the global financial crisis. The new offer replaced the original offer with monthly 20 USD top-ups that a subscriber could use flexibly for voice calls and text messaging, instead of fixed quotas for separate services. The company also reduced the availability of telephone support to members and steered them to solve problems in online discussion forums on Click's website. All existing subscribers were to be transferred to the new mode on their monthly refill day. The impact of the change on audience members was difficult to predict in advance as the company had no experience from earlier transitions. Instead, Members and Operations teams used the available, nearly real-time analytics tools to constantly monitor the member base for patterns that could signal potential problems.

Top-up behavior. The number of top-ups, in other words, the number of subscribers purchasing extra credits, increased as a result of the new offer. This could be expected as the new offer was less generous than the previous one. However, inexplicable patterns also emerged in the data that attracted the attention of the Member Care Manager, as there appeared to be negative top-up events that removed credit from a member's account. The manager discovered that the Technology team had made a change to the CDR format along with the rollout of the new offer, which had not been carried over to the customer insight tool used by the Member Care Manager. This had left parts of the analytics arrangement out of sync with the data, which made it difficult to know how to interpret the plots and measures created by the tool. It took several back-and-forth queries over the following days and tense discussions between Members, Operations, and Technology team employees to establish the facts of the situation. Finally, a careful assessment of the top-up behavior revealed that there had previously been a problem hidden in the processing of top-ups. Negative top-up events resulted from a subscriber trying to top up with a credit card without enough credit. As a result, the money was first credited and then pulled from the account, which turned out to be the correct operation. Eventually, the employees came to realize that subscribers could have previously received free top-ups by paying with a credit card without credit as the retraction mechanisms did not work earlier.

Secondary SIM usage. The data also revealed that a portion of members use their free credit and then disappear from the network until the next refill day. The Member Care Manager interpreted this to mean that some people use Click as their secondary SIM card, which would mean that the number of SIM cards sent to subscribers or even the number of monthly active SIM cards connected to the network are not good measures of the size of the total audience. According to the Member Care Manager, Click has 150,000 primary SIM holders. He also expected secondary SIM usage to increase as a result of the new offer, which would have further consequences for the size of the total audience reachable through Click.

Optimizing the Interaction Between Advertisers and Subscribers

Click does not own the mobile network infrastructure but instead leases transmission capacity, subscriber management, and related network services from a traditional telecommunications operator. As a result, the company incurs a monthly cost for every subscriber whether they receive advertisements or not, whereas inactive subscribers are less problematic for a traditional telecommunications operator who owns its infrastructure. This was illustrated by episodes such as the following:

> The Member Care Manager says that a regular telecommunications operator does not care if the customer is away from the network for couple of weeks. . . . It may be seen as lost revenue, but the inactive customer does not create costs so there is no need to try activate the customer. For us [Click] the customers are an audience we need to connect with.
>
> (Observation log, 24 March 2009)

Click’s business model requires that its subscribers are available to receive advertisements, which is stipulated in a terms and conditions document the subscribers accept upon joining the service: “To remain a Click member you need to keep the SIM Card in a Capable Phone with the correct settings activated. . . . If after being sent notice by Click you continue to receive less than half of the messages sent to you, your Refill may be suspended and your membership may be terminated.” A subscriber is expected to make sure he or she can receive advertisements, and the company reserves a contractual right to terminate the subscription if the subscriber does not fulfill the minimum requirements. The rules had not been enforced before, but they created a space for cost-saving interventions that Members and Operations teams started to explore. The main data source for this purpose was again the central database with CDR, subscriber profile, and other data. The teams used statistical software and visualization tools to mine information that could be used to make interactions between subscribers and advertisers more efficient.

Attempts to optimize the media platform were more idiosyncratic and nonroutine than operations related to producing the audience and understanding subscriber behavior, and they involved the use of more sophisticated analytical techniques. Regular operations had to mainly rely on preprogrammed metrics and pipelines of data replication and aggregation, whereas the optimization efforts typically entailed customized queries to the central database to pull data into general-purpose statistical packages to predict what could make operations more efficient. This required technical knowledge and a good understanding of the massive central database structure. Senior employees in the Operations team were typically the ones who had the required technical skills and could access the database directly. According to an Operations team employee, they were also concerned that the replication of data and preprogrammed analytics could accumulate errors when moving further away from the data source.

The size of the total audience. An employee from the Technology team pointed out that "this is a very good shock when our people see how many active members we really have. About 120,000 [are] really active, while some people speak about over 400,000." The comment illustrates a recurrent theme in internal discussions about how different people in the company had diverging views about such a central figure as the number of active audience members. The number could be inferred neither from the reach of individual advertisements nor from the number of SIM cards sent to subscribers. Also, simply establishing the exact number of currently active subscribers was not easy. An episode shown in the excerpt found in the research design section (see above) from early March 2009 reveals that Click and the partner telecommunications operator calculated the number of active subscribers variously in their systems. In the episode, the Head of Operations and the Member Care Manager discovered to their surprise that the parties came up with the different number of monthly active subscribers. There are several potential reasons for the discrepancy that was in the end settled by deciding on a common algorithm with the telecommunications partner. Yet, a subscriber or even an active subscriber is different from an audience member who is available to receive advertisements, which depends on the time when the advertisements are sent. “It would make sense to send advertisements mostly on weekdays as there is more traffic in the network,” suggested an Operations team employee. Ultimately, the size of the total potential audience that the company could rely upon was not a stable, unequivocal number but depended on the interpretation of several factors that shift constantly.

Removing unprofitable customers. To reduce operational costs, Click decided to terminate subscribers that did not receive any advertisements. These subscribers generated costs without adding anything to the potential audience from which the company carved out actual audiences for individual advertisements. Inactive subscribers would first be warned and then terminated if they did not reappear in the network. This resulted in tedious discussions about how to select the subscribers who would be terminated, revealing the difficulties associated with establishing unambiguous facts on the matter and having employees who perceived audience members differently. People from the local sales office were reluctant to terminate relationships with subscribers, whereas the Members team at the head office perceived the intervention with the subscriber base primarily as a technical operation targeting database entities. As a consequence, a Members team employee sarcastically described suggestions from the sales office as attempts to come up with complex rules that would terminate nobody. At the same time, the operation was urgent and had to be executed early in the month so that the targeted subscribers would not have time to make a CDR and become counted as active subscribers by the telecommunications partner.

ANALYSIS

We found that producing a commodity from data is a complex technical and organizational process that extends beyond the standard view of data as more or less faithful representations of external facts described, for instance, in representation theory (Burton-Jones et al. 2017; Wand and Weber 1995) and the understanding of analytics as data integration and the application of statistical techniques. This is a key message we derived from our empirical narrative, which resonates with earlier studies that have shown how "data must be narrated to be made to work" (Dourish and Cruz 2018, p. 5). To address our first set of research questions concerning data management and its linkages to the wider data environment, we identify three types of practices that are essential in supporting the production of data commodities: *metrics production*, *data-based improvisation*, and *data analytics projects.* These represent different ways in which data sources, analytical tools, and organizational practices are co-configured to support the production of a data commodity and to deal with or alleviate constant threats to the validity of the commodity. The practices are summarized in Table 3.

Datawork Practices

Table 3. Types of datawork practices

Type	**Function**	**Threat to validity**	**Empirical examples**
Metrics production	Produces facts about a data commodity	Inability to maintain the stability of essential metrics in all situations	• The reach of advertising • Response rate
Data-based improvisation	Provides a capacity to react to ongoing changes in a data commodity	Data, analytics tools, and work practices drifting out of sync	• Top-up behavior • Secondary SIM usage
Data analytics projects	Employs advanced analytics to optimize the production of a data commodity	Misunderstanding the underlying infrastructure and data generative processes	• The size of the total audience • Removing unprofitable customers

Metrics production. The reach of advertising and response rates documented in the post-campaign reports are, for all practical purposes, facts that describe the commodity. Even though these may seem statistically trivial figures, our findings reveal them to be replete with inconsistencies and ambiguities that need to be addressed in practice. Using big data as the raw measurements entails assumptions about what and how to measure and may further require establishing the same measure through different routes of calculation. At the same time, the validity of indicators that quantify the commodity along relevant dimensions is threatened by discrepancies that can easily appear in the construction of metrics as a result of changes to the data or the formula used to aggregate the data, or by shifts in the interpretation of these. The production of industry-accepted measures thus calls for constant stabilization. This requires special attention to counter various destabilizing forces such as employee turnover, discrepancies in aggregation practices, changes to data and software systems, alternative interpretations, and differences between contractors.

Data-based improvisation. We observed how a transition to a new consumer offering triggers various reactions from subscribers that Click needs to attend. These are difficult to predict in advance due to a lack of historical data and the evolving nature of the setting. Instead, the employees use analytics tools to constantly observe relevant aspects of subscriber behavior, such as top-up behavior and secondary SIM usage, which allows them to quickly react to emerging issues. A threat to this capacity arises, however, from necessary changes to the data and the systems that can easily put the data, tools, and human practices out of sync. Consequently, employees are not always sure if they can trust the information seen through their systems, causing delays, frictions, and ambiguity in decision making. It is important to stress that keeping individual parts of the analytical arrangement in sync is different from metrics production. Here, stability results from concerted changes to several components as the business evolves, whereas the calculation of metrics describing the data commodity needs to be protected against any changes or deviations across various systems and data collected at different times.

Data analytics projects. Finally, we observed practices such as assessing the size of the total audience and removing unprofitable subscribers from the network that combine both descriptive and predictive analytics as projects of varying size and sophistication. Together with the use of analytics to allow ongoing improvisation in the evolving business, these projects form a basis for making decisions about cultivating and expanding the subscriber base and optimizing interactions between consumers and advertisers. Such work is an important precondition for the type of business in which data production is closely tied to the active participation of users. From this perspective, a threat to the success of data analytics projects emerges from the fact that one must know the infrastructure and each data source well or risk using wrong or misinterpreted data. While featuring sophisticated techniques for extracting information from large volumes of data, such methods and techniques need at the same time to cope with and calibrate their use to the technological, organizational, and industry contingencies that underlie the business.

Building Data Commodities

If data commodities such advertising audiences would simply exist ‘out there’—waiting to be represented by data and analytics—establishing them as commodities would hardly require the kind of interpretive effort and persistent ambiguity management that we find in the case of Click. Important information such as the reach and the response rate of advertising or the size of the total audience cannot be simply read off from the data but must be established against a relevant organizational and industry context. Simple metrics entail complex predilections and negotiations with respect to what facts are captured or reported, how to aggregate data together, and several interpretations concerning what data, and the data-based objects they help construct, exactly refer to in the industry (Dourish and Cruz 2018). We found that the management of ambiguity in the big data age (Martin 2015) requires constant negotiations on the meaning and value of data not only among different teams within the organization but also between the organization and several external actors. The embedment of the making of data commodities in a wider industry context makes the current case different from earlier studies in which issues with data are largely seen as internal to the organization (e.g., Burton-Jones 2014; Kallinikos 1999; Zuboff 1988). In this sense, the study of data commodities requires a perspective that transcends the perception of organizations as bounded and self-contained entities (Winter et al. 2014).

Maintaining data-based objects requires keeping several data sources, analytical tools, and organizational practices in sync, which poses diverse coordinative challenges when parts of the organization and its environment are in flux. The ambivalent nature of data commodities (Kallinikos et al. 2013) is further emphasized by an increasingly complex and rich infrastructure made by different systems, applications, standards, techniques, and tools, whereby the same consumers can give rise to different audiences in different media. Taken together, these observations suggest that the audience as a data commodity is being made possible not just by the media company and its partners but also by the wider data environment in which these actors are embedded. An advertising audience is a data-based object that is projected from relations between data to the reality just as much as that reality is reflected in the data.

These observations address our first set of research questions by showing that advertising audiences or, more generally, data commodities are products of embedded data management and analytics practices. Most of these practices are conditioned and shaped by the problem of how to lend facticity and, ultimately, confer value to an audience commodity that virtually has “no tangible existence outside” (McGuigan 2015, p. 204) the data relations by which it is conveyed and the industrial ecosystem by which and for which it is produced. The apparent simplicity of a great deal of data commodities belies the constant backstage work by which such commodities are produced. The value of data-based objects as commodities depends on persistent *datawork* that repairs, frames, calibrates and amends the sheer production of quantified data and the indicators by which data-based objects are given their qualities.

TOWARD A PROCESS THEORY OF DATA COMMODIFICATION

Working from the original data tokens toward the making of a data commodity is a protracted journey marked by various steps and practices that act upon, qualify, add, verify, and repurpose the data and the metrics provided by algorithms embedded in the systems. Before they are transformed into commodities, data first have to be produced. The production of data entails the encoding of incidents or events 'out there' into what the organization recognizes as data. Data tokens such as CDRs can then serve different purposes by virtue of being amenable to transformative (analytical) operations, resulting in more complex data-based objects such as records of audience members and, ultimately, audiences. The transformations data undergo along their way to becoming data commodities are, to use Knorr-Cetina's words, *decision impregnated* (Knorr-Cetina 1981, 1999; Knorr-Cetina and Bruegger 2002), which is to say that there is no automatic pilot in this journey. Instead, the process is governed by battles with data ambiguities, discordant interpretations on what data convey, and negotiations concerning how to respond to the changing standards of the big data environment.

To address our second set of research questions concerning the nature of transformations that data undergo and how these are linked to the attributes of digital data, we first summarize the stages through which data must pass to become commodities in Table 4 and subsequently discuss the operations that underlie each stage. We then advance what we call a process theory of data commodities that links observed practices and operations to the digital nature of data and their basic attributes of editability, portability, and recontextualizability.

Table 4. The production of data commodities			
Stage	**I Data tokens**	**II Data-based objects**	**III Data commodities**
Illustrative entity	Call Detail Record (CDR)	Records of audience members	Advertising audience
Key organizational operation	Coordinate the production of data	Agree upon metrics and analytical models	Align data-based objects to business objectives
Examples of typical tasks and practices	• Selecting and formatting events into data • Merging data from different sources and filtering and cleaning • Copying data between multiple systems • Adjusting the main data token format across all analytical tools and systems	• Aggregating and structuring data into data-based objects • Defining metrics across organizations and an institutional field • Negotiating between partners on the definition of metrics and ways to implement them • Ensuring that the metrics are calculated the same way by different systems • Monitoring, adjusting, and optimizing metrics and algorithms iteratively	• Emphasizing the defining qualities of the commodity based on metrics and contextual data • Maximizing gain by resizing, cleaning, adding, or merging data and objects • Optimizing the packaging of commodities and learning from user feedback

The Production of Data Tokens

Our findings align with the observation in the literature that data are human-made products that are not found but are actively generated or acquired (Gitelman 2013; Jones 2019; Wang et al. 1996). The process of data production applies both to the generation of records out of events as well as to the repurposing of already existing records such as CDRs. To become data, events yet to be recorded or existing records need to be imagined as data for some purpose (Aaltonen and Penttinen 2021; Alaimo and Kallinikos 2017, 2020; Gitelman 2013). Repurposing data amounts thus to recontextualizing extant records in one's own operations with a view of them serving new ends. Because of the open-endedness of data, the process requires substantial articulation work (Suchman 1996; Star 1995) and coordination across different units or even across organizational boundaries, as the selection and formatting of events is often distributed across devices, systems, and infrastructures. These articulation and coordination efforts acquire characteristics that are bound up with the specific properties of the digital medium. Different systems, in other words, can transform the same events into different data.

The making of data implies the selection of an idiosyncratic contextual event (e.g., a phone call or a text message) and its representation in a format that allows abstracting and generalizing it beyond its context. While the event is defined by its context, a data token is defined by its standardized format and largely the absence of context. Transforming an event into a standardized record, in this sense, establishes the semiotic properties of data and circumscribes the opportunities and limitations of data as resources. The process of capturing an event as a digital data token, together with the affordances of the digital medium, confers to the facts carried by data their properties of editability and portability. Stripped of the contextual cues and of the significance events have for individual users, data also afford recontextualizability. However, a number of trade-offs come with these opportunities.

Editability easily leads to loss of relevance, instability, and epistemic (knowledge) uncertainty that become expressed as data inconsistencies or duplications, missing records, corrupted fields, and other ambiguities that are often solved by adding new data or control systems (Hanseth and Ciborra 2007) and additional datawork. The transmission and proliferation of data can then lead to further duplications, inconsistencies, and problems, which have a significant and often hidden role in steering additional data-related practices. As a result, a considerable amount of work needs to go into coordinated acts of interpretation that remain bounded to the instability and epistemic uncertainty of digital data as bearers of facts (Alaimo et al. 2020; Alaimo and Kallinikos 2020; Zuboff 1988). These interpretation practices are linked to the semiotic nature of digital data as unstable representations (Burton-Jones 2014, p. 92) and are conditioned by the process and medium through which they occur (Bailey et al. 2012).

The Making of Data-based Objects

Once events are transformed into data, a number of operations become possible. Enabled by the editability of data, abstract tokens such as CDR instances or, to refer to other contexts, likes or tags in social media, can be easily brought together and combined with other data into data-based objects. This happens by means of aggregation, that is, putting together and structuring different instances of the same data type and different data types together to construct more complex objects (Alaimo and Kallinikos 2017, 2020). Data-based objects are structured data entities made of variously aggregated data, which can further serve new sets of practices, tools, and data analytics operations. For instance, aggregating data along the time dimension of CDRs creates new data objects that make visible temporal patterns and afford additional possibilities of reading and interpretation, such as frequency, response rate, peak time of response, and so on. Aggregation in this respect is seldom just assembling data together but a fundamental operation of meaning-making that renders certain data attributes visible and relevant while, at the same time, hiding or making others less evident. We use *aggregation* as an umbrella term indicating a number of technical operations and organizational practices that lead to the production of objects that, even if data-based, are treated as 'things' in organizational and industry settings (Desrosières 1998).

The creation of structured objects out of data marks an important step in the production of data commodities. Data-based objects such as audience members, users, or customers afford a different kind of editability and become portable when organizations operating in a particular field agree upon the metrics that describe and stabilize these objects. To be portable across organizations, data-based objects need to be recognized by the actors in the field. To achieve this, organizations need to negotiate the adoption of metrics and their meanings, eliminate ambiguities, and build consensus across the industry or ecosystem in which they operate (Alaimo and Kallinikos 2020; Mackenzie 2012). Shared metrics perform an important role as they confer to data-based objects external stability or facticity that they inherently lack (Alaimo and Kallinikos 2018; Espeland and Sauder 2007; Mackenzie 2012). In this sense, metrics both presuppose and create data-based objects, by lending facticity to entities that have a precarious existence outside the production process that makes them visible (Alaimo and Kallinikos 2018).

When metrics are successfully negotiated in an industry or a field of practice, a complex institutional network is established that provides the definitions and criteria by which data-based objects become portable, that is, shareable in a relevant community. As the calculation of a metric (e.g., the reach or the response rate of advertising) becomes common practice in the field, it acts as the verification of data-based objects and supports their subsequent insertion into marketplaces qua commodities. Negotiating and agreeing upon metrics is thus an essential organizational practice at this stage. The metrics produce exportable objects that make sense in a given domain (e.g., viewable impressions in the programmatic advertising industry or similar artists in the online music platforms). Metrics take the existence of objects as given and make visible the qualities of objects as, for instance, rates, ratio, size, similarity, viewability, and so on. With the stability gained by various metrics or models, data-based objects acquire portability within an industry and thus become shareable objects around which organizational, industry, and market practices can unfold.

Becoming a Data Commodity

The final phase in the production of data commodities happens when data-based objects are turned into goods that are exchanged on a market and thus assigned monetary value. This is variously conditioned by the operations we have analyzed previously and, in many cases, remains only an incidence of a broader and, crucially, recurrent open-ended process of data production, repurposing and aggregation we have just described (see also Alaimo et al. 2020). The outputs of such a process cannot but be commodities of ambivalent and fluid existence, such as digital advertising audiences (Napoli 2003, 2011, 2012), personalized suggestions, rates or popularity indexes and scores, or digital profiles (Ekbia 2009; Kallinikos et al. 2013). Many of the datawork practices we reported earlier are responses to the intrinsic instability of digital data as bearers of facts that have to be settled through work that is embedded in organizational settings and particular industrial contexts.

In contrast to other transformations, the commodification of data-based objects must be performed for specific individuals, stakeholders, or groups while conforming to broader market practices. In this regard, the transformation of data-based objects into data commodities coincides with their recontextualization. It is the interpretation of data-based objects against a new context (i.e., a market opportunity, a client need, a user behavior) that allows their commodification. In many settings and industries, the last phase of producing data commodities, which involves contextualizing and tailor-packaging objects, happens in real time and is increasingly automated (Alaimo and Kallinikos 2020). The automation-driven commodification of data objects implies that a set of operations and organizational practices needs to be in place to standardize and streamline the ad hoc recontextualization of data commodities. Often, recontextualization is the result of preexistent practices and agreements, organizational aims and objectives, and real-time data (on users, clients, or partners). Commodification gives market value to data objects through a standardized, semiautomated and real-time recontextualization. This last aspect largely confers to data commodities their highly contingent character and short life span.

What we have called datawork and the practices of data management, analytics, and commodification (the establishment of their market value) are core activities of an ever-growing big data industry. They are also tightly linked to the representing and signifying power of digital data and their inherent ambiguity. The process of data production and commodification we have described essentially consists of establishing that something as abstract as a data token or a data-based object can be a bearer of facts that express meaning and value. Today, this process is being institutionalized into several practices that take place across an industry in which a great deal of actors constantly negotiate routines, metrics, and standards; exchange data and data-based objects, and agree on the market value of new data commodities. As a result, data that could previously be considered largely context-specific and unambiguous (e.g., health data or hard science data), can now be easily ported, repurposed and repackaged. Along this journey, data can lose their proximity to the facts they originally signified or represented (Martin 2015). In different settings, digital data are increasingly used to refer to entities that lack immediate reality anchorage (e.g., audiences, consumer preferences, and so on). Such data, which are often derived from user-based domains, can undergo heavy manipulation by data management platforms or brokers, and then be valued and exchanged as novel 'facts' (Alaimo and Kallinikos 2017, 2020; Martin 2015).

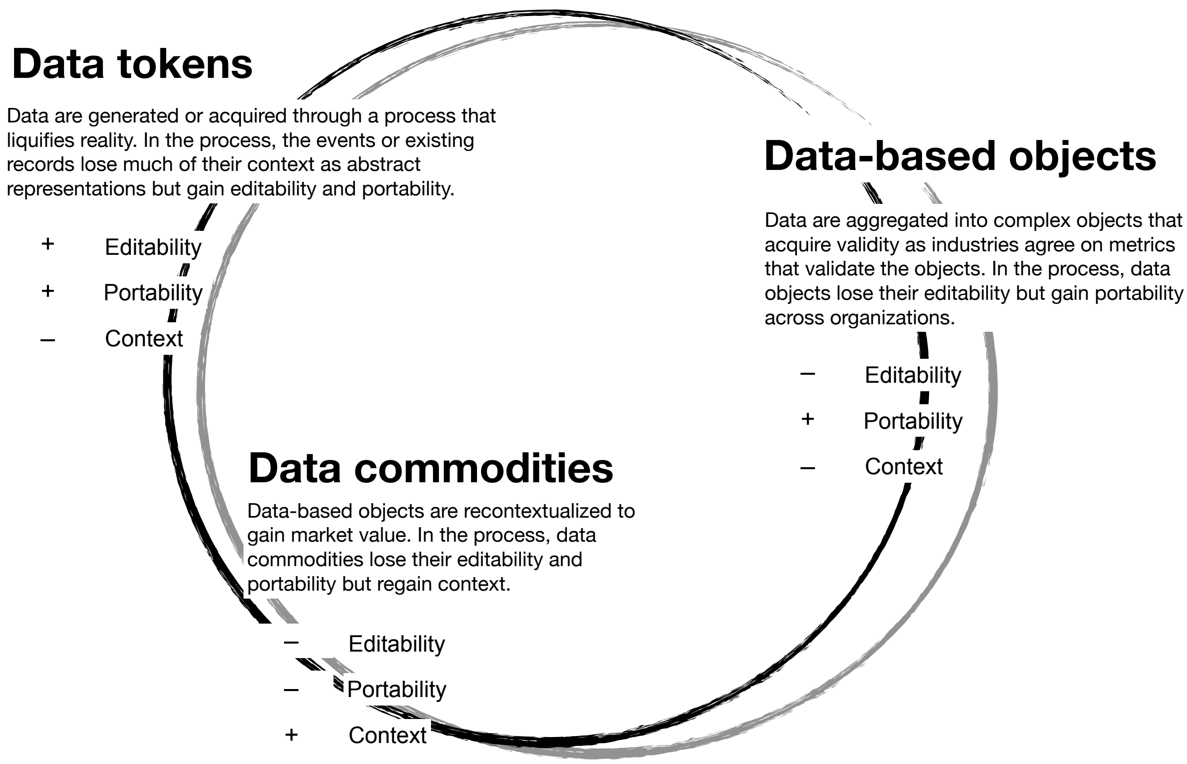

Figure 1. A process view of data commodities

Figure 1 summarizes a process view of data commodities that breaks with a monolithic conceptualization of data as given entities and instead looks at data and their characteristics at the different stages of commodity production. This foregrounds how the editability, portability, and recontextualizability of data and data-based objects are not fixed properties but are established and articulated by relentless datawork that allows the data to signify and acquire value. To begin with, data are not found but are produced, and their journey toward becoming commodities is underpinned by constant and iterative transformations. Data tokens become data-based objects when aggregated and structured for a purpose, whereas objects can revert back to data and be repurposed into new objects. In contemporary digital ecosystems, the delivery of data commodities is increasingly being automated and maintained by the massive production of additional data. In the process, data gain and lose properties depending on how they are managed. When data are aggregated into objects, they present a further set of problems and opportunities that often need industry-level resolution. Finally, data objects are commodified by real-time automated processes that optimize their relevance and package their value on the basis of constant feedback from users or consumers.

CONTRIBUTIONS TO THEORY AND PRACTICE

Our study makes several contributions to the literature. First, we complement and extend current research into the ambiguities of data management and analytics (e.g., Jones 2019; Lycett 2013; Monteiro and Parmiggiani 2019; Passi and Jackson 2018) by identifying three types of datawork practices through which data are made to matter and data ambiguity is managed in organizations. Metrics production, data-based improvisation, and data analytics projects are critical in embedding analytical operations to the model of value creation. The typology is simple enough so that it can help organizing datawork in practice, and shows that understanding the production of data commodities requires going beyond objectivist assumptions underpinning much of the data modelling and data quality literature (Burton-Jones et al. 2017; Hazen et al. 2014; Hirschheim et al. 1995; Wand and Weber 1995).

Second, we have conceived data commodities as valuable data-based objects that are projected from relations between data to the reality as much as that reality is represented by the data. The process of making such commodities is embedded in complex and distributed data environments whose dynamics change traditional assumptions about data and their production, and the role of individual actors in the emerging digital ecosystems. Examining how data are made to matter entails focusing on how the management of data ambiguity and the interpretive work underlying it transcend the confines of individual organizations and render work with data a complex and protracted journey that involve multiple actors in the value chain (Abbasi et al. 2016; Martin 2015; Winter et al. 2014).

Third, building on our findings, we develop a process theory of data commodification that links the management of data and their interpretation practices to the editability, portability, and recontextualizability of digital data. The processual approach captures the different stages through which data are produced, repurposed, aggregated and packaged into goods that have organizational, industry, and market relevance. The framework we advance here can give market participants and practitioners, regulators included, a novel entry point into data management practices by virtue of distinguishing between different stages in the process and showing how each stage entails a different set of practices, tools and strategic decisions. More generally, framing data commodities as processual entities opens opportunities to study practices and technologies along with the big data industry actors that make up the environments in which data commodities develop. The framework further allows looking at data characteristics, and their related risks and opportunities, as dynamically articulated by datawork.

No study is without limitations. The bulk of the empirical evidence we rely upon was collected ten years ago, yet we believe that our analysis of datawork practices and contributions to the making of data commodities are novel and maybe even more timely today due to the rapidly advancing datification of economy. What has changed, however, is the rise of various data aggregators that make a business out of hoarding and packaging data for sale. Studying the making of data commodities in the context of big data industry and the aggregation and commodification practices it enables offers interesting future research opportunities. Also, the case selection in a single case study research design can raise the question of external validity—does knowledge gained by studying a new type of mobile advertising audience apply beyond the specific type of advertising? To this end, we have theorized the process through which data commodities are produced, which can be tested and extended through studies in other settings. Finally, one might argue that observed ambiguities in interpreting data can result from a mere failure to apply processes, methods, and systems that have been developed over several decades to manage and analyze data. While this may be sometimes true, the difficulties can also signal a creative struggle to produce something new out of the data. There is a well-established research tradition that studies issues of the former kind, whereas the current study suggests that the latter represents a largely untapped opportunity to contribute to management theory and practice.

CONCLUSIONS

Our study shows that producing data commodities is a complex sociotechnical practice through which actors seek to disambiguate, interpret, and manage cognitive outputs of the data environments and analytics that increasingly pervade contemporary organizations. To this end, it is important to recognize the embedded nature of the data integration and analytics processes, that is to say, the multivalent datawork practices through which metrics and other sorts of analytical outputs are made to matter in a specific setting. While pursuing this line of inquiry, it is necessary to pay due attention to the semiotic properties of digital data and other recurrent and structural attributes of the data processing arrangements (i.e., big data industry, data commodification) that confer to these practices their distinctive profile (Leonardi and Barley 2010). Ultimately, our research contributes to reinserting the insights of a sociotechnical approach in IS and management into the current age of datafication. A process theory of data allows to investigate data as a dynamic resource whose meaning and value have to be constantly calibrated, streamlined and managed (Aaltonen and Penttinen 2021; Alaimo and Kallinikos 2020; Alaimo et al. 2020; Sarker et al. 2019).

REFERENCES

1. Aaltonen, A.; and Penttinen, E. What Makes Data Possible? A Sociotechnical View on Structured Data Innovations. In, *Proceedings of the 54th Hawaii International Conference on System Sciences (HICSS)*, 2021, pp. 5922-5931.
2. Aaltonen, A.; and Tempini, N. Everything Counts in Large Amounts: A Critical Realist Case Study on Data-based Production. *Journal of Information Technology*, 29, 1 (2014), 97-110.
3. Abbasi, A.; Sarker, S.; and Chiang, R. H. K. Big Data Research in Information Systems: Toward an Inclusive Research Agenda. *Journal of the Association for Information Systems*, 17, 2 (2016), i-xxxii.
4. Abbott, A. *Time Matters: On Theory and Method.* Chicago: The University of Chicago Press, 2001.
5. Alaimo, C.; and Kallinikos, J. Computing the Everyday: Social Media as Data Platforms. *The Information Society,* 33, 4 (2017), 175-191.
6. Alaimo, C.; and Kallinikos, J. Objects, Metrics and Practices: An Inquiry into Programmatic Advertising. In, Schultze, U.; Aanestad, M.; Mähring, M.; Østerlund, C.; and Riemer, K. (eds.) *Living with Monsters? Social Implications of Algorithmic Phenomena, Hybrid Agency, and the Performativity of Technology,* IFIP 8.2 International Federation for Information Processing, IS&O, IFIP Advances in Information and Communication Technology, vol 543, Cham: Springer, 2018, pp. 110-126.
7. Alaimo, C.; and Kallinikos, J. Managing by Data: Algorithmic Categorization and Organizing. *Organization Studies*, online first (2020).
8. Alaimo, C.; Kallinikos, J.; and Aaltonen, A. Data and Value. In, Nambisan, S.; Lyytinen, K.; and Yoo, Y. (eds.) *Handbook of Digital Innovation*, Cheltenham: Edward Elgar, 2020, pp. 162-178.
9. Ali, S. M. F.; and Wrembel, R. From Conceptual Design to Performance Optimization of ETL Workflows: Current State of Research and Open Problems. *The VLDB Journal,* 26 (2017), 777-801.
10. Baesens, B.; Bapna, R.; Marsden, J. R.; Vanthienen, J.; and Zhao, J. L. Transformational Issues of Big Data and Analytics in Networked Business. *MIS Quarterly*, 40, 4 (2016), 807-818.

11. Bailey, D. E.; Leonardi, P. M.; and Barley, S. R. The Lure of the Virtual. *Organization Science*, 23, 5 (2012), 1485-1504.
12. Bansal, S. K. Towards a Semantic Extract-Transform-Load (ETL) Framework for Big Data Integration. In, *2014 IEEE International Congress on Big Data*, IEEE, 2014, pp. 522-529.
13. Barnes, B. E.; and Thomson, L. M. Power to the People (meter): Audience Measurement Technology and Media Specialization. In, Ettema, J. S.; and Whitney, D. C. (eds.) *Audiencemaking: How the Media Create the Audience,* Thousand Oaks: SAGE Publications, 1994, pp. 75-94.
14. Bechmann, A.; and Bowker, G. C. Unsupervised by Any Other Name: Hidden Layers of Knowledge Production in Artificial Intelligence on Social Media. *Big Data & Society,* 6, 1 (2019), 1-11.
15. Brynjolfsson, E.; and McAffee, A. *The Second Machine Age.* New York: W. W. Norton Company, 2014.
16. Burton-Jones, A. What Have We Learned from the Smart Machine? *Information and Organization,* 24, 2 (2014), 71-105.
17. Burton-Jones, A.; Recker, J.; Indulska, M.; Green, P.; and Weber, R. Assessing Representation Theory with a Framework for Pursuing Success and Failure. *MIS Quarterly*, 41, 4 (2017), 1307-1333.
18. Chen, H.; Chiang, R. H. L.; and Storey, V. C. Business Intelligence and Analytics: from Big Data to Big Impact. *MIS Quarterly*, 36, 4 (2012), 1165-1188.
19. Clarke, R. Big Data, Big Risks. *Information Systems Journal*, 26, 1 (2016), 77-90.
20. Desrosières, A. *The Politics of Large Numbers: A History of Statistical Reasoning.* Cambridge: Harvard University Press, 1998.
21. Dourish, P.; and Cruz, E. G. Datafication and Data Fiction: Narrating Data and Narrating with Data. *Big Data & Society,* 5, 2 (2018), 1-10.
22. Eisenhardt, K. M. Building Theories from Case Study Research. *Academy of Management Review*, 14, 4 (1989), 532-550.
23. Eisenhardt, K. M.; and Graebner, M. E. Theory Building from Cases: Opportunities and Challenges. *Academy of Management Journal*, 50, 1 (2007), 25-32.
24. Edmondson, A. C.; and McManus, S. E. Methodological Fit in Management Field Research. *Academy of Management Review*, 32, 4 (2007), 1155-1179.
25. Ekbia, H. R. Digital Artifacts as Quasi-objects: Qualification, Mediation, and Materiality. *Journal of the American Society for Information Science and Technology,* 60, 12 (2009), 2554-2566.
26. Espeland, W. N.; and Sauder, M. Rankings and Reactivity: How Public Measures Recreate Social Worlds. *American Journal of Sociology*, 113, 1 (2007), 1-40.
27. Faulkner, P.; and Runde, J. Theorizing the Digital Object. *MIS Quarterly*, 43, 4 (2019), 1279-1302.
28. Gitelman, L. (ed.) *"Raw Data" is an Oxymoron*. Cambridge: The MIT Press, 2013.
29. Goes, P. B. Big Data and IS Research. *MIS Quarterly*, 38, 3 (2014), iii-viii.
30. Goldfarb, A. What is Different About Online Advertising? *Review of Industrial Organization*, 44, 2 (2014), 115-129.
31. Goldfarb, A.; and Tremblay, V. J. Introduction: The Economics of Internet Advertising. *Review of Industrial Organization*, 44, 2 (2014), 113-114.
32. Grover, V.; Chiang, R. H. L.; Liang, T.-P.; and Zhang, D. Creating Strategic Business Value from Big Data Analytics: A Research Framework. *Journal of Management Information Systems*, 35, 2 (2018), 388-423.
33. Günther, W. A.; Rezazade Mehrizi, M. H.; Huysman, M.; and Feldberg, F. Debating Big Data: A Literature Review on Realizing Value from Big Data. *Journal of Strategic Information Systems*, 26, 3 (2016), 191-209.

34. Hanseth, O.; and Ciborra, C. *Risk, Complexity and ICT*. Cheltenham: Edward Elgar Publishing, 2007.
35. Hazen, B. T.; Boone, C. A.; Ezell, J. D.; and Jones-Farmer, L. A. Data Quality for Data Science, Predictive Analytics, and Big Data in Supply Chain Management: An Introduction to the Problem and Suggestions for Research and Applications. *International Journal of Production Economics*, 154 (2014), 72-80.
36. Hirschheim, R.; Klein, H. K.; and Lyytinen, K. *Information Systems Development and Data Modeling: Conceptual and Philosophical Foundations.* Cambridge: Cambridge University Press, 1995.
37. Jarvenpaa, S. L.; and Markus, M. L. Data Perspective in Digital Platforms: Three Tales of Genetic Platforms. In, *51st Hawaii International Conference on System Sciences (HICSS)*, Hawaii, USA, 2018, pp. 4574-4583.
38. Jones, M. What We Talk about When We Talk about (Big) Data. *Journal of Strategic Information Systems*, 28, 1 (2019), 3-16.
39. Kallinikos, J. Computer-Based Technology and the Constitution of Work: A Study on the Cognitive Foundations of Work. *Accounting, Management & Information Technology*, 9, 4 (1999), 261-291.
40. Kallinikos, J. *The Consequences of Information: Institutional Implications of Technological Change.* Cheltenham: Edward Elgar, 2006.
41. Kallinikos, J.; Aaltonen, A.; and Marton, A. The Ambivalent Ontology of Digital Artifacts. *MIS Quarterly*, 37, 2 (2013), 357-370.
42. Kitchens, B.; Dobolyi, D.; Li, J.; and Abbasi, A. Advanced Customer Analytics: Strategic Value Through Integration of Relationship-oriented Big Data. *Journal of Management Information Systems*, 35, 2 (2018), 540-574.
43. Knorr-Cetina, K. D. *The Manufacture of Knowledge: An Essay on the Constructivist and Contextual Nature of Science*. Oxford: Pergamon Press, 1981.
44. Knorr-Cetina, K. D. *Epistemic Cultures.* Cambridge: Harvard University Press, 1999.
45. Knorr-Cetina, K.; and Bruegger, U. Global Microstructures: The Virtual Societies of Financial Markets. *American Journal of Sociology*, 107, 4 (2002), 905-950.
46. Langley, A. Strategies for Theorizing from Process Data. *Academy of Management Review*, 24, 4 (1999), 691-710.
47. Lee, Y. W.; Strong, D. M.; Kahn, B. K.; and Wang, R. Y. AIMQ: A Methodology for Information Quality Assessment. *Information & Management*, 40 (2002), 133-146.
48. Lehrer, C.; Wieneke, A.; vom Brocke, J.; Jung, R.; and Seidel, S. How Big Data Analytics Enables Service Innovation: Materiality, Affordance, and the Individualization of Service. *Journal of Management Information Systems*, 35, 2 (2018), 424-460.
49. Leonardi, P. M.; and Barley, S. R. Materiality and Change: Challenges to Building Better Theory About Technology and Organizing. *Information and Organization*, 18, 3 (2008), 159-176.
50. Leonardi, P.; and Barley, S. R. What's Under Construction Here? Social Action, Materiality, and Power in Constructivist Studies of Technology and Organizing. *Academy of Management Annals,* 4, 1 (2010), 1-51.
51. Leonardi, P. M.; Nardi, B. A.; and Kallinikos, J. (eds.) *Materiality and Organizing: Social Interaction n a Technological World.* Oxford: Oxford University Press, 2012.
52. Llewellyn, S. Narratives in Accounting and Management Research. *Accounting, Auditing & Accountability Journal*, 12, 2 (1999), 220-236.
53. Lycett, M. 'Datafication': Making Sense of (Big) Data in a Complex World. *European Journal of Information Systems*, 22, 4 (2013), 381-386.
54. MacKenzie, D. Knowledge Production in Financial Markets: Credit Default Swaps, the ABX and the Subprime Crisis. *Economy and Society*, 41, 3 (2012), 335-359.

55. Martin, K. E. Ethical Issues in the Big Data Industry. *MIS Quarterly Executive*, 14, 2 (2015), 67-85.
56. McGuigan, L. Direct Marketing and the Productive Capacity of Commercial Television: T-commerce, Advanced Advertising, and the Audience Product. *Television & New Media*, 16, 2 (2015), 196-214.
57. Monteiro, E.; and Parmiggiani, E. Synthetic Knowing: The Politics of the Internet of Things. *MIS Quarterly*, 43, 1 (2019), 167-184.
58. Müller, O.; Junglas, I.; vom Brocke, J.; and Debortoli, S. Utilizing Big Data Analytics for Information Systems Research: Challenges, Promises and Guidelines. *European Journal of Information Systems*, 25, 4 (2016), 289-302.
59. Napoli, P. M. *Audience Economics: Media Institutions and the Audience Marketplace.* New York: Columbia University Press, 2003.
60. Napoli, P. M. *Audience Evolution: New Technologies and the Transformation of Media Audiences.* New York: Columbia University Press, 2011.
61. Napoli, P. M. Audience Evolution and the Future of Audience Research. *The International Journal on Media Management*, 14, 2 (2012), 79-97.
62. Orlikowski, W. J. Using Technology and Constituting Structures: A Practice Lens for Studying Technology in Organizations. *Organization Science*, 11, 4 (2000), 404-428.
63. Orlikowski, W. J.; and Baroudi, J. J. Studying Information Technology in Organizations: Research Approaches and Assumptions. *Information Systems Research*, 2, 1 (1991), 1-28.
64. Østerlie, T.; and Monteiro, E. Digital Sand: The Becoming of Digital Representations. *Information and Organization*, 30, 1 (2020).
65. Passi, S.; and Jackson, S. J. Trust in Data Science: Collaboration, Translation, and Accountability in Corporate Data Science Projects. In, *Proceedings of the ACM on Human-Computer Interaction*, 2 (2018), article 136.
66. Polkinghorne, D. E. *Narrative Knowing and the Human Sciences*. Albany: State University of New York Press, 1988.
67. Power, D. J.; Cyphert, D.; and Roth, R. M. Analytics, Bias, and Evidence: The Quest for Rational Decision Making. *Journal of Decision Systems*, 28, 2 (2019), 120-137.
68. Ransbotham, S.; Kiron, D.; and Prentice, P. K. Beyond the Hype: The Hard Work Behind Analytics Auccess. *MIT Sloan Management Review Research Report, 2016*.
69. Rieh, S. Y.; and Danielson, D. R. Credibility: A Multidisciplinary Framework. *Annual Review of Information Science and Technology,* 41 (2007), 307-364.
70. Sarker, S.; Chatterjee, S.; Xiao, X.; and Elbanna, A. The Sociotechnical Axis of Cohesion for the IS Discipline: Its Historical Legacy and Its Continued Relevance. *MIS Quarterly*, 43, 3 (2019), 695-719.
71. Sayer, A. *Method in Social Science: A Realist Approach.* London: Routledge, 1992.
72. Schweidel, D. A.; and Kent, R. J. Predictors of the Gap Between Program and Commercial Audiences: An Investigation Using Live Tuning Data. *Journal of Marketing*, 74, 3 (2010), 18-33.
73. Sivarajah, U.; Kamal, M. M.; Irani, Z.; and Weerakkody, V. Critical Analysis of Big Data Challenges and Analytical Methods. *Journal of Business Research*, 70 (2017), 263-286.
74. Strauss, A. L.; and Corbin, J. *Basics of Qualitative Research: Techniques and Procedures for Developing Grounded Theory*. 2nd ed. Thousand Oaks: Sage Publications, 1998.
75. Suchman, L. Supporting Articulation work. In, Kling, R. (ed.) *Computerization and Controversy: Value Conflicts and Social Choices.* San Diego: Academic Press, 1996, pp. 407-423.
76. Star, S. L. (ed.) *Ecologies of Knowledge: Work and Politics in Science and Technology*. Albany: SUNY Press, 1995.
77. Swanson, E. B. How Information Systems Came to Rule the World: Reflections on the Information Systems Field. *The Information Society*, 36, 2 (2020), 109-123.

78. Taneja, H. Audience Measurement and Media Fragmentation: Revisiting the Monopoly Question. *Journal of Media Economics*, 26, 4 (2013), 203-219.
79. Taneja, H.; and Mamoria, U. Measuring Media Use Across Platforms: Evolving Audience Information Systems. *International Journal on Media Management*, 14, 2 (2012), 121-140.
80. Tuomi, I. Data is More than Knowledge: Implications of the Reversed Knowledge Hierarchy for Knowledge Management and Organizational Memory. *Journal of Management Information Systems*, 16, 3 (1999), 103-117.
81. Turow, J. *The Daily You: How the New Advertising Industry is Defining Your Identity and Your Worth*. Yale University Press, 2012.
82. Wand, Y.; and Weber, R. On the Deep Structure of Information Systems. *Information Systems Journal*, 5, 3 (1995), 203-223.
83. Wang, R. Y.; and Strong, D. M. Beyond Accuracy: What Data Quality Means to Data Consumers. *Journal of Management Information Systems*, 12, 4 (1996), 5-33.
84. Webster, J. G.; Phalen, P. F.; and Lichty, L. W. *Ratings Analysis: The Theory and Practice of Audience Research*. 3rd ed. Mahwah: Lawrence Erlbaum Associates, 2006.
85. Winter, S.; Berente, N.; Howison, J.; and Butler, B. Beyond the Organizational 'Container': Conceptualizing 21st Century Sociotechnical Work. *Information and Organization*, 24, 4 (2014), 250-269.
86. Yin, R. K. *Case Study Research: Design and Methods.* 3rd ed. Thousand Oaks: Sage Publications, 2003.
87. Zuboff, S. *In the Age of the Smart Machine: The Future of Work and Power*. New York: Basic Books, 1988.

Appendix 1

Information Systems Serving Analytical Operations at the Research Site

The table lists analytics applications and services used at the research site.

System or application	**Description**
Advertising engine	Custom-built system for delivering and responding to advertisements
Advertising reporting (old)	Commercial application for generating reports for advertisers
Advertising reporting (new)	Custom-built system for reporting advertising performance
Central relational database	SQL queries to a relational database containing CDR data
Customer insight tool	Tailored commercial system for analyzing subscriber behavior
Customer support system	Support request ticket management system
Excel	Spreadsheet application
Google Alerts	Online news tracking service
Minitab	Commercial statistical package
Mobile website usage statistics	Custom script to generate statistics for the mobile website usage
SurveyMonkey	Online survey tool
SiteCatalyst	Website traffic analysis

Appendix 2

Interviewees and Their Organizational Roles

Date	Job title	Symbol	Age	Follow-up interview
2/20/2009	Business Manager, Member Acquisition	BMMA	28	
2/24/2009	Member Care Manager	MCM	36	
2/25/2009	Architect, Advertising	AA	36	
3/6/2009	Architect, Browsing	AB	42	
3/9/2009	User Experience, Director	UED	34	
3/9/2009	Brand Manager	BRM	36	
3/10/2009	QA Manager	QAM	36	
3/17/2009	Service Manager	SM	32	
3/17/2009	Head of Security and IP Architect	HSA	41	
3/18/2009	Chief Operating Officer	COO	N/A	
3/19/2009	Technical Support Manager	TSM	36	
3/19/2009	Senior Software Engineer	SSE	37	
3/23/2009	Business Manager, Advertising	BMA	39	
3/25/2009	Head of Communications and PR	HCPR	45	
3/25/2009	Business Manager	BM	35	
3/25/2009	Head of Member Operations	HMO	49	
3/26/2009	Head of Technology	HT	45	
4/7/2009	Finance Director	FD	37	
4/15/2009	Assistant	AS	46	
4/27/2009	Head of Legal, People, and Culture	HLPC	47	
4/27/2009	Human Resource Manager	HRM	29	
5/5/2009	Head of Strategy and Business Development	HSBD	33	
5/6/2009	Chief Executive Officer	CEO	52	
5/8/2009	Assistant	AS	46	X
5/12/2009	Chief Executive Officer	CEO	52	
5/12/2009	QA Manager	QAM	36	X
5/12/2009	Head of Operations	HO	N/A	
5/13/2009	Business Manager, Advertising	BMA	39	X
5/13/2009	Architect, Advertising	AA	36	X
5/13/2009	Chief Financial Officer	CFO	N/A	
5/14/2009	Member Care Manager	MCM	36	X
5/14/2009	Brand Manager	BRM	36	X
5/15/2009	Business Manager, Member Acquisition	BMMA	28	X
9/16/2009	Head of Brand and Design	HBD	N/A	
11/27/2017	Brand Manager	BRM	44	X (third interview)
11/29/2017	Member Care Manager	MCM	44	X (third interview)
12/1/2017	Business Manager	BM	43	X

Appendix 3
Media Measurement Regimes and Advertising

	Offline advertising	**Internet advertising**	**Mobile advertising**
Delivery channel	E.g., television, radio, print	Web browser, email	Mobile messages
Measurement approach	*Population-centric* panel studies use diaries; fixed and portable measurement devices capture data on people's media consumption.	*Server-centric* harvesting of server log and cookie data from different websites allows the construction of profiles of web users.	*Infrastructure-centric* harvesting of CDRs from the network infrastructure offers a full picture of network subscriber behavior.
Cost of data production	*Expensive*; separate arrangements from the transmission infrastructure are needed to capture the data.	*Cheap* but requires a large advertising network to solicit data for comprehensive consumer profiles.	*Cheap*, as the network infrastructure produces CDRs automatically.
Sampling	*Small but representative* samples provide information about the audience (only) at the aggregate level.	*Large*, *mixed* samples as advertising networks aggregate measurement data across services to build comprehensive picture of audiences.	*Census-like* data provides a complete picture of individual subscriber behavior in the channel.
Data granularity and coupling	*Coarse* data record media usage at the level of channel or content item at most. The data are tightly coupled to a specific measurement objective.	*Detailed but anonymous* data are not prescribed for a particular purpose.	*Detailed personal* data reveal individual behavior and interactions with each advertisement. The uses of data are not specified in advance.
Illustrative metrics	- Ratings point - Opportunity-to-see	- Impression - Click	- Reach - Response rate