

# Datastructuring

## Organizing and Curating Digital Traces into Action

Flyverbom, Mikkel; Murray, John

*Document Version*

Final published version

*Published in:*

Big Data & Society

*DOI:*

[10.1177/2053951718799114](https://doi.org/10.1177/2053951718799114)

*Publication date:*

2018

*License*

CC BY-NC

*Citation for published version (APA):*

Flyverbom, M., & Murray, J. (2018). Datastructuring: Organizing and Curating Digital Traces into Action. *Big Data & Society*, 5(2). <https://doi.org/10.1177/2053951718799114>

[Link to publication in CBS Research Portal](#)

**General rights**

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

**Take down policy**

If you believe that this document breaches copyright please contact us ([research.lib@cbs.dk](mailto:research.lib@cbs.dk)) providing details, and we will remove access to the work immediately and investigate your claim.

Download date: 27. Aug. 2022

# Datastructuring—Organizing and curating digital traces into action

Mikkel Flyverbom<sup>1</sup> and John Murray<sup>2</sup>

Big Data &amp; Society

July–December 2018: 1–12

© The Author(s) 2018

Article reuse guidelines:

sagepub.com/journals-permissions

DOI: 10.1177/2053951718799114

journals.sagepub.com/home/bds



## Abstract

Digital transformations and processes of “datafication” fundamentally reshape how information is produced, circulated and given meaning. In this article, we provide a concept of “datastructuring” which seeks to capture this reshaping as both a product of and productive of social activity. To do this we focus on (1) how new forms of social action map onto and are enabled by technological changes related to datafication, and (2) how new forms of datafied social action constitute a form of knowledge production which becomes embedded in technologies themselves. We illustrate the potential of the datastructuring concept with empirical examples which also serve to highlight some new avenues for research and some empirical questions to explore further. We suggest a focus on datastructuring can ignite scholarly debates across disciplines that may share an interest in the technological configurations, sorting activities, and other socio-material forces that shape digital spaces, but which are rarely brought together. Such cross-disciplinary conceptualizations may give more attention to how information is structured and organized, becomes “algorithmically recognizable”, and emerges as (in)visible in digital, datafied spaces. Such a concept, we suggest, may help us better understand the novel ways in which “backstage datawork” and “data sorting processes” gain traction in political interventions, commercial processes, and social ordering.

## Keywords

Datafication, infrastructures, platforms, knowledge production, content moderation, social ordering

This article is a part of special theme on Knowledge Production. To see a full list of all articles in this special theme, please click here: <http://journals.sagepub.com/page/bds/collections/knowledge-production>.

## Introduction

The ubiquity of digital technologies and processes of “datafication” shape many parts of our everyday lives (Hansen and Flyverbom, 2015; Kallinikos, 2013; Mayer-Scheönberger and Cukier, 2013; West, 2017). The concept datafication captures how human activities are converted to data which can be put to multiple uses (Mayer-Scheönberger and Cukier, 2013). This is reflected in a variety of public (e.g. security, transport) and private (e.g. production, advertising) domains of activity that have come to rely on techniques for sorting and interpreting data. In this way, human activities are rendered as digital traces that can be combined with data aggregated from multiple sources, structured via

a wide range of digital platforms, and presented “as knowledge” by using advanced algorithms and visualization techniques. Such masses of data and advanced ways of sorting, analyzing, and visualizing them constitute the foundations of a more extensive development whereby human actions and social phenomena are shaped; data is repackaged and reflected back to us

<sup>1</sup>Department of Management, Society and Communication, Copenhagen Business School, Denmark

<sup>2</sup>Stockholm Business School, Stockholm University, Sweden

### Corresponding author:

Mikkel Flyverbom, Department of Management, Society and Communication, Copenhagen Business School, Frederiksberg, Denmark.  
Email: [mf.msc@cbs.dk](mailto:mf.msc@cbs.dk)



through digital spaces. These reflections of knowledge inform action and so can be seen to shape social domains. To the degree that we even think of their existence, such data infrastructures come across as technical and relatively neutral.

Despite the rapid spread of these data-based forms knowledge production and dissemination we are only beginning to conceptualize and understand how digital transformations shape how we come to see, know and govern the world around us. Along the lines set out by Alaimo and Kallinikos (2017) and related work on data sorting (Flyverbom and Madsen, 2015; Gillespie, 2017a), we suggest that the “infrastructural, backstage datawork of social media platforms” (Alaimo and Kallinikos, 2017: 2) has largely been invisible, and has important societal consequences. We suggest that research in this area can benefit from a stronger focus on how data are structured, sorted, and curated, and how these socio-technical arrangements shape what becomes visible, knowable, and actionable. We develop this argument by highlighting extant work across diverse literatures which address different component parts of our concern, and by offering “datastructuring” as a conceptualization that allows for the cross-pollination of ideas from these literatures. Addressing these issues is especially salient because digital traces are increasingly viewed as a primary resource for value creation, influence, and knowledge production. Social scientists have a longstanding interest in investigating, understanding and conceptualizing how different information formats, such as narratives and numbers shape how we see, know, and govern the world (Czarniawska, 1997; Desrosières, 2002). In contrast, we know too little about processes by which digital traces are sorted and made actionable, as well as the societal implications of these digital transformations (Alaimo and Kallinikos, 2017; Flyverbom and Madsen, 2015; Gillespie, 2014). To this end we present an invitation to explore the structuring of data and its consequences as an underdeveloped concern in studies of the material ecosystems that dominate the internet and datafication.

By datastructures we mean configurations of digital traces that are organized and ordered in ways that allow for analysis, value extraction and connection to different forms of social activity such as commercial production or political advocacy. Our goal with datastructuring is to foreground the practices through which data get organized in digital spaces, and so highlight some of the intricate ways in which digital traces feed into social processes, are used for the purposes of (often secondary) extraction (Zuboff, 2018) and come to shape human conduct and social ordering (Flyverbom, 2016). Such processes and effects of datastructuring are partly captured by work on content moderation (Gillespie, 2017b), visibility management (Flyverbom et al.,

2016), and data visualization (Halpern, 2015). Such work explores how largely invisible and seemingly technical ways of dealing with data have implications for how individuals, organizations and societies get access to worlds and realities, curate their presence, and carry out interventions.

Through the concept of datastructuring we address issues that are central in communication theory. The advent of extensive digital transformations, including datafication and algorithmic forms of sorting, invites us to return to basic questions about our objects of study. Beyond concerns with the *contents* of communication (e.g. how messages are crafted and made salient) and the *distribution* of information (e.g. how communication processes create relations and gain traction), there is scant attention paid to what we can think of as the *conditions* of communication. This focus on the environments and other forces conditioning communication—how technological developments create new conditions for communication activities—deserves more attention, especially as this is where particularly datafication will have its most far-reaching and fundamental consequences. Put somewhat boldly, datastructuring may become more important than contents or social networks in strategic communication and related attempts to influence customers, publics, and politics. This dynamic around processes of datastructuring and conditions of communication is the central problematic for our paper: the need for new ways of conceptualizing and exploring *how data get organized in digital spaces* and *how such configurations shape knowledge production and social ordering*. Such an agenda is not particularly new, nor very precise. Nevertheless these issues warrant attention, particularly at a moment in time where we see the consequences of largely unregulated tech companies rolling out tools and services with few considerations about socio-political consequences and problematic (ab)uses of these new ecosystems. Moreover, where these issues are touched on in the scholarly literature, they remain piecemeal and disconnected, and are not presented in ways that allow for interdisciplinary and more general engagement by scholars outside more technical and specialized research communities. Our concern is that most analytical vocabularies attend to the individual interests of different disciplines and are rarely brought together. This motivates our search for an umbrella concept which is focused on understanding how data are structured and become resources for knowledge production, advocacy, and other uses beyond the immediate contexts and situations where they are produced. Our focus on datastructuring seeks to bring together insights from studies of infrastructures, platforms and algorithms, and our hope is that this broader research agenda can ignite scholarly debates across disciplines that may

share an interest in the technological architectures and configurations of digital spaces, and in the sorting activities at work.

In pursuit of this ambition, we develop the concept of datastructuring as a starting point for research that looks more closely at how data get structured, sorted and made visible, and how such configurations contribute to various kinds of social ordering. Against this backdrop we illustrate the potential value of a focus on datastructuring by suggesting a number of avenues for empirical research and conceptual articulations in need of attention. By way of illustration, we discuss how datastructuring has ordering effects in areas such as social contents, political influence, and commercial processes.

### *Towards datastructuring: An overview of relevant literatures*

Digital transformations in IT systems, social media, algorithms, and processes of datafication have fundamentally changed how data are produced and circulated in organizational settings, as well as how data are sorted out and turned into knowledge, insights, and intelligence (Amoore and Piotukh, 2015; boyd and Crawford, 2012; Flyverbom and Madsen, 2015). There are growing literatures on the margins of both communication studies and information science which partially address these issues at present. More specifically the literatures on infrastructure, platforms, and algorithms provide entry points to articulate the subtle information-related transformations that the flourishing of digital and datafied spaces entail. Providing a conceptual approach to these changes offers a way to bridge these literatures and contribute to the broader literature on the societal consequences of digital technologies and data. As Amoore and Piotukh (2015: 343) argue, we cannot understand a phenomenon such as “Big Data” without attention to “the little analytical devices without which the giant of Big Data would not be perceptible at all”. In the same manner, we need to understand the minutiae of data processing to understand digital transformations at large.

We do not claim to address a major gap in the literature here. Rather, we argue that that more can be done to launch conversations across disciplines. To pave the way for such conversations, we need a sense of where to look and who to engage with. With this article, we highlight some emergent scholarly discussions of digital transformations and processes of datafication that articulate the minutiae of what happens inside social media and related digital spaces. We seek to contribute to emergent literatures that—from very different starting points and theoretical

perspectives—highlight the technical, infrastructural and design-based conditions for the circulation and information and the production of knowledge in digital spaces. In doing so we seek to spur theoretical conversations and empirical explorations about the roles of technical, human, design-based and conceptual features of social media, and related digital spaces in the shaping of knowledge production. That is, how do features such as algorithms, metadata, and other technical ways of sorting and organizing digital information constitute assemblages that shape how we come to see, know, and govern the world? We propose the term “datastructures” to capture such infrastructural and algorithmic conditions for communication and knowledge production, and “datastructuring” to capture the forms of social action through which we can examine the underlying range of design choices and technical elements in digital spaces, such as algorithms, tools, search functions, recommendation systems, tags, likes, friends, profiles and so on. These are important to bring into view because they structure flows of information and come to guide our attention in ways that are largely invisible to users and most others.

Our focus on these infrastructural conditions draws on insights from a variety of sources and a heterogeneous set of scholarly traditions. In big fields such as communication theory and information science there are a range of approaches that address not only the contents of communication but also the conditions and other forces that shape communication. For example, by extending Bateson’s (1972) concept of “metacommunication”, Jensen and Helles (2017) point to the value of approaches that move beyond communication per se. As they put it, we need to “capture some of the implicit, yet essential conditions that make communication possible in the first place” (Jensen and Helles, 2017: 19). These are the types of concepts that we want to work with analogically in specific sub-fields. The information systems literature also addresses the questions about how data are structured and condition knowledge production (Abbasi et al., 2016; Baesens et al., 2016; Lycett, 2013). Such work seeks to characterize the nature of data, such as their volume, velocity, variety, and value (Laney, 2001; Lycett, 2013) and the quality of data (Baesens et al., 2016). But even this work is mainly focused on the systems that produce data or the processes whereby humans or organizations make sense of data, and less on how data are structured in ways that shape what emerges as seeable and knowable. Work on internet governance has taken some steps in this direction with more encompassing notions of governance which have highlighted how multiple forces and forms of action contribute to social ordering (Flyverbom, 2011, 2016; Hofmann et al., 2016), and so pushes beyond a limited focus on human actors or

organizations as sources of governance to include a focus on material forces. These types of activities are captured in work that has examined how social media platform policies, design choices, and business models shape communication and other social processes (e.g. DeNardis and Hackl, 2015), broadly what Gillespie (2017b) has termed “governance by platforms”, i.e. how internet companies, social media and other actors in digital spaces “police the content of their sites and the behavior of their users”. Such work extends and reignites earlier, influential arguments about the need to understand how multiple forces, including laws, norms, markets, and architecture and technical codes, shape the governance of the internet and the impact of the internet on social transformations (Lessig, 1999). The direction of this work establishes a key pillar of our argument that digital spaces contribute to processes of social ordering, which may be grasped in the ways that technical and infrastructural configurations come to shape conduct.

In order to bring these questions to the fore, we need approaches that explore the intersection of how digital, datafied spaces are structured, and the forms of conduct and ordering they afford. Within the limited scope of this paper, we focus on discussions foregrounding how data are sorted and organized in ways that shape social ordering. Our review is not exhaustive, and does not highlight empirical findings throughout these literatures. Instead, we focus on relevant work in the specific literatures on infrastructures, platforms and algorithms which draw on broader work in communication, information science and beyond. This helps us carve out a more limited conceptual and empirical domain in need of attention, and to articulate “datastructuring”.

### *Infrastructure studies*

Some attempts to understand the internet as a material form have drawn on physical metaphors. One such metaphor has been “infrastructure”, which serves to highlight how the internet facilitates the exchange of information in digital spaces, just like other underlying and taken-for-granted systems for delivering basic services such as the electricity grid, water, and sewerage pipes (Larkin, 2013). Particularly scholars in Science and Technology Studies (STS) have long argued that material and seemingly neutral institutional and technical arrangements have political and other social effects (Winner, 1980). Work in this area suggested that we should pay more attention to all sorts of invisible infrastructures because they do important work when it comes to conditioning social orders (Bowker and Star, 1999; Hughes, 1983). Also, infrastructure studies remind us that once systems are in place, they come across as natural and given. This is why we need to

question their design and politics when they are in the making, and also why studying them is so difficult. More recent work in this area has focused on information infrastructures and their consequences for our “ways of knowing” (Bowker et al., 2010), and related work has given attention to ways in which information gets classified, categorized, and sorted out (Bowker and Star, 1999; Flyverbom and Madsen, 2015).

Based on insights from this body of literature, we are better equipped to explore the research question driving this paper, namely how datastructuring shapes knowledge production and organizational, commercial, and political processes. For the purpose of our approach, these discussions of infrastructures help us understand and articulate how information travels, is rendered accessible, and embedded into social interaction in somewhat indirect and unexpected ways. We thus attend to the myriad of largely invisible ways that infrastructures forge connections between particular and fixed paths, and re-shape possibilities for social action.

### *Platform studies*

Another physical metaphor which has become central to discussions of digital spaces in recent times is “platform” (Bucher and Helmond, 2017). Platforms are not merely technical—“a programmable infrastructure upon which other software can be built and run” (Gillespie, 2017c), but more extensive phenomena that shape our lives in complex ways. The emergence of digital platforms as spaces for commerce and exchange have been addressed by a growing number of scholars interested in historical, legal, and operational aspects (Gillespie, 2017b; Lobel, 2016; Plantin et al., 2016; van Dijck, 2013). Such work highlights how platforms and their makeup create novel conditions for interaction, communication, and sales. Platforms provide “an architecture from which to speak or act, like a train platform or a political stage” (Gillespie, 2017c). Platforms come across as neutral, or as public benefits provided by internet companies as fair and impartial conduits for user activity, an attractive space for advertisers, and not in need of regulation. But platforms also structure data and digital spaces in material and technical ways that have societal consequences (Alaimo and Kallinikos, 2017; Helmond, 2015). Claims to neutrality serve to downplay and obscure more problematic aspects, such as that these digital spaces “organize, structure, and channel information, according both to arrangements established by the platform (news feed algorithms, featured partner arrangements, front pages, categories) and arrangements built by the user, though structured or measured by the platform (friend or follower networks, trending lists)” (Gillespie, 2017c).



Such studies illustrate the shift from social media sites, understood as places for people to share content, to social media platforms, i.e. digital spaces where (re)programming across sites is possible (Helmond, 2015). This has consequences for how data is structured, and processes of making data “platform-ready” also affect the makeup of digital spaces and social ordering more broadly. As Helmond (2015: 1) suggests, the introduction of new architectures, such as “application programming interfaces” (APIs) create new “data pours [that] not only set up channels for data flows between social media platforms and third parties but also function as data channels to make external web data platform ready”. Studies in this area also show us that structuring information is not only technical, but also involves various forms of human labor. Sociological investigations of digital platforms have highlighted the manual and (often distressing) human labor that goes into running, moderating, and cleaning up social media (Roberts, 2016). This growing and important body of work on content moderation brings us much closer to the largely hidden forces at work when information enters or disappears from social media platforms. By giving attention to various forms of content moderation, such work has shown that platforms are edited spaces, with similarities to other spaces such as newspapers where information gets circulated and comes to have socio-political effects.

These discussions of the processes whereby information gets sorted and curated speak directly to our focus. From this work, we take insights about the way human, technical, and material configurations in digital spaces condition particular kinds of outcomes or forms of ordering. Data does not travel or arrive in fixed form or encoded with meaning, but nor are they without shape or raw. As Bowker (2013) has put it, “raw data is an oxymoron”, and we are interested in how infrastructures and platforms “cook” data in particular ways and what this means for what ends up on our plates. Platform studies point to the way data are structured, how the supporting structures are constituted by data as well as how these data inputs shape the structure itself. In this sense the structure can be shaped or manipulated by adding data, and the datastructures at work have consequences for outputs and user experience. It is not just that the path of travel is constantly re-shaped by data attached to the structure, and it is not just that the path is re-shaped towards the particular contexts where it provides a match, it is that the path is re-shaped at the same time as the path-context is re-shaped. Traversing a datastructure, any sort of message or recommendation arrives in some way pre-conditioned. To paraphrase Habermas (1971), reason resides in the conditions under which what is said can be expressed. Conditions provide a kind of social proof

which dovetail with our concerns about datastructures and social ordering.

### Algorithm studies

The workings of digital spaces are also shaped by the software that structures and sorts information. These data sorting mechanisms include algorithms and other automated ways of dealing with the scale and complexity of data (Gillespie, 2014; Kitchin, 2017). While the most widely known work in this area has focused on the opacity of algorithms and described them as “black boxes” (Pasquale, 2015), others have sought to understand and describe the workings of algorithms more actively. Recent work has highlighted some of the difficulties in getting to know how algorithms operate, and suggested that simply enforcing transparency is not as simple as it seems (Ananny and Crawford, 2018). One reason for this is that it is hard to understand the workings of algorithms in separation from the data they sort out. Also, simply seeing what a technical system does is not the same as knowing how it works, and all attempts to create transparency involve complex human technical work (Hansen and Flyverbom, 2015). Such accounts offer us insights about the workings of algorithms as editing mechanisms, as well as how data can be made recognizable by algorithms and how algorithms create forms of social ordering such as inclusion and exclusion (Gillespie, 2014). As Gillespie (2014: 1) puts it: algorithms are emerging as a “key logic governing the flows of information on which we depend”. Studies of algorithms highlight how the editing and visualization of datafied phenomena come to shape our understanding of the world and our place in it.

The kind of work done by algorithms and the way such automated operations shape data configurations are key concerns in the research agenda that we propose. Algorithms and other socio-material forces at work inside digital spaces shape what we come to see, know and act on, and these forces deserve more attention. To articulate the possible contribution offered by a focus on datastructuring, this section has situated the argument in relation to more well-established discussions about the workings of digital spaces. While research emerging under these headings—as well as attempts at bringing them together (Ananny and Gillespie, forthcoming; Plantin et al., 2016)—certainly pave the way for more nuanced understandings of how digital spaces contribute to social ordering, they do so at a relative distance from the minutiae and complexities of what we term datastructuring. That is, conceptualizations of infrastructures, platforms, and algorithms focus mostly on the actors or organizations *behind* or *underneath* data structuring practices and have little to say about the *actual*, *inside* processes

whereby data get organized and structured. The goal with this paper is to suggest a research orientation that can help us articulate what happens inside these digital spaces.

From these literatures, we develop four broad conceptualizations of the features of datastructures: First, datastructures are *socio-material entanglements*. By this we mean that they are the result of technical as well as human activities, and that these features are entangled in complex ways (Leonardi, 2011; Orlikowski and Scott, 2008). Algorithms, ways of linking data and other structures are designed along the lines of both human and technical logics, and—once in place—these are hard to disentangle and even speak about. A second feature of datastructures is that they *recursively inform*. This means that they are encoded features of digital spaces that revolve around integrating and responding to new data inputs. We may think of this as recursive abilities that allow a wide range of data about users and their digital traces to be folded back into data, in an endless cycle of “informing” happening inside digital infrastructures (Zuboff, 1985). This recursive nature and the ability to integrate new data gives datastructures a capacity to grow and to “learn”. A third feature of datastructures is their *flexible reusability*. Data about users can be flexibly classified and reclassified. Because of this flexibility, datastructures can be put to different uses, and the same data can be repurposed endlessly. This feature creates a range of possibilities for the digital spaces that can make use of them, and also have consequences for the way users provide and encounter data. A fourth feature of datastructures is that they are *conditioning forces*. Datastructures have effects that go far beyond their primary functions, such as making social media platforms run smoothly, providing data-based services, and curating contents in ways that make them relevant for users and valuable for companies. They also shape processes of knowledge production and social ordering in multiple and complex ways. This feature invites us to explore how design choices and sorting mechanisms in digital spaces come to shape social life simply by organizing information in particular ways. It also further highlights the need for novel conceptual vocabularies and research methods that capture the concrete workings and societal consequences of what may come across as very mundane and technical processes.

Explorations of datastructuring may start from fairly basic and open-ended questions such as these: How do some kinds of information become visible and accessible at the expense of others? How do various actors rely on and “game” datastructures to guide our attention and convince us in subtle ways? And how do largely invisible forms of datastructuring shape organizational processes, social practices, and societal orders?

## Datastructuring at play

Datastructuring helps to bridge the concerns and limitations that we have identified in relevant literature and is attuned to the idea of shifting the attention to the *conditions* of communication. Datastructuring addresses these conditions by unpacking the ways that the connections between data and users that are established through information technologies also recast users and data—that is, have ordering effects. Through the following illustrations, we set out to articulate how datastructures entangle data-technology-users in ways that render datastructuring as a form of social action. This takes place when datastructures are used to commercial and political ends, or when datastructures come to act as drivers of various forms of radicalization or the formation of new tastes in cultural products.

We illustrate by offering three examples of datastructuring—social datastructuring, political datastructuring, and commercial datastructuring. The domains we have chosen will be immediately recognizable to scholars doing work within or across the theoretical areas that we discussed above, and also constitute common-sense parts of modern societies. It should be stressed that we do not offer these illustrations as exhaustive empirical accounts, but rather as illustrations which set out a starting point or invitation to further research. We also need to stress that our presentation is not exhaustive of the relevant domains where datastructuring could offer insight. The conceptual and empirical distinctions we make are analytical and heuristic. Still, we believe that they add up to something that is broad enough to make our point about the possible value of exploring datastructuring as a key component of knowledge production and social ordering, and establish the value of thinking across cases and disciplines with concepts that are flexible and inclusive enough to capture what may otherwise seem to be unrelated phenomena.

### Datastructuring social content

One obvious domain where datastructures contribute to social ordering is where most journeys into digital spaces start: sites such as search engines, social media, and video sharing platforms that organize and give access to content. In order to deal with masses of data in real time, automated operations are needed. Such sites rely on similar ways of organizing data, such as through metadata, thumbnails, search functions, friends and connections, recommendations, hash-tags, likes, profiles, and so on. These datastructuring elements can be aggregated and disaggregated with data from multiple sources, and guide the attention of publics, i.e. what we see, know and come to act on. This

is touched on in research exploring the role of internet search engines when it comes to shaping public opinion and political affairs (Dutton et al., 2017) and in controversies surrounding Facebook (e.g. experiments in social influence and political mobilization, [Bond et al., 2012], experiments with users' emotions [Kramer et al., 2014], and Twitter (e.g. the spread of fake news, [Vosughi et al., 2018]. These discussions are concerned with the social effects of filtered, biased and otherwise structured information produced through data sorting and analysis techniques used by internet companies. Our purpose here is to extend this reasoning to focus on datastructuring in terms of how, through organized activity, these systems can be deliberately manipulated towards the achievement of specific ends.

A somewhat notorious case encapsulates datafied social action which in turn shapes the possibilities of others' social action, namely the case of "spreading santorum". During an interview in 2003, Republican senator Rick Santorum articulated his support for a narrow definition of marriage as a union "between a man and a woman" in controversial language (Brewer, 2008: 67). The comments quickly sparked an ordinary political dispute where Santorum was challenged by gay activists and others. But Santorum's comments also gave rise to a different kind of response, which highlights our argument that datastructuring involves covert processes in which data gets structured and comes to have effects that go far beyond the content or shape of messages or direct impact on publics. Dan Savage, a sex columnist, editor of an alternative newspaper, and LGBT activist, set out to produce something more durable than a short outcry against Santorum's comments (Gillespie, 2017a). Based on suggestions and votes by his readers, Savage announced that the new meaning of "santorum" would be "that frothy mixture of lube and fecal matter that is sometimes the byproduct of anal sex" (Gillespie, 2017a: 66). The campaign involved setting up the website [www.spreadingsantorum.org](http://www.spreadingsantorum.org) and engaging activists and others in linking to the website, announcing the term widely and feeding it to digital spaces and algorithms. That is, the social action was to teach datafied information to forge specific pathways, which performed a social act of making a political statement through datastructures. Moreover, to contest the neologism and recolonise the meaning of his own name, Santorum was compelled to engage in the same form of social action, i.e. through the datastructure. Some 15 years later, searches for "santorum" continue to produce results that seem detrimental to the political aspirations of the senator, and hint at the political potential of what Gillespie (2017a) calls algorithmically recognizable communication. Such practices of feeding search engines particular kinds of data have been referred to as

"googlebombing", and are not new. Our argument is that the case of "spreading santorum" illustrates how the social act of making a political claim took a new form (embedded in hyperlinks which forged a datastructure), came to shape what we know (Santorum remains notorious for comments made almost two decades ago), and also shaped the possible forms of others' social actions (Santorum and supporters can only contest the datastructure by trying to forge new links within the datastructure). This illustrates the broader argument of the paper, that datastructures profoundly condition and shape political activities by turning meta-communication into communication (Boellstorff, 2013), or, by turning the form into the action (Easterling, 2015). Processes of information sorting have long been recognised as of central importance in societal affairs, including how we view the world (Lippmann, 1922) and how relations between knowledge and power become institutionalized (Foucault, 1977), but processes of digitalization and datafication produce new dynamics and approaches.

### *Datastructuring political issues*

Datastructures also increasingly shape how people engage in political activity. This includes how people come to know about and express opinions on political issues such as electoral candidates, proposed policies and so on. Our suggestion is that contents, messages, and explicit framings of positions may become less important than the underlying conditions that give people access to particular kinds of information. This is where our concept of datastructuring hopefully offers a new way to think about political influence and social ordering. We highlight this potential for new understandings of politics in digital spaces by looking at the recent outcry about Facebook and Cambridge Analytica.

The case of Cambridge Analytica provides some insight into the use of data in the Trump campaign (The Observer, 2018), but what appears to be accepted at face value is that data-based micro-targeting and persuasive messaging works by design. Micro-targeting implies that telling people with the intention to vote for Trump that they should vote for Trump is straightforward. But the argument is that Cambridge Analytica persuaded targeted audiences to switch their voting intentions, i.e. to attach what they already think about the issues to a different political candidate. We need a way to understand the role of data and particularly datastructuring in persuasion—a hidden grammar that works as a kind of digital rhetoric.

A concept of datastructuring with elements of recursivity and flexibility, and which attends to conditioning, helps to make sense of this case. At the outset, there is



the analyst working with voter profile data aggregated across platforms to construct discrete audiences. If we start with just one discrete audience and one message (Bloomberg (2016) reported that the campaign claimed to have created 100,000 distinct pieces of creative content), a data-based message is sent to one data-constructed audience which sets off a recursive loop between sender and receiver: did they click, did they share, who did they share it with, etc. For the sake of the example, the recursivity provided by our data-based message means we now have three discrete audiences: those who received the message from the sender directly and signaled support, those who received the message from the sender and signaled ambivalence or hostility, and a new audience of prospective supporters consisting of those who received the message indirectly from their contacts. Flexibility means we can target now three discrete audiences with different messages, setting off a cycle of whereby it is now possible to test the message variants themselves in an endlessly recursive sequence of increasingly precise message targeting.

This is where the conditioning element of datastructuring as a concept might usefully provide insight. Datastructuring does not just connect target messages to audience segments, it communicates with information structures in ways that condition how the message is delivered. For example, it was observed during the campaign in question that Trump appeared to have the support of hundreds of thousands of Twitter bots that amplified his campaign messages through different forms of duplication, the creation and repetition of subtle variations, and seemingly endless retweeting (BBC, 2016). These repetitions are data-based messages in a different form which communicate with the infrastructure of the platform (here the trending algorithm on Twitter), and condition the message by creating a context for its reception—the message from Trump is delivered to the audience together with a projection of the audience reception in which it appears that other voters have signaled agreement with the message. Could it be that some voters identified with what they believed was the audience response more than the message itself, and a sense of safety in numbers was what persuaded them to change voting intentions? Could it be that some voters identified with the urgency in what they believed was the audience response, and a sense of frenzy was what convinced them to translate their voting intentions into the action of voting? With datastructuring, the conditioning element of the data-based messages appears to offer a plausible way to analyze persuasion and forms of social action, including Twitter bots as a form of programmed and automated social actions. At minimum, it seems to be worth working with a concept that also provides or captures the role of data communicating with information

infrastructures. Moreover, moving towards such a concept is only the beginning of the inquiry, it opens up numerous questions such as are messages and their conditioning separable? Do they have a positive (mutually reinforcing) or a negative (mutually undermining) relationship? What factors (e.g. data quality, analyst quality, candidate quality, audience biases, message fit, etc.) do these relationships depend on? If people are persuaded, what parts message and what parts conditioning are doing the persuading? Do these messages cascade and spill over between groups? Is it possible to create a mass of political significance? Such questions bring us back to our concern with understanding new forms of social action through datastructures, their capacity to shape others' forms of social action, as well as how they guide our attention and influence what we know.

### *Datastructuring commercial processes*

Shopping online, reading the news in digital form, or listening to music via streaming services is largely dependent on platforms that structure data in particular ways. Such activities require multiple systems for aggregating, quantifying, and connecting different types of data, and the ability to identify and construct “quantifiable users that are made commensurable to other users” (Alaimo and Kallinikos, 2017: 4). At the same time, digital spaces are constructed in ways that allow for flexible reuse and recursive informing. These datastructuring features take the shape of, for instance, ways of turning user data into profiles that can guide the targeting of advertisements. It may also involve the reliance on tools for measurement and analyses that give valuable insights about traffic and behavior for the purpose of engaging users further, for example in the way recommendation systems in YouTube or Netflix create particular trajectories for users. Or the use of systems that repurpose reviews or recommendations as ways to sell products or develop trust relations between users. Such datastructures are at play in most digital spaces and take a number of shapes. But for the purpose of this paper, illustrations from news production allow us to highlight some salient features of datastructuring with relevance for other domains.

News production is increasingly inseparable from digital technologies and social media, and the way data gets structured in digital spaces has far-reaching consequences for the operations and business models of this industry. Newspaper articles are published in digital formats and distributed via the internet, and many people primarily encounter news stories via Facebook, Google, and Twitter. These developments set new conditions for the production and circulation of news.

Most news publishers have to consider how to make their subscription and business models viable in environments where copying and distributing is easier than ever, and where most users expect content to be free. But these processes of digitalization are only the most obvious and visible ways that technologies shape news production, and in the background a much more fundamental transformation plays out. When a news story is published on a web site, it is not simply a digital copy of what was previously printed on paper. It is also a digital object that relies on a wide range of digital resources that are repurposed from other users and interactions, and may recursively informate—allowing for insights and learning from users and how they engage with news. Empirical investigations of tens of thousands of online news stories show that commercial news stories often involve between 100 and 250 partners and services that contribute and extract digital resources (Flyverbom, 2016; Lindschow, 2016). To publish a news story, most media production houses rely on tools and services that facilitate editing, the selling of ads or the measurement of traffic and clicks. The digital traces produced by users are an increasingly important resource that helps internet companies know users, gain insights about customer preferences and design new products and markets. This has consequences for business models, the production of knowledge and broader forms of social ordering. US internet companies are reshaping the news and advertising industries via their access and abilities when it comes to extracting digital traces and turning them into valuable insights about peoples' preferences, needs and habits. But beyond these instrumental uses of data, internet companies are also increasingly taking charge of how information gets structured and made visible and which speak directly to the issue of datastructuring. Internet companies increasingly provide and control the ecosystems in which content is produced, circulated, and consumed. This allows them to extract value, but also structure how information and digital traces circulate and become visible and valuable for commercial activities such as advertising. In the context of news production, the difference between journalistic principles such as relevance or societal importance are challenged by other principles such as popularity, measured in amounts of click, shares, etc. These developments create new conditions and criteria that shape not only the daily work of journalists but also the value chains and business models of media producers of all sorts.

Datastructures in news production mediate the connection between journalists and audiences in ways that have commercial, epistemological, and political significance. This can also be exemplified in one particular symptom of changing news production—the rise of

“fake news”. Datastructuring enables news producers to glean increasingly fine-grained data on what audiences want to read. With audience traffic driving advertising revenue, some commercial actors naturally search for ways to meet audience demand for particular stories. Because digital spaces facilitating news production and distribution are set up to recognize quantifiable levels of activity and recursively informate, they tend to work from relatively crude forms of categorization and commensuration. Often, they interpret high degrees of activity (clicks, reads, shares, likes and comments) as indications of something having value. Put slightly differently, they equate popularity with quality, and this creates new dynamics in areas such as news production. Compared to traditional journalistic criteria for the selection and valuation of news, datastructuring delivers more crude forms of editing and sorting. This means that some news producers deliberately prioritize commercial ends ahead of journalistic means, with little concern for whether reported events are accurately described. Datastructuring in the context of news production is suggestive of how groups with different political identities have come to seemingly live in different realities, which provides significant obstacles when searching for common grounds for democratic deliberation. Cast in broader terms, many kinds of datastructuring in commercial contexts come to shape how people encounter and are shaped by products, other consumers, companies and more fuzzy phenomena such as buying habits, taste, and socialization. These examples all speak to our overall point that datastructuring conditions social ordering in complex ways that deserve more scholarly attention.

## **Discussion and conclusion: Datastructures as epistemic architecture**

These three forms of datastructuring—social, political, and commercial—illustrate our argument that knowledge production and social ordering also come into existence through datastructures. These (and other) domains are worth exploring because they highlight that datastructures have social, organizational, and political ramifications. There are different logics and developments at work in search and information provision, in political interventions and corporate advocacy, and in value creation and value extraction in the news industry and other commercial sectors affected by digital transformations. But for now and with the limited empirical illustrations we have provided, we leave those important questions for future research to explore. In these and other contexts, the point is that we should shift our focus from contents and other immediately visible features towards the more invisible

conditions through which data gets organized and are made amenable to other uses, such as data-driven marketing, affecting public opinion, and so on.

We have proposed “datastructuring” as an emergent field of research and suggested possible avenues for research in this area. More specifically, we have offered a conceptualization of a field of inquiry with a focus on both how digital spaces structure information and how humans and technologies are involved and shape these datastructures. In terms of empirical sites and analytical paths to explore, there are numerous issues in need of further research. For instance, we still know relatively little about the moderation and curation practices of social media platforms at the level of datastructuring. Also, we need more research on the strategic ways in which actors seeking to influence public opinions and regulatory priorities can feed and manipulate digital spaces with the provision of particular types of information and attempts to optimize how certain data becomes algorithmically recognizable (Gillespie, 2017a). While we may have access to the “community guidelines” and other rules for digital platforms, we lack more general overviews of the forms of editing, information architectures, and choreographies of visibility and invisibility that characterize these spaces. To fill this gap we have provided a theoretical conceptualization of how digital traces are organized, recognized, and visualized across a wide range of spaces, and suggested how future research can address such questions about “datastructuring”. By offering an umbrella concept for various approaches to the study of organized data and information control, we have sought to show how some things become visible at the expense of others, how material forces and technological designs shape what comes to count as knowledge, and how the structuring of information guides our attention. This focus on datastructuring is one way to move away from the simplistic calls for algorithmic transparency (Pasquale, 2015) and into more nuanced attempts to understand the systems and assemblages of algorithms, data, and human choice in datafied spaces (Ananny and Crawford, 2018).

In our conceptualization we take theoretical inspiration from various sources that support our view of structured data as *productive* of specific realities—they do not just *selectively* “disclose” what is already there (Hansen and Flyverbom, 2015). Material reconfigurations such as datastructures do not only transmit or support the sharing of specific truths, but are generative factors that define what matters (and what does not), and guide our attention. In this manner, we seek to account for datastructures as over-looked socio-material forces of social ordering in the algorithmic age.

We consider the broader focus on datastructuring as a nascent field of inquiry. This is an important

endeavor because we lack more general conceptualizations of the contours and dynamics of the datastructures that result from processes of digitalization and datafication, and the particular “possibilities for action” (Leonardi, 2011: 153) they allow for. These spaces are not just conduits for communication and interaction, but constitute more extensive environments and forms of metacommunication that condition particular forms of communication (Jensen and Helles, 2017; Meyrowitz, 1997). Thus, in contrast to most accounts of data, our focus is not on the contents, meanings or interpretations—i.e. the substance and the groups that receive information. Rather, we are interested in the machineries, infrastructures, and other socio-material arrangements that facilitate the management of visibilities. This is particularly timely in an age where (big) data is often assumed “to speak for itself” and the forms of intelligence they generate are easily taken as truths.

Datastructuring is an increasingly important condition for communication and knowledge production. As such, we may think of it as a novel and overlooked type of “epistemic architecture”—structured and structuring conditions that work as conduits for and barriers to information (Costas and Grey, 2016: 115). With the concept of datastructuring, we want to highlight how design choices, ways of sorting data, and other dimensions of digital spaces create novel conditions for knowledge production and communication in general. Understood as epistemic architecture, these seemingly mundane and technical features come to life as key drivers and dynamics at work in digital spaces.

Ultimately such discussions are important because in contemporary information spaces it becomes increasingly difficult to establish and identify what we can know and how we can know it. The developments we highlight in this paper demand that we revisit questions about what it means to have “free will”, to make “informed decisions” or to “produce knowledge”. Such seemingly straightforward issues need renewed attention in times marked by rapidly transforming epistemic architectures and information ecosystems. We look forward to identifying and engaging in further explorations of these developments and hope that others will join us in this endeavor.

### Declaration of conflicting interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

### Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

## References

- Abbasi A, Sarker S and Chiang RHL (2016) Big Data research in information systems: Toward an inclusive research agenda. *Journal of the Association for Information Systems* 17(2): i–xxxii.
- Alaimo C and Kallinikos J (2017) Computing the everyday: Social media as data platforms. *Information Society* 33(4): 175–191.
- Amoore L and Piotukh V (2015) Life beyond big data: Governing with little analytics. *Economy & Society* 44(3): 341–366.
- Ananny M and Crawford K (2018) Seeing without knowing: Limitations of the transparency ideal and its application to algorithmic accountability. *New Media & Society* 20(3): 973–989.
- Baesens B, Bapna R, Marsden J, et al. (2016) Transformational issues of big data and analytics in networked business. *MIS Quarterly* 40(4): 807–818.
- Bateson G (1972) *Steps to an Ecology of Mind*. Chicago: University of Chicago Press.
- BBC (2016) Trump's twitter debate lead was 'swelled by bots'. Available at: <http://www.bbc.com/news/technology-37684418> (accessed 29 August 2018).
- Bloomberg (2016) Inside the Trump bunker, with days to go. Available at: <https://www.bloomberg.com/news/articles/2016-10-27/inside-the-trump-bunker-with-12-days-to-go> (accessed 29 August 2018).
- Boellstorff T (2013) Making big data, in theory. *First Monday* 18(10). Available at: <http://firstmonday.org/ojs/index.php/fm/article/view/4869/3750%2330a> (accessed 7 November 2017).
- Bond RM, Fariss CJ, Jones JJ, et al. (2012) A 61-million-person experiment in social influence and political mobilization. *Nature* 489: 295–298.
- Bowker G (2013) Data flakes: An afterword to "Raw Data" is an oxymoron. In: Gitelman L (ed.) *"Raw Data" is an Oxymoron*. Cambridge, MA: The MIT Press, pp. 167–172.
- Bowker GB, Baker K, Millerand F, et al. (2010) Toward information infrastructure studies: Ways of knowing in a networked environment. In: Hunsiger J, et al. (eds) *International Handbook of Internet Research*. London: Springer, pp. 97–117.
- Bowker GB and Star SL (1999) *Sorting Things Out: Classification and Its Consequences*. Cambridge, MA: The MIT Press.
- boyd d and Crawford K (2012) Critical questions for big data. *Information, Communication & Society* 15(5): 662–679.
- Brewer PR (2008) *Value War: Public Opinion and the Politics of Gay Rights*. Plymouth: Rowman & Littlefield, pp. 67–68, 86.
- Bucher T and Helmond A (2017) The affordances of social media platforms. In: Burgess J, Poell T and Marwick A (eds) *The SAGE Handbook of Social Media*. London/New York, NY: SAGE Publications.
- Costas J and Grey C (2016) *Secrecy at Work: The Hidden Architecture of Organizational Life*. Stanford: Stanford University Press.
- Czarniawska B (1997) *A Narrative Approach to Organization Studies*. London: Sage.
- DeNardis L and Hackl AM (2015) Internet governance by social media platforms. *Telecommunications Policy* 39(9): 761–770.
- Desrosières A (2002) *The Politics of Large Numbers: A History of Statistical Reasoning*. Cambridge, MA: Harvard University Press.
- Dutton W, Reisdorf B, Dubois E, et al. (2017) Internet search and politics: The uses and impacts of search in Britain, France, Germany, Italy, Poland, Spain, and the United States. *Quello Center Report*, Michigan State University.
- Easterling K (2015) *Extrastatecraft: The Power of Infrastructure Space*. London: Verso.
- Flyverbom M (2011) *The Power of Networks: Organizing the Global Politics of the Internet*. Cheltenham: Edward Elgar Publishing.
- Flyverbom M (2016) Disclosing and concealing: Internet governance, information control and the management of visibility. *Internet Policy Review* 5(3): 1–15.
- Flyverbom M, Leonardi P, Stohl C, et al. (2016) The management of visibilities in the digital age. *International Journal of Communication* 10(2016): 98–109.
- Flyverbom M and Madsen AK (2015) Sorting data out: Unpacking big data value chains and algorithmic knowledge production. In: Süssenguth F (ed.) *Die Gesellschaft der Daten: Über die digitale Transformation der Sozialen Ordnung*. Bielefeld: Transcript Verlag, pp. 123–144.
- Foucault M (1977) *Discipline and Punish: The Birth of the Prison*. New York, NY: Vintage Books.
- Gillespie T (2014) Social relevance of algorithms. In: Gillespie T, Boczkowski PJ and Foot KA (eds) *Media Technologies: Essays on Communication, Materiality, and Society*. Cambridge, MA: MIT Press.
- Gillespie T (2017a) Algorithmically recognizable: Santorum's Google problem, and Google's Santorum problem. *Information, Communication & Society* 20(1): 63–80.
- Gillespie T (2017b) Governance of and by platforms. In: Burgess J, Poell T and Marwick A (eds) *SAGE Handbook of Social Media*. London: Sage.
- Gillespie T (2017c) The platform metaphor, revisited. Available at: <http://culturedigitally.org/2017/08/platform-metaphor/> (accessed 29 August 2018).
- Habermas J (1971) *Knowledge and Human Interests*. Translated by Jeremy J Shapiro. Boston, MA: Beacon.
- Halpern O (2015) *Beautiful Data: A History of Vision and Reason since 1945*. Durham: Duke University Press.
- Hansen HK and Flyverbom M (2015) The politics of transparency and the calibration of knowledge in the digital age. *Organization* 22(6): 872–889.
- Helmond A (2015) The platformization of the web: Making web data platform ready. *Social Media + Society* 1(2): 1–11.
- Hofmann J, Katzenbach C and Gollatz K (2016) Between coordination and regulation: Finding the governance in Internet governance. *New Media & Society* 19(9): 1406–1423.
- Hughes TP (1983) *Networks of Power: Electrification in western society, 1880–1930*. Baltimore, MD: John Hopkins University Press.
- Jensen KB and Helles R (2017) Speaking into the system: Social media and many-to-one communication. *European Journal of Communication* 32(1): 16–25.



- Kallinikos J (2013) The allure of big data. *Mercury Magazine* 2(3): 40–43.
- Kitchin R (2017) Thinking critically about and researching algorithms. *Information, Communication and Society* 20(1): 14–29.
- Kramer ADI, Guillory JE and Hancock JT (2014) Experimental evidence of massive-scale emotional contagion through social networks. *Proceedings of the National Academy of Sciences of the United States of America* 111(24): 8878–8790.
- Laney D (2001) *3D data management: Controlling data volume, velocity, and variety*. Technical report, META Group, 6 February. Available at: <https://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf> (accessed 29 August 2018).
- Larkin B (2013) The politics and poetics of infrastructure. *Annual Review of Anthropology* 42: 327–343.
- Leonardi P (2011) When flexible routines meet flexible technologies: Affordance, constraint, and the imbrication of human and material agencies. *MIS Quarterly* 35(1): 147–167.
- Lessig L (1999) *Code and Other Laws of Cyberspace*. New York, NY: Basic Books.
- Lindskow K (2016) *Exploring Digital News Publishing Business Models: A Production Network Approach*. Frederiksberg: Copenhagen Business School (PhD Series; No. 07.2016).
- Lippmann W (1922) *Public Opinion*. New York, NY: Harcourt, Brace and Company.
- Lobel O (2016) The law of the platform. *Minnesota Law Review* 101(1): 87–166.
- Lycett M (2013) ‘Datafication’: Making sense of (big) data in a complex world. *European Journal of Information Systems* 22(4): 381–386.
- Mayer-Scheonberger V and Cukier K (2013) *Big Data: A Revolution That Will Transform How We Live, Work, and Think*. London: Eamon Dolan/Houghton Mifflin Harcourt.
- Meyrowitz J (1997) Shifting worlds of strangers: medium theory and changes in “them” versus “us”. *Sociological Inquiry* 67(1): 59–71.
- Orlikowski WJ and Scott SV (2008) Sociomateriality: Challenging the separation of technology, work and organization. *Academy of Management Annals* 2(1): 433–474.
- Pasquale F (2015) *The Black Box Society: The Secret Algorithms That Control Money and Information*. Cambridge, MA: Harvard University Press.
- Plantin JC, Lagoze C, Edwards P, et al. (2016) Infrastructure studies meet platform studies in the age of Google and Facebook. *New Media & Society* 20(1): 293–310.
- Roberts ST (2016) Commercial content moderation: Digital laborers’ dirty work. In: Noble SU and Tynes B (eds) *The Intersectional Internet: Race, Sex, Class and Culture Online*. New York, NY: Peter Lang.
- The Observer (2018) ‘I made Steve Bannon’s psychological warfare tool’: Meet the data war whistleblower. Available at: <https://www.theguardian.com/news/2018/mar/17/data-war-whistleblower-christopher-wylie-faceook-nix-bannon-trump> (accessed 29 August 2018).
- van Dijck J (2013) *The Culture of Connectivity: A Critical history of social media*. Oxford: Oxford University Press.
- Vosughi S, Roy D and Aral S (2018) The spread of true and false news online. *Science* 359(6380): 1146–1151.
- West SM (2017) Data capitalism: Redefining the logics of surveillance and privacy. *Business & Society*. Epub ahead of print 2017. <https://doi.org/10.1177/0007650317718185>.
- Winner L (1980) Do artifacts have politics? *Daedalus* 109(1): 121–136.
- Zuboff S (1985) Automate/Informate: The two faces of intelligent technology. *Organizational Dynamics* 14(2): 5–18.
- Zuboff S (2018) *The Age of Surveillance Capitalism: The Fight for the Future at the New Frontier of Power*. Berlin: Campus Verlag.