THE LONDON SCHOOL OF ECONOMICS AND POLITICAL SCIENCE

Datafication

LSE Research Online URL for this paper: http://eprints.lse.ac.uk/104417/

Version: Published Version

Article:

Mejias, Ulises A. and Couldry, Nick (2019) Datafication. Internet Policy Review, 8 (4).

Datafication

Ulises A. Mejias

Communication Studies, State University of New York at Oswego, United States,
ulises.mejias@oswego.edu

Nick Couldry

Media, Communications and Social Theory, London School of Economics & Political Science,
United Kingdom

Abstract: Datafication is not just the making of information, which, in one sense, human beings have been doing since the creation of symbols and writing. Rather, datafication is a contemporary phenomenon which refers to the quantification of human life through digital information, very often for economic value. This process has major social consequences. Disciplines such as political economy, critical data studies, software studies, legal theory, and—more recently— decolonial theory, have considered different aspects of those consequences to be important. Fundamental to all such approaches is the analysis of the intersection of power and knowledge.

This article belongs to Concepts of the digital society, a special section of Internet Policy Review
guest-edited by Christian Katzenbach and Thomas Christian Bächle.

INTRODUCTION

The term "datafication" implies that *something* is made into data. What that something is, and what the processing comprises, are matters that need to be put into context. The term "data", however, is relatively clear, at least in its contemporary usage. Data is the "material produced by abstracting the world into categories, measures and other representational forms [...] that

constitute the building blocks from which information and knowledge are created" (Kitchin, 2014, p. 1). While, in principle, *any* thing or process (from a sun or rain pattern, to a beating heart, to a lesson delivered in a class) can be made into data, our focus in this short essay will be on processes of datafication that create digital data out of human life. Since most writers on data also care about what happens to human life, the term "datafication" has quickly acquired an additional meaning: the wider *transformation of human life* so that its elements can be a continual source of data. The beneficiaries of this are very often corporations, but also states and sometimes civil society organisations and communities.

The term "datafication" was introduced in a 2013 review of "big data" processes across business and the social sciences (Mayer-Schönberger and Cukier, 2013, chapter 5): "to datafy a phenomenon is to put it in quantified form so that it can be tabulated and analyzed" (2013, p. 78). Datafication, the authors argued, involves much more than converting symbolic material into digital form, for it is datafication, not digitization, that "made [digital] text indexable and thus searchable" (2013, p. 84). Through this process, large domains of human life became susceptible to being processed via forms of analysis that could be automated on a large-scale. The dynamic that drives datafication as a social process then becomes apparent: the drive to "render [...] human behavior... into an analyzable form" in a process that in the review mentioned above was already called "the datafication of everything" (2013, p. 93-94).

It was not long before critical perspectives on datafication began to appear. As our initial definition of "data" makes clear, data do not naturally exist, but only emerge *through* a process of abstraction: something is taken from things and processes, something which was not already there in discrete form before. Lisa Gitelman (2013) sums up this point in the title of a well-known edited collection: *Raw Data is an Oxymoron*. Indeed, implicit in the very notion of data (or what is given as fact, from the Latin *data*) are the notions of selection and transformation: "data are [...] elements that *can* be abstracted from [...] phenomena" (Kitchin, 2014, p. 2). Kitchin even argues that "data" should be replaced with another Latin term, *capta*—what is captured—to refer to how, practically, data is harvested from life. José van Dijck, surveying various terms that emerged around data processes, also offers a critical interpretation of datafication as "a means to *access [...]* and *monitor* people's behavior" (Van Dijck, 2014, p. 1478). She proposes that practices of datafication are becoming "an accepted new paradigm for *understanding [...]* social behavior" (2014, p. 1478, added emphasis). Such understanding involves a vision of "processes of datafication as a new way of interpreting the world". Pushing the argument further, Shoshana Zuboff argues that what we are living through is a new stage of "surveillance capitalism" (2019) in which human experience becomes the raw material that produces the behavioural data used to influence and even predict our actions.

We can approach the study of digital data as a complex matrix of actors and structures, which different disciplines can help us analyze at multiple levels. In terms of actors, we have corporations, states, and various civic (activists, journalists, etc.) and even non-state (terrorists, hackers) actors, all of which can produce, collect and analyse data for different purposes. Here the focus can range from the big corporate players responsible for the bulk of datafication in our lives—Facebook, Apple, Microsoft, Google, and Amazon in the West, and their Chinese counterparts Baidu, Alibaba, Tencent and Xiaomi)—to smaller players across what can be called the "social quantification sector" (Couldry and Mejias, 2018), including hardware, software, platforms, data analytics, data brokerage firms, and even spammers (depending on which country we examine, this sector has more or less close relations with how government at various levels seeks to extract data for monitoring its citizens; China is one country where those relations are particularly close, cf. Chen and Qiu, 2019). Datafication can obviously benefit some

of these actors, but it can also be used to discriminate against others on the basis of race, class, etc. (cf. Gandy, 1993; Peña Gangadharan, 2012). In terms of structures, data can flow within various architectures which can include platforms, services, apps, databases, and hardware devices. To make sense of this complexity, various research disciplines can help us zoom in or out on different intersections of players and infrastructures. For instance, software or platform studies can address issues of technological configuration and affordances, while a critical political economy approach can address issues of commodification and exploitation. Most of these approaches attempt to explain in some way how big data is "made" in terms of its relationship to time, context, and power (Boellstorff, 2013).

Next, we consider the specific elements that make up datafication, and the perspectives from which different disciplines have approached datafication's consequences, with specific emphasis on datafication by corporations for economic profit.

ELEMENTS OF DATAFICATION

The production of data cannot be separated from two essential elements: the external *infrastructure* via which it is collected, processed and stored, and the *processes of value generation*, which include monetisation but also means of state control, cultural production, civic empowerment, etc. This infrastructure and those processes are multi-layered and global, including mechanisms for dissemination, access, storage, analysis and surveillance that are owned or controlled mostly by corporations and states.

Put another way, datafication combines two processes: *the transformation of human life into data through processes of quantification, and the generation of different kinds of value from data.* Despite its clunkiness, the term datafication is necessary because it signals a historically new method of quantifying elements of life that until now were not quantified to this extent.

The process of quantifying life itself requires various components and conditions. First, as we already identified, it involves mechanisms of data collection. This can take many forms, but very often involves an app or platform that collects wide-ranging data about users, aggregates and analyses the data, and generates micro-targeted marketing data and predictive insights about behaviours. Some platforms such as Facebook have acquired the power to incorporate links to their mechanisms of data gathering within other platforms, turning Facebook itself in all its manifestations into a 'data infrastructure' (Nieborg and Helmond, 2019). The process is then monetised by using such data to sell products or services to the users, or by selling the data to parties wishing to influence or persuade users towards various goals. But that infrastructure also involves prior conditions: the condition of encouraging people to use the app or platform, that is, organising their habits so that life actions previously performed elsewhere (such as communicating with friends, sharing cultural products, hailing a taxi, etc.) become actions performed via the app. Even more importantly, the process of quantification involves *abstraction* via the process of turning the flow of social life and social meaning into streams of numbers that can be counted. This form of abstraction involves many subtle transformations, both cognitive and evaluative, as management theorists Cristina Alaimo and Jannis Kallinikos describe (2017). The transformations *of* social life that are inherent to datafication are so many, and so consequential for our orientation to the social domain, that Alaimo and Kallinikos write of a "computed sociality" (2017, p. 177; see also Van Dijck, 2013, p. 5, on "platformed sociality").

Even though these processes are relatively new, the basic idea of datafication—that the flow of

human life could be converted into discrete data—has a long history.

DATAFICATION: FROM PAST TO PRESENT

Datafication is implicated in more than just social media apps and content sharing platforms. The first domain of datafication was business, not social life. Even today, the amount of data generated by commerce exceeds the amount of data generated by the datafication of human life (Chairman's Letter in IBM, 2018). Key areas of business, such as logistics—the management of the flow of goods and information—have matured into complex practices thanks to datafication. The monitoring of continuously connected data flows to organize all aspects of production and distribution across space and time within global commodity chains could not be achieved without datafication (Cowen, 2015).

But there are many other ways in which aspects of the social world came to be counted or quantified during modernity, as a way of making it more 'legible' for governing (Poovey, 1998, chapters 2 and 7; Scott, 1990). One of particular importance is social network analysis, where applications of network science to social domains have contributed to the evolution of datafication. Social graphs and network visualisations have allowed corporations to extract information from the flow of life for descriptive and predictive use, aided by the incorporation of "smart" devices into these social circles (the so-called Internet of Things), which record not just interactions between people, but between people and things, or between things themselves.

Issues of power permeate these apparently neutral forms of datafication. The reason derives from the underlying way in which data is produced so that it can be counted. In a network, nodes only recognise other nodes, and if something is not represented as a node it does not exist. Likewise, a process or entity can only be represented *in* a network if it can be described in terms of the relations that the network can count or process. Something that cannot be codified as a potential network member cannot be accounted for. This process of nodocentrism (Mejias, 2013) is similarly implicit in the social modelling that renders social flux into data-driven computer processes (Rieder, 2012). When such schemes are applied, the result is the transformation of the very ways in which the social world is accounted for, as various sociologists have noted (Fourcade and Healey, 2013; Espeland and Sauder, 2007). The question of who is doing this codifying of life into datafied realities acquires extreme importance at this point.

Yet the effects of power that are intrinsic to datafication are often made invisible. Paradoxically, much-used metaphors that equate datafication to other extractive processes help to further obscure, not uncover, these power relations. Consider the saying that "data is the new oil", something that can be naturally extracted or mined since it exists in the "ground" of social life. As legal scholar Lauren Scholz notes, this metaphor "sidesteps evaluation of any misappropriation or exploitation that might arise from data use" (Scholz, 2018, p. 2). This understanding of datafication as somehow a natural process is surprisingly common, as evident in this sentence from an information booklet distributed by the UK's Royal Society: "Machine learning is a brand of artificial intelligence that allows computer systems to *learn directly* from examples, data and experience" (2019, n.p.). The idea of direct learning from data is regarded by many critical data scientists as mythical; it is part of a discourse which critical disciplines have attempted to debunk, as we will see in the next section.

CONTROVERSIES OVER DATAFICATION

Important controversies over social justice have emerged about how datafication is applied by corporations or states in particular sectors (from credit ratings to social services) to discriminate against individuals particularly from disadvantaged classes and ethnic populations (e.g., Gandy, 1993; Eubanks, 2017; Benjamin, 2019). More broadly, disciplines like political economy, legal studies, and decolonial theory approach the social quantification sector's work from different angles, each drawing on critical data studies.

POLITICAL ECONOMY

Marxist critiques of data production have mostly analysed the power dynamics inherent to datafication by focusing on a traditional interpretation of labour relations, looking at the "labour" that users perform by interacting with digital media and generating data (Fuchs and Mosco, 2017). Outside the Marxist tradition, similar critiques of digital labour and data production have emerged (cf. Scholz, 2016), while management scholar Shoshana Zuboff has advanced the thesis that the large-scale collection of personal data by corporations represents an aberrant form of capitalism (Zuboff, 2015, 2019). Common to these approaches is the fact that, as a social process, datafication is linked to the generation of profit—whether through data's sale as a commodity or data's incorporation as a factor of production (Sadowski, 2019, alternatively formulates data itself as 'capital').

However, recent critical work on datafication looks beyond the idea of labour. One approach is to consider the economic form constituted by the platforms across which so much data is generated and collected. Platforms represent much more than a commercial label for computing interfaces, as Tarleton Gillespie first noted (2010). They are a fundamental new kind of multi-sided market focused on datafication, a market that brings together platform users who generate data, data buyers (advertisers and data brokers), and platform service providers who benefit from the release, sale, and internal use of data (Rieder and Sire, 2014; Cohen, 2018).

Another approach interprets datafication via a rereading of Marx to argue that the most fundamental characteristic of datafication is not labour, but the abstracting force of the commodity, that is, the very *possibility* of transforming life processes into "things" with value through abstraction (Couldry and Mejias, 2018, 2019; Sadowski, 2019). This interpretation frames datafication as a social process configured around new relations ("data relations") designed to optimise the generation of data from social life (compare to Zuboff, 2015, 2019).

LEGAL STUDIES

Legal theory offers an alternative critique of datafication, arguing that datafication threatens the basic rights of the self. This is already suggested in the first sentence of the General Data Protection Regulation (GDPR): "the protection of natural persons in relation to the processing of personal data is a fundamental right" (Recital 2). The risks from the collection of personal data for individual autonomy have been predicted for at least two decades (cf. Schwartz, 1999; Cohen 2000). Legal theorist Julie Cohen in particular has argued for the importance of holding onto the concept of privacy in some form as a defense versus the chilling effects of continuous data collection and processing (Cohen, 2013). The processes of datafication are so wide-ranging, however, that others have raised questions about the usefulness of the term 'privacy' itself (Barocas and Nissenbaum, 2014). In a world where datafication seems continuous and multi-layered, there is clearly a need for a more contextual approach to the norm of privacy

(Nissenbaum, 2013).

Lately, questions have emerged about the implications of datafication—and artificial intelligence based on processing data—for the concept of autonomy (Hildebrandt, 2015). The datafication enabled by things like self-tracking devices, psychometric algorithms, and workplace tracking systems arguably interferes with the minimal integrity of the self *as a self* (Couldry and Mejias, 2019), which can be understood as the very basis of autonomy. Similar concerns have been expressed in terms of attempts by marketers and others to influence behaviour through data analytics (cf. Rouvroy, 2015, on "data behaviorism"). This line of critique argues that we are, through datafication, becoming dependent on (external, privatised) data measurements to tell us who we are, what we are feeling, and what we should be doing, which challenges our basic conception of human agency and knowledge.

Nonetheless, datafication creates practical openings for proposals for regulation. One such opening revolves around the question of who owns the data. There are competing interests set up by datafication, which means regulatory nuances have to be worked out. On one side, there are the interests of the individual who generates data or owns a device that produces the data; on the other, there are the interests of the owners of the infrastructure through which data flows and is collected (the social quantification sector). The latter usually ask the former to forgo any ownership rights to their data as a condition for using their infrastructure, sometimes framing access to the infrastructure as a "free" service that offsets the surrendering of property rights. Regulators, mostly in the EU through efforts such as the GDPR, are starting to intervene in this relationship to uphold some minimal rights for the individual.

Legal critiques sometimes imply an even broader question: how is it that human life came to be datafied—treated as an open domain for data extraction—in the first place (Cohen, 2018)? This is better understood in a longer historical perspective, which decolonial critiques provide.

DECOLONIAL THEORY

If datafication within capitalism is a process of abstracting and extracting life across various spaces to generate profit (with ancillary benefits for governments), then where does the wealth generated by this extraction go, and why? In order to examine the geography and politics of datafication (Thatcher et al., 2016), a connection to historical colonialism might be instructive.

Datafication can be understood as itself a colonial process, not just in the metaphorical sense of saying things like "data is the new oil", but quite literally as a new mode of *data colonialism* (Couldry and Mejias, 2019) that appropriates human life so that data can be continuously extracted from it for the benefit of particular (Western, but also increasingly global capitalist) interests. Instead of territories, natural resources, and enslaved labour, data colonialism appropriates social resources. While the modes, intensities, scales and contexts of data colonialism are different from those of historic colonialism, the function remains the same: to dispossess.

Within this wider perspective, datafication can be analysed as a continuation of the *coloniality of power* (Quijano, 2007), a form of domination in both social and cognitive domains (de Sousa Santos, 2016). A war for the social resources of the world is currently being waged between the social quantification sectors of China and the United States, principally (Couldry and Mejias, 2019). This "land grab" employs a whole arsenal of quantification weapons, from artificial intelligence, facial recognition, and new e-commerce models, to cyberwarfare, chip manufacturing, and multinational agreements regulating intellectual property. It is important to recall that, historically, information and communication technologies enabled the

administration and surveillance of colonised territories, as well as the propagation of narratives that legitimised extraction and dispossession. Datafication continues and extends these functions.

CONCLUSION

The analytical value of the term "datafication" lies in its ability to name the processes and the frameworks by which a new form of extractivism is unfolding in our times, via the appropriation of data about our lives. Corporations are the main actors in, and beneficiaries of, this process, with government in many countries having a strong stake in the process as well. Assuming that the problem is not with data per se (there are indeed consensual community projects for data collection), but with how and by whom it is systematically collected and used, a key question becomes how to halt the social quantification sector's expansion across social space. How do we stand *outside* datafication, when it seeks to capture the entirety of social space and time?

The term datafication itself can suggest practical ways to do this. By naming a process (data*fication*), we also invoke its limits. Just like the colonial project involved the separation of the world into centres and peripheries, datafication as a form of rationality also creates peripheral (or paranodal, cf. Mejias, 2013) things that cannot be quantified, and so, in principle, *cannot* be datafied.

Various forms of resistance—from the ineffective but strategic opting out of individual platforms, to a larger awareness of ourselves as the objects of datafication—can contribute to creating challenges and alternatives to the growth of datafication. Whether such resistance becomes successful in halting certain aspects of datafication remains uncertain, but it is surely one of the major social questions of our time.

REFERENCES

Alaimo, C., & Kallinikos, J. (2017). Computing the Everyday: Social Media as Data Platforms. *The Information Society*, *33*(4), 175–191. doi:**10.1080/01972243.2017.1318327**

Baack, S. (2015). Datafication and empowerment: how the open data movement re-articulates notions of democracy, participation and journalism. *Big Data & Society*, *7*(1). doi:**10.1177/2053951715594634**

Barocas, S., & Nissenbaum, H. (2014). Big Data's End Run Around Anonymity and Consent (pp. 44–75). In J. Lane, V. Stodden, S. Bendo, & H. Nissenbaum (Eds.), *Privacy, Big Data and the Public Good*. New York: Cambridge University Press.

Benjamin, R. (2019). *Race After Technology*. Cambridge: Polity.

Boellstorff, T. (2013). Making Big Data, In Theory. *First Monday*, *18*(10). doi:**10.5210/fm.v18i10.4869**

Chen J., & Qiu, J. (2019) Digital Utility: Datafication, regulation, labor, and DiDi's platformization of urban transport in china. *Chinese Journal of Communication*, *12*(3), 1-16. http://doi.org/10.1080/17544750.2019.1614964

Cohen, J. E. (2000). Examined Lives: Information Privacy and the Subject as Object, *Stanford Law Review*, *52*(5), 1373–1438. Retrieved from https://scholarship.law.georgetown.edu/facpub/810/

Cohen, J. E. (2013). What Privacy Is for. *Harvard Law Review*, *126*(7), 1904–1933. Retrieved from https://harvardlawreview.org/2013/05/what-privacy-is-for/

Cohen, J. E. (2018). The Biopolitical Public Domain: The Legal Construction of the Surveillance Economy. *Philosophy & Technology*, *31*(2), 213–233. doi:**10.1007/s13347-017-0258-2**

Couldry, N., & Mejias, U. A. (2019). *The Costs of Connection: How Data is Colonizing Human Life and Appropriating it for Capitalism*. Redwood City: Stanford University Press.

Couldry, N., & Mejias, U. A. (2018). Data Colonialism: Rethinking Big Data's Relation to the Contemporary Subject. *Television & New Media*, *20*(4). doi:**10.1177/1527476418796632**

Cowen, D. (2014). *The Deadly Life of Logistics*. Minneapolis: University of Minnesota Press.

Espeland, W., & Sauder, M. (2007). Rankings and Reactivity: How Public Measures Recreate Social Worlds. *American Journal of Sociology*, *113*(1), 1–40. doi:**10.1086/517897**

Eubanks, V. (2017) *Automating Inequality*. New York: St Martin's Press.

Fourcade, M., & Healy, K. (2013). Classification Situations: Life-Chances in the Neoliberal Era. *Accounting, Organizations and Society*, *38*(8), 559–572. doi:**10.1016/j.aos.2013.11.002**

Fuchs, C., & Mosco, V. (Eds.). (2017). *Marx in the Age of Digital Capitalism*. Leiden: Brill.

Gandy, O. (1993). *The Panoptic Sort*. Boulder: Westview Press.

Gillespie, T. (2010). The politics of 'platforms'. *New Media & Society*, *12*(3), 347–364. doi:**10.1177/1461444809342738**

Gitelman, L. (Ed.). (2013). *"Raw Data" is an Oxymoron*. Cambridge, MA: The MIT Press.

Hildebrandt, M. (2015). *Smart Technologies and the End(s) of Law*. Cheltenham: Edward Elgar Publishing.

IBM. (2018). *Annual Report*. Available at: https://www.ibm.com/annualreport/assets/downloads/IBM_Annual_Report_2018.pdf Last accesed: 26 November 2019.

Kelly, K. (2016). *The Inevitable*. New York: Penguin.

Kitchin, R. (2014). *The data revolution: big data, open data, data infrastructures & their consequences*. London: Sage.

Mayer-Schönberger, V., & Cukier, K. (2013). *Big Data: A Revolution That Will Transform How We Live, Work and Think*. London: John Murray.

Mejias, U. A. (2013). *Off the Network: Disrupting the Digital World*. Minneapolis: University of Minnesota Press.

Nieborg, D., & Helmond, A. (2019). The political economy of Facebook's platformization in the mobile ecosystem: Facebook Messenger as a platform instance. *Media culture & Society*, *41*(2), 196–218. doi:10.1177/0163443718818384

Nissenbaum, H. (2010). *Privacy in Context*. Redwood City, California: Stanford University Press.

Pentland, A. (2014). *Social Physics*. New York: Penguin.

Peña Gangadharan, S. (2012). Digital Inclusion and Data Profiling. *First Monday, 17*(5). doi:10.5210/fm.v17i5.3821

Poovey, M. (1998). *A History of the Modern Fact*. Chicago: Chicago University Press.

Rieder, B. (2012). What Is in PageRank? A Historical and Conceptual Investigation of Recursive Status Index. *Computational Culture: A Journal of Software Studies*, (2), 1–28. Retrieved from http://computationalculture.net/what_is_in_pagerank/

Rieder, B., & Sire, G. (2014). Conflicts of Interest and incentives to bias: A Microeconomic critique of Google's tangled position on the Web. *New Media & Society*, *16*(2), 195–211. doi:10.1177/1461444813481195

Rouvroy, A. (2012). The End(s) of Critique: Data Behaviourism versus Due Process . In M. Hildebrandt & E. de Vries (Eds.), *Privacy, Due Process and the Computational Turn* (pp. 143–167). London: Routledge.

Royal Society, The. (2019). *Machine Learning: the power and promise of computers than learn by example* [Report]. London: The Royal Society. Retrieved from https://royalsociety.org/topics-policy/projects/machine-learning/

Sadowski, J. (2019). When data is capital: Datafication, accumulation, and extraction. *Big Data & Society*, *6*(1). doi:10.1177/2053951718820549

Santos, B. de S. (2016). *Epistemologies of the South: Justice Against Epistemicide*. London: Routledge. doi:**10.4324/9781315634876**

Scholz, L. H. (2018). *Big Data is not Big Oil: the Role of Analogy in the Law of New Technologies* [Research Paper]. Tallahassee: Florida State University College of Law. Retrieved from **https://ssrn.com/abstract=3252543**

Scholz, T. (2016). *Uberworked and Underpaid*. Cambridge: Polity.

Schwartz, P. (1999). Internet Privacy and the State. *Connecticut Law Review, 32*, 815–859. Retrieved from **https://works.bepress.com/paul_schwartz/10/**

Scott, J. C. (1990) *Seeing Like a State: How Certain Schemes to Improve the Human Condition Have Failed*. New Haven; London: Yale University Press.

Thatcher, J., O'Sullivan, D., & Mahmoudi, D. (2017). Data Colonialism Through Accumulation by Dispossession: New Metaphors for Daily Data. *Environment and Planning D: Society and Space*, *34*(6), 990–1006. doi:**10.1177/0263775816633195**

van Dijck, J. (2014). Datafication, Dataism and Dataveillance: Big Data Between Scientific Paradigm and Ideology. *Surveillance & Society, 12*(2), 197–208. doi:**10.24908/ss.v12i2.4776**

van Dijck, J. (2013). *The Culture of Connectivity*. Oxford, UK: Oxford University Press.

Zuboff, S. (2015). Big other: surveillance capitalism and the prospects of an information civilization. *Journal of Information Technology*, *30*(1), 75–89. doi:**10.1057/jit.2015.5**

Zuboff, S. (2019) *The Age of Surveillance Capitalism*. London, UK: Profile Books.