Elmasri, R. A., & Navathe, S. B. (2017). *Fundamentals of database systems* (7th ed.). Harlow: Pearson.
Härder, T. (2005). DBMS architecture – still an open problem. In G. Vossen, F. Leymann, P. C. Lockemann, & W. Stucky (Eds.), *Datenbanksysteme in business, technologie und web, 11. fachtagung des gi-fachbereichs "datenbanken und informationssysteme" (dbis), karlsruhe, 2.-4. märz 2005* (Vol. 65, pp. 2–28). GI. Retrieved from http://subs.emis.de/LNI/Proceedings/Proceedings65/article3661.html.
Heath, T., & Bizer, C. (2011). *Linked data: Evolving the web into a global data space*. San Rafael: Morgan & Claypool Publishers.

Datacenter

► Data Center

Data-Driven Discovery

► Data Discovery

Datafication

Clare Southerton
Centre for Social Research in Health and Social Policy Research Centre, UNSW, Sydney, Sydney, NSW, Australia

Definition

Datafication refers to the process by which subjects, objects, and practices are transformed into digital data. Associated with the rise of digital technologies, digitization, and big data, many scholars argue datafication is intensifying as more dimensions of social life play out in digital spaces. Datafication renders a diverse range of information as machine-readable, quantifiable data for the purpose of aggregation and analysis. Datafication is also used as a term to describe a logic that sees things in the world as sources of data to be "mined" for correlations or sold, and from which insights can be gained about human behavior and social issues. This term is often employed by scholars seeking to critique such logics and processes.

Overview

The concept of datafication was initially employed by scholars seeking to examine how the digital world is changing with the rise of big data and data economies. However, as datafication itself becomes more widespread, scholarship in a range of disciplines and subdisciplines has drawn on the concept to understand broader shifts towards rendering information as data for pattern analysis, beyond online platforms. The concept of datafication, in its present form, emerged in the last 5 years with the growth of data analytics, being popularized by Viktor Mayer-Schönberger and Kenneth Cukier's (2013) book *Big Data: A Revolution That Will Transform How We Live, Work, and Think*, who describe its capacity as a new approach for social research. While datafication is distinct from digitization, as Mayer-Schönberger and Cukier point out (2013), digitization is often part of the process of datification. So too is quantification, as when information is "datafied," it is reduced to elements of the information that can be counted, aggregated, calculated, and rendered machine-readable. As such there are significant complexities that are lost in this process that renders qualitative detail invisible, and indeed this critique has been substantially developed in the literature examining the logics of big data (see, e.g., Boyd and Crawford 2011; Kitchin 2014; van Dijck 2014). Furthermore, a range of methodological and epistemological issues are raised about the insights of data drawn from new data economies, in which there are a range of existing inequalities, as well as huge value to be found in encouraging participation in digitally mediated social interaction and practices of sharing personal information online (see, e.g., Birchall 2017; van Dijck 2013; Zuboff 2015).

The Datalogical Turn

As more and more aspects of social life have begun to generate digital data, the possibility of analyzing this data in order to produce opportunities for profit has substantially changed the nature of how digital infrastructures are oriented. In particular, the capacity that exists now to analyze large data sets, what we call 'big data', and the ability to draw together insights from multiple data sets (e.g., search engine data, social media demographic information, viewing history on YouTube etc.) has significantly changed how online platforms operate. Data scientists seek to produce findings on a wide range of issues by examining the data traces that individuals left behind. The big data analytics have been used in the private sector for a range of purposes including for filtering digital content or products in the online marketplace in the form of recommendations, and, most prominently, through targeted advertisements. In addition, datafication has been identified by some as an opportunity to gain unprecedented access to data for social research. Sometimes called "the datalogical turn" or "the computational turn," recently greater attention has been paid to the sociological insights offered by these large datasets. There has also been unease in the social sciences surrounding the use of big data, particularly social media data, to analyze culture and address social problems. Media scholar José van Dijck (2014) argues that datafication has become a pervasive ideology – "dataism" – in which the value and insights of aggregated data are seen as implicit. This ideology also places significant trust and legitimacy in the institutions that collect this data, despite their financial interests. Indeed, such an ideology is clear in the claims made early in the big data revolution about the exponential capacity of big data to explain social life, with some proponents of big data analysis proclaiming the "end" of social theory. These claims were made on the basis that theorizing *why* people act in certain ways, indeed the very questions that formed the basis of much social scientific inquiry, was rendered irrelevant by big data's ability to see patterns of actions on a mass scale. In essence, datafication was seen as a way to bypass the unnecessary complexity of social life and identify correlations, without the need for meaningful explanation. Many of these early claims have been tempered, especially as the predictive power of big data has failed to deliver on many of its utopian promises.

The Datafication of Social Life

The data that is aggregated by data scientists is predominantly made possible by the data traces that online interactions generate which can now be collected and analyzed, by what we might term the "datafication of social life." As social media platforms have come to host more of our daily interactions, these interactions have become parcels of data in the form of comments, likes, shares, and clicks. Alongside these interactions, our digital lives are also comprised of a range of data-gathering activities: web browsing and using search engines, interactions with advertisements, online shopping, digital content streaming, and a vast array of other digital practices are rendered as pieces of data that can be collated for analysis to identify trends and, likely if commercialized, further opportunities for profit (Lupton 2019). Even beyond the activities users undertake online, the geo-locative capacities of digital devices now allow the collection of detailed location data about their user. This datafication of social life has significantly changed the organization of digital platforms as profit can now be drawn by the collection of data, and as such dataveillance has become embedded into almost all aspects of digital interaction.

Beyond online interactions and social media use, recent years have seen datafication and data analytics spread to a range of fields. Scholars have identified the datafication of health, both in the trend of individual self-tracking technologies, such as fitness trackers and smart watches, and the ways in which clinical practice has become increasingly data-driven, especially when it comes to how governments deal with medical information (Ruckenstein and Schüll 2017). So too has education been impacted by datafication,

as children are increasingly monitored in schools by RFID in uniforms, facial recognition-enabled CCTV, and online monitoring of classwork (Taylor 2013). Scholars have also drawn attention to the forms of dataveillance impacting childhood beyond education, through parenting apps and child-tracking technologies. The spread of datafication points to the power of the pervasive ideology of datafication van Dijck (2014) described, whereby objective truth is to be found by rendering any social problem as digital data for computational analysis.

Critiques of Datafication

The logics of datafication have been substantially critiqued by social scientists. Privacy and surveillance scholars have highlighted widespread issues surrounding the way datafication facilitates the collection of personal information passively, in ways that platform users may not be aware of, and data is stored for a wide range of future uses that users cannot meaningfully consent to. As datafication spreads into more areas of social life, notions of consent become less helpful as users of digital platforms and datafied services often feel they do not have the option to opt out. Furthermore, large-scale data leaks and hacks of social media platforms demonstrate the fragility of even high-standard data protection systems.

In addition to privacy concerns, datafication can reproduce and even exacerbate existing social inequalities. Data-driven risk evaluation systems such as those now routinely employed by financial service providers and insurance companies can perpetuate discrimination against already marginalized communities (Leurs and Shepherd 2017). Furthermore, such discrimination is masked by the mythology of objectivity, insight, and accuracy surrounding these systems, despite their often opaque workings. While discrimination is certainly not new, nor does it arise solely as a product of datafication and systems drive by big data, however these systems facilitate discrimination in a manner that eludes observation and dangerously legitimizes inequalities as natural, evidenced by data, rather than a product of implicit bias.

Scholars have also raised concerns about the datafication of social science and the ways computational methods have impacted social research. Computational social science has been accused of presenting big data as speaking for itself, as a kind of capture of social relations rather than constituted by commercial forces and indeed by the new forms of digital sociality this data emerges from. For example, using social media data as a way to gauge public opinion often inappropriately represents such data as representative of society as a whole, neglecting important differences between the demographics of different platforms and the specific affordances of digital spaces, which may give rise to different responses. Similarly, these large data can establish correlations between seemingly disparate variables that, when presented as proof of a causal relationship, can prove misleading (Kitchin 2014). This is not to suggest scholars disregard this data, but rather caution must be employed to ensure that such data is appropriately contextualized with theoretically informed social scientific analysis. Qualitative differences must be examined through attention rather than smoothed out.

Conclusion

The process of datafication serves to transform a wide range of phenomena into digitized, quantifiable units of information for analysis. With the mass infiltration of smart technologies into everyday life and as more social interaction is filtered through social media platforms and other online services, data is now generated and collected from a diverse array of practices. Consequently, datafication and computational social science can offer significant insights into digitally embedded lives. However, as many scholars in the social sciences have argued, inevitably this process is reductive of the complexity of the original object and the rich social context to which it belongs.

Further Reading

Birchall, C. (2017). *Shareveillance: The dangers of openly sharing and covertly collecting data*. Minneapolis: University of Minnesota Press.

boyd, d., & Crawford, K. (2011). *Six Provocations for Big Data*. Presented at the a decade in internet Time: Symposium on the Dynamics of the Internet and Society, Oxford. https://doi.org/10.2139/ssrn.1926431.
Kitchin, R. (2014). Big data, new epistemologies and paradigm shifts. *Big Data & Society, 1*(1), 1–12.
Leurs, K., & Shepherd, T. (2017). Datafication & Discrimination. In M. T. Schäfer & K. van Es (Eds.), *The Datafied society: Studying culture through data* (pp. 211–234). Amsterdam: Amsterdam Press.
Lupton, D. (2019). *Data selves: More-than-human perspectives*. Cambridge: Polity.
Mayer-Schönberger, V., & Cukier, K. (2013). *Big data: A revolution that will transform how we live, work, and think*. London: John Murray Publishers.
Ruckenstein, M., & Schüll, N. D. (2017). The Datafication of health. *Annual Review of Anthropology, 46*(1), 261–278.
Taylor, E. (2013). Surveillance schools: A new era in education. In E. Taylor (Ed.), *Surveillance schools: Security, discipline and control in contemporary education* (pp. 15–39). London: Palgrave Macmillan UK.
van Dijck, J. (2013). *The culture of connectivity: A critical history of social media*. Oxford: Oxford University Press.
van Dijck, J. (2014). Datafication, dataism and dataveillance: Big data between scientific paradigm and ideology. *Surveillance & Society, 12*(2), 197–208.
Zuboff, S. (2015). Big other: Surveillance capitalism and the prospects of an information civilization. *Journal of Information Technology Impact, 30*(1), 75–89.

Data-Information-Knowledge-Action Model

Xiaogang Ma
Department of Computer Science, University of Idaho, Moscow, ID, USA

Synonyms

DIKW pyramid; Information hierarchy; Knowledge hierarchy; Knowledge pyramid

Introduction

Facing the massive amounts and various subjects of datasets in the Big Data era, it is impossible for humans to handle the datasets alone. Machines are needed in data manipulation, and a model of Data-Information-Knowledge-Action will help guide us through the process of applying big data to tackle scientific and societal issues. Knowledge is one's expertise or familiarity with a subject under working. Knowledge is necessary in the process to generate information from data about a certain issue, and then take actions. New knowledge can be generated on both the individual level and the community level, and certain explicit knowledge can be encoded as machine readable knowledge bases and be used as tools to facilitate the process of data management and analysis.

Understand the Concepts

The four concepts data, information, knowledge and action are often seen in the language people used in problem tackling and decision-makings for various scientific and societal issues. Data are the representation of some facts. We can see data of various topics, types, and dimensions in the real world, such as a geologic map of United Kingdom, records of sulfur dioxide concentration in the plume of Poás Volcano, Costa Rica, the weekly records of the sales of cereal in a Wal-Mart store located at Albany, NY, and all the Twitter tweets with hash tag #storm in January 2015. Data can be recorded on different media. The computer hard disks, thumb drives, and CD-ROMS that are popularly used in nowadays are just part of the media, and the use of computer readable media significantly promote and speed-up the transmission of data.

Information is the meaning of data as interpreted by human beings. For example, a geologist may find some initial clues for iron mine exploration by using a geologic map, a volcanologist may detect a few abnormal sulfur dioxide concentration values about the plume of a volcano, a business manager may find that the sales of the cereal of a certain brand have been rising in the past three weeks, and a social media analyst may find spatio-temporal correlations between tweets with hash tag #storm and the actual storm that happened in northeast United States in January, 2015. In the context of Big Data, there are massive amounts of data available but most of them could be just noise and are irrelevant to the subject under working. In this situation, a step of