iS
CHANNEL

Research Commentary: Data as Facts of the World

Florian Allwein
PhD Candidate in Information Systems and Innovation
Department of Management
London School of Economics and Political Science

KEYWORDS

Data
Information
Knowledge
Raw Data
Big Data

ABSTRACT

Information Systems research so far has not come up with a clear conceptualization of the central term of 'information'. A similar claim can be made about the term 'data', which is used inconsistently, but often taken to mean raw information. The field would benefit from clear definitions that distinguish these terms from each other. This paper argues for seeing data as facts of the world and information as data stored and processed in information systems. It aligns this with the ontology of critical realism and outlines how this view can guide future research. Information systems are seen as efforts to capture the facts of the world from the domain of the actual and store them in the domain of the empirical in order to make them accessible for analysis. This view is especially useful in the context of big data research, but also helps to justify why explanatory research in the social science tradition is still essential.

Introduction

As the field of Information Systems (IS) research is trying to develop original theories (Grover & Lyytinen 2015), there is surprising ambiguity over some of its key terms. For example, the field has not yet come up with a clear conceptualization of the central term 'information' (McKinney & Yoos 2010). A similar claim can be made about the term 'data', which is used inconsistently, but often taken to mean raw information. This is unfortunate as the term has become more widely used in IS research recently, especially due to the phenomenon of "big data" (Mayer-Schönberger & Cukier 2013; Constantiou & Kallinikos 2014; Goes 2014). This paper argues that the field would benefit from clear definitions that distinguish these terms from each other. It discusses the use of the terms 'information' and 'data' in the IS literature before proposing an alternative view that sees data as facts of the world and information as data stored and processed in information systems. It aligns this with the ontology of critical realism and outlines how this view can guide future research.

Conceptualizing data and information

As McKinney & Yoos (2010) show, "'Information' is poorly defined in the Information Systems research literature, and is almost always unspecified, a reflexive, all-purpose but indiscriminant solution to an unbounded variety of problems." (p. 329). They present four views of information, the most common being the token view, in which information "is synonymous with data: both refer to tokens manipulated by processes" (p. 331). It is interesting to note that even in a popular book on Big Data, the authors use the term 'information' to define data, as Mayer-Schönberger & Cukier (2013) define big data as "[t]he ability of society to harness *information* in novel ways to produce useful insights or goods and services of significant value" (p.2, my italics).

As the field of IS research is increasingly concerned with data, a clear definition of this term would be desirable. Looking at some basic definitions for these terms, we find marked differences: The Oxford English Dictionary defines data as "symbols on which operations are performed by a computer" and information as "that which is obtained by the processing of data". On the other hand, in a typical introductory IS textbook, Laudon (2014) defines data as "[s]treams of raw facts" (p. 609) and information as "[d]ata that have been shaped into a form that is meaningful and useful to human beings" (p. 612).

Recent IS papers on data apply definitions like "data as the measurement or description of states" (Kettinger & Li 2010, p.409) or "[d]ata is the underlying resource for [business intelligence]" (Lycett 2013, p.381). Kettinger & Li (2010) make an explicit attempt to conceptualize the terms. Motivated by concerns about the conceptual foundations of the field, they discuss different views on information. They acknowledge that the "'data --> information --> knowledge' hierarchy is very popular in the IS field" (p. 409), but point out that a "completely satisfying solution in defining these concepts and their relationship allowing for consistent and generalized use" (p. 409 f.) has not been found. They find that "[d]ata (…) have been generally defined as the measure or description of objects or events" (p.

Corresponding Author
Email Address: F.O.Allwein@lse.ac.uk (F. Allwein)

411) and "[i]nformation is usually defined as data processed into a form that has meaning to the user" (p. 412). Thus we see that there is a common notion of information being generated out of data. Data is generally seen as abstract quantities or characters. This is also reflected in a recent book on the data revolution: "Data are commonly understood to be the raw material produced by abstracting the world into categories, measures and other representational forms – numbers, characters, symbols, images, sounds, electromagnetic waves, bits – that constitute the building blocks from which information and knowledge are created" (Kitchin 2014, p.1). Overall, two main views of data can be distinguished:

- Data as raw information – this is the view shared by the OED, in Laudon's textbook and many recent information systems papers on data (e.g. Aaltonen & Tempini 2014; Constantiou & Kallinikos 2014).

- Data as synonymous to information – this is what (McKinney & Yoos 2010) call the "token view" of information.

A different view is proposed by Agarwal & Dhar (2014, p.444), who argue that "economic and social transactions are moving online, allowing for the digital capture of big data", which seems to imply that data exists outside the digital.

In summary, we see that there is significant disagreement in the IS field on what the two central terms of 'information' and 'data' mean and how they are distinguished. Moreover, this paper argues that the views outlined here are not sufficient. The token view appears unsatisfying as there is no reason to use these central terms interchangeably, or synonymously. The "raw information" view is more convincing and has been applied in much relevant research. However, it leaves central questions unanswered, e.g. how exactly data is distinguished from information. Would a raw stream of structured data classify as information? How about a confusing display of data from a database on an old ERP system? Is it structured enough to be called information?

A different view

This paper proposes a different view that was initially inspired by a website dedicated to "surviving and thriving in the age of information overload" (Ingebrigtsen 2016). This is how it is outlined there:

> Data is/are the facts of the World. For example, take yourself. You may be 5ft tall, have brown hair and blue eyes. All of this is "data". You have brown hair whether this is written down somewhere or not.
>
> In many ways, data can be thought of as a description of the World. We can perceive this data with our senses, and then the brain can process this. (…)
>
> Information allows us to expand our knowledge beyond the range of our senses. We can capture data in information, then move it about so that other people can access it at different times.

This can be related to the notion of data existing outside the digital, as formulated by Agarwal & Dhar (2014). Moreover, a similar idea has been developed earlier by Checkland & Holwell (1998):

> there is a distinction to be made between the great mass of facts and the sub-set of them which we select for attention, those to which we pay heed. The obvious word for the mass of [f]acts is 'data' (p. 89).

Consequently, it is argued here that data should be seen as facts of the world – which also fits the literal translation of the Latin data, 'that which is given'. Whereas Checkland & Holwell coin the term "capta" for the sub-set of data that is captured, this paper argues for calling them "information". This means that we should talk about information (rather than data) stored and processed in information systems.

The benefit of this view is that it distinguishes data from information both conceptually and epistemologically. Data is seen as something that exists in the world, whereas information is processed data (as in the "raw information" view outlined above). This is interesting as it seems to relate to the ontology of critical realism (e.g. Mingers 2004; Wynn & Williams 2012), especially the stratified view of reality as the domains of the real, the actual and the empirical. The real contains generative mechanisms but cannot be observed. The actual contains all events that occur, while the empirical contains the subset of the actual that is observed. Thus, it is argued that events in the domain of the actual usually create a data trail and that information systems are created to capture this data, turn it into information (in the domain of the empirical) and manage and manipulate this information.

Discussion

As the literature review has shown, the view proposed here contrasts with the views commonly held in Information Systems literature. Yet it has some distinct advantages that make it worth considering. First of all, it would solve the dilemma of the token view of information – the fact that the central terms of 'information' and 'data' are used interchangeably. Given the centrality of these terms for the field, it seems important to come up with definitions that clearly distinguish between them, and to use them consistently. Lee (2010) shows how the token view of information is "indispensable to IS practice" (p. 344) and "foundational to the other views" (p. 344 f.). He nevertheless argues that researchers should adopt other views more often. Either way, they should make an effort to define how they use these terms, e.g. using the typology by McKinney & Yoos (2010) as a starting point.

Critical realism has been increasingly taken up in IS research (e.g. Bygstad 2010; Henfridsson & Bygstad 2013; Zachariadis et al. 2013). Thus, the view of data as facts of the world presents an opportunity to strengthen the conceptual basis of IS research on data and information while at the same time aligning to this growing ontological tradition. Indeed, the parallels of this view to critical realism are interesting. In such a view, the appeal of "big data" technologies would be that they extend the scope of the domain of the empirical – as more and more data can be captured from the domain of the actual, big data tools turn this into accessible and useful information. This would apply to e.g. management information systems capturing real-time production data and turning it into information to present in a dashboard, but also for the "quantified self" movement, where individuals gain insight into their habits, e.g. by counting their daily steps (data) and storing them as information in web-based information systems in order to analyse and share it.

Information systems can thus be seen as efforts to capture "the facts of the world" from the domain of the actual and store them in the domain of the empirical in order to make them accessible for analysis. This also points to a reason why we still need social sciences and qualitative methods even in a world of big data: as the domain of the real cannot be directly observed, big data based approaches have no access to it either. The only way to research generative mechanisms in the domain of the real remains to hypothesize them and research their explanatory potential, e.g. using the method of retroduction as described by various researchers (e.g. Bygstad & Munkvold 2011; Volkoff & Strong 2013).

Given the on-going debate on the identity crisis in Information Systems research (Baskerville & Myers 2002; Benbasat & Zmud 2003; Grover & Lyytinen 2015), it appears the field would benefit from more variety in research and a distinctive profile based on its own theories. Coming up with a set of clear concepts is an important step in this direction. It is hoped that this paper will encourage a discussion along these lines.

References

Aaltonen, A. & Tempini, N., 2014. Everything counts in large amounts: A critical realist case study on data-based production. Journal of Information Technology, pp.1–14.

Agarwal, R. & Dhar, V., 2014. Big Data, Data Science, and Analytics: The Opportunity and Challenge for IS Research. Information Systems Research, 25(3), pp.443–448.

Baskerville, R.L. & Myers, M.D., 2002. Information Systems as a Reference Discipline. MIS Quarterly, 26(1), p.1.

Benbasat, I. & Zmud, R.W., 2003. The identity crisis within the is discipline: defining and communicating the discipline's core properties. MIS Quarterly, 27(2), pp.183–194.

Bygstad, B., 2010. Generative mechanisms for innovation in information infrastructures. Information and organization, 20(3/4), pp.156–168.

Bygstad, B. & Munkvold, B.E.B., 2011. In Search of Mechanisms. Conducting a Critical Realist Data Analysis. ICIS, pp.1–15.

Checkland, P. & Holwell, S., 1998. Information, systems, and information systems, Wiley.

Constantiou, I.D. & Kallinikos, J., 2014. New games, new rules: Big data and the changing context of strategy. Journal of Information Technology, 30(1), pp.44–57.

Goes, P.B., 2014. Big Data and IS Research. MIS Quarterly, 38(3), pp.iii–viii.

Grover, V. & Lyytinen, K., 2015. New State of Play in Information Systems Research: The Push to the Edges. MIS Quarterly, 39(2), pp.271–296.

Henfridsson, O. & Bygstad, B., 2013. The Generative Mechanisms of Digital Infrastucture Evolution. MIS Quarterly, 37(3), pp.907–931.

Ingebrigtsen, N., 2016. The Differences Between Data, Information and Knowledge. Infogineering. Available at: http://www.infogineering.net/data-information-knowledge.htm [Accessed February 22, 2016].

Kettinger, W.J. & Li, Y., 2010. The infological equation extended: towards conceptual clarity in the relationship between data, information and knowledge. European Journal of Information Systems, 19(4), pp.409–421.

Kitchin, R., 2014. Data revolution: Big data, open data, data infrastructures & their consequences, Los Angeles: Los Angeles : Sage Publications.

Laudon, K.C. & Laudon, J.P. (Jane P., 2014. Management information systems: Managing the digital firm Thirteenth., Harlow, Essex: Pearson Education.

Lee, A.S., 2010. Retrospect and prospect: information systems research in the last and next 25 years. Journal of Information Technology, 25(4), pp.336–348.

Lycett, M., 2013. "Datafication": making sense of (big) data in a complex world. European Journal of Information Systems, 22(4), pp.381–386.

Mayer-Schönberger, V. & Cukier, K., 2013. Big Data: A Revolution that will transform how we live, work, and think, Boston: Houghton Mifflin Harcourt.

McKinney, E.H. & Yoos, C.J., 2010. Information about information: a taxonomy of views. MIS Quarterly, 34(2), pp.329–344.

Mingers, J., 2004. Real-izing information systems: critical realism as an underpinning philosophy for information systems. Information and Organization, 14(2), pp.87–103.

Volkoff, O. & Strong, D., 2013. Critical realism and affordances: theorizing IT-associated organizational change processes. MIS Quarterly, 37(3), pp.819–834.

Wynn, D. & Williams, C.K., 2012. Principles for Conducting Critical Realist Case Study Research in Information Systems. Management Information Systems Quarterly, 36(3), pp.787–810.

Zachariadis, M., Scott, S. & Barrett, M., 2013. Methodological Implications of Critical Realism for Mixed-Methods Research. MIS Quarterly, 37(3), pp.855–879.