Research Commentary: Data as Facts of the World

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

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 & Zervas 2012). This paper argues that the use of the terms ‘information’ and ‘data’ in the IS literature before proposing an alternative view that distinguishes 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 reflective, 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 “[t]he 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; Goos 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.

References


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 collected to produce the raw material produced by abstracting the world into categories, measures and other representational forms such as numbers, 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 on what the two central terms of ‘information’ and ‘data’ mean and how they are distinguished. Moreover, this paper argues that the views outlined above are not sufficient. The token view of data 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.

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 term ‘datum’ (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 comprises of the real, the actual and the empirical. The real contains generative mechanisms but cannot be observed. The actual contains all events that occur, whereas the empirical comprises 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 turn the data into information in the domain of the empirical.

This approach to capturing 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. Last (2010) shows how the token view of information is “indispensable” to IS research (p. 344 f.) 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; Zacharias 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 our disciplines to this growing ontological tradition. Indeed, the parallels of this view to critical realism are interesting. In both views, 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 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 to test their explicable potential, e.g. using the method of retrodiction as described by various researchers (e.g. Bygstad & Munkvold 2011; Volkoff & Strong 2013).

Given the ongoing debate on the identity crisis in Information Systems research (Baskerville & Myers 2002; Benbasat & Zmud 2003; Grover & Lytrivitis 2015), it appears the field would benefit from more variance in research. How do we capture influential profile based on our 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


