

'DATAFICATION': MAKING SENSE OF (BIG) DATA IN A COMPLEX WORLD

Professor Mark Lycett

School of Information Systems, Computing and Mathematics

Brunel University

Uxbridge

Middlesex

UB8 3PH

United Kingdom

Pre-correction version

Full paper/citation: Lycett, M. (2013), Datafication: Making Sense of (Big) Data in a Complex World, *European Journal of Information Systems*, **22** (4), pp. 381-386.

INTRODUCTION

Business Intelligence (BI) is a topic of growing importance for both industry and academia. Though still viewed from its technology roots, it is slowly broadening to encompass the data infrastructure, applications, tools and best practices related required for the effective capture, representation and delivery of data to inform decision-making and action. The lines between enterprise and social intelligence are also becoming increasingly blurred, as action from decision-making is oriented at influencing people's (future) behaviour. From an industry perspective, BI is consequently seen as a fruitful foundation for innovation, competition and productivity. From an academic perspective, the richness and importance of the area for research is becoming increasingly apparent (Chen *et al*, 2012; Sharma *et al*, Forthcoming) .

Data is the underlying resource for BI (accepting limitations of data protection and privacy). Arguably, it is the increasing availability of data (so called 'big data' ultimately) that provides the impetus for BI, most typically characterised by industry commentators in the '3 Vs' as follows:

- *Volume* proposes that there is key benefit in being able to process large amounts of data – the underlying analytic thesis being that more data beats better models. Key considerations here relate to scalability, distribution, the ability to process etc.
- *Velocity* proposes that the data flow rate is important, not least in relation to the feedback loop to action. Key considerations here include the granularity of data streams, understanding what can be discarded and the latency acceptable in relation to data, decision-making and action-taking.
- *Variety* proposes that data is messy in reality, coming from many sources in many different forms – often unstructured, error ridden and inconsistent in nature. Key considerations here include the degree of information loss in cleanup, semantic integration and versatility in representation.

More recently, *value* has emerged as a fourth and, perhaps, integrating 'V' – doing something valuable with the data is important! As an example, the core finding of a recent joint MIT Sloan Management Review/IBM Institute for Business Value report was that top performing organisations cite (effective) analytics as a key differentiator (Lavelle *et al*, 2010). Aside from highlighting a widespread belief that analytics offers value, the report found that top performing organisations made decisions based on rigorous analysis at more than double the rate of lower performing organisations (taking note that competitive position was self-rated). Increasingly, however, analytic insight is used within top performing organisations to guide both future strategies *and* day-to-day operations.

The primary barrier to achieving the competitive advantage that (big) data can offer was 'lack of understanding of how to use analytics to improve the business' – the capture, represent and deliver side of the equation. Undeniably, the way that we both deliver and consume products and services is increasingly instrumented; but instrumentation alone is not enough to deliver value. Consequently, consideration is required on what is necessary to deliver value and the challenges and opportunities that delivering value holds. 'Datafication' provides the lens for that consideration as, within the 'ether', it is increasingly being used to characterise the reliance of enterprises on data (and their data infrastructures), the democratisation of data and, of focus here, the process of turning that data into something of value.

DATAFYING THE WORLD: DEMATERIALISATION, LIQUIDITY AND DENSITY

Datafication can be conceptualised via three innovative concepts that allow the logic of value creation to be rethought – dematerialisation, liquification and density (Normann, 2001). Dematerialisation highlights the ability to separate the informational aspect of an asset/resource and its use in context from the physical world. Liquification highlights the point that, once dematerialised, information can be easily manipulated and moved around (given a suitable infrastructure) allowing resources and activity sets that were closely linked physically to be unbundled and 'rebundled' – in ways that may have traditionally been difficult, overly time-consuming or expensive. Density is the best (re)combination of resources, mobilised for a particular context, at a given time and place – it is the outcome of the value creation process.

There is no mistaking that IT provides an important driving force in this logic of value creation, as it provides the infrastructure and artefacts that liberate us from constraints related to when things can be done (time), where things can be done (place), who can do what (actors) and with whom it can be done (configurations/constellations) (Normann, 2001). In addition, this logic of value creation (alongside information technology) provides us with some degrees of freedom from 'frozen knowledge': Normann (*ibid*) argues that physical products are efficient because they are reproducible and predictable, but that they are instruments in which activity and knowledge is frozen at the point of production – the accumulation of past knowledge and activities in effect. He further argues that the distinction between goods and services is misleading, proposing '*offerings*' as a richer conception that are "a reconfiguration of a whole process of value creation, so that the process – rather than the physical object – is optimized in terms of relevant actors, asset availability and asset costs" (*ibid*, p.115). Readers familiar with service-dominant logic will see clear and tangible parallels (e.g., Vargo and Lusch, 2004; Lusch *et al*, 2007)

In this school of thinking, offerings are the input to (rather than the output of) the value creation process, which is primarily defined from the perspective of value-in-use – the interplay between the offering and the customer. It is the dynamics of this interplay that is perhaps most interesting as the value creation context of today (driven by information technology) allows a much denser reconfiguration of resources into *co-created* value patterns and, indeed, a greater (more-individualised) variety of patterns. As these ideas are quite abstract, however, let us examine them in the context of Netflix for want of an example.

NETFLIX AS AN ANECDOTAL EXAMPLE

Netflix is today a provider of on-demand Internet streaming media operating in over 40 countries with some 33 million streaming members. Historically, however, its operation was more physical in nature with its core business in mail order-based disc rental (DVD and Blu-ray). Crudely speaking, the operating model here is that a subscriber creates and maintains a queue (an ordered list) of media content they wish to rent (e.g., a film). With a limit on the overall number of discs, content can be retained for as long as a subscriber wishes. To rent a new disc, however, the subscriber mails a previous one back to Netflix, who then forwards the next available disc in the subscriber's queue. The business goal of the disc rental model is therefore to help people fill their queue.

Dematerialisation is present in the disc rental model. Standardised and structured metadata has evolved around the disc content itself alongside data associated with the subscriber queue for example (see Bloch *et al*, 2009). Aside from the business concept itself (the unbundling of physical content from shops), liquification is further present in-and-around the understanding and management of queues; improving demand management/distribution (e.g., via a practice called throttling) and, importantly, recommendation as a means of developing a relationship between the provider and subscriber. From the subscriber's perspective, the queue is the primary manifestation of density, as it mobilises the resources for a given context, time and place. In addition, the notion of value-in-use is present as it is the queue and queue content that is the focus (the interaction between provider and subscriber) not the physical product per se (i.e., the disc). Bounds are still placed on each of these concepts however, as selection is distant in time from viewing and there is no feedback during viewing.

It is fair to say that there is a step-change in all aspects in the streaming incarnation of the Netflix business, where IT infrastructure and artefacts fully liberate media content from its physical manifestation (e.g., the disc and its postal delivery). With streaming, subscribers can sample videos before settling on a particular one, they can consume several videos in one sitting and Netflix can observe viewing statistics to a much finer degree (and in real-time to a greater extent). Much more data is dematerialised in the streaming model therefore. In addition, data sources have evolved to be many and varied – including the catalogue data (more than a 1000 facets are now associated with a title), search terms, stream queues and plays, interactions and external sources such as film reviews and social data. Further, liquification is more present in the increasing pervasiveness of recommendation in the streaming model. The removal of time and distance from the business model has increased the potential for interaction between provider and subscriber via *dynamic* personalisation (by household, genre etc.), explanation of content to promote trust, rating, ranking and review and social influence stemming from what connected friends have watched or rated.

In daily terms, the Netflix dematerialisation has some 30 million daily plays and 3 million odd searches to inform the dynamics of recommendation. What that offers via dematerialisation and liquidity combined has allowed an interesting manifestation of density, via Netflix's recent move from streaming content to producing it. Statistical analysis of years' worth of user behaviour was used to inform content rather than recommendation, presenting Netflix with an interesting intersection of genre, actors and director. The result of that data intersection was their recent remake of the House of Cards television series – a political thriller starring Kevin Spacey, directed by David Fincher.

SENSE-MAKING: THE CHALLENGE OF DATAFICATION

Netflix provides an example of a more provocative argument related to datafication (Anderson, 2008). At its strongest, this argument is that semantic or causal analysis is not required; technology now provides us with the means to spot the patterns, trends, and relationships in political, economic, social and environmental relationships without hypotheses or models to guide the journey. A 'Kelvinian' age of measurement is upon us and correlation trumps causation in effect. More tempered, Anderson's argument is that the datafied world forces us to view data mathematically first and establish the context for it later.

Let us reflect on this. If dematerialisation, liquification and density provide conceptual foundations for datafication, analytics is arguably the engine – key to doing something valuable with the data. Early treatises on the importance of analytics for competition (e.g., Davenport and Harris, 2007) are now echoed empirically (Lavalle *et al*, 2010) and arguments for analytics as a fundamental research area in its own right are increasing (e.g., Chen *et al*, 2012). In and of itself, however, this engine is (potentially) useless unless the outcomes can be incorporated into complex decision-making (Shah *et al*, 2012) and empower actions that can provide value. It may be the case that the context for value comes later – indeed, the Netflix move into content production could be seen as post-hoc – but let us not lose sight of Deming's ethos.

The point made is that datafication is an information technology driven *sense-making process*. In the organisational literature, sense-making refers to processes of organising using the technology of language (e.g., labelling and categorizing for instance) to identify and regularise memories into plausible explanations and whole narratives (Brown *et al*, 2008). Sense-making concerns itself with how people generate what they interpret in terms of: (a) The nature of how and why aspects are singled out from the stream-of-experience; and (b) how interpretations are made explicit through concrete activity.

Framed in this manner there are some interesting challenges that arise outside of those typically defined in-and-around business intelligence. For brevity, we highlight selected aspects that show the fundamental importance (and constraints) of sense-making and provide interesting avenues of research:

- *Conceptualisation and codification*. Dematerialisation is fundamentally about abstraction and some thought needs to be given to what counts as a 'dot' in the first place. The core point here is that while data evoke frames of reference, it is (pre-existing) frames of reference that select and connect data (Klein *et al*, 2006). In fixing the frame of reference in

any formalisation the outcomes are that: (a) Knowledge of the world is de-contextualised and 'fixed' when the world is emergent; (b) the intended meaning has to be recovered and re-contextualised upon use; where (c) making the best of that recovery in a given context may be artificially constrained by the meaning originally imposed (in that perception and action are unduly constrained) (Tuomi, 1999). As many of the sources on which (big) data analysis draw are explicitly (or implicitly) based on abstractions of the world (e.g., framed as entity-attribute-relationship, object-property-behaviour etc), this cannot be ignored.

- *Algorithmic treatment.* The algorithms that clean data at the point of capture, find patterns, trends and relationships in its volume, velocity and variety are closed in their nature. This is of import because they do not only extract and derive meaning from the world; increasingly they are starting to shape it (see Slavin, 2011 for expansion and interesting examples). As Anderson (2008) notes, in many cases that shaping is semantically blind – i.e., Google are happy to match ads to content without 'knowing' anything about either. Netflix provides a salient example of algorithmic shaping effects – according to their figures 75% of content choice is now influenced by recommendation. Though algorithms are 'doers and not informed sceptics', the shaping power inherent in their design should clearly not be underestimated. The emergent shaping from collaborating algorithms and the outcomes of conflicting algorithms all lack deep understanding however.
- *Re-representation of the world.* Unsurprisingly, the sophistication of data visualisation is increasing alongside the need to present more complex data in more aesthetically pleasing and informative ways – both quickly and clearly. Aside from sophistication, trends in the visualisation and presentation of data are toward it being dynamic and interactive and self-service in nature (from a business intelligence perspective at least). Wilkinson (2005) argues that there is an over-arching grammar of graphics whereby the meaning of a (statistical) graphic is determined by the mapping produced by the function chain linking data and graphic. As with algorithms, in practice, the degrees of freedom we have over that function chain are limited in current technology. Self service and interaction are positive, but both aspects alongside function chains need, for example, to be informed by how people actively navigate and search through information structures, what 'information' people choose to consume and what conceptual models people induce about their environmental/virtual landscape in action. This understanding is limited at the moment (Pirolli, 2007; Pirolli, 2009).

CONCLUSION (AND POTTED IMPLICATIONS)

This article has considered what is necessary for data analytics to deliver value alongside some of the challenges and opportunities that delivering value holds. Datafication has provided the focus of discussion; the concepts of dematerialisation, liquification and density proposed as the foundations of understanding datafication and analytics a key means of deriving value. In constructing this narrative, we have made the argument that datafication is an information technology driven *sense-making process*. There are two aspects of the sense-making that have particular salience in relation to the mandates of this work – plausibility and enacting. Plausibility is embodied in accounts (Maitlis, 2005), which construct order among sets of entities (events, people, actions and things for example), making tangible a perceived reality that can be enacted. The brief examples of issues around conceptualisation, algorithmic treatment and re-representation are intended to show that accounts are closed as datafication stands.

Understanding this is important as it should be clear that datafication will unavoidably omit many features of the world, distort others and potentially add features that are not apparent in the first instance. Outcomes will unavoidably channel users towards some kinds of inferences and/or actions more readily than others (see Thaler and Sunstein, 2008 for good examples), funnel creativity in certain directions or, at worst, stifle it (see Leonard, 2013). As with graphics, perhaps what we need is a grammar of sense-making to cater for different frames of reference, allow moves between them, that they can be revised or reconsidered and the accuracy, samples, biases and quality be better understood. The notion that some accounts may be better than others and/or that their plausibility be compared may be positive for novel densities.

In conclusion, it would be foolish to suggest that datafication does not hold the potential for generating significant insight and novel densities – there are already many examples where the value is clear and new business opportunities have emerged. Equally, it is perhaps dangerous to ignore the sense-making aspects of the process – densities do not emerge from data alone.

REFERENCES

ANDERSON C (2008) The End of Theory: The Data Deluge Makes the Scientific Method Obsolete. *Wired* 16(07), http://www.wired.com/science/discoveries/magazine/16-07/pb_theory.

BLOCH M, COX A, CRAVEN MCGINTY J and QUEALY K (2009) A Peek into Netflix Queues. http://www.nytimes.com/interactive/2010/01/10/nyregion/20100110-netflix-map.html?_r=2&, accessed 5th April 2013.

BROWN AD, STACEY P and NANDHAKUMAR J (2008) Making sense of sensemaking narratives. *Human Relations* 61(8), 1035-1062.

CHEN H, CHIANG RHL and STOREY VC (2012) Business Intelligence and Analytics: from Big Data to Big Impact. *MIS Quarterly* 36(4), 1165-1188.

DAVENPORT TH and HARRIS JG (2007) *Competing on analytics: The new science of winning*. Harvard Business School Press, Harvard, MA.

KLEIN G, MOON B and HOFFMAN RR (2006) Making sense of sensemaking 2: A macrocognitive model. *IEEE Intelligent Systems* 21(5), 88-92.

LAVALLE S, HOPKINS MS, LESSR E, SHOCKLEY R and KRUSCHWICZ N, Eds. (2010) *Analytics: The New Path to Value*. *Research Report*. MIT Sloan Management Review, Boston, MA.

LEONARD A (2013) How Netflix is Turning Viewers into Puppets. http://www.salon.com/2013/02/01/how_netflix_is_turning_viewers_into_puppets/, accessed 6th April 2013.

LUSCH RF, VARGO SL and O'BRIEN M (2007) Competing through service: Insights from service-dominant logic. *Journal of Retailing* 83(1), 5-18.

MAITLIS S (2005) The social processes of organizational sensemaking. *Academy of Management Journal* 48(1), 21-49.

NORMANN R (2001) Reframing business: When the map changes the landscape. John Wiley & Sons, Chichester, Sussex.

PIROLI P (2007) Information foraging theory: Adaptive interaction with information. Oxford University Press, Oxford.

PIROLI P (2009) Powers of 10: Modeling Complex Information-Seeking Systems at Multiple Scales. *Computer* 42(3), 33-40.

SHAH S, HORNE A and CAPELLÁ J (2012) Good data won't guarantee good decisions. *Harvard Business Review* 90(4), 23-25.

SHARMA R, MITHAS S and KANKANHALLI A (Forthcoming) Special Issue on Transforming Decision-Making Processes: the Next IS Frontier. *European Journal of Information Systems* .

SLAVIN K (2011) How Algorithms Shape Our World.
http://www.ted.com/talks/kevin_slavin_how_algorithms_shape_our_world.html, accessed 11th April 2013.

THALER RH and SUNSTEIN CR (2008) Nudge: Improving Decisions about Health, Wealth and Happiness. Yale University Press, Newhaven and London.

TUOMI I (1999) Data Is More Than Knowledge: Implications of the Reversed Knowledge Hierarchy for Knowledge Management and Organizational Memory. *Journal of Management Information Systems* 16(3), 103-117.

VARGO SL and LUSCH RF (2004) Evolving to a new dominant logic for Marketing. *Journal of Marketing* 68(1), 1-17.

WILKINSON L (2005) The Grammar of Graphics. Springer, New York.