Spring 3-19-2013

Theorizing Data, Information and Knowledge constructs and their inter-relationship for effective Data Analytics

Martin Douglas
Cranfield University - School of Management, martin.douglas@cranfield.ac.uk

Joe Peppard
Cranfield University - School of Management, joe.peppard@cranfield.ac.uk

Theorizing Data, Information and Knowledge constructs and their inter-relationship

Martin Douglas & Joe Peppard
Cranfield University, School of Management, UK
Email: martin.douglas@cranfield.ac.uk

Abstract

Good explanatory constructs for Data, Information and Knowledge are central to the Information Systems (IS) field in general, and in particular to theorising how best to generate insight from Data. The central role of Knowledge within such theory has been highlighted recently, as well as the importance of Learning and Research frames (for Data Analytics). Building on these ideas, this paper briefly reviews several related literatures, for relevant ideas to enrich IS theory building. A consensus is found as to the complex, socially constructed nature of Knowledge or Knowing, and the importance of human sensemaking for theorizing how new insight is generated. The paper argues for an intuitive conceptual and practical distinction between Data (which exists as an independent, reified resource), and Information and Knowledge (both of which are embodied or embrained). It briefly outlines how the ideas identified can contribute to theorizing, highlighting specific areas for further inter-disciplinary research.

Keywords: Data, Information, Knowledge, Theory, Analytics, Learning

1.0 Introduction

1.1 A 'Big Data' imperative for better theory

Conceptual clarity about Data, Information, Knowledge and their interaction has long been recognised as fundamental to Information Systems as a discipline (Checkland & Holwell: 1998; Davis & Olson: 1984), although achieving a consensus within the field has proved elusive (Kettinger & Li: 2010, Checkland & Holwell: 1998). The opportunity to exploit the recent, rapid growth in Data (Kettinger & Marchand: 2011, Davenport: 2009, Davenport, Harris, De Long & Jacobson: 2001, Marchand, Kettinger & Rollins: 2001) brings renewed interest and urgency to this issue, driven by the question of how best to generate insight (i.e. new Information and Knowledge) from Data. Indeed, this may come to be seen as an increasingly important dynamic capability for organisations.

While this growth in Data (often termed 'Big Data') has prompted many initiatives, implementing a variety of Data Analytics technologies (Ranjan & Bhattnagar: 2011, Bose: 2009), many result in mixed outcomes, i.e. 'a wealth of Data but a poverty of insight' (Marchand & Peppard: 2013, Yeoh & Koronios: 2010, Wixom & Watson:

2001, Cooper, Watson & Wixom: 2000). While projects typically focus on technical implementation, many researchers argue that human and social factors are likely to be more important (Marchand & Peppard: 2013, Yeoh & Koronios: 2010, Hopkins, Lavalle & Balboni: 2010, Wang & Wang: 2008, Nemati & Barko: 2003, Marchand et al: 2001). This may point to a lack of understanding and framing problems, i.e. good theory.

1.2 Shortcomings in dominant IS concepts and theory

Many models within common use in IS research and practice, relate Data, Information, Knowledge and insight in a clear hierarchy and present moving between each as relatively straightforward and linear, although some IS researchers suggests this is a more complex, interdependent process (Kettinger & Li: 2010). In recognition of the need for better theory in this area, Kettinger & Li (2010) propose a Knowledge-based theory of Information, extending Langefors' Infological Equation. Their theory asserts that Information is a function of the interaction of Data and Knowledge. They see Data, Information and Knowledge as distinct, presenting a reductionist, positivist formulation of these constructs and their relationship, mainly grounded in codified aspects of Knowledge, Data and Information.

However, they acknowledge human differences in meaning attribution and the importance of a social dimension, although these don't feature prominently in their theory. Their paper also doesn't really engage with social constructionist or Sensemaking perspectives, and while it proposes an evolutionary mechanism for new Knowledge creation it doesn't offer a compelling explanation of how such 'natural variation' or generating alternative ideas occurs.

1.3 Addressing the social deficit in IS concepts and theory

This paper seeks to contribute a social perspective to IS theorising in this area, to complement the current, dominant IS view outlined above, with which to enhance our ability to generate insights from Data. The paper uses the conceptualisation offered by the soft systems strand within IS, as a familiar, social constructionist starting point. Building on Kettinger & Li's (2010) argument that Knowledge is critical to generating new insight, as well as the similarity of generating insight to research and learning (Marchand & Peppard's: 2013, Wang & Wang: 2008), the bulk of the paper reviews these adjacent areas of literature, with a particular emphasis on their social constructionist strands, to see what they can contribute to IS theorising.

Based on this review the paper identifies several important ideas for a socially constructed framing of these concepts and theory. In particular, it argues for the importance of making an intuitive conceptual and practical distinction between Data (which exists as an independent, reified resource) on one hand, and Information and Knowledge (both of which are embodied or embrained) on the other hand. Finally, the paper highlights and argues for the importance of further inter-disciplinary engagement and research across IS, Knowledge Management and Organisational Learning in order to further develop theory about generating insight from Data.

2.0 A social constructionist starting point within IS

In Checkland & Holwell's (1998) review and reflection of the IS field, they discerned no consensus as to concepts of *Data*, *Information* and *Knowledge*. They criticise traditional input-process-output thinking within IS as founded on rationalist, positivist traditions of management research, underpinned by a resource based view of the firm and Information. However, they do identify an important partial consensus that 'Data is transformed into Information when meaning is attributed to it' (p.95), which implies a uniquely human activity, i.e. Information cannot exist independently of humans. Their chief criticism is that the clusters of ideas in use within the field fail to make a clear distinction between Data available or observable versus selected Data for attention (which they term CAPTA).

They go on to present a more compelling starting point for theorizing, summarized in Figure 1 below, explaining how these key concepts are linked, incorporating ideas of human cognition, as well as the importance of context, interest and existing Knowledge as important in relation to the process of attributing meaning to facts.

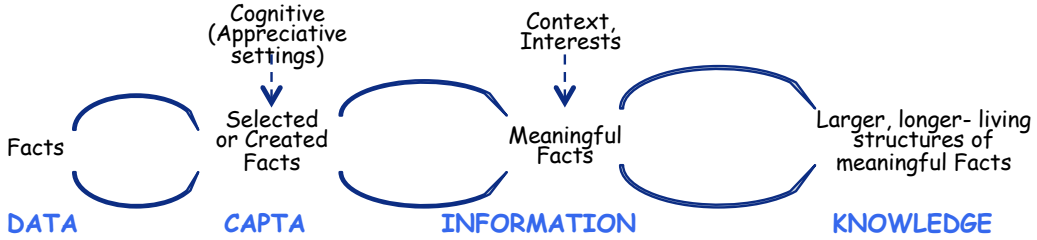

Figure 1. The links between Data, Capta, Information and Knowledge (Checkland & Holwell: 1998: p. 90)

Checkland and Holwell (1998) argue that the social relationship context is central to meaning attribution and Information use. They go on to illustrate (Figure 2) how any Information user perceives the real world, either directly, via formal Information systems or un-designed (informal) Information systems. In all cases a cognitive filter

is involved when someone perceives various sources of sensory Data. Meaning is also attributed to this Data in relation to their internalised memory, Knowledge and values. While not illustrated, the importance of relationships, collective sense-making and seeking of consensus on goals is also stressed, as opposed to what is characterised as straightforward framing of Information use to support goal-seeking decision-making.

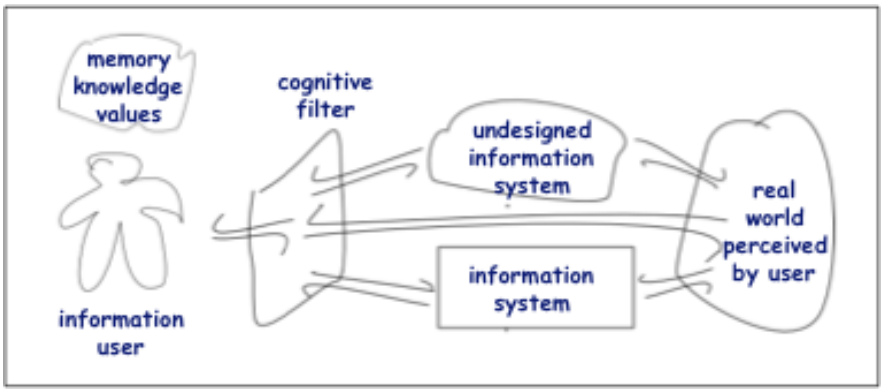

Figure 2. Information system context based on Land (Checkland & Holwell: 1998: p98)

2.1 Information and Knowledge as a continuum

The idea presented above: of Information as '*meaningful facts*' versus Knowledge as '*larger, longer-living structures of meaningful facts'* represents an important insight. It implies a continuum between Information and Knowledge rather than discrete concepts, with increasing complexity in Information relationship structures, as well as increasing permanence, as differentiating dimensions as you move from Information to Knowledge. Firstly, this means that the common term *insight* (not well defined but widely used in connection with Data Analytics) could equally apply to both without being problematic. Secondly, this means that what we know about *Knowledge* may also be true and relevant for our thinking about *Information*. This is consistent with more recent arguments for information being viewed as a subset of Knowledge (Boell & Cecez-Kecmanovic: 2010).

While the idea of relating new Data or facts to existing Knowledge is implicit in their explanation, Checkland and Holwell (1998) do not explore this in detail, nor do they really address the process of transforming Information into Knowledge, i.e. how this happens and what factors may be important in this process. With this in mind, we consider what other disciplines concerned with Knowledge creation (as a phenomenon) may have to contribute to related IS theorising.

3.0 How other disciplines theorize creating Knowledge

This section briefly introduces various disciplines interested in creating Knowledge or insight, then goes on to briefly outline central ideas and debates in those disciplines that focus on how *situated individuals* and *groups* or *teams* generate insight or Knowledge. Several fields were identified and are presented in Figure 3 below.

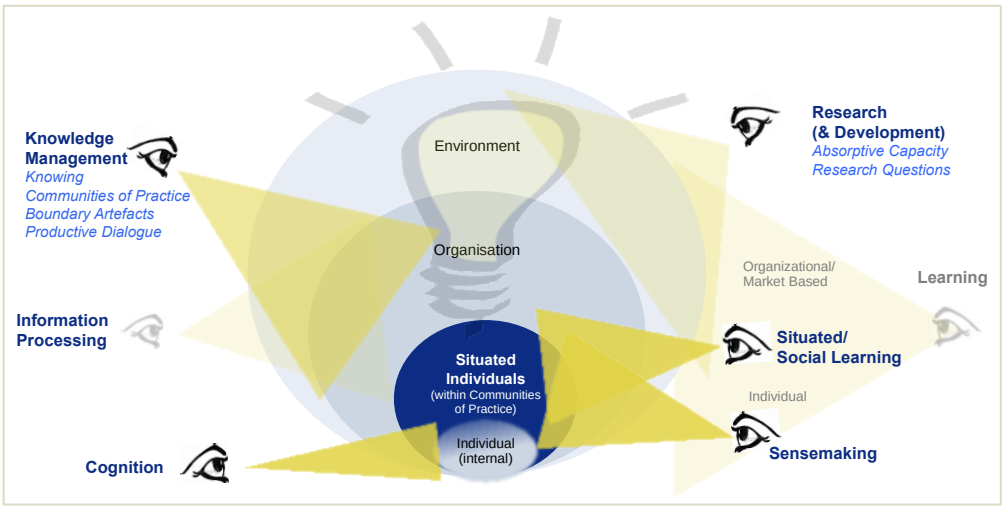

Figure 3 – Various disciplinary perspectives on generating insights (from Data)

Many of these disciplines have distinct research purposes and focus on different aspects of creating Knowledge, some more closely than others, at different levels of analysis, from different perspectives, and often in particular contexts (e.g. Research & Development). Several overlap or represent strands within broader fields. This is tentatively depicted in Figure 3 above, by the focus and strength of the 'beams' used for each.

Given Research & Development's (often external) focus on the phenomenon at an organisational level or unit of analysis, this field has not been reviewed in detail. In the following subsections, we now briefly summarize the chief ideas and debates emerging from each field (in turn) that might be useful for IS theorizing.

3.1 Cognition

This field has long been recognised in IS as important with Davis and Olson (1984) highlighting issues of cognitive bias (e.g. anchoring, etc) in their conceptual foundation for the field. Grounded in Psychology, this discourse is increasingly enriched by insights from neurology. It is focused at the individual level of analysis and focuses on how individuals internally process external stimuli and Data in relation to pre-existing mental models of reality, as well as how this influences the meaning

they attach to such Data and how this impacts on their behaviour or action. It recognises conscious and unconscious (or tacit) processing.

Evidence is accumulating that cognitive theory is consistent with a socially constructed view of Sensemaking, Knowledge Management and Learning (D'Eredita & Barreto: 2006). With a particular focus on *tacit* Knowledge accumulation, they review the Cognition literature and highlight the following three cognitive assumptions, for which they argue there is considerable empirical evidence:

- Constructing and relating tacit Knowledge is episodic in nature, based on individual instances or memories (the number of instances rather than their duration being important)
- Formulating new episodes is dependent on attention (i.e. we filter out what is perceived as less relevant Information during sensemaking)
- Relating current to past episodes depends on what cues, stimuli and related sensemaking and action/responses are attended to in drawing previous episodes from memory

D'Eredita & Barreto (2006) go on to conclude the following:

- tacit Knowledge is episodic in nature and based on accumulated experience,
- experience represents the sense that is made of current activity and experience by relating it to prior episodes or instances, and
- organizational tacit Knowledge results from active collaboration by individuals to construct meaning or episodes by relating current experience to previous episodic experiences

Kahneman (2011) also provides further support for this view, providing a challenge to simplistic, exclusively rationalist approaches and assumptions in connection with Data processing, decision-making, action and Learning.

3.2 Knowledge Management

Based on a review of the Knowledge Management literature, generating new Knowledge represents one of two broad research themes, the other focusing on the nature, classification and situation of various types of Knowledge (not covered in detail here). As in Information Systems, two broad schools of thought can be discerned in Knowledge Management (D'Eridita & Barreto: 2006) which seems to reflect a split along ontological lines, crystallising in a focus on treating Knowledge as an asset (i.e. as a resource) that can be exploited by the one school, versus a focus on *Knowing* as an activity or process by the other (Blackler: 1995, 1993). Both recognise the importance of tacit Knowledge, although they conceptualise this very differently, with important implications for how they believe new Knowledge may be created.

Criticism of the widely cited Resource view

The widely cited, resource focused school (Nonaka, Toyama & Konno: 2000, Nonaka: 1994) believe that creating new Knowledge is fundamentally about the interaction between tacit and explicit Knowledge. Nonaka (1994) identify four patterns of such interaction: Socialization, Combination, Externalization and Internalization, positing a continuous 'spiral model' for creating Knowledge, starting with individuals in an 'interaction community' or group (citing communities of practice as an example), then progressing to organizational and inter-organizational levels. Nonaka characterizes Knowledge creation as essentially about converting tacit Knowledge, mainly to explicit Knowledge, that can then be codified and shared as a resource. He distinguishes Knowledge creation from Learning, although his argument here is not clear: he doesn't seem to exclude action-based or social Learning and may simply be pointing to a concern about more traditional Learning focused on acquiring existing codified or abstract Knowledge.

Nonaka's notion of *Externalization* and *conversion* from tacit to explicit has drawn significant criticism (Tsoukas: 2005, Seely Brown & Duguid: 2000, Blackler: 1995). They point to a misunderstanding of the nature of tacit Knowledge. Tsoukas (2005) emphasises the complex nature of Knowledge, and its implicit tacit human dimensions, criticising commonly circulated definitions such as Nonaka's for adopting a very narrow Cartesian view of Knowledge and cognition and not revealing a useful enough conception of its constituent components and how these interrelate. Taking Polanyi as his starting point, he argues for his emphasis on the personal nature of Knowledge, i.e. 'All Knowing is personal Knowing' (Polanyi quoted by Tsoukas & Vladimirou: 2001: p.974).

Based on a close reading of Polanyi's (1966) work, Tsoukas identifies the following essential elements of tacit Knowledge:

- a coherent object of focus or phenomenon,
- comprising subsidiary elements, integrated subconsciously, and
- a person linking and integrating these components in pursuit of a purpose (realised in a focus for attention), using a semantic capacity and ontology to give meaning to the coherent whole.

He observes that tacit and explicit Knowledge are intertwined and inseparable, therefore he first argues that it is impossible to *convert* tacit to explicit Knowledge and, second, that any explicit Knowledge will have associated tacit predicates that are

inferred, based on experience, in light of a relevant action context, purpose and values.

In spite of the criticism, Nonaka's work still clearly points to the importance of tacit Knowledge and of the following factors or dimensions for Knowledge creation:

- its action orientation or purpose,
- its situation within a specific context and 'interaction community' or community of practice
- the importance of reflection and sensemaking activities, and
- its social nature and the associated importance of dialogue, language and metaphor for collective Learning, sensemaking and dissemination to occur.

Preferred emphasis on Knowing

By contrast, the social constructionist characterization, as outlined by Blackler (1995), emphasizes the process or activity of *Knowing*, rather than abstracted Knowledge as a resource, characterising *Knowing* as:

- Mediated
- Situated
- Provisional
- Pragmatic, and
- Contested.

Tsoukas (2005) also stresses the 'ineffable' nature of tacit Knowledge. He argues that the knower, focusing their attention on a focal target or purpose, is only peripherally aware of subsidiary particulars that may be relevant to their purpose or focal attention. Subsidiary particulars are assimilated through experience and practice and are interiorised over time, forming an 'unarticulated background' which influences and frames action but cannot be focused on *during* action. Instead, he argues that particulars can only be focused on during reflection on the activity with a view to drawing attention to features of our action that may have escaped our attention during action (which act as cues for interpretation and sensemaking). He therefore argues for the centrality of reflecting on practice and drawing attention to particulars or features of a phenomenon within a particular action context in order to generate new Knowledge or insight.

Given the time-bound, contextual, recursive and socialised nature of Knowledge, Tsoukas (2005) argues for the importance of what he terms narrative Knowledge, embedded in practice and constantly evolving through dialogue, reflection and practice, which he feels is likely to be neglected in institutional settings. He goes on to point out several paradoxes created by consistently privileging abstract, universal propositional Knowledge and its related simplifying, rules-based approach to

management. Instead, he sees both of these types of Knowledge as relevant and on a continuum, where propositional Knowledge and rules (grounded in tacit or implied predicates) are created to provide a consensus for action by providing a measure of certainty. He sees narrative Knowledge as having the advantage of recognizing the narrator, the context and its reflexivity, the narrator and characters' motives or purposes, and the particular temporal context of the Knowledge (i.e. not seeking universality). In doing so he stresses the critical role and use of language and dialogue, in order to facilitate make increasingly fine distinctions about a phenomenon, within a recognised action context. He regards this as a defining characteristic of Knowledge (at individual and organization levels) and argues for the importance of questions of epistemology both at the individual and organizational levels.

The importance of dialogue

Tsoukas (2009) finds widespread support for the importance of social practices and social interaction for new Knowledge to 'emerge', agreeing with Nonaka's idea of creating new Knowledge through dialogue and the importance of using metaphoric language to facilitate this. Turning to Dialogue and creative cognition research, he theorises and richly illustrates how dialogue can give rise to new Knowledge. In essence, he distinguishes *productive* dialogue (contrasted with *calculated*), describing it as collaborative exchanges to address mutually perceived 'strangeness' to generate new concepts or distinctions. When new distinctions are inter-subjectively accepted, these then represent new Knowledge, which gradually gains wider acceptance and becomes part of what he calls 'the inherited background', which forms the accepted Knowledge context for future action and dialogue.

As part of this work on Dialogue, Tsoukas (2009) points to the possible role and importance of what he terms *Boundary Artefacts* to facilitate productive conversations between actors or participants, by acting as 'an across-boundaries shareable framework, tool, object, or tangible demonstration' (p952). This seems a particularly useful concept for multidisciplinary teams (from different communities of practice) interacting to develop new insights. Tsoukas (2009) calls for more research on the dialogical creation of Knowledge between different communities of practice.

3.3 Learning

Learning is fundamentally about how people (and through them, teams and organizations) acquire existing and new Knowledge: consciously, through directed Learning or research activity, and unconsciously, through observation, action, participation and experience.

Within Organizational Learning, Easterby-Smith & Lyles (2003) identify four different Learning perspectives and related psychological groundings, including a social constructionist or social Learning perspective, and recognize clear overlaps between Knowledge Management and Organizational Learning. Elkjaer (2003), in her related review of the field, contrasts social Learning theory with *individual* Learning theory, which she argues emphasizes the enhancement of individual cognitive frames and privileges abstract Knowledge acquisition (e.g. conceptual Bodies of Knowledge) over that emerging from practice. She sees social Learning theory's starting point as our everyday lived experience. She equates *social Learning theory* with several other terms: *situated Learning*, *practice-based Learning* and *Learning as cultural processes*. She describes social Learning as ubiquitous and integral to human activity, and related to the purpose of becoming a practitioner (with its associated emphasis on identity formation and the influence of social and related power structures). She goes on to characterize what it is and how it occurs as follows:

> *'a social learning theory emphasizes informality, improvisation, collective action, conversation and sense making, and learning is of a distributed and provisional nature' (Elkjaer: 2003: p.44)*

As such, the aim of social Learning is less about acquiring existing Knowledge and address known or explicitly defined problems, and more about addressing unknown issues and address what she terms 'mystery'.

The importance of Communities of Practice

Knowledge Management has already highlighted Communities of Practice as an important context for socially situated Knowing and as a mechanism for generating new Knowledge. Wenger (1998) offers a broad conceptual framework for understanding and analysing situated Learning as a process of social participation within a 'community of practice'. He considers dialogical interaction central to such Learning and also acknowledges that the degree to which a practice community is reflective about its practice (which varies across different communities) is a very important characteristic in determining the kind of Learning it engages in. He sees

meaning as the ultimate product of Learning, and argues that it is contextual and located in a process of *negotiation* within a community of practice.

Importantly, he introduces and argues that it involves the interaction of two constituent processes (a complementary duality): *reification* and *participation*. He stresses the importance of identity in the negotiation of such meaning within a practice community's more formal structural elements (through membership), and explains how this leads to *economies of meaning* (through ownership of meaning, recognising power and institutionalisation). Based on this he argues for the importance of three processes for both identity formation and negotiating meaning: Engagement, Imagination and Alignment. He argues that they are also important considerations when formulating a design to facilitate emergent Learning.

Ongoing debate within Learning

Elkjaer (2003) examines some key challenges and debates within the field, starting with the tensions between individual and social Learning approaches touched on earlier: where social Learning theory argues for taking a more situated or contextual approach, individual Learning theory emphasises the knowledgeable, mobile individual.

She also discerns two very different aims for Learning – the first, a purposeful acquisition of explicit, abstract Knowledge, whereas the second focused on acquiring practitioner skills and gaining identity. She argues that people, self-evidently, engage in both types of Learning and persuasively argues for a synthesis of the two approaches, turning to Dewey and his ideas of inquiry, reflection and experience as a route to such a synthesis, which also addresses the inseparability of identity, practice and Knowledge (abstract and tacit).

3.4 Sensemaking

This is an area of research cited by several of the social constructionist perspectives outlined already as influential in providing underpinning ideas and constructs for their work. Weick (1995) steers clear of providing a neat or simple definition of Sensemaking, opting instead to provide a rich exposition of 'the seven distinguishing characteristics that set Sensemaking apart from other explanatory processes such as understanding, interpretation, and attribution' (p. 17), with which it might otherwise easily be confused or equated. He explains Sensemaking as a process that is:

- 'Grounded in identity construction

- Retrospective
- Enactive of sensible environments
- Social
- Ongoing
- Focused on and by extracted cues
- Driven by plausibility rather than accuracy' (Weick: 1995: p.17)

Weick's Sensemaking work contributes several key concepts and considerations in relation to how insights may emerge, in particular:

- The importance of ***enactment*** for meaning and the extraction of cues
- The distinction between ***uncertainty*** and ***ambiguity*** and its implication that more Data is only useful when addressing issues of uncertainty rather than ambiguity
- The idea of ***minimal sensible structures connecting cues*** with ***pre-existing frames*** in order to create meaning
- The impact of ***arousal*** on perceptions of context and its likely adverse impact on sensemaking (which may help explain the problem of Information or Data overload).

His work focuses largely at the level of the situated individual or group, essentially making sense of their context (most often organizational), attributing meaning to it in order to inform action. He makes an explicit connection to Lave and Wenger's (1991) work on situated Learning and goes on to describe Sensemaking's possible broader adoption as a *perspective*, as 'a frame of mind about frames of mind that is best treated as a set of heuristics rather than as an algorithm' (Weick: 1995: p. xii).

The importance of IS for sensemaking

Given the pervasiveness of Information Technology (IT), Weick (1995) argues for the need for more interpretive research of IS in relation to sensemaking. He identifies several concerns in relation to IT and how these may impact on sensemaking and the key ideas and constructs outlined above. These centre on the limitations of the rationalist, algorithmic IT approaches to anticipate all situations in a complex setting and their inability to facilitate re-framing and identifying new, relevant cues.

As an important example of such work, he cites Orlikowski (1991), who draws on structuration theory to offer a socially constructed explanation of IT systems and how they are used. The ideas of institutionalisation and use she explores are consistent with Weick (1995) and Wenger's (1998) characterisation of systems as reifications of practice. Subsequently, her work in this area has gone on to focus on issues of '*entanglement'* involved in tool and systems use and how these impact on framing and generating new Knowledge (Orlikowski: 2007, 2006 & 2002, 2000).

The most important idea to emerge from Orlikowski (2007, 2006, 2002, 2000, 1991) and Weick (1995), in relation to generating insight or Knowledge from Data, relates

to their characterization of systems as an institutionalisation (or reification) of the designer's thinking at the time of designing the system, although subject to subsequent reinterpretation by practitioners in using it. The extent to which these become fixed and inflexible are at the root of Weick's (1995) framing and sensemaking concerns. Similar concerns may arise for Data design and use, in terms of framing the phenomenon it purports to describe, e.g. which elements or dimensions are relevant, thereby bounding the nature of the questions that can be asked of such Data and what new Knowledge can be generated.

4.0 Implications for IS concept development and theorising

This section starts by recognising the fundamentally different starting point for theorising in IS, versus the disciplines reviewed above, in order to identify where and how these other disciplines can most usefully contribute. It then goes on to explain how they can be used to enrich IS theorising and concept development, grouping these contributions into two main areas:

- Refining concepts of Data, Information & Knowledge (4.2)
- Improving theory about generating insight from Data (4.3)

The paper's focus on Data Analytics as a context is reflected in the examples used throughout, as well as the narrow interaction focus of the second contribution area. Contributions to understanding other interactions are also likely but are not explored.
Finally, the section identifies several areas where inter-disciplinary research and collaboration may be particularly useful.

4.1 Different starting points for theorising

The review of adjacent disciplines' concepts and theory revealed strikingly different starting points for their theorising, compared to IS, which reflects the different challenges they have historically sought to address.
In the case of IS, the starting point has been automated Data and the challenges associated with capturing, organising, storing, processing, and transmitting such Data, reflected in 'Information Theory', with its semiological focus, and in the early term 'Electronic Data Processing' for the field (Boell & Cecez-Kecmanovic: 2010, Davis & Olson: 1984). Over time the field has broadened to encompass a broader scope: including Information and Information Systems (rather than just automated Data processing, related software and hardware) and a broader set social challenges rather

than purely technical (e.g. value and benefits). This is evident in Checkland & Holwell's (1998) Figure 1, which adopts Data as its starting point, as well as their subtle distinction of CAPTA from Data and Information, while being relatively less clear in their conceptualisation of Knowledge.

By contrast, the adjacent disciplines (particularly Knowledge Management and Learning) have tended to focus almost exclusively on conceptualising Knowledge, how to create new Knowledge or insight, and its mainly human or social transmission. Latterly, these fields have recognised the potential of Information Technology, as an enabler (Easterby-Smith & Lyles: 2003).

The above suggests that adjacent disciplines such as Knowledge Management and Learning are likely to be stronger than IS in their concepts for Knowledge and theory about generating insight, while IS concepts and theory about Data may be stronger. With this in mind, we turn to how they can contribute to extend and enrich our IS theorising.

4.2 Data, Information & Knowledge concepts

Another striking observation, when reviewing the adjacent fields, is the consensus and support for the inherently socially-constructed nature of Data, Information and Knowledge as phenomena.

These fields stress the *embodied, situated nature* of Knowledge and Learning, which *starts with socially situated individuals* attributing *meaning* within a particular, related *organisational action context* (often within communities of practice); where *meaning is enacted* and *framed* by *purpose*, via *attention* to *extracted cues*, which are then related to and dependent on *prior Knowledge and experience*. *Tacit and explicit Knowledge* dimensions are seen as *complementary and interdependent*. Knowledge emerges as *reified* or *institutionalised* by negotiating *economies of meaning*, arguing that such *codified* Knowledge can be viewed as Data, with tacit predicates. Its processual nature is emphasised introducing the notion of *Knowing* as preferable.

This consensus provides considerable support for existing initiatives in IS to conceptualise Information from a social constructionist perspective, for example in terms of identifying attributes using a socio-material lens (Boell & Cecez-Kecmanovic: 2010). Ideas and concepts from these adjacent disciplines may help simplify and extend this emerging IS thinking in two important ways:

- By facilitating a much richer concept of Data as a socio-material, reified phenomenon, quite distinct from the purely embodied, situated phenomena of Information and Knowledge.
- Many ideas, attributes and concepts about Knowledge could be adopted for Information. This may prompt a shift in emphasis and focus towards a dynamic, processual view of Information and Knowledge within IS.

These are briefly explained and illustrated below in the context of Data Analytics.

A richer concept of Data, distinct from Information & Knowledge

We have already argued that Information and Knowledge are inherently similar, both with an embodied or embrained nature, both centred on meaning attribution, so thinking of them as occurring on a continuum (or Information being a subset of Knowledge) seems more useful than as discrete concepts. This conception also rejects a simple rationalist, resource based view of them as phenomena.

Turning to captured Data though (or CAPTA), a resource based view seems more intuitive, given that it can exist physically and independently of a human observer, sensemaker or learner, i.e. has *materiality*. Figure 4 below, seeks to extend Checkland and Holwell's (1998) earlier illustration in Figure 3, to more clearly unpack some of the complexity of Data, highlighting its social communication and tacit elements. Automation is not depicted, which would further complexify the picture (e.g. unstructured automated data).

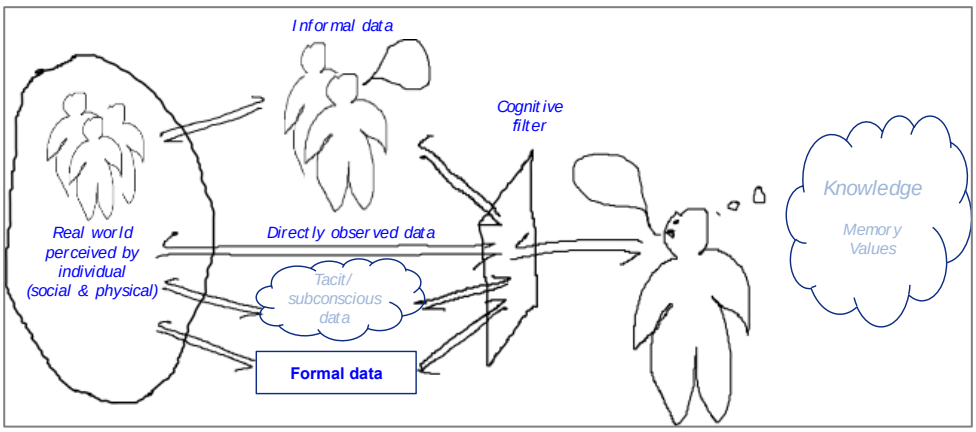

Figure 4. Extended illustration of Data use, versus Information & Knowledge

The tacit element introduced has a role both in interpretation and sensemaking of directly observed phenomena, communication and in interpreting reified or formal Data. Different levels of formalisation or complexity of Data presuppose very different levels and relative contributions of specialised technical and contextual

Knowledge (e.g. highly structured Data, versus relatively less structured verbal and non-verbal exchanges). At its most complex and structured, Data would encompass codified Knowledge, which exists independently of a sensemaker or learner. This highlights that Data can vary across several important dimensions: levels of complexity, structure and relationships, and Data's inevitable (inherent) tacit Knowledge predicates.

These aspects aren't adequately reflected and addressed in our current IS concept of Data. Although Kettinger & Li (2010) clearly recognise the importance of Knowledge to generate Information, this is typically framed as applying (rational) algorithmic logic (or codified, reified Knowledge) to Data. Especially in the context of automation, this simply produces more Data, which still requires meaning to be attributed to it by users.

This is where extending Orlikowski's (1991) conceptualisation of software systems, using structuration theory, can make a significant contribution to our thinking. In addition to software system's reified logic elements, its associated Data can similarly be thought of as a reified 'snapshot' of what designers identify as relevant dimensions to capture about a phenomenon (e.g. customer related fields, etc.). This neatly connects to Checkland and Holwell's (1998) concept of CAPTA. It also allows for the subsequent, unintended evolution in Data capture and use. This offers rich explanatory power, and explicitly recognises the social dimension of Data design and its ongoing use. Orlikowski's (2007, 2006 & 2002, 2000) subsequent work on '*entanglement'* is also useful to highlight Data's framing impact on users and organisations, institutionalising thinking (and related Knowledge) about a pheonomenon (e.g. customer), potentially introducing inflexibility and bounding our thinking. For example, the absence of social relationship or network Data fields within CRM software solutions reflect designers not anticipating the introduction of online social networks or adequately identifying social relationship Data as important (e.g. family, friends, etc).

Data Analytics introduces a further level of complexity, as Data used is often divorced from its source applications (or contexts), often integrating Data from different sources. This is where the literature on Research philosophy and method can also make a significant contribution, by highlighting Validity and Epistemological considerations: in terms of the purpose or (research) questions being posed, associated claims being made using the Data, and how well Data describes the phenomena of

interest. Data that purports to capture social (versus physical) phenomena prompt very different Validity criteria.

A rich, social conceptualisation of Data, developed along the lines outlined above, will greatly enhance our ability to understand and theorise about generating insight from Data.

Adopting Knowledge concepts for Information

Having considered Data, we now turn to contributions to conceptualising Information. If we accept the similarity or commonality of Information and Knowledge argued for earlier, then much or most of the Knowledge concepts and theory can be adopted for Information. In particular, the characteristics identified by Blackler (1995) represent an excellent starting point for thinking about Information as Mediated, Situated, Provisional, Pragmatic, and Contested. We would also anticipate similar tacit and explicit Information dimensions and interaction (probably overlapping and interacting simultaneously with more structured Knowledge). IS theorising and research could then focus on how some of these characteristics may vary along the proposed continuum of increasing complexity, structure and relationships.

The emphasis on *Knowing*, as a dynamic, emergent phenomenon, may also contribute towards a subtle but important shift in IS research towards greater emphasis and focus on the dynamic, situated, emergent dimensions of Information. This is also where Sensemaking and Cognition can contribute, enriching our appreciation of purposeful, situated *enactment* of meaning, and stressing the importance of both context and memory in determining focal attention, cue extraction and attributing meaning, by connecting these to relevant prior Knowledge and experience. This connects with the idea of a path dependency on prior Knowledge, highlighted by Learning and in work on *Absorptive Capacity* within Research & Development (Cohen & Levinthal: 1990).

A notion and term *Informing*, particularly enriched as described above, may be very useful within IS. From a research perspective, this may encourage and theoretically inform more immersed, longitudinal research about the social dimension of Information and Data use, as well as related systems design and adoption. This will be of particular value where these systems are specifically aimed at generating Data in order to *Inform* and generate new insight, which we turn to next.

4.3 Generating insight from Data

Adjacent fields shed significant light on the specific social processes involved in generating insight, an area highlighted earlier as relatively underdeveloped within IS. Learning, in particular, highlights the need to consider theorising at different levels of analysis (i.e. individual, group and organization), and to integrate theory across them. For example, questions arise about the potential need to distinguish individual sensemaking activities and Data use, from similar activities occurring within groups. The latter are likely to be far more complex, involving questions of shared meaning and communication. The relative role and balance of cognitive versus social factors may also vary at different levels. However, it is apparent that many questions about how insight is created are far from settled in these fields.

A good starting point for IS could be to build on existing efforts to theorise Data Analytics by Wang & Wang (2008), who make an explicit connection to Knowledge Management and Learning, proposing an iterative model illustrated in Figure 5 below.

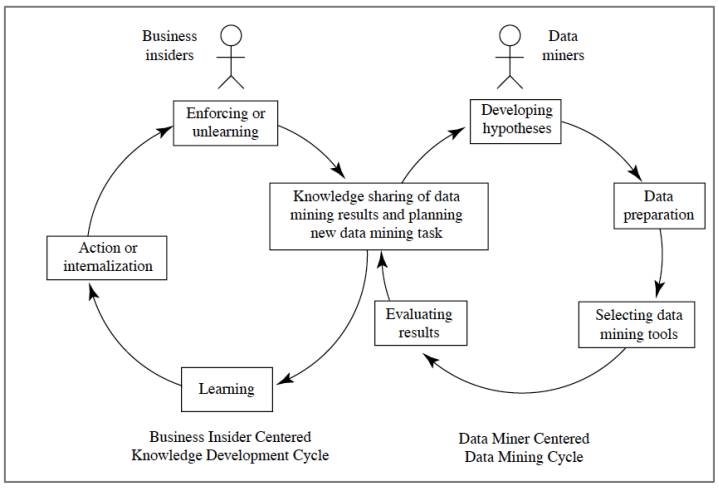

Figure 5 - Two cycles of knowledge development through Data Mining (Wang & Wang: 2008: p.627)

This reflects a fairly simplistic, rational view of learning from Data, without recognising any of the social complexity highlighted by adjacent fields in relation to Data selection, definition, achieving shared meaning or ultimate use, nor reflecting tacit elements. This points to the first contribution to our theorising: adding a social dimension to such a model.

Social processes of Reflection and Dialogue

As highlighted earlier, there is considerable consensus within Knowledge Management as to the importance of tacit knowledge, and social processes to generate new insight. Its action orientation and purpose are important for framing and

enactment of meaning or learning; reflection, language and dialogue are central and these are typically situated within an 'interaction community' context.

Sensemaking (Weick: 1995) contributes a framework and several concepts that may be useful as a theoretical research lens for examining the social processes at work when participants *frame* and *enact* meaning in relation to Data, especially in the face of *arousal*, which he argues narrows the participants' attention to peripheral, potentially important contextual *cues*. Another pertinent contribution is the importance he places on correctly identifying whether the sensemaking 'problem' is one of *Ambiguity* or *Uncertainty*, arguing that the latter benefits from more, relevant Data whereas the former does not. This distinction has important practical implications for framing Data Analytics initiatives to ensure they address realistic problems or questions.

In addition to his concept of developing 'an articulated background' of tacit knowledge which is important for cue extraction, Tsoukas' (2009) work on *productive dialogue* and associated *Boundary Artefacts* (to facilitate these, especially across different disciplines) is also likely to be particularly useful to our IS theorising about generating insight from Data. For example, the existence, role and use of documents or artefacts such as Data Models, Design Diagrams, Draft Report Designs and Visualisations could all represent *Boundary Artefacts*, helping develop shared understanding as to requirements during design, as well as shared meaning from the results of Data Analysis.

As highlighted in the discussion on conceptualising Data, how Data is defined or selected, in terms of relevant dimensions to capture and how they should be captured (and coded where necessary), is not trivial and fundamentally socially constructed. Kettinger & Marchand (2011) have already highlighted that *Sensing* Data requirements is an activity that is not appreciated or well understood by managers (Kettinger & Marchand: 2011). This may reflect the inherently social and unstructured nature of this activity, so the introduction of richer social theory and explanatory concepts here could advance theory and practice significantly.

Learning within and across Communities of Practice

Secondly, Wang and Wang's (2008) model, in identifying the interaction of Data miners and Business insiders as important, points to the likely contribution of Community of Practice frameworks and related situated learning theory (Wenger:

1998). Data Analytics teams can typically include technical IS developers and technicians, as well as various functional specialists (e.g. Marketing, Forensics, Product Development, etc.), depending on the nature, scope and scale of a Data Analytics project, highlighting their multi-disciplinary nature, which brings together different perspectives, a priori Knowledge and experience.

Wenger's (1998) framework addresses learning within and across such practice communities (or disciplines). This complements Tsoukas' (2009) approach, sharing his emphasis on the role of social, dialogical processes and reflection to generate insights, as well as concepts such as *Boundary Artefacts*. It also extends these to address issues such as Identity, the *duality* of *reification* and *participation*, and the inevitable negotiation involved in creating codified Knowledge.

Combining this framework, focused at the level or unit of a group, with *Sensemaking*, which is often used at the level of the individual, could also provide a useful way of triangulating findings in multi-level research, by using them for a priori coding of qualitative Data related to participating individual and group level outcomes and processes.

A Research Paradigm

Finally, *Research* (as a field) has a valuable contribution to make, as a potential broad characterisation of the process, a relevant Community of Practice to emulate, and in its formalised approaches and techniques. These approaches facilitate both *exploratory* and *directed* inquiry, adopting multiple *research paradigms*, analysing *Qualitative* and *Quantitative* Data and carefully evaluating results using appropriate *Validity* criteria to justify related Knowledge claims. These are likely to become increasingly important for Data Analytics, in order to avoid a simplistic positivist paradigm and a quantitative technique bias dominating the practice of Data Analytics, which fails to recognise its inevitable (often tacit) epistemological and ontological assumptions, particularly for inherently social phenomena (e.g. customer preferences). For example, analysing unstructured Data could benefit from specialised methods, techniques and underlying epistemology for textual analysis. A focus on *Research Questions* represents a further important contribution. Blaikie (2007) argues that *Research Questions* evolve from *what*, to *why* and ultimately *how* questions, and are refined as a richer understanding is gained of a phenomenon. Implicit or explicit Research Question refinement is likely to emerge from the learning cycles illustrated

in Figure 5 above, and an increasingly rich description of relevant Data (e.g. field dimensions), reflecting Tsoukas' (2005) essential notion of Knowledge as the ability to draw ever-finer distinctions about a phenomenon.

4.4 Areas for collaboration with other disciplines in theory-building

Earlier sections have highlighted several areas where adjacent fields can contribute greatly to IS theorising and research. Collaborating in these areas to build and test theory will benefit all fields involved. Given the mainly theoretical nature of much work within Knowledge Management and Organisational Learning, they will benefit from empirical research to test and refine or extend their theory, concepts and frameworks in different contexts.

What will also be apparent, are the significant remaining gaps in our understanding across all fields in connection with how to generate insight. While this phenomenon is clearly important to several fields, they often characterise it slightly differently in relation to particular problems and research questions arising in their fields (e.g. Research & Development and Absorptive Capacity). While this has led to different descriptions, language and constructs to describe the phenomenon and its related dimensions, hampering cross-fertilisation across disciplines, some researchers in these fields have already identified clear overlaps and synergies between fields.

This is particularly true of Organisational Learning and Knowledge Management (Easterby-Smith & Lyles: 2003, Vera & Crossan: 2003), which have identified areas of relative research strength, as well as areas of overlap, calling for further inter-disciplinary research, for instance about Situated Learning and Knowing in Communities of Practice, where research could contribute to both fields, and to investigate how current Knowledge impacts on future Learning. There has also been some recognition of the overlaps between Knowledge Management and Cognition (D'Eredita & Barreto: 2006), particularly in terms of Cognition's support for the social constructionist, *Knowing* perspective within Knowledge Management.

There has been relatively less recognition of overlaps between Knowledge Management and Learning with IS, except to recognise Technology as an important *enabler* (Hayes & Walsham: 2003, Alavi &Tiwana: 2003). That may be shifting with the recognition of the importance of Knowledge and Learning to Data Analytics (Marchand & Peppard: 2013, Wang & Wang: 2008) and more generally (Kettinger & Li: 2010). Generating new insights from Data seems to represent an important, special

case of Learning or creating Knowledge, differentiated by its explicit Data focus as well as Data's likely framing impact. Therefore, Data clearly lies at the intersection between IS and these fields. As it represents an area of conceptual strength for the field, we can make a significant contribution, working together with these adjacent fields to enhance theory in this area.

5.0 Conclusion

This paper has identified and introduced several useful concepts and theory from other fields that focus on creating Knowledge or insight, which will be useful for IS theorising. It has found a wide consensus for the importance of a social framing of Data, Information and Knowledge, and for the social processes involved in creating insight. These other fields start with an interest in Knowledge and theorize from this concept as a starting point, which complements IS thinking, which has traditionally started theorizing from Data.

The paper has argued for the importance and usefulness of distinguishing Data, as an independent, reified resource, on the one hand, from Information and Knowledge on the other (characterized as embodied or embrained and occurring on a continuum), because different issues and challenges are likely to arise in connection with managing them, associated with different solutions and interventions. However, for the distinction to be useful, we will need to promote and employ much greater discipline when using the terms Data and Information (in particular), as they are currently often used interchangeably.

The contributions identified also offer some preliminary ideas to IS practitioners as to particular social aspects of Data Analytics initiatives that may need more emphasis and attention, including:

- explicit consideration of framing initiatives and questions, adopting a broad Learning and Research framing for such initiatives
- inter-disciplinary team composition and achieving shared meaning across disciplines
- recognizing the limits and potential biases inherent in simply recruiting analytical skills (although these are necessary)
- working more closely and holistically with Learning and Knowledge professionals, and with general management to build related, broader skills and capabilities

Finally, the paper identifies several areas for inter-disciplinary engagement and research, especially at the intersection of IS, Knowledge Management and Organizational Learning fields, around a reinvigorated socially-constructed concept of Data.

References

Alavi, M. & Tiwana, A. (2003) *Knowledge Management: The Information Technology Dimension*, in Handbook of Organizational Learning and Knowledge Management (Eds. Easterby-Smith, M. & Lyles, M.A.), Blackwell Publishing, Malden.

Blackler, F. (1993) *Knowledge and the Theory of Organizations: Organizations As Activity Systems and the Reframing of Management*, Journal of Management Studies, 30(6), pp.863-884.

Blackler, F. (1995) *Knowledge, Knowledge Work and Organizations: an Overview and Interpretation*, Organization Studies, 16(6), pp.1021-1046.

Blaikie, N. (2007) Approaches to Social Enquiry: Advancing Knowledge (2nd Ed), Polity Press, Cambridge, United Kingdom.

Boell, S. Cecez-Kecmanovic, D. (2010) *Attributes of Information*, In Proceedings of 16th Americas Conference on Information Systems (AMCIS 2010), Paper 129, Lima, Peru.

Bose, R. (2009) *Advanced Analytics: Opportunities and Challenges*, Industrial Management & Data Systems, 109(2), pp.155-172.

Cohen, W.M. & Levinthal, D.A. (1990) *Absorptive Capacity: A New Perspective On Learning And Innovation*, Administration Science Quarterly, 35(1), pp.128-152.

Cooper, B., Watson, H. & Wixom, B. (2000) *Data Warehousing Supports Corporate Strategy at First American Corporation*, MIS Quarterly, 24(4), pp.547-567.

Checkland, P. & Holwell, S. (1998) Information, Systems and Information Systems – making sense of the field, John Wiley & Sons, Chichester, England.

Davenport, T.H., Harris, J.G., De Long, D.W. & Jacobson, A.L. (2001) *Data to Knowledge to Results: Building an Analytic Capability*, California Management Review, 43(2), pp.117-138.

Davenport, T.H. (2009) *Make Better Decisions*, Harvard Business Review, 87(11), p.117-123.

Davis, G.B. & Olson, M.H. (1984) Management Information Systems: Conceptual Foundations, Structure, and Development (2nd Ed), McGraw-Hill, pp.199-268.

D'Eredita, M.A., Barreto, C. (2006) *How Does Tacit Knowledge Proliferate? An Episode-Based Perspective*, Organization Studies, 27(12), pp.1821-1841.

Easterby-Smith, M. & Lyles, M.A. (2003) The Blackwell Handbook of Organizational Learning and Knowledge Management, Blackwell Publishing, Oxford.

Elkjaer, B. (2003) *Social Learning Theory: Learning as Participation in Social Processes*, in Handbook of Organizational Learning and Knowledge Management (Eds. Easterby-Smith, M. & Lyles, M.A.), Blackwell Publishing, Malden.

Hayes, N. & Walsham, G. (2003) *Knowledge Sharing and ICTs: A Relational Perspective*, in Handbook of Organizational Learning and Knowledge Management (Eds. Easterby-Smith, M. & Lyles, M.A.), Blackwell Publishing, Malden.

Hopkins, M.S., Lavalle, S. & Balboni, F. (2010) *The New Intelligent Enterprise 10 Insights: A First Look at the New Intelligent Enterprise Survey on Winning*

with Data 10 Data Points: Information and Analytics at Work, MIT Sloan Management Review, 52(1), pp.22-31.
Kahneman, D. (2011) Thinking, Fast and Slow, Allen Lane (Penguin Books), London.
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 (March), pp.409–421.
Kettinger, W.J. & Marchand, D.A. (2011) *Information Management Practices (IMP) from the Senior Manager's Perspective: an Investigation of the IMP Construct and its Measurement*, Information Systems Journal, 21, pp.385-406.
Lave, J. & Wenger, E. (1991) Situated Learning: Legitimate Peripheral Participation, Cambridge University Press, Cambridge.
Marchand, D.A., Kettinger, W.J. & Rollins, J.D. (2001) Information Orientation – The Link to Business Performance, Oxford University Press, Oxford.
Marchand, D.A. & Peppard, J. (2013) *Why IT Fumbles Analytics*, Harvard Business Review, 91(1), pp.104-112.
Nemati, H.R. & Barko, C.D. (2003) *Key Factors for Achieving Organizational Data-Mining Success*, Industrial Management & Data Systems, 103(4), pp.282–292.
Nonaka, I. (1994) *A Dynamic Theory of Organizational Knowledge Creation*, Organization Science, 5(1), pp.14-37.
Nonaka, I. Toyama, R. & Konno, N. (2000) *SECI, Ba and Leadership: a Unified Model of Dynamic Knowledge Creation*, Long Range Planning, 33(1), pp.5-34.
Orlikowski, W.J. (1991) *The Duality of Technology: Rethinking the Concept of Technology in Organizations*, Organization Science, 3(3), pp.398-427.
Orlikowski, W.J. (2000) *Using Technology and Constituting Structures: A Practice Lens for Studying Technology in Organizations*, Organization Science, 11(4), pp.404-428.
Orlikowski, W.J. (2002) *Knowing in Practice: Enacting a Collective Capability in Distributed Organizing*, Organization Science, 13(3), pp.249-273.
Orlikowski, W.J. (2006) *Material Knowing: the Scaffolding of Human Knowledgeability*, European Journal of Information Systems, 15(5), pp. 460-466.
Orlikowski, W.J. (2007) *Sociomaterial Practices: Exploring Technology at Work*, Organization Studies, 28(9), pp.1435-1448.
Polanyi, M. (1966) The Tacit Dimension, The University of Chicago Press, Chicago and London.
Ranjan, J. & Bhatnagar, V. (2011) *Role of Knowledge Management and Analytical CRM in Business: Data Mining based Framework*, The Learning Organization, 18(2), pp.131-148.
Said, E. (2011) The Value of Market Research Information: How do Clients of Market Research Services Contruct Value from their Usage of Market Research Information?, MResearch dissertation, Cranfield University, Cranfield.
Seely Brown, J. & Duguid, P. (2000) The Social Life of Information, Harvard Business School Press, Boston, Massachusetts.
Tsoukas, H. (2005) Complex Knowledge: Studies in Organizational Epistemology, Oxford University Press, Oxford.
Tsoukas, H. (2009) *A Dialogical Approach to the Creation of New Knowledge in Organizations*, Organization Science, 20(6), pp.941-957.

Tsoukas, H. & Vladimirou, E. (2001) *What is Organizational Knowledge?* Journal of Management studies, 38(7), pp.973-993.
Vera, D. & Crossan, M. (2003) *Organizational Learning and Knowledge Management: Toward and Integrative Framework*, in Handbook of Organizational Learning and Knowledge Management (Eds. Easterby-Smith, M. & Lyles, M.A.), Blackwell Publishing, Malden.
Wang, H. & Wang, S. (2008) *A Knowledge Management Approach to Data Mining Process for Business Intelligence*, Industrial Management & Data Systems, 108(5), pp.622-634.
Weick, K. (1995) Sensemaking in Organizations, Sage Publications, Thousand Oaks, California.
Wenger, E. (1998) Communities of Practice: Learning, Meaning & Identity, Cambridge University Press, New York.
Wixom, B.H. & Watson, H.J. (2001) *An Empirical Investigation of the Factors Affecting Data Warehousing Success*, MIS quarterly, 25(1), pp.17–41.
Yeoh, W. & Koronios, A. (2010) *Critical Success Factors for Business Intelligence Systems*, Journal of Computer Information Systems, 50(3), pp.23-32.

Appendix 1 – Key Contributions identified to aid Conceptualisation

Concept/ Idea/ Framework	**Contributing Literature** (Key Authors)	**Implication for IS Theorising**
Socially constructed nature of knowing and learning	**All fields reviewed**	• More emphasis on social nature of Information & Knowledge generally • Need to re-conceptualise Data in particular
Information-Knowledge Continuum	**Information Systems** (Checkland & Holwell: 1998)	• Concepts and characterisation of Knowledge are likely to apply to Information as well • Also serves to highlight a potential distinction between them and Data as a phenomenon
CAPTA as a selection of observable facts	**Information Systems** (Checkland & Holwell: 1998)	• Consistent with social constructionist view and reconceptualization of Data • Also a potential starting point for understanding managerial challenges with *Sensing* activities
Socio-material framework for Technology	**Organisational Learning** (Orlikowski: 1991)	• Prompts a similar conceptualisation for Data, using structuration concepts, with powerful explanatory power for design reification and unintended subsequent capture and use
Social characteristics of *Knowing*	**Knowledge Management** (Blackler: 1995)	• An equivalent notion on *informing* with similar characteristics may be useful • Approaches to improve knowing may also improve *informing*
Tacit-Explicit Knowledge, Codified & experiential, narrative knowledge complementarity Tacit Knowledge forms an unarticulated background for cue extraction and relating them to prior knowledge and experience	**Knowledge Management** (Tsoukas: 2005, Polanyi: 1966)	• Recognition & sensitivity of knowledge predicates inherent in all Data • Provides the basis and argument for codified knowledge to be thought of as Data • Consistent with reification and structuration ideas asserted elsewhere • Framing of focal attention and cue extraction highlighted as important (links to CAPTA idea)
Economies of Meaning	**Knowledge Management** (Wenger: 1998)	• Highlights the situated, negotiated quality of Knowledge, introducing power and identity as important factors or considerations

Appendix 2 – Key Contributions identified about generating insight

Concept/ Idea/ Framework	**Contributing Literature** (Key Authors)	**Implication for IS Theorising**
Importance of context and purpose for attention & enactment of meaning	**All fields reviewed**	• The importance of clarity of purpose and related consensus for Data Analytics initiatives • Recognition of the likely diversity of purpose, perspectives and prior knowledge/experience within multi-disciplinary Data Analytics teams • Recognition that diversity may facilitate identifying a wider range of cues and meaning
Importance of Productive Dialogue & Language	**Knowledge Management** (Tsoukas: 2009)	• An important social process to focus on when researching Data Analytics • Rich theory and concepts to use during Qualitative Fieldwork and coding
Role of Boundary Documents	**Knowledge Management/ Organisational Learning** (Tsoukas: 2009 & Wenger: 1998)	• Important artefacts to focus on in Data Analytics research, as a participant tool for mediating between different disciplines or communities of practice, to generate shared meaning • Instances of reified knowledge in their own right
Defining characteristics	**Sensemaking** (Weick: 1995)	• Potential coding approach for qualitative research at the individual level of analysis
Uncertainty & Ambiguity		• Important aspect of problem framing for Data Analytics initiatives to pay attention to • Indicator for when a Data-driven strategy is likely to be appropriate for an initiative
Narrowing impact of Arousal (i.t.o. peripheral attention)		• An important factor when considering questions or issues of Data overload in Data Analytics, as well as during framing the purpose or problem, selecting or defining Data and enacting meaning
Episodic memory-based nature, particularly of tacit knowledge	**Cognition** (D'Eredita & Barreto: 2006)	• Supports situated sensemaking and *knowing* theories of learning and generating new insight • Supports Tsoukas' ideas of the role of an unarticulated background • Supports Weick's sensemaking assumptions
Knowledge as ability to make ever-finer distinctions	**Knowledge Management** (Tsoukas: 2005)	• Dimensions of data are likely to be refined over time to accommodate finer distinctions • This needs to be anticipated during systems and Data design
Communities of Practice as situated contexts for learning Peripheral engagement Shared language and economies of meaning Identity & Negotiation Boundary	**Situated Organisational Learning** (Wenger: 1998)	• Potential to view Data Analytics both as a practice in its own right, as well as initiatives that cross practice areas/disciplines • Rich set of explanatory concepts for understanding, researching and describing Data Analytics initiatives (theoretically) • Stresses the importance of focusing on issues of Identity and Power in researching Data Analytics • The usefulness of Boundary documents and reified knowledge have already been highlighted

Documents		
Concept/ Idea/ Framework	**Contributing Literature** (Key Authors)	**Implication for IS Theorising**
Reification and Participation duality	**Situated Organisational Learning** (Wenger: 1998)	• An important area to pay attention to in researching Data Analytics • Consistent with earlier ideas of reification and structuration
Research Framing	**Learning/ Research Philosophy** (Blaikie: 2007)	• Research Questions as implicit or explicit purposes and objectives for Data Analytics initiatives • Importance of Validity criteria for Data Analytics initiatives – e.g. does Data capture all relevant dimensions of the phenomenon of interest • Potential for bounding or framing what can be known or discovered (based on epistemological and ontological assumptions)
Path dependency of new knowledge	**Learning/ Research & Development** (Cohen & Levinthal: 1990)	• Raised within Learning and Absorptive Capacity literatures as a potentially important limitation on discovering new knowledge • Aligned to ideas of cue extraction from an existing unarticulated background (Tsoukas), as well as sensemaking's focus on relating cues to prior knowledge (supported by Cognition)
Entanglement	**Organisational Learning** (Orlikowski: various)	• Potential framing impact of tool (and Data) use within Data Analytics initiatives • Useful theory and concepts for research, to provide theoretical explanations, and to identify relevant factors