Chapter 31
Shifting Perspectives:
A Process Model for Sense Making Under Uncertainty

Geoffrey Hill
Kent State University, USA

Pratim Datta
Kent State University, USA

William Acar
Kent State University, USA

ABSTRACT

This paper proposes that, in the context of generating actionable knowledge, uncertainties pertaining to big data streams should be recognized, categorized and accounted for at the appropriate level of knowledge management process models. Arguing that sensemaking from big data sources is a complex series of processes extending beyond just the application of sophisticated analytics, this paper proposes a big data reengineering (BDR) framework to guide requisite categorization, contextualization and remediation processes. The authors discuss the characteristics that uncertainty presents to organizations using big data streams as potential knowledge sources – surfacing relationships between the underlying knowledge flows and uncertainty and presenting typologies that categorize the effects of several common sources of uncertainty. These typologies also serve to provide guidance to transformation agent(s) regarding appropriate actions ultimately aimed at the generation of actionable knowledge.

INTRODUCTION

Why has a substantial increase in the amount of available data, generated by big data initiatives, not led to a comparable increase in actionable knowledge generation? A 2011 article from MIT Sloan Management Review reported upon a survey of nearly 3,000 executives, managers and analysts surmises:

Big data is getting bigger. Information is coming from instrumented, interconnected supply chains transmitting real-time data about fluctuations in everything from market demand to the weather. Additionally, strategic information has started arriving through unstructured digital channels: social media, smart

DOI: 10.4018/978-1-5225-7501-6.ch031

phone applications and an ever-increasing stream of emerging Internet- based gadgets. It's no wonder six out of 10 respondents said their organization has more data than it knows how to use effectively. (LaValle, Lesser, Shockley, Hopkins, & Kruschwitz, 2011, p. 29)

The promise of sustainable advantages, novel insights and actionable knowledge from this deluge of data has led 70% of enterprise organizations to deploy, or be actively planning to deploy, big data initiatives at an average cost of $8 million (Columbus, 2014). Unfortunately, due to technological dependencies inherent to big data, sensemaking projects are at times implemented as being predominantly a technical concern of data mining instead of a more holistic approach of knowledge discovery (Piatetsky-Shapiro, 1990) and sensemaking (Weick, 1979; 1995). Knowledge discovery and sensemaking is the "… need to understand any intended change in a way that 'makes sense' or fits into some revised interpretive scheme or system of meaning" (Gioia & Chittipeddi, 1991, p. 434). Conversely, data mining, a means for finding useful patterns in data, has been described as a single step in the *process* of knowledge discovery (Fayyad, Shapiro & Smyth, 1996). These same authors further contend that the blind application of data mining methods is a dangerous methodology that can easily lead to the "discovery" of spurious relationships and patterns that suffer from temporal instability (Fayyad, et al, 1996). Therefore, although such spurious patterns in the data may be exploitable for short term advantage, a focus upon data mining may serve to blind practitioners to potential sustainable advantages that may be realized from a more comprehensive approach.

The data mining perspective views knowledge extraction as a technical exercise relating to algorithms, capacity acquisition and processing power that expects knowledge to be generated as function of an organization's commitment and resource allocations in accordance with such theories as the resource based view of the firm (Coase, 1937; Wernerfelt, 1984; 1995). Continual increases in processing performance as suggested by Moore's Law (Moore, 1965) in conjunction with distributed processing frameworks such as MapReduce (Dean & Ghemawat, 2008) and Hadoop (ASF, 2014) provide the underlying technological capacity and raw processing power to harness ever increasing data requirements. Additionally, a variety of unstructured data mining techniques such as Sentiment Analysis (SA), Named Entity Recognition (NER), Natural Language Processing (NLP) and Entity Extraction (EE) have been successfully developed and implemented with the intent being to generate actionable business intelligence from big data. Unfortunately, when these technologies, techniques and methods are applied in the absence of rigorous theoretical underpinnings, the deluge of data often confounds, frustrates and overwhelms sensemaking efforts even in the face of lavish spending.

Projects that were once presented as models of big data sensemaking, such as Google's Flu Tracker (GFT) and the Distributed Common Ground System – Army (DCGS-A), are instead serving as examples of the inadequacies of some contemporary implementations. The GFT initially was able to provide an extremely accurate flu prediction model based upon Google's search query data. Unfortunately, the results have not stood the test of time and GFT's current predictive results rank below that of even simplistic lagged-models (Butler, 2013). DCGS-A was envisioned and actively developed as an intelligence component that gathers, analyzes and shares intelligence information as a common platform spanning all echelons of users (U.S. Army, 2014). At a current cost of approximately $5 billion, the system encompasses an array of software, sensors and data storage that suffers from instability and "overall network operational readiness issues" leading the project to cancel a benchmark readiness test scheduled for the fall of 2014 (Dilanian, 2014).