FINAL DRAFT – 10th Feb 2014

Forget information overload….

….the real challenge is content intelligence

Research Summary, MindMetre, February 2014

<u>Management Summary</u>

- According to research among senior information management professionals in the US and Europe, 85% of large organizations are creating more unstructured data than ever before, and it is critical to mine that content for the value it holds
- Unstructured data, such as customer correspondence, market intelligence, internal communications, product information, R&D reports, field and case notes, service information, customer/citizen feedback, and many other categories, holds enormous amounts of commercially valuable information and insights.
- According to 89% of respondent organizations, gaining greater insight into that unstructured data is key to their organization's commercial advantage and competitive edge
- So what's the main barrier to tapping into the competitive intelligence contained in that unstructured data?
- Received wisdom would say the key hurdle is the sheer volume of unstructured data being generated. According to this study, that is not the case.

- Only a third (34%) of organizations see sheer volume as the main challenge. The majority (71%) see the steepest hurdle as the fact that this data is fragmented and dispersed, in disparate formats

- And the next most cited obstacle (by 56% of respondents) is that information is either not meta-tagged/marked-up, or is ineffectively tagged, so it's difficult for the system to find what people are looking for

- Therefore, in order to improve performance and productivity, organizations need to invest in systems that can automatically and accurately categorize, meta-tag and mark-up unstructured data so that its value can be harnessed

- Equally, these systems need to be able to perform these tasks while leaving the data in its original locations, avoiding the expense of formatting and ingesting huge volumes of disparate information into a single hub

Volume is not the Big Story

Received wisdom would have it that the challenge of Big Data is all about dealing with volume. Certainly, big volumes there are. But, to be precise, the real challenge is finding what your need, when you need it. The objective is to make it easy, quick and efficient for people, externally or internally, to find the information they need amongst the vast reams available:- to differentiate the relevant needle from the irrelevant haystack. In fact, the Big Data challenge is discerning the 'small data' someone needs to do their job better.

The haystack is indeed vast. Nine tenths of all the data in the world has been created since 2010[1]. We process over 1 billion 'information transactions' every day that passes[2]. Every minute, we send 204 million emails and transmit a total of 640,000GB of data[3].

The major consultancies and analysts have established an approach to dealing with these vast volumes of data, known as the "3 Vs"[4]. The three Vs stand for *volume*, *velocity*, and *variety*. Notably, these three aspects of Big Data management are all focused on dealing with the sheer scale and heterogeneity of huge data flows (all of which are indeed challenges). However, the research results published in this paper reveal that a fourth V needs adding – *viability*. How does the user access the data that is viable and relevant for their task in hand? The answer to this question puts that last piece of the jigsaw in place to make Big Data truly deliver a return on its (equally big) investment.

Big Content and Content Intelligence

In a world of catch phrases, it was inevitable that Big Data should have developed an offshoot that specifically addresses the need for data viability, and that it should have been given its own epithet – namely, ***Big Content.*** One leading analyst talks about, "the notion of Big Content as shorthand for incorporating unstructured content into the Big Data world in a systematic and strategic way[5]." Such incorporation is achieved by making that unstructured content searchable and findable, be it emails, documents, plans, presentations, images, technical descriptions, research findings, field notes, sales contact management comments, social media participation, or any other data that is not pre-organised into database fields. This is becoming known as ***Content Intelligence***.

The task of trying to make the whole world of unstructured data *intelligent* is a mammoth task indeed. Most organizations are concentrating on the unstructured data over which they have more control – namely that resident in their own organization, or managed through their own channels to partners and customers (such as a loyalty scheme, a customer service portal, a reseller management system, a stakeholder portal, and so on). Few corporations – either in the public or private sector – yet understand the sheer richness of the 'knowledge assets' that they hold; even fewer can yet access that knowledge and deploy it to gain competitive advantage, operational excellence or improved public service delivery.

---

1 Science Daily, Big Data, for Better or Worse: 90% of World's Data Generated Over Last Two Years, 22 May 2013
2 Forbes CIO Network, Big Data: Getting Ready for the 2013 Big Bang, 15 Jan 2013
3 Source: Intel
4 Gartner blog, Deja VVVu, 14 Jan 2012
5 Darin Stewart, Gartner Blog Network, Big Content is still Big Data, 24 May 2013

The aims of harnessing of unstructured data within the corporation, or within its network tends to focus on a few key objectives:-

- Make the total sum of corporate knowledge and experience available to employees so they can do their job better and faster
- Eliminate the duplication of work
- Understand and repackage the corporation's 'knowledge assets' into new products or services
- Enable quicker/lower cost production of proposals and tender responses
- Obtain better strategic planning and management information
- Understand and manage risk better (often with pre-emptive action)
- Capture and act upon the intelligence and insight held in unstructured customer or citizen feedback

In summary, Big Data (dealing with the sheer volumes) needs to be joined by Big Content (capturing the reams of unstructured data), which in turn has to be imbued with Content Intelligence (to make relevant content easily and quickly accessible to those who need it, and stop work being inefficiently repeated).

Harnessing the Competitive Power of Content – The Research Findings

In order to gauge corporate views of the relationship between Big Data, Big Content and Content Intelligence, MindMetre (www.mindmetreresearch.com) polled opinion from senior information professionals at over 300 large organizations across the US and Europe. The key research findings are as follows.

Confirming the rising tide of unstructured material that might contain important information and insights, 85% of respondents were of the opinion that large organizations are creating more unstructured data than ever before. Respondents also confirmed that this unstructured data might comprise market intelligence, internal communications, product information, R&D reports, field and case notes, service information, internal and external presentations, customer/citizen feedback, technical plans, drawings and reports, and many other categories.

So what value do organizations place on all that unstructured data? A great deal, it would seem. According to 89% of respondent organizations, gaining greater insight into that unstructured data is key to their organization's commercial advantage and competitive edge. In other words, nine out of ten surveyed organizations feel that they will slip behind their competitors if they cannot find affordable and effective ways of tapping into the competitive intelligence held in their unstructured data.

So what's the main barrier to unlocking the competitive advantage contained in that unstructured data? Our research findings make one thing quite clear – it not the sheer volume of data, per se, that is the problem. Only a third (34%) of organizations see sheer data volumes as the main challenge. The majority (71%) see the steepest hurdle as the fact that this data is fragmented and dispersed, in disparate formats. And the next most cited obstacle (by 56% of respondents) is that information is not meta-tagged/marked-up, or is ineffectively tagged. Lack of tagging/mark-up/categorisation makes it difficult for people to find what they are looking for quickly, easily and

accurately; it also makes it virtually impossible to check whether work has been done before, and could easily be repurposed or updated rather than starting from scratch. Effective categorisation therefore brings a dual advantage: corporate knowledge and competitive insights can be tapped to release value; and existing work processes can be made radically more efficient.

Therefore, in order to improve performance *and* productivity, organizations need to invest in systems that can automatically and accurately categorize or meta-tag unstructured data so that its value can be harnessed. While sheer volume is not the greatest challenge, the task of manually categorising vast volumes of unstructured data is clearly neither affordable nor humanly possible. Equally, these automated meta-tagging/mark-up systems need to be able to perform the task while leaving the data in its original locations, avoiding the expense of formatting and ingesting huge volumes of disparate information into a single hub. The emerging class of solutions able to fulfil both these criteria would appear to be fulfilling a well-recognised demand from our panel of senior information professionals at larger corporations.

A Side Note on Sharepoint

On a final note, previous research from MindMetre has shown that MS SharePoint is clearly establishing itself as the principal gateway to this mass of enterprise content, with 53% of surveyed companies using SharePoint as their platform of choice for managing and utilizing content. According to Microsoft itself, more than three-quarters of Fortune 500 companies use SharePoint in some capacity.

The MindMetre research shows the main motives behind the adoption of MS Sharepoint are: eliminating work duplication; improving collaboration; and ensuring regulatory compliance. Interestingly, though, the research also shows that many users feel aspects of SharePoint functionality require enhancing to meet their needs – particularly when it comes to enterprise search. A key area of disappointment was the amount of content SharePoint actually makes readily findable: a quarter of SharePoint users believe its search facility allows them to access less than half of their organization's internal information. Overall, disappointment with the application's ability to enable access to internal information was a concern for 78 per cent of respondents – a worrying statistic, given that this is a key motivation behind many companies' implementation of the platform.

However, this research should be viewed in context. MS Sharepoint was conceived as an enterprise information management platform – used to manage, process, store and share documents. It was not intended to be a pure search facility and, given the range of implementations through which it is deployed, some gaps in functionality are inevitable. The new class of automated classification, tagging/mark-up and search solutions are providing a symbiotic enhancement to the Sharepoint platform to address this need.

Conclusions

It is clear from this research summary, that if organizations were to leverage their unstructured information assets more effectively, it would bolster product development, market intelligence, customer relationship management, new business pitches, risk management, the monetisation of enterprise knowledge and a host of other initiatives. However, these research findings also highlight that harnessing the unstructured information generated within an organization, its channels and its networks, remains a major challenge for many.

Dealing with the sheer volumes of data is an issue, but it is not the principal one. Big Data (volume) needs to be enhanced with attention to Big Content (unstructured data), itself made accessible and findable through Content Intelligence (accurately, automatically categorised and tagged content).

Making this content findable for staff, partners, clients, investors, regulators and other stakeholders is perceived as crucial for businesses to maintain competitive edge. Achieving the unprecedented precision and speed of information retrieval that full content intelligence can make possible will enable organizations to capitalise on the potential of Big Data like never before.

MindMetre Research, February 2014