Research Brief

Aberdeen Group
A Harte-Hanks Company

February 2013

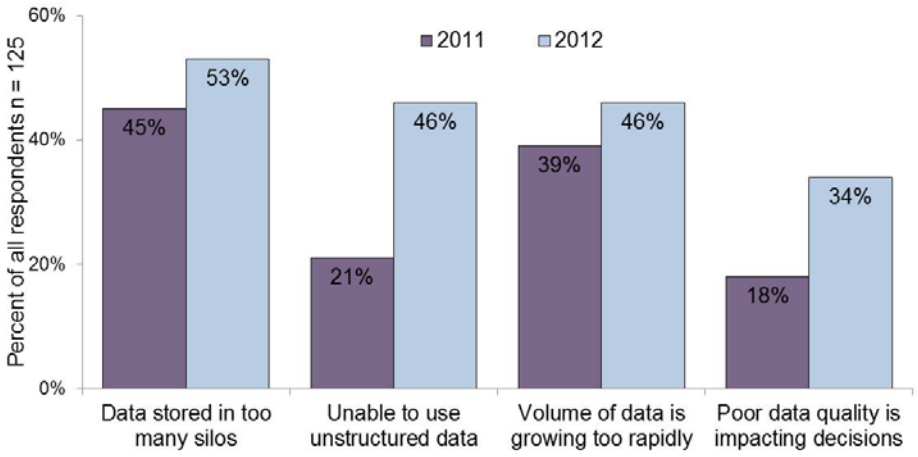
Content Analytics: Helping the Best-in-Class Drive Superior Customer Service

While many tools exist to govern and analyze business data stored neatly in relational databases, handling unstructured, text-based content can be much more challenging. With the rapid explosion in data volumes, however, effectively managing and understanding this content is now a top priority for many organizations. Aberdeen's January 2013 research on *Data Management for Big Data* collected information from 125 organizations worldwide and identified the Best-in-Class practices for unstructured data management. This study further examined organizations using content analytics, and how these tools relate to performance increases in sales effectiveness, customer satisfaction, and customer retention.

Business Context: The Datacenter Blind Spot

Most Business Intelligence (BI) solutions are designed to process historic, structured information and present these facts in a simple graphical interface. This is often enough to understand the mechanics of how a business operates — dollars sold, clients served, products shipped, etc. While this information is essential to smooth operations, too often companies continue to ignore a giant part of their business data, which is information that can provide valuable context and insight into the reason behind the facts.

Figure 1: Top Business Pressures for Data Management



Source: Aberdeen Group, January 2013

Research Brief

Aberdeen's Research Briefs provide a detailed exploration of a key finding from a primary research study, including key performance indicators, Best-in-Class insight, and vendor insight.

Definitions

√ **Content analytics** refers to the process of aggregating, categorizing, summarizing, exploring, and contextualizing unstructured data in all its forms.

√ **Unstructured data** refers to data stored in files, documents, presentations, spreadsheets, web pages, email messages, instant messages, images, audio files, video files, etc. While each of these formats do indeed have "structure," conventional use of the term unstructured data is intended to distinguish from data stored in *structured* formats (e.g. in databases).

Aberdeen's *Content Management* research (February 2012) reported that across all industries, over 51% of business data was considered unstructured, or stored in a format not meant for relational databases and most BI tools. The inability to properly manage or analyze this data is a top tier business pressure in Aberdeen's January 2013 *Data Management* study (Figure 1 above).

What is most surprising is how quickly this issue has become a major problem. Concern over unstructured data has more than doubled in the last year alone; in 2011 only 21% of organizations listed it as a top pressure. Part of the reason for this increased attention is likely due to the rapid evolution of Big Data technologies. One of the key components of a Big Data strategy is tackling the wide variety of data formats available today; essentially to address the unstructured data problem that has persisted for years. Furthermore, many recent technological developments in the Big Data space have been focused on unstructured data, such as the Hadoop distributed file system, advances in not-only-structured-query-language (NoSQL) databases, and tools to collect and analyze data from social media. All this hype has done a good job at getting the attention of CIOs, and companies are paying attention to the large blind spot of unstructured data sitting in their datacenters.

However, despite the increased attention, most organizations lack the basic capabilities for collecting, managing, classifying, and analyzing unstructured data. Only 27% of all companies in Aberdeen's January 2013 *Data Management* study reported being able to consistently use and analyze a significant portion of their unstructured data. The Best-in-Class organizations (see Page 3 sidebar), were almost twice as likely to find value in this data, with 48% indicating success at handling unstructured formats. Furthermore, Aberdeen's research study identified a number of technology solutions that these top performing organizations had implemented in order to achieve this success, and how these solutions ultimately impacted their business performance.

Definitions

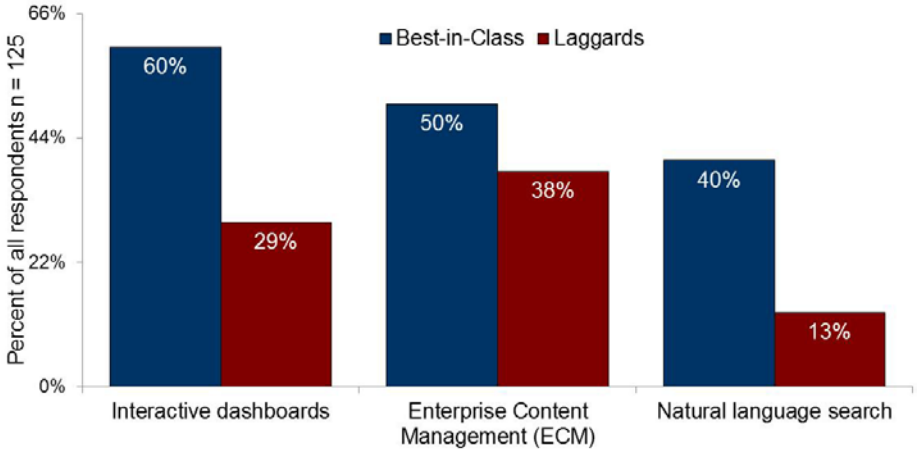
√ **Big Data** refers to the problems of capturing, storing, managing, and analyzing massive amounts of various types of data. Most commonly this refers to terabytes or petabytes of data, stored in multiple formats, from different internal and external sources, with strict demands for speed and complexity of analysis.

How to Find Value in Unstructured Data

The first step in making unstructured data accessible and usable is providing a standard platform for storage and classification. Far too often employees have to navigate server directories, folders, and email archives to get to a particular file or document they need. In order to avoid this problem, Best-in-Class companies are more likely to have implemented **enterprise content management (ECM) systems**. These solutions provide the ability to consistently capture, classify, index, and distribute almost any type of unstructured content. Whether it is images, documents, or emails, this data can be collected from multiple sources, tagged with essential metadata (i.e. time stamps, author, keywords, file type) and stored in designated repositories. It can then be distributed to other enterprise applications or collaboration portals, putting it directly into the hands of the end-user. As Figure 2 shows below, ECM installations are popular systems used by many organizations; even Laggards report 38% adoption. However, the Best-in-

Class take steps to get the most out of their investment either by ensuring their ECM solution has certain critical features, or implementing additional tools to add these capabilities.

Figure 2: Essential Tools for Content Analytics



Source: Aberdeen Group, January 2013

One of the best technologies to combine with an ECM system is a strong enterprise search engine. Similar to popular search websites like Google or Bing, this interface takes full advantage of the classification, metadata tagging, and indexing abilities of ECM to quickly scan unstructured data and pull out files that meet the search criteria. A more advanced form of this, typically a simple keyword search, is **natural language search**. This method allows users to use questions and phrases in their query and, instead of relying on basic metadata, the search engine tries to understand the context and nature of the request. For example, when asking "What store sold the most widgets in November," a keyword search would hone in on words like "store," "widgets," and "November" and return all related files. Natural language search would understand verbs like "sold," the context of "most," and return more relevant, targeted results. These tools are used by two out of every five Best-in-Class companies, while only 13% of Laggards have adopted these solutions.

If ECM provides the foundation for unstructured content, and search allows for easy navigation of this data, then the final step in the analytic process is presenting the information to the end-user. This is the bread-and-butter of BI dashboards, which have been popular for many years because of their ability to display large amounts of transaction data in a graphic interface. In conjunction with ECM and search, an **interactive dashboard** solution can allow an employee to submit queries and explore data from both structured and unstructured formats. After accessing metadata descriptions and crawling through the actual text content of files, the dashboard can display large amounts of information through different visualization techniques. Some examples include tracking the mentions of a particular word or

Maturity Class Definition

The January 2013 report on *Data Management for BI* defined the maturity classes with the following metrics:

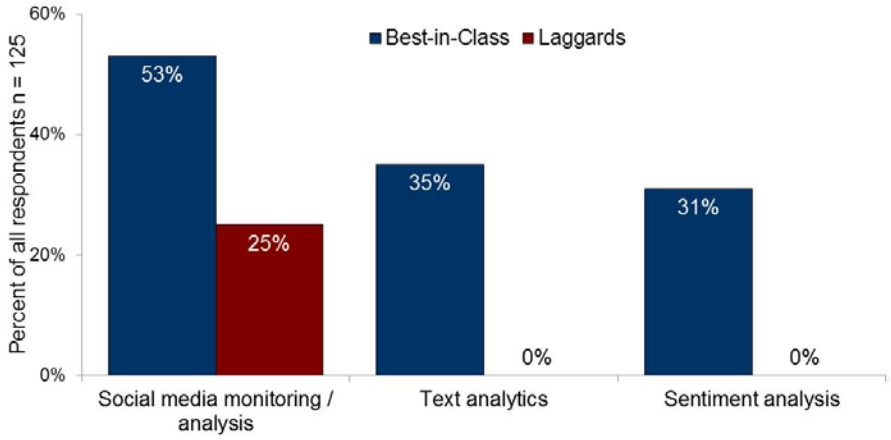
- √ **Time to integrate data sources**. Best-in-Class organizations took 9 days, while Laggards took 137 days — 15 times longer.
- √ **Accessible business data.** The Best-in-Class improved the accessibility of their data by 35% last year. Laggards saw a 10% reduction in the same metric.
- √ **Data accuracy**. Best-in-Class companies reported 93% of their data to be reliable, while Laggards could only report 57%.
- √ **Delivery of information to decision-makers**. The Best-in-Class met the demand for actionable information 91% of the time, while Laggards met their delivery window only less than half the time (47%).

phrase over a certain length of time, or creating a word cluster map to identify new buzzwords or common complaints sent in by customers.

Taking Content Analytics to the Next Level

Once an organization has tackled the basics of managing and analyzing their unstructured content, they can look towards more specific tools and methods of content analysis. Aberdeen's research has identified several areas where Best-in-Class organizations are significantly more advanced than their peers, displayed below in Figure 3. Unsurprisingly, this begins with the ability to monitor and analyze perhaps the most important repository of data on customer behavior developed in the last 10 years: social media.

Figure 3: Advanced Tools for the Best-in-Class



Source: Aberdeen Group, January 2013

It is no secret that social media sites contain valuable information on customer purchase patterns, reviews, brand loyalty, and buying habits. In fact, 77% of organizations in Aberdeen's January 2013 study on *Data Management* had either invested in tools to access social media data, were evaluating these solutions, or planned on investing in the next two years. However, collecting this data and successfully analyzing it are two very different tasks. The sheer amount of social media data and how quickly it is generated presents a significant strain on an IT infrastructure. Then, once it is all gathered, it takes powerful tools to sift through vast quantities of useless information and isolate the few nuggets of valuable data. After all, while getting insight from data is great, getting it fast means the organization can act on it quickly and drive even more business value. More than half (53%) of the Best-in-Class have these **social media monitoring** solutions in place, which is twice the amount of Laggards.

Likewise, the Best-in-Class are early adopters of two other, similar technologies. **Text analytics** takes the concept of natural language searching to the next level. Not only do these tools endeavor to understand context and meaning in text, but they use advanced models to develop and

Data Sources to Target

The following data sources were listed as Important or Very Important to Big Data analytic initiatives:

- √ **94%** needed transactional application data
- √ **90%** needed other internal sources of structured data, like master records
- √ **76%** needed external unstructured data from customers or business partners
- √ **74%** needed internal sources of unstructured data (i.e. emails, documents and reports)
- √ **67%** needed internet-generated data like clickstream and web traffic data
- √ **53%** needed rich media files (i.e. audio, video, images)

"The biggest challenge we have is finding a resource that can help manage unstructured data."

~ Marketing Manager, Small Transportation / Logistics Company, North America

refine their pattern recognition. This allows for the fast identification of commonalities between different files, content outliers that are unique and interesting, summarization of larger files, or trends for predictive analytics. One specific brand of text analytics revolves around understanding emotion and intensity, which is often applied to social media data or customer complaints. **Sentiment analysis** identifies particular words, phrases, and context that indicate positive or negative emotions. By linking these emotional words with mentions of products, brands, or problems, this analysis can identify which products are well received or which issues irritate a company's customer base. Advanced tools even try to learn from history by correlating older material with a given message in order to "read between the lines." They try to identify familiar patterns, brief references to other events, and implied sentiment, even if everything isn't expressly spelled out. Given how quickly bad reviews or feedback can go viral, these tools can often provide companies with just enough advance warning to address problems before they get out of hand. Roughly one third of the Best-in-Class have deployed both text and sentiment analysis, while none of the Laggards in this study had a mature-enough content analysis infrastructure able to support these tools.

Content Analytics Drives Better Business Performance

Implementing and managing a content analytics platform can be a considerable investment for organizations, but Aberdeen's research shows that there are significant benefits from doing so. A strict comparison of 30 organizations that reported the ability to analyze most of their unstructured data, and 52 organizations that could not analyze any, presents a clear performance gap.

Table 1: Benefits of Analyzing Unstructured Data

Performance Metrics	Analyzing Unstructured Data	Not Analyzing Unstructured Data
Amount of all business data able to be analyzed	27%	15%
Ability to deliver analysis in a timely fashion	76% of the time	66% of the time
Hours per week spent by employees looking for data / answers	1.9 hours	5.9 hours

Source: Aberdeen Group, January 2013

Organizations with content analytics were much more successful at addressing the datacenter blind spot, reporting the ability to analyze almost twice the amount of data than other organizations. Despite this, there is still room to improve for the best companies, but they have identified the most valuable sources of unstructured data and taken steps to start analyzing them. They were also much more efficient at delivering insight quickly and

Rapid Growth of Data

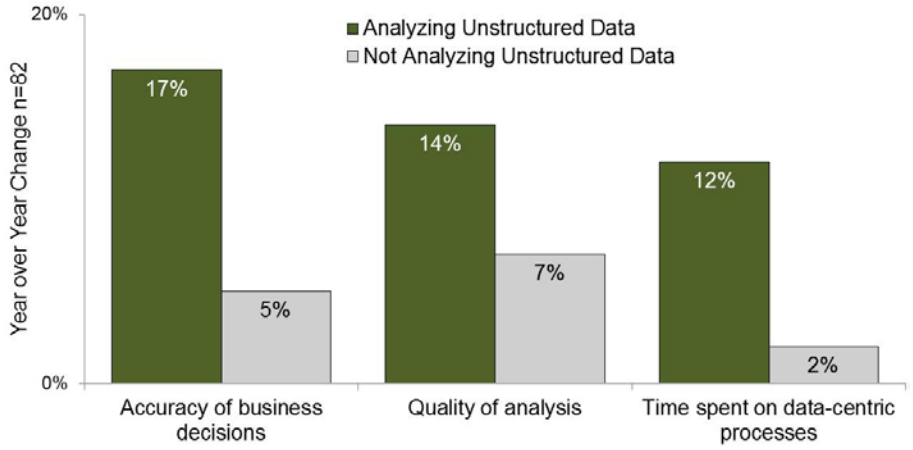
Business data is increasing at a rapid pace, and Aberdeen has tracked the trend over the past several years:

- √ **29% growth** year over year was reported in December 2009
- √ **30% growth** was reported in February 2011
- √ **38% growth** was reported in January 2012
- √ **55% growth** was reported in the most recent survey in January 2013

managing employee time. When complicated requests or queries were submitted, they were able to search through unstructured content and more consistently deliver answers within an appropriate time frame — three quarters of the time, compared to only two thirds for companies without content analytics. Finally, employees at organizations without content analytics spent more than three times longer per week searching for information they needed to do their jobs. This inefficiency amounts to each knowledge worker wasting 200 more hours per year, or five full work weeks.

For companies with content analytics, the streamlined searching and fast response to queries had a trickle-down effect on several universal business metrics. In the last twelve months, these companies reported a 17% improvement in the accuracy of their business decisions, and a similar improvement in the quality and relevance of their overall analytics program. Furthermore, their data-centric business processes — like handling customer complaints, processing invoices, managing inventory, or closing the books — became 12% more efficient. Companies without content analytics reported 2- to 6-times less improvement (Figure 4).

Figure 4: Content Analytics Streamlines Business Processes

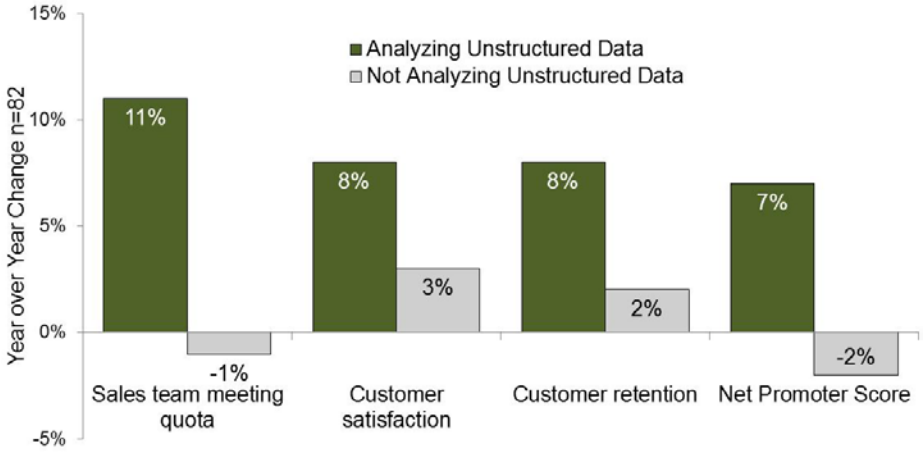


Source: Aberdeen Group, January 2013

Ultimately, the benefit of these efficiency improvements and analytic quality is transferred to a company's customers. Content analytics helps companies understand what customers want, what concerns them, and what they are most likely to buy. When an employee in sales, product development, marketing, or customer service is armed with this information, they can tailor their deliverables to specific customer needs and take proactive steps to address potential problems before they become deal-breakers. As such, companies using content analytics reported an 11% improvement year-over-year in the number of sales reps that met their annual quota (Figure 5). Furthermore, customers that did business with these companies were more satisfied with their products or services, and less likely to take their money

to a competitor. This in turn improved the company's overall net promoter score, boosting the word-of-mouth, grass roots recommendations that can drive new business.

Figure 5: Content Analytics Improves Customer Service



Source: Aberdeen Group, January 2013

Given that companies without content analytics reported only incremental improvements or even performance decreases in these areas, this cements content analytics as a vital contributor to improving positive brand recognition and a trustworthy reputation.

Key Takeaways

The amount of data that businesses access and store is growing at an incredible rate, but more than half this data is unstructured, often lying unused or underused within the datacenter. In fact, Aberdeen's research shows that the average organization can only analyze 20% of their data, which is largely due to the difficulty of managing and analyzing the different formats of unstructured content. Yet the benefits of a successful content analytics program are obvious, with companies using these solutions reporting a workforce that is 3-times more efficient at finding data, a sales team increasing annual quota attainment by 11%, and customer satisfaction and retention improving by 8% year-over-year. For organizations looking to start a content analytics program, or improve their existing capabilities, Aberdeen recommends investigating the following capabilities:

- **Build a solid foundation**. Managing unstructured content isn't easy, but solutions like enterprise content management are the go-to tools of more than half of the Best-in-Class. Make sure to address the full lifecycle of this information, including data capture, classification and indexing, and integration with enterprise applications and BI tools.

Definitions:

√ **Net promoter score (NPS)** is calculated by taking all satisfied customers willing to refer a company and subtracting customers that are unsatisfied and unwilling to give a referral. For instance, if 75% of an organization's customers would give a referral, and 25% would not, their NPS would be 50%.

- **Make your content discoverable**. One of the biggest challenges with leveraging unstructured content is how to find the one file you need when you need it. Tagging unstructured data with metadata keywords and descriptions is a great first step. Then implement enterprise search tools to quickly scan through your data repositories. Natural language search tools in particular show a strong correlation to Best-in-Class performance, with 40% of the top performers using these solutions, compared to only 13% of Laggards.
- **Explore advanced analytic options.** First, identify what types of unstructured content hold the most value for your organization. Are you interested in mining social media information to discover trends in customer behavior? Do you rely on large amounts of text documents and forms in your critical business processes? Would you like to discover which complaints upset your customers the most? These different types of data and business goals each require specialized tools, but they can all contribute to better business performance. The Best-in-Class were more than twice as likely as Laggards to adopt social media monitoring, text analytics, and sentiment analysis tools.
- **Put your data in the hands of business users**. Whether it is your sales force, marketing team, business directors, or C-level decision makers, make sure the insight gathered from unstructured data is presented to the right people. Companies using content analytics reported increased sales performance and customer service, but that only happens if the right people can access the information. Interactive dashboards and other data visualization techniques are popular methods for addressing this problem, with 60% of the Best-in-Class using them to display both structured and unstructured content.

Telephone: 617 854 5200
Fax: 617 723 7897

For more information on this or other research topics, please visit www.aberdeen.com.

Related Research	
Big Data Trends in 2013; February 2013	*High Performance Organizations Empower Employees with Real-Time Mobile Analytics*; April 2012
Data Management for BI: Getting Accurate Decisions from Big Data; January 2013	*Mobile BI 2012: Accelerating Business on the Move*; March 2012
The Big Data Imperative: Why Information Governance Must be Addressed Now; December 2012	*The Little Elephant in the Big Data World: Hadoop 1.0 Goes Live*; March 2012
Big Data for Small Budgets; December 2012	*The Tortoise and the Hare: A Tale of Document Management in Insurance and Banking*; March 2012
Go Big or Go Home? Maximizing the Value of Analytics and Big Data; September 2012	*Divide & Conquer: Using Predictive Analytics to Segment, Target and Optimize Marketing*; February 2012
The State of Big Data: Video Benchmark; July 2012	*Handling Paper in a Digital World: The Impact of Document Management*; February 2012
Agile or Fragile? Your Analytics, Your Choice; July 2012	*Operational Intelligence - Part 1: Driving Performance with Tactical Visibility*; February 2012
Beyond Agile Analytics: Is Agile Data Integration Next; June 2012	*In-memory Computing: Lifting the Burden of Big Data*; January 2012
Managing the TCO of BI: The Path to ROI is Paved with Adoption; May 2012	
Enabling Access to Big Data with Data Integration; April 2012	

Author: Nathaniel Rowe, Research Analyst, Enterprise Data Management (nathaniel.rowe@aberdeen.com)