



Data Smog - or the Art of Organizing & Analyzing Information

"Data smog" a common way to describe the massive amount of data companies and organization handle, review and store. Institutions, companies, organization and individuals alike are dealing with vast amount of information, raw data, e-mails, documents and files on a daily basis. Most of this data is stored, because it might be of interest later, because it is required by law or because it is supposed to be stored and preserved.

Each piece of data and file that is being saved can lead to overload resulting in vast amount of data that cannot and will not be used or read. Combining various kinds of data formats, as might be the case in databases, can be overwhelming. But a right combination between what data is stored, how it is stored and how it is analyzed can result in a ton if insight.

This paper analyzes that too much data can lead users struggling to pull insight from the data because most of their time is devoted to just putting all their different data types together. In addition to that, this research will demonstrate that vast amount of data cannot be analyzed properly, therefore leading to false information, misinformation as well as lack of data retention.

There were times, not too long ago, when digital information was rare and precious: a formal photo or a faint recording of someone's voice. Today, digital data is everywhere, hard drives and media storage are getting bigger and cheaper. All you can do today is, how to avoid leaving a recording behind as you go about your daily business. The average american is caught at least a dozen times a day on surveillance cameras in the airport, banks, intersections or on the dashboard of police cruisers. Even businesses log keystrokes of their computers and customer cervice centers and 911 record all phone calls. According to DigiMine CEO Usama Fayyad, a computer scientist, calculates that the data storage curve is now rocketing upward at a rate of 800 percent per year. "It makes Moore's law look like a flat line," he says. "Companies are collecting so much data they're overwhelmed."

In the data-rich world, the data needs to be organized for easy access, due to the growing amount of data that is collected and produced that is needed to get a job done. Managing which includes indexing and linking documents and other information sources is the first and very important step. But organization of data for easy searches and access is another required and very important step. Making sure that information is formatted, accessible and archived is a relative new technology field called knowledge management.

One main problem, which exists at every level from small business to great archival institutions, is the data or knowledge management. Data is supposed to be arranged and managed in some sensible fashion, which will allow for easy management,



retrieval and searches. Although there are many knowledge and data management software solution on the market, not all provide an easy and efficient management and search interface solution. The ones that do are demanding on technical knowledge for installation, maintenance, budget and enormous amount of manpower to organize, link and maintain the data.

After 9/11, the US government has passed many new laws that allow for collecting, saving and analyzing large amounts of data to prevent similar incidents from happening in the future. As an example, in 2004, federal agents have collected the names of 260,000 people staying in Las Vegas hotels during New Year. Not only is this type of data collection or "vacuuming" very expensive, but it certainly misses the objective of such a project. The main issue with such a very broad data collection is deciding which data needs to be analyzed and interpreted.

Not only after 9/11 but even before the incident, the government has collected so much data, that intelligence organizations cannot possibly analyze it. The failure to prevent 9/11 was a failure of analysis and not a collection and technical failure. In order to analyze data properly, one has to find, what to look for. Unfortunately this can be an impossible task and substantial amount of good data can go unread and unanalyzed.

With more collected data, the amount of "needles" is the same, but you have a larger "haystack" to find them in. History has proven that collecting raw data can become very counterproductive. East German government has collected in 1980s and the years before an enormous amount of data on about 4 millions East Germans, which was about a quarter of their entire population. But with so much technology and manpower in data collection, they neglected data interpretation and did not foresee the peaceful overthrow of the communist government.

Collecting enormous amounts of data and relying on technology to sift through it and sound an alarm is not a feasible solution, which often results in many false positives. Just alone in the past years, these false alarms did not only cost millions of dollars, unduly scared people, trample on individual rights but also inure people to the real threats. But finding meaning among large amounts of irrelevant data, requires people to analyze the data, identify the connections, and access various information from different branches. In addition to that, a well functional data or knowledge management system needs to be in place, where the data has already been stored, organized and is being properly maintained.

People and organizations cannot keep up with the large amount of data and information that is being collected. Too much data for research and work is being produced on a daily basis by technological innovations. Managers and decision makers are having a great difficulty in navigating through a constantly increasing and exploding data volume. Despite many technology uses as well as useful knowledge management solution, this task is not getting easier. The situation is only getting worse.



According to a recent survey of managers from large UK and US businesses with annual revenues of more than \$500 million, 47 percent of Information Technology (IT) managers were affected negatively by the information management minefield and large amounts of data. Half of all surveyed managers blamed their organization for not doing a good job to disseminate relevant information through the departments. "The findings show that companies are failing to get the right information to their employees," said Royce Bell, chief executive of Accenture Information Management Services (AIMS).

This growing issue is becoming a burden not only for large corporations and institution, but affects small business as well as the individual. Dealing with large amounts of data on a daily basis does not only cause a negative impact on the work efficiency, it also has a very negative impact on retention of the data. This has the same effect as having not enough data available.

Although technology allows for easy data collection, more irrelevant data is being collected, disseminated and stored, than is really needed. The current and technological analysis process produces many false positives causing more damage than valuable intelligence and information. There's too much reliance on computers and automation and less human analysis. There's plenty of raw material, but not enough thoughtfulness. This information data overflow is becoming a burden on the entire society.

Our society and each company needs to decide on what information and which data source provides the factual information needed. Not all data is relevant and each source provides a different quality. There is no perfect solution for this growing problem, but a start is to consolidate and streamline the data and information sources. By doing so, companies will enable their workers to focus on their work, rather than researching data to do their work. It will also improve delivery, integration, archiving and retention of valuable and quality data.

By Dasha Deckwerth

Stealth - International Intelligence Security Service Inc.

www.stealth-iss.com

About Stealth - ISS® Inc.

Stealth - ISS®, headquartered in Tampa, FL., is a privately owned Information Technology security consulting company with main focus on regulatory compliance, security integration, security consulting and managed security services for both government and commercial customers.



Founded in 2002, the company has earned an outstanding reputation for professional security services including vulnerability assessments and regulatory compliance such as Sarbanes-Oxley, HIPAA, NIST and ISO standards. Stealth – ISS® also recently expanded its data center services providing faster and secure hosting and colocation solutions for customers worldwide. Stealth - ISS® has partnerships with several leading manufacturers of various security applications and products and was awarded the NATO BOA agreement in 2003.

About the Author

Dasha Deckwerth is the CEO at Stealth - ISS, with extensive experience in international business and computer security. Prior to her position as CEO, Dasha had gained extensive international business experience in various European, Asian and Central American countries and later became the VP of Marketing and Business Development at Stealth - ISS® in Berlin, Germany. She also worked on several projects as security and regulatory compliance consultant in the commercial sector as well as for various NATO countries and government agencies. Dasha's current focus includes managed data center services, knowledge management, regulatory compliance applications and services and security implementations and consulting. Mrs. Deckwerth holds a B.A. in International Relations and Foreign Affairs from Eckerd College, is currently pursuing an MBA in IT Management from Touro University and speaks six languages.