International Journal for Multidisciplinary Research (IJFMR)
E-ISSN: 2582-2160 ● Website: www.ijfmr.com ● Email: editor@ijfmr.com

A Study to Describe the Needs of Big Data in Media

Sudhakara.K[1], Sameena Kausar Farheen[2], Kalpana M N[3], Shradhanjali Mishra[4], Dr.Lokesh.A[5]

[1,2,3,4]Assistant Professor,Dept.of AI&ML,S.E.A College of Engineering and Technology
[5]Professor, Dept.of AI&ML,S.E.A College of Engineering and Technology,Bangalore.

Abstract:
The big thing for media companies are "Big Data" strategies. In this paper, we describe the need for big data in the media industry. The reams of data from media require Big Data which deals with variety of strategies and tactics that involve massive data sets, and technologies. The Big Data trend has impacted all industries, including the media industry, as new technologies are being developed to automate and simplify the process of data analysis, and as throngs of data analysts are being trained and hired to meet the demand for the analysis of these data. Big Data strategies can include audience analytics to enable a better understanding and targeting of customers; tools to understand public and private databases for journalistic storytelling; tools to manage and search the exploding amount of video, social media and other content; tools to target advertising and ad campaigns; tools to automate the production of text and video stories, tools to identify waste and enable efficiencies; and much more for newspapers, television, magazines and Internet-only publishers.

Keywords: media practises; audience anlytics; data journalism; advertising and ad compaigns; big data strategy development and media organisations;

1. INTRODUCTION

In 2014, media companies around the world are morphing the Big Data hype of 2013 into strategies and actions. The opportunities for employing Big Data strategies are many: to better understand cross-platform audiences, create powerful data journalism stories, streamline business processes and identify new products and services to offer customers.

The media industry can think of Big Data as the Four Vs, including volume of data; velocity of data, meaning it needs to be analyzed quickly especially news; in a variety of structured and increasingly unstructured data formats; which all have potential value in terms of high quality journalism and business insights and revenue. Little data has the capacity for storage that is measured in gigabytes or smaller and can be contained on a personal computer. Big Data is too big to fit on a personal computer, and can be stored on the cloud or other big storing system, as most Big Data would be measured in terabytes, petabytes, zetabytes and beyond. To illustrate the point about the differences in storage requirements for big and little data, a seven minute high-definition video requires one gigabyte of storage. However, one petabyte, which equals one million gigabytes, could store 13.3 years of

continuously running high-definition videos. Google and its video website, YouTube, process more than 24 petabytes of Big Data per day.

2. EASE OF USE

A. *Structured vs. unstructured data*

In particular, unstructured data, including video, audio, email, research, social media feeds and more, which represent 80 percent of all Big Data, will be a challenge from which to extract actionable business knowledge. Structured data includes transactions, log data and spreadsheets. External data for media companies typically includes social media, audio, photos and video, all unstructured data, while internal data includes log data, transactions and emails.

B. *What is the Big Data payoff?*

According to Bain & Company's 2013 report, *Big Data:* The Organizational Challenge [1], a business that leads the way in using data strategically are:

- 5x as likely to make decisions faster than market peers.
- 3x as likely to execute decisions as intended.
- 2x as likely to be in the top quarter of financial performances within their industries.
- 2x as likely to use data very frequently when making decisions.

The payoff for media companies specifically are many, including engaging the audience more deeply with more targeted news and advertising, more relevant and socially engaging content, more discoverable and compelling videos and photos, and most of all, the ability to compete with other sophisticated online media companies that are frequently ahead of traditional media companies with advanced technology strategies.

C. *Media industry's impact on Big Data*

Media companies collect reams of data every minute from every area of their organizations: advertising/sales, readership/usership/membership, content, accounting and more. Media companies also produce stockpiles of data in the form of videos, photos, text and graphics. These data represent about 70 percent of the Internet's data storing and sharing, which is growing exponentially, according to Mary Meeker's annual Internet Trends report, published in May 2014. This digital "universe" grew 50 percent from 2012 to 2013, and is expected to skyrocket 40 percent year-over-year, according to IDC Digital Universe. Digital content universe generated by consumers Video and photo generation, consumption and sharing and social media usage made up the bulk of online content in 2013 as shown in Figure 1.

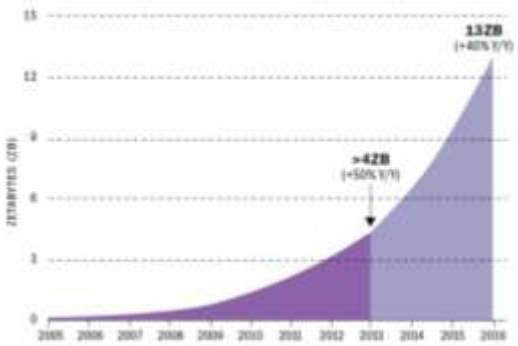

Figure 1. Digital content universe generated by consumers

The digital universe, up until about 2005, was measured in petabytes. Since then, the increasingly rapid

global creation, consumption and sharing of these data have propelled the universe into the zetabyte stratosphere. One million gigabytes equals one petabyte. One million petabytes equals one zetabyte. Just a sampling of data points shows the prolific contribution of Big Data from media companies to the Internet universe.

Figure 2. Big Data: What is it, what kind of data and how much.

Many of these data points are growing exponentially as well, such as the daily number of tweets, the number of videos uploaded, consumed and shared, the number of Facebook posts, and the number of WeChat sessions and many more as shown in Figure 2. While collecting and producing data are the first steps in the development of a Big Data practice, the acts of "analyzing" and making the data "actionable" are the new means for media companies.

D. *How did we get here?*

The Big Data revolution [2] did not happen by accident. Prices for digital media storage and bandwidth, explosion of digital devices including smart phones and tablets, and the exponential growth of audience accessed digital media have whipped up the perfect storm to create this surge in Big Data strategies and implementations.

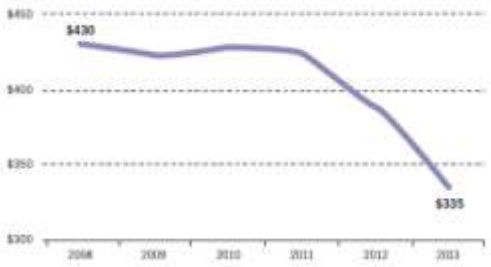

Figure 3. Global smartphone costs

Between 2008 and 2013, average global Smartphone costs have dropped from $430 to $335, a 22.1 percent decrease as shown in Figure 3, according to Deloitte, as reported in Meeker's annual report. Some Smartphone manufacturers are producing sub-$100 smart phones to enable affordability in the developing world, which is helping to drive average global prices downward.

Figure 4. Virtuous cycle of Big Data content

These steep decreases have given rise to the virtuous cycle of Big Data content as shown in Figure 4 more affordable connectivity for consumers, a proliferation of devices enabling cheap access to multimedia news and information, and the emergence of data-driven technologies. This perfect storm makes possible the creation of infinite new business and consumer-facing tools that enable automation, insights, information and efficiencies.

3. BIG DATA FOR MEDIA

The Big Data conferences produced some invaluable lessons for media companies with aspirations in engaging in Big Data strategies. Here's what the speakers at the Big Data for Media conference on 8 May 2014 had to say:

Huffington Post CEO Jimmy Maymann: "It's all about data." HuffPo is using Big Data and small data to improve the user experience. They use Big Data to improve UX from real-time dashboards, social trends, recommendations, moderation and personalisation. The Figure 5 gives the Huffington Post statistics in 2014. They use small data to improve UX from reporting, headline optimisation, SEO, content efficiency and consumer research.

Figure 5. The Huffington Post statistics

Popular news and entertainment social sharing site BuzzFeed earns its large traffic figures by publishing viral content, said Ky Harlin, the site's chief data scientist. Harlin's team identifies trending stories and their unique characteristics in order to duplicate success in the future. The Figure 6 gives its statistics in 2014.

Figure 6. Buzzfeed statistics

Dunnhumby/Tesco's global head of data Matthew Keylock shows how dunnhumby's 10+ years of collecting Big Data in order to improve customer loyalty and sales applies directly to the nascent Big Data strategies in the media industry. Tesco's clubcard scheme enables a customer view of retail and other data, which inspires greater customer understanding, which in turn drives better business decisions, which grows loyalty and brand value [3]. The Figure 7 gives the Tesco big data strategy and organisation centres around customer and the percentage of Tesco customers and sales is as shown in Figure 8 Media companies can learn from the systematic approach from dunnhumby, the force behind the Clubcard.

Figure 7. Tesco Bigdata strategy and organisation centres around customer

Figure 8. Percentage of Tesco customers and sales

Circa technologists use Big Data to serve bite-sized chunks of only new news that readers are following, instead of expecting story followers to re-read content they have already consumed. Each atomized chunk is called a card. "We strive to present each article as atomic content. It's also context. The Figure 9 gives the Circa statistics in 2014. Here is how we construct mobile news on Circa. We don't use auto summarization; we have actual journalists." Each story is grouped as a collection of cards, for example, a headline, a title image, data points, an image with a caption, maps, text and videos. And they are all interconnected. Each have their own links, he said.

Figure 9. Statistics of Circa

Tom Betts, head of customer analytics and research at the Financial Times, said they build customer "signatures" of each customer's digital consumption and use the information to understand customer content preferences, increase the relevance of their communications, personalise their offerings, and deploy intelligence to customer touchpoints, such as customer service, website, mobile apps and third parties. In 2014, the Financial times and ft.com statistics was as shown in Figure 10.

Figure 10. Financial Times and ft.com

CNN International vice president and general manager for digital, Peter Bale, spoke about how Big Data is used as an early warning system for breaking news, how CNN listens to its vast audience members using technology that summarises how viewers are consuming news in real-time, and how major data sets are distilled and presented as data journalism stories [4]. The Figure 11 shows the CNN's statistics.

Figure 11. CNN (Cable News Network)

BBC's statistics is as shown in Figure 12. BBC head of visual journalism Amanda Farnsworth underscored the importance of data journalism to provide insightful, personally meaningful and shareable visual explanations on the BBC's biggest and most significant stories.

Figure 12. BBC statistics

Wilfried Runde, head of innovation projects in new media for Deutsche Welle, emphasised why all newsrooms should work with data: status, skills and scenarios [6]. The ability to create powerful journalism with large datasets accentuates the real value of journalism to the public and underscores the importance of news organisations that take on such insightful projects.

Bertrand Pecquerie, CEO of the Global Editors Network, talked about a variety of finalists for GEN's annual Data Journalism [5] Awards, in the following categories: best story on a single topic, best data-driven investigation, best data-driven visualisation, best application or website, best team or newsroom portfolio, best entry from a small newsroom, jurors' choice and public's choice.

Channel 4 head of data planning and analytics, Sanjeevan Bala, is working to innovate the television business model using data by connecting to their millions of registered viewers, which allows them to segment viewers into groups, create personalised emails, suggest tailored content recommendations and serve relevant advertising.

Archant regional newspapers' digital director, Paul Hood, called the prevention of data leakage priority 1, as targeting advertising through third-party ad networks becomes more pervasive. Data collected from users on media companies sites should belong to those media companies, not third parties like Google or Apple, Hood said. The Figure 13 gives us the statistics of Archant in 2014. There are techniques to stop the data leakage, and strategies to leverage the data for media companies who own it.

Figure 13. Archant statistics

Magazine data guru Pegg Nadler identified successes and failures in Big Data strategies among U.S.-based magazine companies representing 250 titles. Ultimately, these magazine companies are in the early stages of their strategies, with the ultimate goal to make their sensor data, images, databases, location-based data, email, HTML, social and clickstream data more actionable in the future.

Head of trends for GlobalWebIndex Jason Mander underscored the importance of social media trends around the world, and how knowledge about these trends can help inform Big Data strategies for media companies.

Sacramento Bee director of research Darrell Kunken explored how the Bee and its parent company, McClatchy, have developed their Big Data strategy to include technology systems, tracking issues, business models and cultural issues, and how they are working with consultants and academics to refine their plans for the future. The Figure 14 gives the statistics of Sacramento Bee in the year 2014.

Figure 14. Sacramento Bee statistics

4. INSIGHTS OF BIG DATA

A variety of studies about Big Data have been produced in 2014 by some of the best research companies in the world. The top barriers to successful use of Big data is as shown in Figure 15. As the Big Data industry develops, key building blocks are being explored, such as how companies are employing Big Data strategies, how consumers perceive and respond to rising privacy issues as Big Data implementations grow, and which stages companies are in as they plan and execute their Big Data strategy trajectories.

Figure 15. Top barriers to successful use of Big data

According to some of the most forward-thinking experts specialising in Big Data structure, such as Gartner and IBM, a data governance plan must be put into place in order for Big Data strategies to be successful. Data governance plans comprise the technical and security policies and procedures that become the foundation of the Big Data strategy at any company.

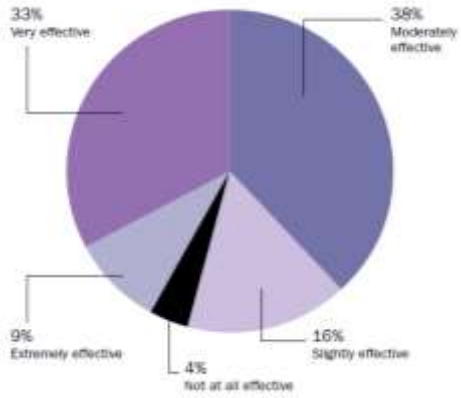

Figure 16. Critical data for decision-making

InformationWeek in 2014 reported on its Big Data survey, done in 2013, which reflected trends in the development of data governance plans at companies around the world. Budget constraints are the top barrier to the successful use of Big Data (38 percent), followed by lack of business interest (13 percent), and more important priorities for their IT departments, tied with lack of Big Data management tools (11 percent), according to the survey. The critical data for decision-making is as shown in Figure 16. More than 40 percent of the respondents called their data analysis practices either "limited" or "abacus-like." One of the key tools for assessing and leveraging Big Data is a data dashboard that could display data trends from a variety of company sectors, such as CRM, finance, online audience, sales and social media sentiment. The most popular application of a data dashboard is an audience usage dashboard. Ninety percent of executive respondents said they plan to use dashboards for key metrics by more than 20 percent of the employees, according to the InformationWeek survey. Particularly useful additions to the dashboard are sales team metrics, including outbound calls, sales rep site visits and proposals generated, activity levels from business'[7] CRM systems, and online usage patterns such as social media, paid content subscription trends, story engagement, content popularity trends, e-commerce trends and more.
The InformationWeek report recommended that the next step after the collection of meaningful Big Data is to identify trends in the raw data in order to make decisions across the business [8]. According to the survey, 38 percent of the respondents say they have been moderately effective in applying findings of Big Data collection to making decisions, while 33 percent said they were very effective and 16 percent said they were slightly effective.

5. CONCLUSION

For newspapers, television, magazines and Internet-only publishers, Big Data strategies can include audience analytics to enable a better understanding and targeting of customers; tools to understand public and private databases for journalistic storytelling; tools to manage and search the exploding amount of video, social media and other content; tools to target advertising and ad campaigns; tools to automate the production of text and video stories, tools to identify waste and enable efficiencies; and much more.

6. REFERENCES

1. Wendy N. Wyatt (ed.) *"The Ethics of Journalism: Individual, Institutional and Cultural Influences", 2014*.

2. Raymond Kuhn and Rasmus Kleis Nielsen (eds) *"Political Journalism in Transition: Western Europe in a Comparative Perspective"*, 2014.
3. Nigel Bowles, James T. Hamilton, David A. L. Levy (eds) *"Transparency in Politics and the Media: Accountability and Open Government"*, 2014.
4. James Painter *"Poles Apart: The International Reporting of Climate Scepticism"*, 2011.
5. Lara Fielden *"Regulating for Trust in Journalism: Standards Regulation in the Age of Blended Media"*, 2011.
6. David A. L. Levy and Robert G. Picard (eds) *"Is there a Better Structure for News Providers? The Potential in Charitable and Trust Ownership"*, September 2011.
7. David A. L. Levy and Rasmus Kleis Nielsen (eds) *"The Changing Business of Journalism and its Implications for Democracy"* .
8. Tim Gardam and David A. L. Levy (eds) *"The Price of Plurality: Choice, Diversity, and Broadcasting Institutions in the Digital Age"*, 2008.