



REUTERS  
INSTITUTE for the  
STUDY of  
JOURNALISM

# REPORT

## Big Data for Media

Martha L. Stone

November 2014



## Table of Contents

Introduction.....	1
Chapter 1 Big Data: The Next Opportunity.....	1
Chapter 2 Case Studies from the Big Data Conference 2014.....	5
The Huffington Post.....	5
Buzz Feed.....	7
<i>Financial Times</i> and FT.com.....	7
dunnhumby.....	9
<i>Sacramento Bee</i> .....	10
Archant.....	12
BBC.....	13
CNN.....	13
Chapter 3 The Issues.....	14
Chapter 4 Insights on Big Data from Beyond the Media Sector.....	17
About the Author.....	32

## Introduction

“Big Data” strategies are the Next Big Thing for media companies.

Big Data is an umbrella term for a variety of strategies and tactics that involve massive data sets, and technologies that make sense out of these mindboggling reams of data. 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.

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.

In response to this rapidly accelerating trend, Martha Stone and the Reuters Institute embarked on a two-year study to identify the opportunities for Big Data strategies for media companies, and to chronicle the best practices in Big Data for media companies.

Dozens of interviews with technologists, data scientists, publishers, broadcasters, academics and other Big Data stakeholders were conducted between Autumn 2012 and Spring 2014. The most important element of the study was to produce two ‘Big Data for Media’ conferences in London: one in June 2013 and another in May 2014. Dozens of experts, publishers and academics spoke at the one-day conferences about the state of the art Big Data for Media strategies, chronicled in this report.

In just two years, the Big Data for Media landscape has changed drastically. In 2012, when this research was conceived, the buzzword “Big Data” was recognised as a trend in the media industry, but few media executives knew how to define the catchphrase, let alone knew what Big Data could mean for their companies. In 2014, as media companies around the world have quickly ramped up their understanding and deployed their Big Data strategies, many media companies are well on their way to taking advantage of technologies to better understand the unrelenting avalanche of data surging through their companies every day. While media industries are learning a lot from each other about Big Data, they are also increasingly drawing insights from other sectors beyond the media. Chapter 2 on p. 5 draws on some of the most useful insights from beyond the media sector which can help inform company strategies.

---

## Chapter 1 Big Data: The Next Opportunity

---

**In 2014, media companies around the world are morphing the Big Data hype of 2013 into strategies and actions. The opportunity 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.**

---

### Big Data for media

So, what is Big Data, as it relates to media companies?

The media industry can think of Big Data as the Four Vs, including **volume** of data; **velocity** of data, meaning it needs to be analysed 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.

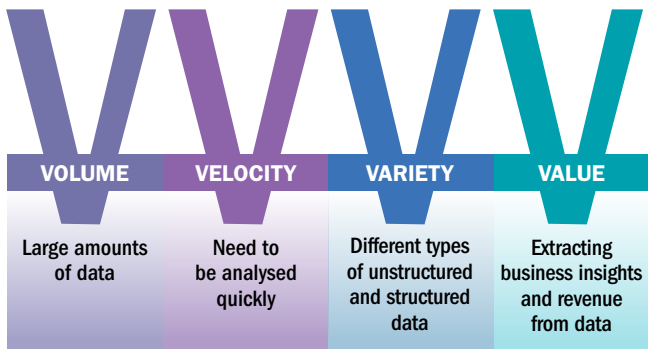
There are a variety of definitions for Big Data, including being a catch-all for the opportunities presented by the exponential growth of data in the media sector, including structured, internal data available through media companies’ own databases, as well as unstructured data on a multitude of digital channels, including video, audio, photos and reams of social media text. “Little”

data and Big Data have distinctly different characteristics. 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, zettabytes 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, processes more than 24 petabytes of Big Data per day.

## Big Data: The four Vs

Volume, Velocity, Variety and Value



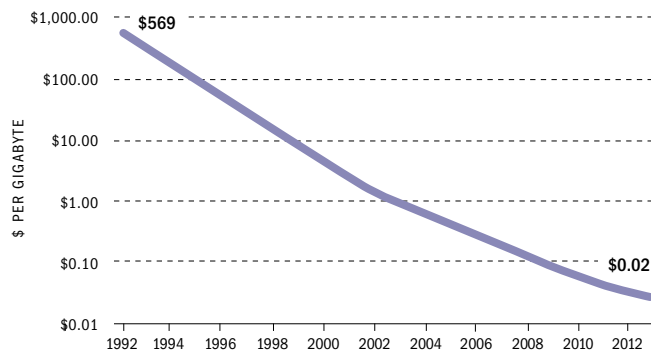
© World Newsmedia Network 2014

## Structured vs. unstructured data

The amount of work required for making sense of data is growing at warp speed. 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.

## Global data storage costs, 1992-2013

In US\$ per gigabyte of storage



Source: Deloitte, May 2014, as reported by Internet Trends 2014 by Mary Meeker for Kleiner Perkins Caufield & Byers

© World Newsmedia Network 2014

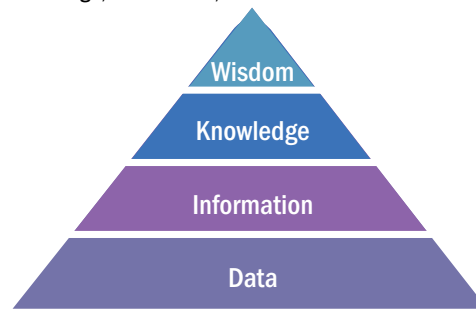
Big Data is a huge opportunity for media companies, but media companies in general are still far away from extracting maximum benefit from a Big Data strategy. If a Big Data strategy could be broken down like a Maslow's Hierarchy of Needs pyramid, media companies' level would fall somewhere in the middle, between the information and knowledge stages.

In order to reach the wisdom stage, media companies need to embrace the opportunities from a Big Data strategy, invest in technologies and training in order to leverage the investment, and finally, use this newfound wisdom to improve processes such as customer service and products.

<sup>1</sup> Bain & Company, Travis Pearson and Rasmus Wegener, [http://www.bain.com/Images/BAIN\\_BRIEF\\_Big\\_Data\\_The\\_organizational\\_challenge.pdf](http://www.bain.com/Images/BAIN_BRIEF_Big_Data_The_organizational_challenge.pdf), 2013

## 'Maslow's' hierarchy of Big Data stages

Wisdom, Knowledge, Information, Data



Sources: WNMN and Freeworld Media  
© World Newsmedia Network 2014

Although massive amounts of data have been stored and accessed for decades, the Big Data buzz has been rapidly growing in business lexicon for only a few years. Some of the drivers of the buzz are the expanding mass of data and the general business awareness that the Big Data disruption must be addressed.

## What is the Big Data payoff?

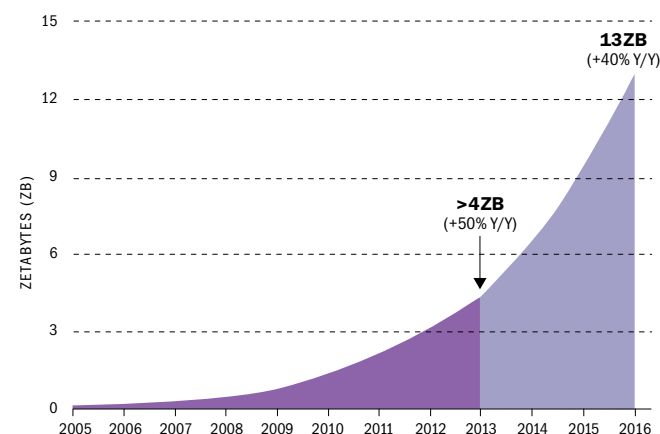
According to Bain & Company's 2013 report, *Big Data: The Organizational Challenge*<sup>1</sup>, businesses that lead the way in using data strategically are:

- **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.

## 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



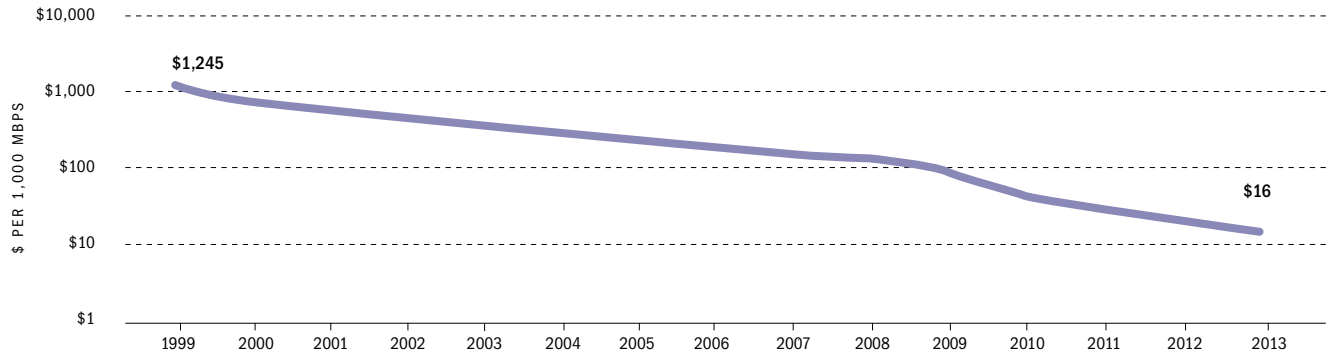
Note: 1 petabyte = 1 million gigabytes. 1 zetabyte = 1 million petabytes

Source: Deloitte, May 2014, as reported by Internet Trends 2014 by Mary Meeker for Kleiner Perkins Caufield & Byers

© World Newsmedia Network 2014

## Global bandwidth costs, 1999-2013

In US\$ per 1,000 megabytes per second



Source: Deloitte, May 2014, as reported by Internet Trends 2014 by Mary Meeker for Kleiner Perkins Caufield & Byers  
© World Newsmedia Network 2014

## Media industry's impact on Big Data

Media companies collect reams of data every minute from every area of their organisations: 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.

To put Big Data and little data in perspective, storage units for a laptop computer are in gigabytes, e.g., little data. 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 show the prolific contribution of Big Data from media companies to the Internet universe. 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.

While collecting and producing data are the first steps in the development of a Big Data practice, the acts of *analysing* and making the data *actionable* are the new mantras for media companies.

## How did we get here?

The Big Data revolution did not happen by accident. Prices for digital media storage and bandwidth, explosion of digital devices including smartphones and tablets, and the exponential growth of audience-accessed digital media have whipped up the perfect

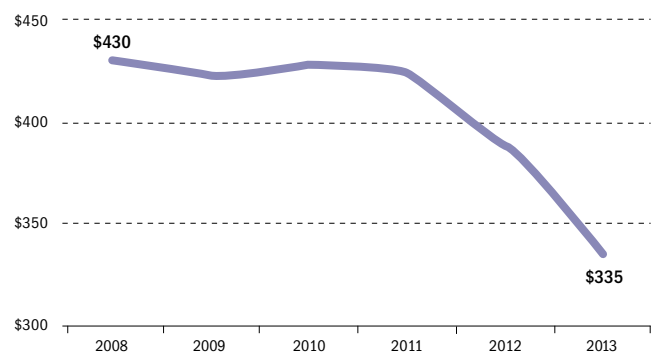
storm to create this surge in Big Data strategies and implementations.

Global data storage costs have plummeted 100 percent from 1992 to 2013, from US\$569 to \$0.02 cents per gigabyte of storage, according to Deloitte research, as reported in Mary Meeker's 2014 *Internet Trends*<sup>2</sup> report.

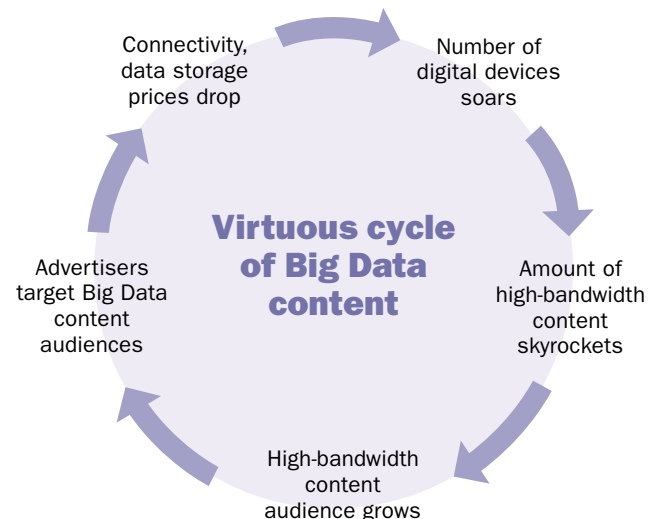
Meanwhile, global bandwidth costs dove 99 percent from 1999 to 2013, from \$1,245 to \$16 per 1,000 megabytes per second, according to Deloitte.

## Global smartphone costs, 2008-2013

Average price per unit, in US\$



Source: Deloitte, May 2014, as reported by Internet Trends 2014 by Mary Meeker for Kleiner Perkins Caufield & Byers  
© World Newsmedia Network 2014



Source: World Newsmedia Network 2014  
© World Newsmedia Network 2014

<sup>2</sup> Kleiner, Perkins, Caufield, Byers, Mary Meeker, 2014, <http://www.kpcb.com/internet-trends>



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

These steep decreases have given rise to the virtuous cycle of Big Data content: 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.

## Big Data: What is it, what kind of data and how much?

Video, social media, other digital platforms



### Video

Streaming video takes up more than 1/3 of the Internet traffic during normal television watching hours

72 hours of video are added to YouTube every minute

864,000 hours of YouTube video are uploaded each day

22 million hours of TV and movies are watched on Netflix each day

Zynga processes 1 petabyte of videogame content per day



### Social media

More than 1.4 billion online consumers are spending 22 percent of their time on social platforms

172 million individuals visit Facebook each day

4.7 billion minutes spent on Facebook each day

532 million Facebook statuses updated each day

250 million photos uploaded to Facebook each day

30+ billion pieces of data are added to Facebook each month

40 million Twitter individual users each day

50 million tweets per day

32 billion searches performed on Twitter per month

22 million LinkedIn individual users each day

20 million Google+ individual users each day

17 million Pinterest individual users each day

2 million blog posts are written each day



### Other digital platforms

1.3 exabytes of data sent and received by mobile Internet users each month

Average teenager sends 4,762 text messages per month

More iPhones are sold than babies born each day

294 billion emails are sent each day

72.9 products ordered per second on Amazon

18.7 million hours of music is streamed on Pandora each day

1,288 new apps are available to download each day

More than 35 million apps are downloaded each day

By 2018, there will be a demand for about 450,000 data scientists in the U.S., leaving a shortage of more than 150,000 positions

Sources: Mushroom Networks, Deep Blue Analytics, MBAOnline, IBM, Gartner  
© World Newsmedia Network 2014

## About this study

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. They use small data to improve UX from reporting, headline optimisation, SEO, content efficiency and consumer research.
- 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. Media companies can learn from the systematic approach from dunnhumby, the force behind the Clubcard.
- 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.
- 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.
- 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.
- Wilfried Runde, head of innovation projects in new media for Deutsche Welle, emphasised why all newsrooms should work with data: status, skills and

scenarios. 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 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. (see Chapter 3's sub-section on the winners on page 14).
- 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. There are techniques to stop the data leakage, and strategies to leverage the data for media companies who own it.

- 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.

---

## Chapter 2 Case Studies from the Big Data Conference 2014

---

### THE HUFFINGTON POST

**Global headquarters:** New York City  
**Unique users:** 90 million users  
**Monthly visitors:** 22 million  
**Monthly video views:** 110 million  
**Average stories published per day:** 1,600  
**Average comments per day:** 1,300  
**Editorial workers:** 350  
**Bloggers:** 70,000, including Hillary Clinton, Deepak Chopra and Desmond Tutu

The HuffingtonPost.com, with 22 million monthly visitors to 11 global editions in 2014 and 16 editions planned by the end of 2015, is one of the most trafficked and fastest growing news and entertainment websites in the world.

Huffington Post CEO Jimmy Maymann lives by the mantra, "It's all about the data." HuffPo uses Big Data to optimise content, authenticate comments, ensure efficacy of native advertising, regulate advertising placement and create passive personalisation.

### Optimising content

Maymann said Big Data allows for a more accurate analytical approach to decision making to improve user and advertiser experience. A real-time statistical dashboard and analytics platform powers HuffPo's entire editorial process, he said.

For example, HuffPo headline writers test the success of headlines in driving readership with "A-B testing." Maymann credits the A-B testing with significantly higher traffic on stories.

Maymann also emphasised the importance of delivering content at the right time and on the right device to the right audience. Through the analysis of Big Data on its parents' section, they determined that parents typically access content on their mobile as opposed to their desktops, are active online on weeknights after kids' bedtimes and engaged during weekend mornings. In order to reach this audience, content is typically published when parents are typically online, which boosts traffic to stories of interest to parents, Maymann said. One of the HuffPo's parent blogs titled "10 Ways Living with a Toddler is Like Being in Prison," received 24,000 views in 7 hours, and 41 percent of the views were driven by social media referrals.

### Authenticating comments

HuffPo content garnered 300 million reader comments in 2013. The high volume but sometimes low quality of comments drove Maymann to seek a Big Data analysis to learn how to improve the user experience through authenticating readers' comments.

HuffPo conducted a conjoint analysis, that is, a statistical technique used to determine how people value different features of a product or service. In the case of HuffPo comments, the conjoint analysis was employed to determine the quality of comments coming from

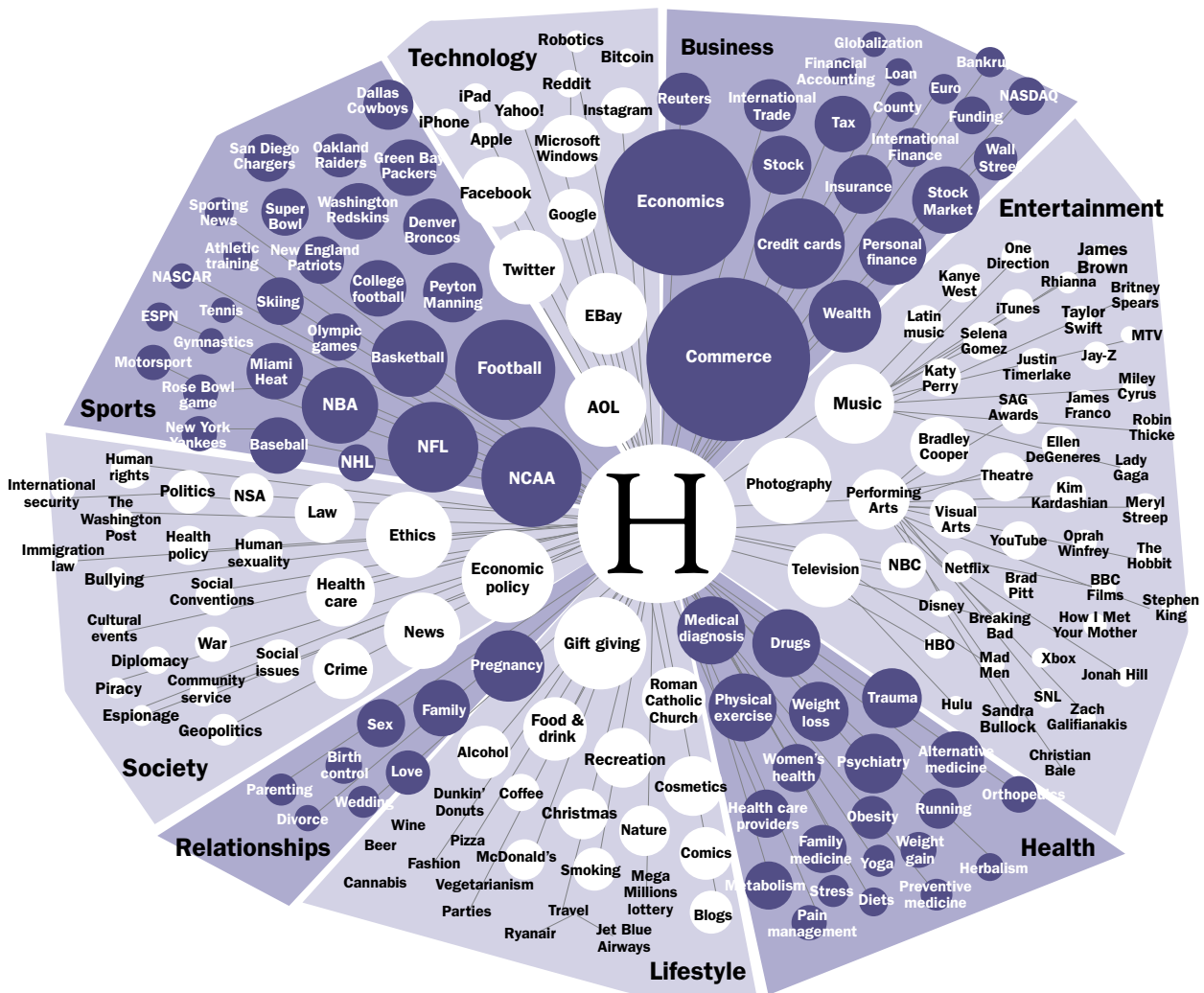
an anonymous person or those who have identified themselves either by name or by avatar and from specific geographies. The analysis determined that the quality of the comments were markedly higher from those who were regionally closer to those surveyed, and from those who were not anonymous. The findings drove HuffPo to require commentator registration and disallow anonymous postings, Maymann said.

“We are getting rid of the bashers, and now have higher quality of people commenting, even though we took some heat in the press for [ending anonymous comments],” Maymann said.

## Proving the efficacy of new revenue models

HuffPo is using Big Data to determine how successful new advertising revenue models are. The company developed the Partner Studio, an arm of the advertising department that handles native advertising, or content marketing. HuffPo partners with major brands to create sub-websites that are content oriented, leveraging HuffPo’s writers, editors and designers.

“We meet the client’s needs while maintaining the HuffPost’s voice and standards,” Maymann said.



Topics Dec. 2013 HuffPo visitors were interested in

By employing Big Data analytics, the HuffPo determined the site’s native ads get a two times higher brand recall over only sponsored content, and a 1.3 times higher brand recall over standard media.

## Adapting technology to help advertisers meet regulations

Pharmacy advertisers face legal restrictions around sensitive content, so HuffPo is using Big Data to flag possible violations of proximity between pharmaceutical ads and related editorial content. The word search function identifies sensitive words that violate the pharma regulations. These flagged words trigger the

blocking of advertisements in proximity to the content, Maymann said.

## Passive personalisation

As a subsidiary of AOL, HuffPo is using the technology of one of its mother company’s newest assets, Gravity, a tech company that identifies individual user interests through browsing behaviours. Gravity allows the HuffPo to connect the right content to the right people through “passive personalisation.”

“Interest graphs are derived from algorithms that identify user-level interest data. This is a snapshot of all



the interests all of the December visitors are interested in; you can see how much depending on size of circles. What is driving engagements on sites? This tells us what we should double down on,” he said.

The Gravity technology also informs HuffPo about where individuals go to access content, and is helping them decide how to improve the user experience for the front page, and other landing pages where users enter the site. Maymann said users now access between 10 and 12 articles per session on average.

“If we can give them 15, that would be very valuable,” he said.

## BUZZFEED

**Global headquarters:** New York City  
**Unique users in May 2014:** 151,000, including online and mobile  
**Unique users:** 95,000 on mobile only

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. For example, photos of food are popular, as are photos with guns, and the colour red and women in bikinis tend to rank higher with traffic than others.

“We use two methodologies to power our analysis, identify characteristics with predictive relationship to virality: quantitative and descriptive,” Harlin said. “We have to understand how the spread of content differs by [social] network.”

“Machine learning predicts social hits: we know what’s viral before it takes off. Regression analysis and machine learning are approaches for data analysis,” he said.

The key takeaways are that data can be used to optimise content for sharing through the life of every article, Harlin said. Before publishing, data can help determine what to write about and how to present it. After publishing, data can help optimise the article’s promotion as well as the article itself, he said.

## CIRCA, DATA-DRIVEN MOBILE CONTENT APP START-UP

**Headquarters:** San Francisco  
**Venture capital to date:** US\$3.4 million

Big Data is an important tool for understanding what content and how to serve that content to mobile users, said Arsenio Santos, co-founder and CTO of Circa.

“What is the problem: Plenty of news. We believe in mobile first. People don’t engage on mobile news. At first we thought it was size of screen,” Santos said. “It’s less about real estate and more about the time.”

In order to bridge the barrier of time, 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 atomised chunk is called a card.

“We strive to present each article as atomic content. It’s also context. Here is how we construct mobile news on Circa. We don’t use auto summarisation; 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.

Metrics are an important element for journalists. “You have to decide how much to give to journalists. You want to give them a nice balance of analytics that matter most to you: users, pages, sources, editorial statistics, story sharing. The metrics that matter include explicit and implicit user behaviour,” he said. “We tend to look at follows and shares more than anything else.”

Santos said Circa has learned lessons from its data-driven news business. First, breaking news is broken, Santos said. News outlets overuse breaking news alerts and disappoint readers and viewers. “There’s an escalating arms race for urgency and attention. We use breaking news sparingly, sometimes we don’t do it for weeks.”

Santos urged publishers to re-think their content, change the way to re-formulate it, and make it more efficient and dynamic. Use Big Data to assist you in streamlining news feeds, as well as creating actionable audience metrics for your team, he said.

## FINANCIAL TIMES AND FT.COM

**Global headquarters:** London  
**Total paid subscriptions (2014):** 652,000  
**Average daily readership:** 2.1 million  
**Total digital subscriptions:** 415,000  
**Growth rate of digital subscriptions, 2013-14:** 31%, or 99,000 subscriptions  
**Total print subscriptions:** 237,000  
**Digital-only subscription:** US\$325/year  
**Non-paid registrants:** 4.5 million  
**Number of FT journalists worldwide:** 600

Data has had a powerful impact on the *Financial Times* over the past year, said Tom Betts, head of customer analytics and research for the *FT*. The credit goes to

the collection of registration data through its metered paywall. The collected data allows the *FT* to serve the customer better, create targeted advertising campaigns and create new products based on information collected on background and areas of interest to its readers.

“Five or six years ago we started a new media model, charging for access through a metered system. When we started doing that, it was primarily to build a revenue stream online, but probably what was more important over time was the data and customer insight that that gave us. That’s what transformed the business,” said CEO John Ridding, according to an April 2014 Mashable article.

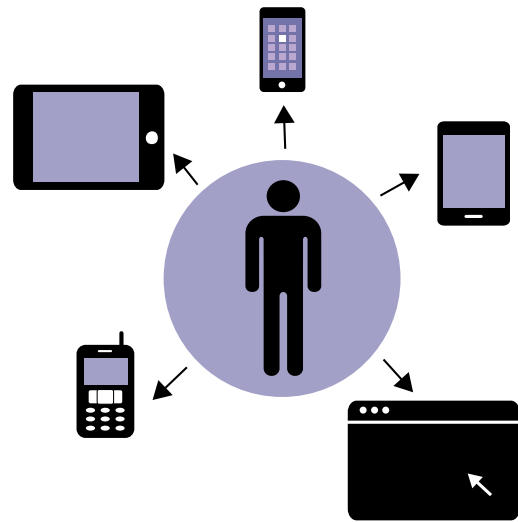
Betts said news consumption patterns are shifting, and now subscriptions to FT.com outstrip subscriptions to the *Financial Times* newspaper. The collection of data is helping to grow digital subscriptions by leveraging a better understanding of the customer.

With the collected data, *FT* builds signatures of digital consumption on their readers’ consumption of content from the companies, markets, world, comments, management, personal and weekend sections of the *FT* in order to:

- Understand customer content preferences
- Increase the relevance in *FT*’s communication to customers
- Personalise content
- Deploy intelligence to customer touchpoints, including customer service, website, mobile apps and third parties, such as advertisers, in order to target campaigns

The addition of mobile channels including tablet and smartphone platforms added a level of complexity to understand each *FT* customer. For example, weekend content might be consumed on a smartphone or tablet on the weekend, while management or finance content might be consumed on the desktop in the office on a weekday.

## Putting customers at the centre



“A multichannel view gives us unique context about customer needs and behaviour,” Betts said.

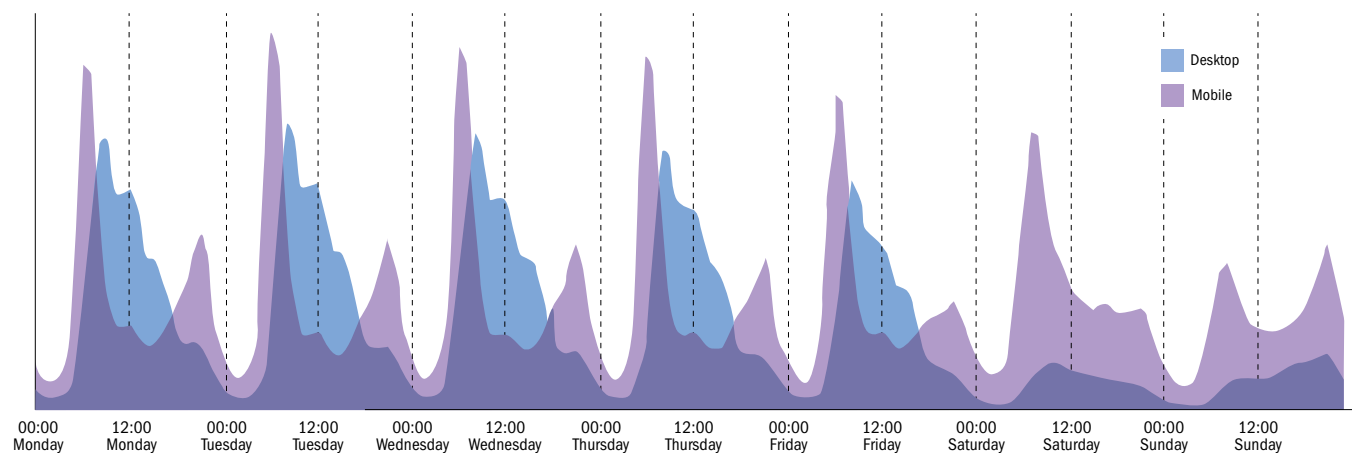
The profiles are useful to create personas for which to build new *FT* products, for example, a retail executive with plenty of discretionary income for luxury and travel, or a young middle manager with high ambitions for getting promoted through the ranks.

“Baked into this segmentation is engagement, not just engaging the most likely subscribers,” Betts said. “[Segmentation] is helping us with product development. ‘I am going to build a product for persona A,’ helps us focus. We will define people in this persona group. They will help us develop and shape the products we want to build.”

Big Data about *FT* subscribers’ user habits have been graphed as time of day consumption over the space of a week.

## Financial Times access by PC and mobile

Time of week and day access of FT content on mobile and print channels



Source: Financial Times 2014  
© World Newsmedia Network 2014

“What are the implications for a daily newspaper with 126 years of history? That brings ways of working and process. We now have to look at the way things are and evolve. Should we publish in the evening [to meet the demands of the reader]?” Betts said. “There is a disconnect when you are publishing and when they are consuming. We must think more like a broadcaster with the programming mindset. We need to optimise to meet these demands.”

Betts’ final recommendations for media companies embarking on a Big Data strategy are:

- Bring your data to life with customer-centricity
- Obsess about data capture
- Don’t be dogmatic about data use and tools
- Break free, break misconceptions and get stuck in!

## Resistance to embracing Big Data practises

Like so many digital opportunities before it, such as the Internet, social media and mobile, the Big Data opportunity is sometimes treated with skepticism and resistance.

Only a year ago, media industry insiders regarded Big Data as the big buzz. Circa 2014, media companies around the world are implanting a variety of strategies using Big Data technologies and initiatives, in order to leverage trends in audience behaviours and pinpoint stories best told through the prism of data sets that are enormous but distilled to provide meaningful messages.

## dunnhumby winning with retail Big Data

Matthew Keylock, dunnhumby’s global head of data analytics, described how to “win” with Big Data by putting the customer first. And customers they have: The global customer science company has 400 million around the world, representing US\$300 billion in retail spend.

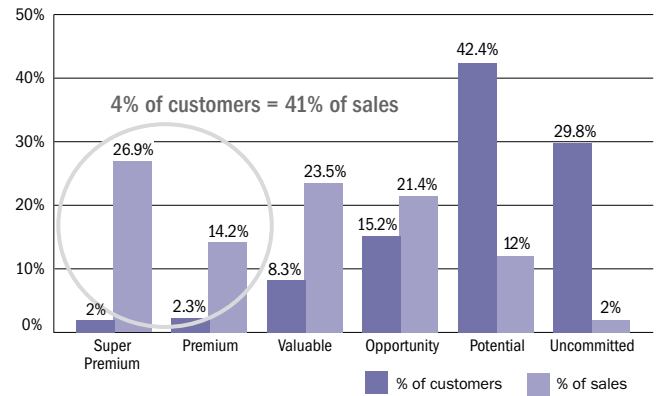
dunnhumby has more than a decade of experience collecting and making actionable reams of Big Data about customer spending at companies such as Tesco and Kroger, in the UK and U.S., respectively. Their success serves as a beacon to inspire and replicate at media companies who wish to engender great customer insight and loyalty among readers, viewers and users.

Keylock described the methodical ways dunnhumby collects, analyses and makes actionable their data from loyalty cards such as Tesco Clubcards, Kroger Plus Cards and the Macy’s loyalty card.

Activating Big Data allows companies to understand who their best customers are and how they behave. For Kroger, a grocery store chain in the United States, the collection of data from their customer loyalty card revealed a staggering fact: that their top 4.3 percent of customers were responsible for 41 percent of the sales. These revelations drove Kroger to understand the kind of

## Tesco’s best customers

Percentage of Tesco customers and sales



1 Super Premium = 351 Uncommitted Customers

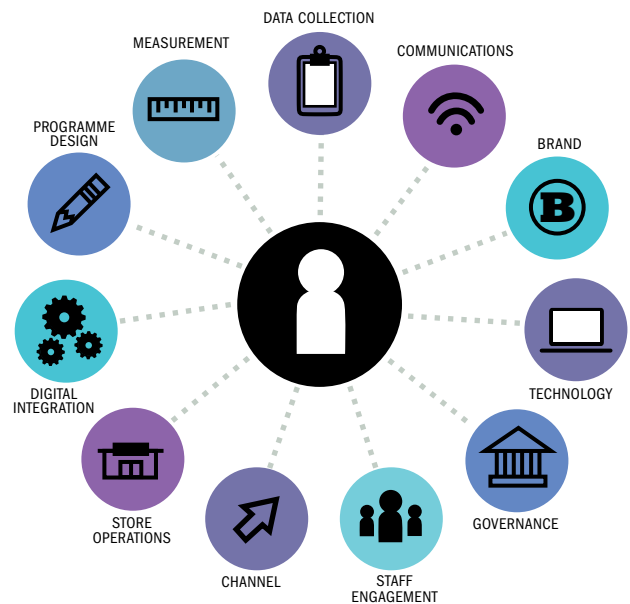
Source: Tesco, dunnhumby  
© World Newsmedia Network 2014

buying patterns their best customers possessed, so they could serve them better.

“If you don’t engage with your best customers, they won’t want to engage with you. Every decision either grows or erodes loyalty,” Keylock said.

For example, further study revealed the best customers tended to respond to promotions, they tended to buy fuel, foreign products and fresh products. They were less likely to be price sensitive, buy kids products and buy items with a long shelf or freezer life.

## Tesco Big Data strategy and organisation centres around customer



Source: Tesco, dunnhumby  
© World Newsmedia Network 2014

“If you don’t engage with your best customers, they won’t want to engage with you. Every decision either grows or erodes loyalty.”

— Matthew Keylock, global head of data analytics, dunnhumby, a subsidiary of Tesco

These learnings changed the way Kroger packaged, priced, promoted, and positioned products, and inspired the way they targeted relevant communications to these customers. These Big Data analyses improved operations and service for customers because of the new information they gathered on these important and other customers.

“See what motivates [each customer]. Build out DNA for each customer. This makes your business a better business,” he said.

Another outcome of data collection is the bringing together of operations regarding Big Data. Each department understands and participates in their part of the Big Data-driven operation.

One of the key outcomes of the collection of data from the Tesco Clubcards is the enablement of targeted marketing messages and special offers sent to 9.5 million members. Tesco reports the unique offers get a 70 percent response, including a 12 percent to 17 percent coupon redemption, which translates to a four-to-one return on investment for brands, and billions of dollars in incremental sales, Keylock said.

In fact, Kroger credits the loyalty programme for outperforming the grocery industry in the United States. Especially significant is that while grocery chains were losing money during the recession of 2009, Kroger was in the black.

Keylock said Big Data can help companies get a 360-degree view of each consumer, that is, their influences, behaviours and motivations. Influences include social media, media data, response data and network and sentiment data. Behaviours include retail data, payment data, tracking and clickstream data, mobile data and location data. Motivation includes lifestyle, household size, contact details, favourite products and reviews and consumer attribute data.

These reams of Big Data allow dunnhumby and its clients, Tesco and Kroger, to create a personalised experience for each customer.

Keylock said these retail Big Data strategies are very applicable to media companies, particularly as it relates to the drivers of behaviours for media users. Here are some of the customer-focus related questions to consider when planning your Big Data strategy, focused on audience, Keylock said:

- Do I know my audience as individuals?
- Is my media measured right? What is the return on investment?
- Which content attracts niche customer groups?
- How should I price and sell my channel better?
- Why is my media not personalised yet?
- Which ads destroy the customer experience?
- How can I help my advertisers make sure every campaign is better than the last one...so they will invest more with me?

Big Data strategies represent huge opportunities for media companies. Here are some tactics to consider, Keylock said:

- Build a loyalty scheme for your audience
- Benefit from other data to marry with your data
- Drive targeting and measurement with these data
- Use your data to influence beyond your existing business

## SACRAMENTO BEE, PART OF MCCLATCHY CORP.

**Headquarters: Sacramento, Calif., USA**

**Founded: 1857**

**Sacbee.com average monthly visitors (2013):  
3.2 million**

**Average monthly page views: 26 million**

**Newspaper circulation (2013): 193,570  
weekdays; 311,329 Sundays**

### Sacramento Bee's actionable Big Data strategy

The *Sacramento Bee* has taken a systematic approach to Big Data strategy, according to Darrell Kunken, head of research for the *Bee*. With the goal to position itself for the future, the American newspaper company is taking the Big Data challenge from four perspectives:

- Technological systems
- Customer analytics and tracking
- Business models
- Culture issues

Each of these areas represent tactical issues, such as data visualisations of audience and financial data, and enterprise issues, such as the company's "individual content profiling system," which is focused on learning more about digital visitors and how to deliver a better user experience to them. "It's all about our content," Kunken said.

### Data visualisation

Kunken said there is a companywide move to identify areas impacted by data, and to apply the tools to visualise the data in order to make it useful to the company. Among the areas to deploy data visualisation tactics are auto dealer sales, shopper trends, small and mid-size customers and path-to-purchase data for advertising sales materials.

The *Bee* also maps where current subscribers and former subscribers live, and CPM reporting for business management support materials. Audience analytics assists the news department in understanding what,

when and how long readers engage with stories. All of these visualisation tactics go a long way to inform business decision-making at the *Bee*, Kunken said.

The *Bee* uses a data visualisation technology called Tableau to visualise data for car dealers in the Sacramento market to show them how they rank among competitors in vehicle sales. Tableau also is used to map former subscribers and their reasons for cancelling their subscriptions. This analysis prompts analysis and action for the circulation department.

Big Data has allowed the *Bee* to develop an elaborate path-to-purchase visualisation tool to help advertisers target demographic and special interest groups, like gardeners or home improvement enthusiasts.

The *Bee* also works with local businesses' data and can analyse it to show the owners their best customers, their demographic breakdown and their geographies. One example of this is the breakdown shown for the *Bee*'s advertiser, the B-Street Theater.

## Needs analysis to consumer path-to-purchase tool

Target audience selection

### Select Target Audience

Select your target age, industry, lifestyle, income, and family status below.

Industry    
 Lifestyle    
 Income    
 Age    
 Children

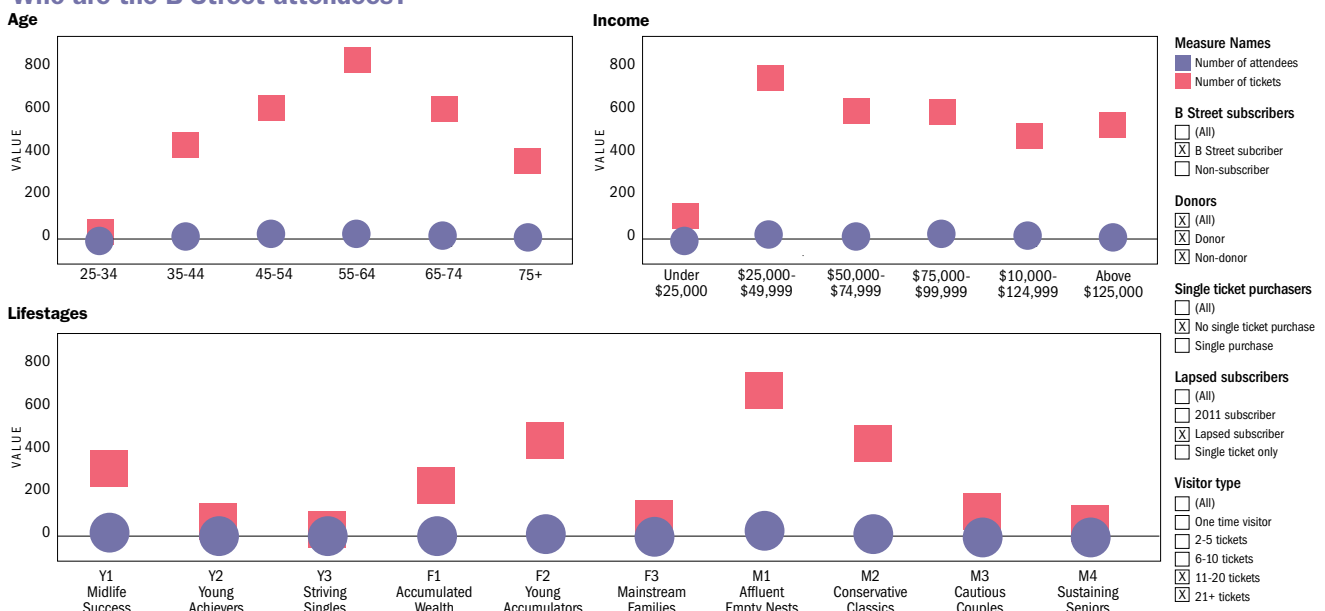
#### Target's lifestyle and media use

Lifestage	Name	Description	Avg. index	Media Use
Y1	MIDLIFE SUCCESS	Childless singles and couples in their thirties and forties. The wealthiest of the Younger Years class are college-educated residents who make six-figure incomes.	108	Daily print, Sunday print
F1	ACCUMULATED WEALTH	Accumulated Wealth contain the wealthiest families, mostly college-educated, white-collar Baby Boomers living in sprawling homes.	100	Sacbee.com
F2	YOUNG ACCUMULATORS	Adults typically have college educations and work a mix of white-collar managerial and professional jobs.	103	Mobile Newspaper
F3	MAINSTREAM FAMILIES	Middle- and working-class child-filled households with at least one child under 18, living in modestly priced homes.	76	Yahoo
M1	AFFLUENT EMPTY NEST	Feature upscale couples who are college educated, hold executive and professional positions and are over 45.	117	Email coupon users
M2	CONSERVATIVE CLASSICS	College educated, over 55 years old and upper-middle-class, offer a portrait of quiet comfort.	93	Mail coupon users

## Big Data strategy and organisation centres around customer

*Sacramento Bee*'s customer, B Street Theater, analysed profiles of its best customers through *Sacramento Bee*'s Big Data technology

### Who are the B Street attendees?



Source: Sacramento Bee  
© World Newsmedia Network 2014



In the editorial department, Big Data are being used extensively for ordinary stories and investigative stories. Stanford-trained data reporter Phillip Reese became a 2013 Pulitzer Prize finalist for his data-driven investigative story on patient busing in Nevada. Reese also maintains the *Bee's* DataCenter, [sacbee.com/datacenter](http://sacbee.com/datacenter), where the *Bee* stores stockpiles of data sets and data-driven stories of interest to *Bee* users.

Further, the *Bee* is embarking on a new editorial mission, an enterprise project that aims to analyse user Big Data to personalise and push desired content to users, and to monetise that personalised content.

“Consumers need to understand what you do with their data,” Scott Howe, CEO of Acxiom, said at a Newspaper Association of America conference in March 2014. “Seventy-five percent of consumers would be happy to exchange their data for added value.”

Kunken said readers were asked what they most wanted from the *Bee*, and the vast majority wanted to be informed. What the *Bee* wants from readers and visitors in return is frequency, sharing, satisfaction from their visit and engagement with the *Bee*, he said.

In order to achieve their objectives, key leaders are working with Stanford University journalism and computer technology students to create a tracking and profiling system in order to customise the user experience on [sacbee.com](http://sacbee.com). The system adds meta-tags to online articles and connects that information with readers of news for news publications.

The technology scans articles for content types and matches that with the needs of users, enabling the serving of content and advertising to targeted individuals. The content targeted can be segmented in many ways, including by device, behaviours, influential users, super news consumers, content-specific (business or sports, for example), and ads of interest.

The *Bee* is working with a variety of technology providers to enable this targeting, including Cxense, Neodata, Lotame, Jaspersoft, Gigy and Tableau. The organisation is changing to accommodate the new approach to business. Programmers have been added in the newsroom, research analysts are becoming data miners and data scientists, and top directors of the company are working hand-in-hand to meet the challenges of the new Big Data imperative.

The newsroom director of innovation, director of audience, director of digital and director of research all work together to achieve their objectives of transparency, users' full control and privacy for users' data. They also strive to achieve a stronger engagement with the users by providing more value and variety in content.

## ARCHANT

**Headquarters:** Norwich, United Kingdom

**Newspapers:** 4 regional papers in the UK, with 1.6 million weekly readers

**Magazines:** 80 regional and lifestyle titles

**Websites:** More than 200 sites, with 6 million unique monthly viewers

Serving relevant and compelling content on every digital channel puts customers at the centre of Archant's strategy, said the group's digital director, Paul Hood. A variety of challenges face traditional publishers as they try to monetise their growing digital operations.

Digital display advertising is forecast to grow from US\$40.2 billion in 2013 to \$57.6 billion in 2015, according to ZenithOptimedia. However, publishers get only a small percentage of that revenue, as the four top advertising companies, Google, Facebook, Microsoft and Yahoo!, realised 39.9 percent of that revenue in 2013, and are expected to garner 43.2 percent by 2015 in the United States, according to eMarketer.

Further, advertisers buy digital ad impressions at a variety of price points, from £2 to £20 per thousand impressions (CPM); however, traditional media sales channels don't capture all the money. Hood said ad sales directly from traditional publishers to advertisers might fetch £20 CPM, while automated, third-party ad networks might only drive a CPM of 50 pence. Consequently, an increasing amount of advertisers are opting for third-party ad networks, run by the likes of Google and AOL.

In addition, these third parties have superior technology and better reach than traditional media companies, Hood said.

### Data leakage

Third-party ad networks process millions of ads and embedded ad-serving instructions each day, making them enormous Big Data operations. As more advertisers are opting for third-party ad placement, publishers are increasingly concerned about hemorrhaging customer data into remnant ad networks. Hood said publishers should embed native ad formats into their sites and audit third-party tags on their sites.

“Read the fine print in the third-party ad contracts,” he said.

Hood also shared the verbiage for a publisher contract with a third-party ad network that serves ads to more than 9 million websites per day:

“We may deploy a cookie on our own behalf or on behalf of one of our data partners, to record information about how an End User uses the web. We may use the Data to target advertising toward the End User or authorize others to do the same. ... If you do not cause us to discontinue the placement of our cookies, you grant us a non-exclusive, perpetual, worldwide and irrevocable right and license to collect, use and disclose the Data as provided in our Privacy Policy and to allow our third party data partners to do the same.”

Publishers must be proactive and tell the third-parties to discontinue the use of the cookies. “Minimising data leakage is the first step to reclaiming market share,” Hood said.

## BRITISH BROADCASTING CORPORATION

**Headquarters:** London

**BBC News Online traffic, March 2014:**  
96 million users

**Tweets of BBC stories, March 2014:** 2.71 million, the most shared news brand in the world

**Mobile usage:** 88 percent growth year-over-year to 24 million browsers per month in March 2014, eclipsed two-to-one by desktop access

## Data journalism

For the BBC, data journalism is visual journalism.

“We want to use our skill and creativity to engage and inform our audiences on the biggest, most significant stories, providing insightful, personal and shareable visual explanations,” said Amanda Farnsworth, editor for visual journalism at BBC News, describing the BBC’s mission for visual journalism.

Farnsworth described one of the most successfully visual journalism stories as the “Great British Class Calculator,” a survey urging participants to learn to which British social class they belong. The huge popularity of the survey and follow-up stories confirmed British societal obsession with class is still alive and well. After the story was launched and promoted on BBC TV channels, BBC.co.uk page views spiked into hundreds of thousands of users consistently every hour.

Visual journalism stories with personal relevance to users are also popular. Stories reaching out to users to ask how much they pay at the petrol pump, and calculators showing how the new national budget will impact a person’s budget create a buzz and play a role in boosting traffic, she said.

Farnsworth’s department was launched in 2014, and continues to develop its impact across BBC departments. As data journalism continues its influence in the BBC, more data projects are expected to be tackled, she said.

Farnsworth recommended some tactics for other journalism operations to grow their data journalism departments:

- If you’re unsure about a project, dedicate a week or two of research where you explore the data before committing resources.
- When combining data sets make sure they work together.
- A global data set will be relevant to more users and will get more traffic.

## CABLE NEWS NETWORK

**Headquarters:** Atlanta, USA

**CNN News Online traffic, March 2014:** 2.8 billion page views, 1 million app page views, and 274 million global video starts across platforms.

**Mobile usage (March 2014):** 1.6 billion page views

## CNN’s approach to data journalism

At CNN, Big Data is used in three ways: as an early warning system for breaking news, as a window into how its vast audience is using technology and how they are consuming news in real time, and as a way to distill major data sets and present them in data journalism stories, said Peter Bale, CNN International vice president and general manager for digital.

CNN and its partner Twitter work with Dataminr, a New York-based technology firm that transforms Twitter’s “firehose” of 500 million tweets per day into a stream of actionable alerts.

Dataminr is also used as an “early warning system” of news, Bale said. The information technology company uses an algorithm that identifies, qualifies and alerts clients to key information in real-time. It is also used to find needles in a haystack, that is, the genesis of news, and can be used to verify information before publishing.

CNN also uses a technology to tap into the “wisdom of the crowds” in order to achieve fair and balanced journalistic stories. CNN has partnered with Jana, a Boston-based technology company, to create opinion polls related to news events, most recently surveying about 19,000 panelists in 20 countries about life after Nelson Mandela’s passing.

Before the election of a new pope in 2013, CNN International used Jana’s technology to poll 20,000 people in 11 countries in sub-Saharan Africa about their opinions on the possibility of an African being named pope. The survey results were used on air and online,

providing a unique perspective from a seldom-tapped region of the world.

For their data journalism strategy, Bale said a key take-away is that “what works online may not work on TV.” Data allows the journalists to show online, not tell, as

on TV. For example, in coverage of Davos 2014, CNN created a series of maps showing key issues discussed at the summit. On a story about U.S. war casualties, CNN created a satellite map populated by data pinpointing where concentrations of servicemen were from, particularly the East Coast of the United States.

---

## Chapter 3 The Issues

---

### Training data journalists

Wilfried Runde, head of innovation projects for new media at Deutsche Welle in Germany, advocates the addition of data journalism practises in newsrooms worldwide.

Two of the key challenges to any newsroom in getting started with data journalism practises are resources and training. Runde recommended a new data journalism organisation providing free online data journalism training courses, called Data Driven Journalism, [Datajournalismcourse.net](http://Datajournalismcourse.net).



Instructors for the course are:

**Stephen Doig**, Pulitzer Prize-winning data journalist and data journalism professor at Arizona State University, United States

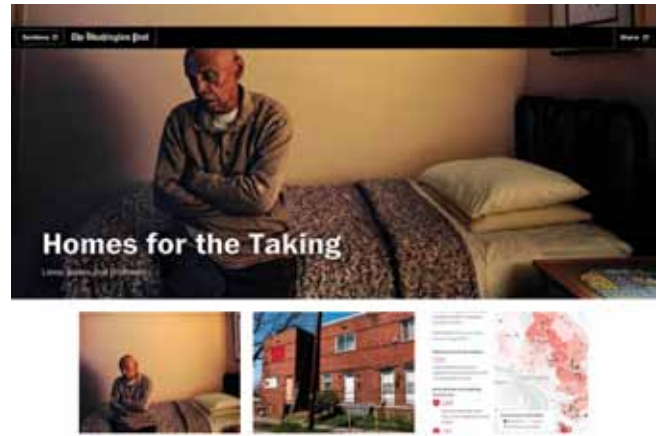
**Simon Rogers**, former *Guardian* journalist, current data editor at Twitter and author of the book “Facts are Sacred: the Power of Data”

**Paul Bradshaw**, professor, Birmingham City University, United Kingdom

**Nicolas Kayser-Bril**, co-founder and CEO of Journalism ++, data journalism startup

**Alberto Cairo**, professor, University of Miami, United States

The five modules for the inaugural course in May, June and July 2014 were the following. The course is expected to be repeated at a later date.



**Module 1 — Data journalism in the newsroom**

**Module 2 — Finding data to support stories**

**Module 3 — Finding story ideas with data analysis**

**Module 4 — Dealing with messy data**

**Module 5 — Telling stories with visualisation**

### Data journalism award winners

The Global Editors Network, based in Paris, awarded eight winners from a field of 75 nominated data journalism projects in a global competition for 2014. Google and the Knight Foundation sponsored the third annual competition, which concluded at an awards ceremony in Barcelona in June 2014.

Details of each winning data journalism project can be found here: [globaleditorsnetwork.org/programmes/data-journalism-awards/](http://globaleditorsnetwork.org/programmes/data-journalism-awards/)

The winners were:

- Best Story on a Single Topic: “The Migrant Files,” Detective.io
- Best Data-driven Investigation: “Homes for the Taking,” The Washington Post
- Best Data Visualisation: “Reshaping New York,” The New York Times
- Best Application or Website: “Open Statement of Assets from Argentina’s Main Public Officials,” La Nacion
- Best Individual Portfolio: Chad Skelton
- Best Team or Newsroom Portfolio: NZZ
- Best Entry from a Small Newsroom: Kiln
- Jurors’ Choice: ProPublica
- Public’s Choice: “In Climbing Income Ladder, Location Matters,” The New York Times

## Data-driven automation in journalism

In a profession where journalists typically train for years before being published by a major publication, it may seem unfathomable to have machines automatically write articles, create videos and recommend story placement. However, the emergence of Big Data tools for journalism have made it possible to do all of the above, and effectively.

Consider the following technology companies in the content automation space:

- Narrative Science and LocalLabs, both based in Chicago, data-to-content generation
- Yseop, based in Dallas, automated personalised and contextual content
- Automated Insights, based in Durham, N.C.; and Cxense, based in Oslo, audience insights and recommended content placements
- Wibbitz and Woch.it, both based in Israel, text-to-video automation





## Narrative Science: Turning data to a story

Countries with smartphone penetration below 45 percent in 2013

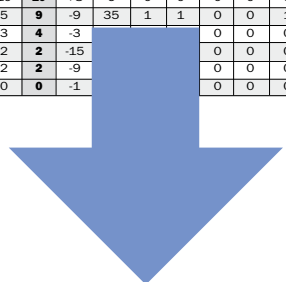
Rio Grande Valley											Laredo										
#	NAME	G	A	PTS	+/-	S	PIM	#	NAME	G	A	PTS	+/-	S	PIM						
5	Rioux, Nicolas	0	1	1	-1	0	2	4	Mariani, Kyle	0	0	0	0	0	0						
9	Lee, Aaron	0	0	0	-1	1	0	5	Moore, Greg	0	0	0	0	0	0						
10	Campos, Brandon	0	0	0	-1	0	2	8	Masters, Rusty	0	0	0	0	0	0						
11	Mikkelsen, AJ	0	0	0	0	2	0	9	Sparling, Ryan	0	0	0	0	0	0						
14	Pearson, Zac	0	0	0	-2	5	0	10	Boisvert, Erik	0	0	0	0	0	0						
15	Dixon, Jason	0	0	0	-1	1	0	12	Boyd, Matt	0	0	0	0	0	0						
19	Ischi, Phillip	0	0	0	-1	2	0	13	Harcharik, Jose	0	0	0	0	0	0						
21	Klempa, Tomas	1	1	2	-1	1	0	15	Pierce, Matt R	0	0	0	0	0	0						
24	Fontaine, Charles	0	0	0	-1	0	2	17	Inman, David	0	0	0	0	0	0						
26	Wilson, Mike	0	0	0	-1	0	0	19	Styffe, Justin	0	0	0	0	0	0						
28	Barr, Tyler	0	0	0	0	1	2	20	Hau, A.J.	0	0	0	0	0	0						
30	Murray, John	0	0	0	0	1	0	23	Trimm, Neil	0	0	0	0	0	0						
34	Rogers, Wylie	0	0	0	0	0	0	27	Stitt, Tyler	0	0	0	0	0	0						
51	Balan, Scott	0	0	0	0	2	0	28	Rivet, Adam	0	0	0	0	0	0						
55	Lineker, Craig	0	1	1	-2	0	0	31	Jung, Torrie	0	0	1	0	0	0						
67	Startsev, Kirill	0	0	0	-1	1	0	33	Rondeau, Ryan	0	0	0	0	0	0						
71	Gendur, Dan	1	1	2	-1	7	4	44	Hanson, Matt	0	0	2	0	7	0						
82	Boogaard, Aaron	0	0	0	-1	0	0														

Laredo Bucks																
#	PLAYERS	P	GP	G	A	PTS	+/-	PIM	PPG	PPA	SHG	SHA	GWG	SHOTS	SPCT	
8	Masters, Rusty	F	43	12	21	33	-15	22	5	4	1	1	5	126	.095	
23	Trimm, Neil	F	28	7	22	29	-6	6	1	11	0	1	1	72	.097	
28	Rivet, Adam	D	43	8	19	27	-11	42	3	6	2	3	1	88	.091	
9	Sparling, Ryan	F	43	9	17	26	-8	37	2	3	1	1	0	78	.115	
15	Pierce, Matt R	F	29	7	18	25	-5	27	2	8	0	1	0	75	.093	
12	Boyd, Matt	F	42	10	11	21	-9	43	1	4	2	1	2	64	.156	
10	Boisvert, Erik	F	42	7	11	18	-4	18	2	4	0	0	1	83	.084	
44	Hanson, Matt	D	43	7	10	17	-26	79	3	4	1	1	2	117	.060	
17	Inman, David	D	33	9	7	16	+2	44	5	3	0	0	1	87	.103	
13	Hau, A.J.	F	32	7	7	14	-4	56	1	1	0	0	1	66	.066	
20	Harcharik, Joseph	F	40	4	9	13	-10	6	1	2	0	0	1	61	.125	
10	X Griffith, Todd	F	12	5	7	12	-10	39	1	3	0	0	0	40	.087	
19	Styffe, Justin	F	38	6	5	11	-10	133	0	0	2	0	1	69	.083	
	X Reese, Jason	F	17	4	7	11	-8	11	1	1	0	1	0	48	.000	
27	Stitt, Tyler	D	42	0	10	10	+8	6	0	0	0	0	0	23	.078	
4	Mariani, Kyle	D	42	4	5	9	-9	35	1	1	0	0	1	51	.100	
28	Barr, Tyler	D	14	1	3	4	-3						0	0	10	.000
30	Murray, John	D	17	0	2	2	-15						0	0	17	.000
34	Rogers, Wylie	D	9	0	2	2	-9						0	0	9	.000
5	Moore, Greg	D	15	0	0	0	-1						0	0	8	.000

GOALIES						GOALIES					
#	NAME	DEC	MIN	SHOTS	SAVES	#	NAME	DEC	MIN	SHOTS	SAVES
30	John Murray	LOSS	58:51	25	22	31	Torrie Jung	WIN	60:00	24	22
34	Wylie Rogers (b)					33	Ryan Rondeau (b)				



## Automating text stories using Big Data technologies

Narrative Science's technology saves time and money for media companies, Larry Birnbaum, the start-up's chief science advisor, told audiences at both Northwestern University's Big Data conference in Qatar in November 2013 and the Global Editors Network conference in Barcelona in June 2014. Narrative Science's algorithms use structured data such as databases with sports scores, stock market performance data, Twitter feeds and audience usage data, and parlays them into narrative stories. For example, the company can take sports statistics for a baseball game, complete with names of players and statistics about their performance during the game, and transform that information into a story ready for publication within minutes.

Some may object to this process, insisting that journalists must check facts, and Birnbaum agrees. He said the need for journalists still exists, especially for high-level journalism, but not necessarily for basic sports and business stories.



### Rio Grande Valley rolls over Laredo, 6-3

The Rio Grande Valley Killer Bees were firing on all cylinders against the Laredo Bucks, and when the final buzzer sounded Killer Bees emerged with a 6-3 win.

Zac Pearson was the hero of the night for Rio Grande Valley, as he tallied two goals and one assist in the first period. Pearson scored the first of his two goals at 5:23 into the first period, and then scored the second goal at 12:44. He picked up the assist on the second goal at 12:44.

The Rio Grande Valley Killer Bees were firing on all cylinders against the Laredo Bucks, and when the final buzzer sounded Killer Bees emerged with a 6-3 win. The game was a 6-3 win for Rio Grande Valley. There were 14 penalties in the game. The first penalty was at 12:44. The game was the first period, which saw 18 minutes of penalty time combined between the two teams.

The Killer Bees' goal total was higher than their season average. Rio Grande Valley averages two goals per game. The Killer Bees could not stay out of the penalty box, as the team accrued 17 minutes in penalties during the game. The leading offender was Jason Beeman, who totaled five minutes in penalty time with one major. With 48 shots on target during the contest, Rio Grande Valley exceeded the 22 shots it averages per game this year.

Rio Grande Valley additionally got points from Aaron Lee, who had one goal and one assist, Marshall, who registered one goal and two assists, and Dan Gendur, who racked up one goal and one assist. Dan Nicholls also scored for Rio Grande Valley. Others to record assists for Rio Grande Valley were AJ Mikkelsen, who had two and Adam Bartholomay and Marc-Andre Carre, who each chipped in one.

Laredo was often in penalty trouble, as it ended with six minors and one major for 17 minutes in penalty time. The leading offender was Justin Styffe, who totaled nine minutes in penalty time with two minors and one major.

Source: Narrative Science  
© World Newsmedia Network 2014



## Automating video stories using Big Data technology

Woch.it enables video creation on the fly, with text-to-video automation. The system fetches video and photos from vast, licensed databases, such as the Associated Press, Reuters and Getty, and also searches for relevant social media and information graphics content. Media companies also can plug in their own libraries of content.

In minutes, the video is generated, and a voiceover can either be added by Woch.it or by the media company. The technology promises to produce video on demand, reduce production costs, generate video ad revenue and maintain high editorial standards. While the automation streamlines video-making from hours to minutes, the technology gives publishers control over video selection and voiceovers, for example.

## Chapter 4 Insights on Big Data from Beyond the Media Sector

### Big Data studies

A variety of studies about Big Data have been produced in 2014 by some of the best research companies in the world. 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.

### InformationWeek 2014 study

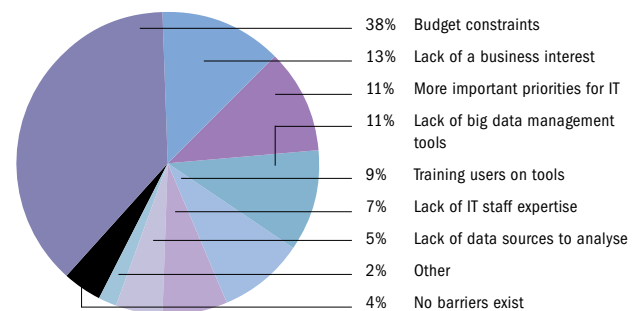
According to some of the most forward-thinking experts specialising in Big Data structure, such as Gartner and IBM, a data governance plan must be put into place in order for Big Data strategies to be successful. Data governance plans comprise the technical and security policies and procedures that become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

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. 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’ CRM systems, and online usage patterns such as social media, paid content subscription trends, story engagement, content popularity trends, e-commerce trends and more.

### Top barriers to successful use of Big Data

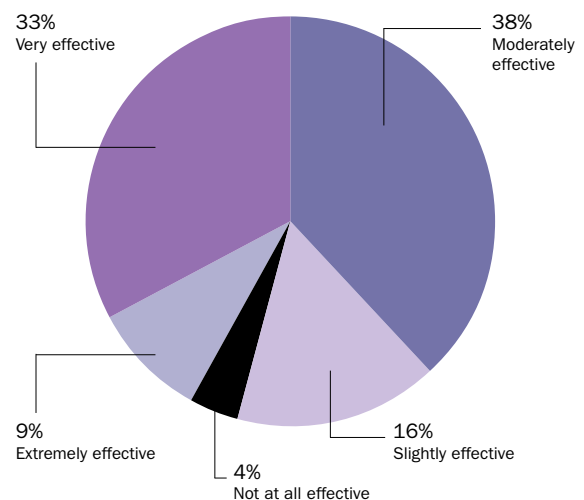
In percentage of respondents



Source: InformationWeek, “Blinded by Big Data,” 2013  
© World Newsmedia Network 2014

### Critical data for decision-making

The degree to which organisations are effective at identifying critical data and using it for decisions

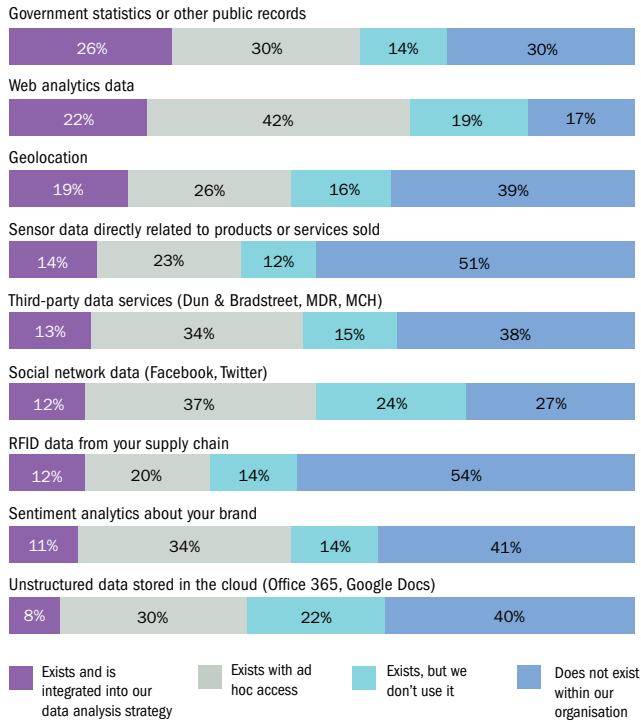


Source: InformationWeek, “Blinded by Big Data,” 2013  
© World Newsmedia Network 2014

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. 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.

## External Big Data sources

The degree to which external data sources are used as part of the respondents' data analysis and reporting programmes



Source: Information Week, "Blinded by Big Data," 2013  
© World Newsmedia Network 2014

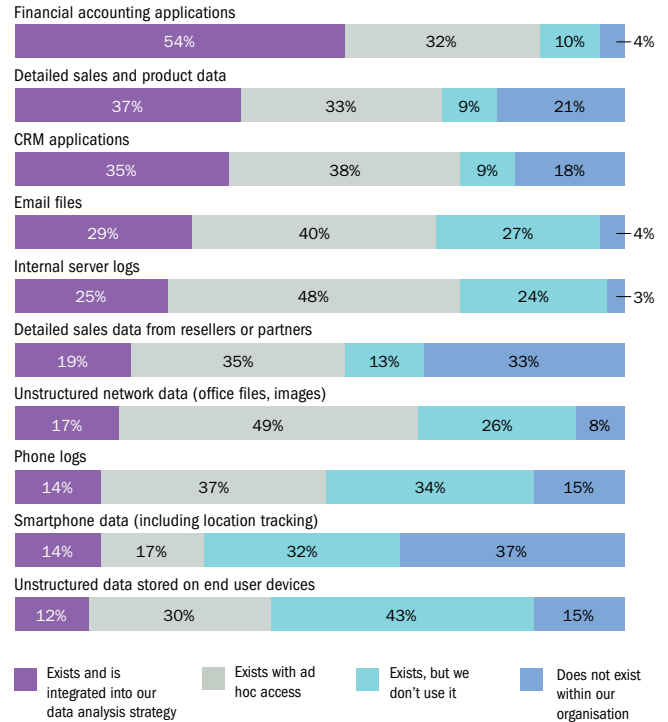
More companies are marrying their own data with that of external Big Data sources, according to the survey. In particular, governmental statistics and public records (26 percent); Web analytics data (22 percent) and geo-location data (19 percent) are some of the most popular data to be currently integrated into the respondents' data practices, according to the study.

Among the most applied internally available data used in respondents' Big Data strategies are financial accounting applications (54 percent); detailed sales and product data (37 percent); CRM applications (35 percent); email files (29 percent); and internal server logs (25 percent).

Developing "actionable" insights about the reams of Big Data generated at companies is a resounding theme in the many studies published in 2013 and 2014. It's the logical next step to gathering stockpiles of data across business enterprises. The No. 1 strategy made in response to Big Data collection is to increase the capacity of hardware or personnel in order to analyse Big Data (47 percent), according to KPMG's "Going Beyond the Data: Achieving Actionable Insights with Data and Analytics." The survey of CFOs and CIOs also found that attaining management buy-in for Big Data collection (39 percent) and honing the parameters of data collection (27 percent) were top Big Data strategies.

## Internal Big Data sources

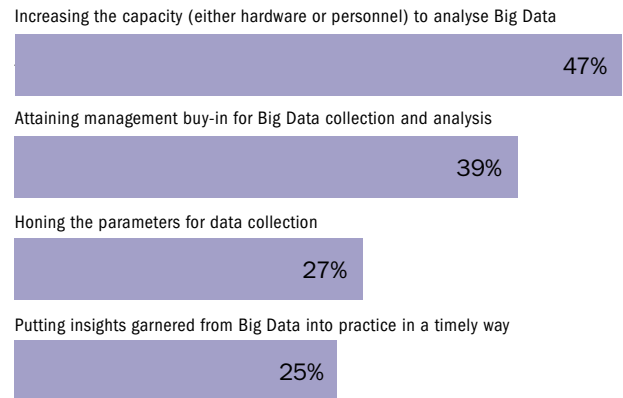
The degree to which internal data sources are used as part of the respondents' data analysis and reporting programmes



Source: Information Week, "Blinded by Big Data," 2013  
© World Newsmedia Network 2014

## Business strategies made in response to Big Data

% of global CFO and CIO respondents

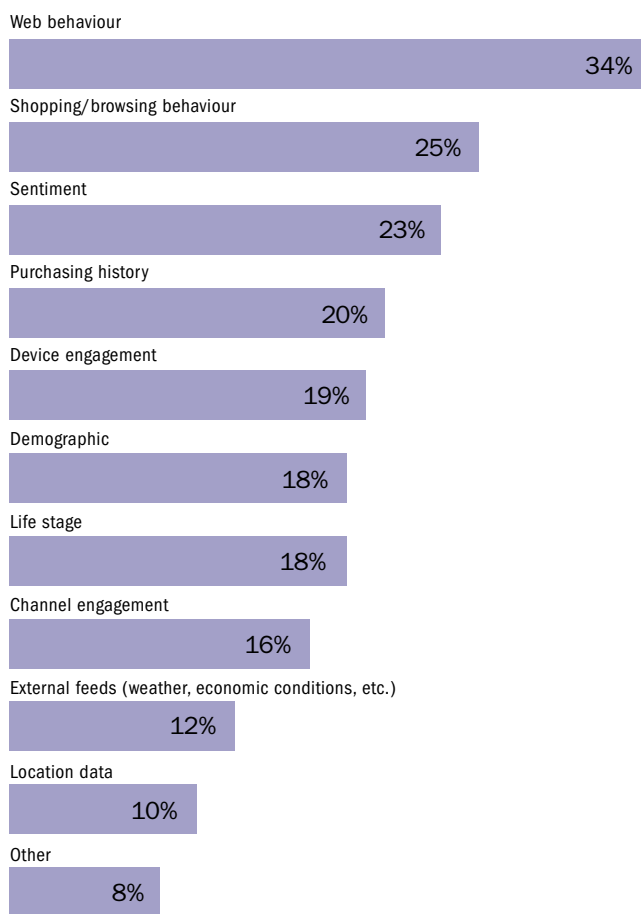


Source: KPMG, "Going Beyond the Data: Achieving Actionable Insights with Data and Analytics," January 2014, eMarketer  
© World Newsmedia Network 2014

Leveraging Big Data for actionable insights isn't always as straightforward as it sounds. According to StrongView's "2014 Marketing Trends Survey," a significant number of marketers surveyed found leveraging Big Data in their email campaigns to be challenging. One-third of the respondents said they reported a disconnect between the collection of Web behaviour data and making it actionable in an email campaign, while one-fourth found it a challenge to

## Inability to leverage Big Data insights in email marketing

% of respondents reporting disconnect between available data and inability to make it actionable in an email campaign



Source: StrongView, "2014 Marketing Trends Survey," December 2013, eMarketer © World Newsmedia Network 2014

leverage shopping and buying behaviour in a campaign. Data collected about sentiment (23 percent), purchasing history (20 percent) and device engagement (19 percent) also proved to be challenging to leverage in the marketers' email campaigns.

## Big Data and privacy concerns

While Big Data audience analytics data collection presents an enormous opportunity for companies to better understand their audiences and consumers, the collection of consumer information can be problematic if not handled properly. Every country has different regulations applying to data collection, and the use of these data to content targeting and advertising. Corporate policy making regarding the use of these data will be key as Big Data strategies progress among media companies.

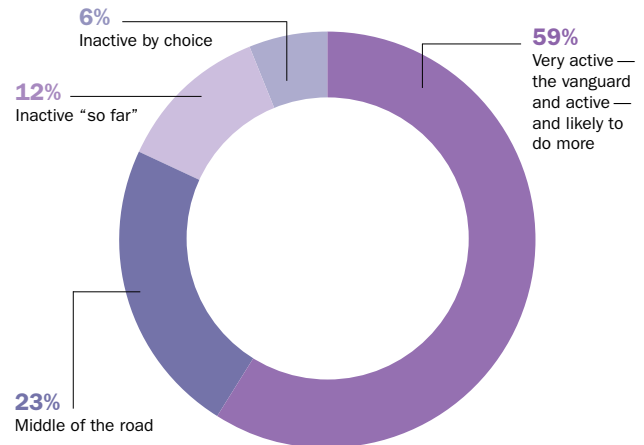
The number of companies gathering data for audience analytics around the world continues to grow, according to a global study by Forbes Insights and Turn research in Sept. 2013. Three out of five (59 percent) business

executives surveyed say their companies are active in collecting, analysing and acting on consumer data. Further, 42 percent said they are likely to do considerably more in the future.

Meanwhile, one-fourth (23 percent) of the executives said their companies are "middle of the road" in the collection of data and customer analytics, followed by 12 percent who are inactive so far, and 6 percent who are inactive by choice.

## Global companies' data gathering and analytics

In % of responses



Source: Forbes Insights, "The Promise of Privacy: Respecting Consumers' Limits While Realizing the Marketing Benefits of Big Data," 2014 © World Newsmedia Network 2014

Meanwhile, 85 percent of the executives surveyed agree or strongly agree that significant breaches of customer data security would do great harm to customer relationships, according to the business survey. In fact, the larger the company, the more likely they perceived the risk of damage to reputation and customer relationships if there was a privacy breach regarding customer data.

The Forbes Insights study was a combination of three surveys of executives with knowledge of their Big Data operations, business-to-consumer consumers and business-to-business consumers.

While 81 percent of the business executives think consumers are highly concerned regarding privacy, only 47 percent of the business-to-consumers surveyed in a related

## Concern over privacy issues

In % of responses

	Company executives	B2C consumers	B2B buyers
Very concerned	48%	24%	11%
Concerned	33%	23%	20%
<b>TOTAL</b>	<b>81%</b>	<b>47%</b>	<b>31%</b>

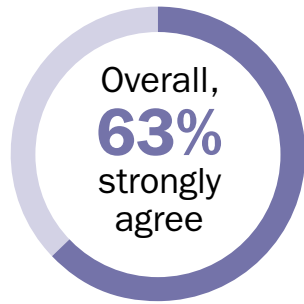
Source: Forbes Insights, "The Promise of Privacy: Respecting Consumers' Limits While Realizing the Marketing Benefits of Big Data," 2014 © World Newsmedia Network 2014

## Perceived risks of privacy breaches

In % of responses

Overall, 85% agree (22%) or agree strongly (63%) that privacy breaches are bad for relationships, reputations and the bottom line.

But the larger the company, the more likely it is to agree strongly:



Under \$1 billion



\$1 billion – \$4.9 billion



\$5 billion – \$999.9 million



\$10 billion or more



Source: Forbes Insights, "The Promise of Privacy: Respecting Consumers' Limits While Realizing the Marketing Benefits of Big Data," 2014  
© World Newsmedia Network 2014

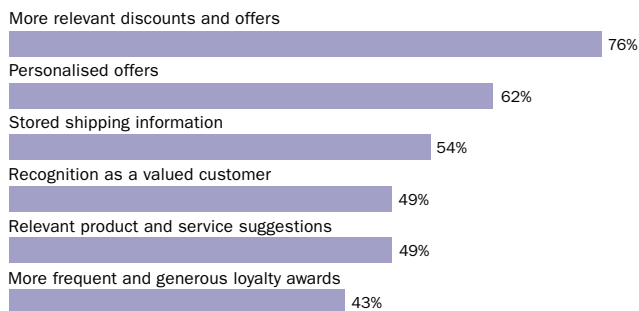
study said they were concerned about privacy issues. Thirty-one percent of business-to-business consumers surveyed were concerned about privacy issues. Meanwhile, consumers are more concerned about data security issues than businesses, according to the study.

Personal data collection has recognised benefits, according to consumers in the study. Three quarters (76 percent) of those surveyed said data collection translates to more relevant discounts and offers, while 62 percent recognise data collection makes personalised offers possible, and 54 percent say it enables stored shipping information.

Half of the respondents (49 percent) said personal data collection enables recognition as a valued customer and also enables relevant product and service suggestions, while 43 percent said this collection enables more frequent and generous loyalty awards.

## Consumers notice benefits of personal data collection

In % of responses



Source: Forbes Insights, "The Promise of Privacy: Respecting Consumers' Limits While Realizing the Marketing Benefits of Big Data," 2014  
© World Newsmedia Network 2014

Business-to-business consumers recognise more than any other group the importance of giving up personal data as a tradeoff for benefits. Eighty-seven percent of the B2B respondents said personal data collection yields better products to meet their needs, and also can yield money, while 88 percent said it can enable better services and 86 percent said it could improve timing.

Executives and consumers, meanwhile, were less convinced of the value of data collection. Sixty-six percent of the executives and 62 percent of the B2C respondents said data collection yields better products, while 64 percent of executives and 67 percent of consumers believe it yields better services.

## Importance of incentives to give up personal data

In % of responses

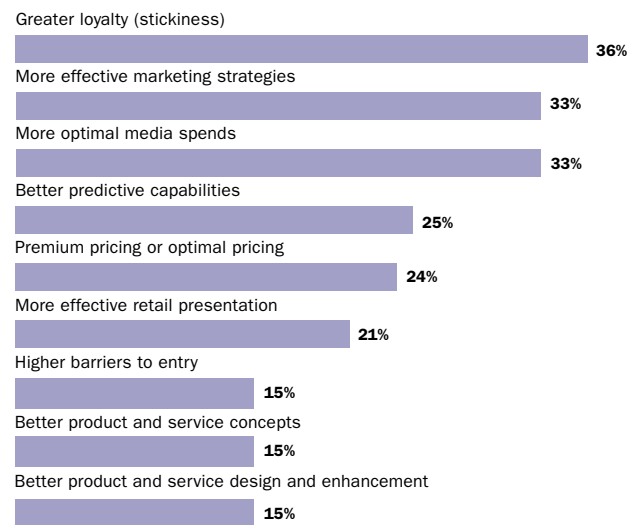
	Executives believe*	B2C values	B2B values
Better products (to meet their needs)	66%	62%	87%
Better services	64%	67%	88%
Money	60%	87%	87%
Time	57%	68%	86%
Improved retail experience	53%	68%	N/A
Improved online experience	48%	7%	71%

\*Percentage of executives believing this form of value accrues to customers  
N/A (Survey did not ask B2B about retail experience)

Source: Forbes Insights, "The Promise of Privacy: Respecting Consumers' Limits While Realizing the Marketing Benefits of Big Data," 2014  
© World Newsmedia Network 2014

## Benefits of personal data collection for company executives

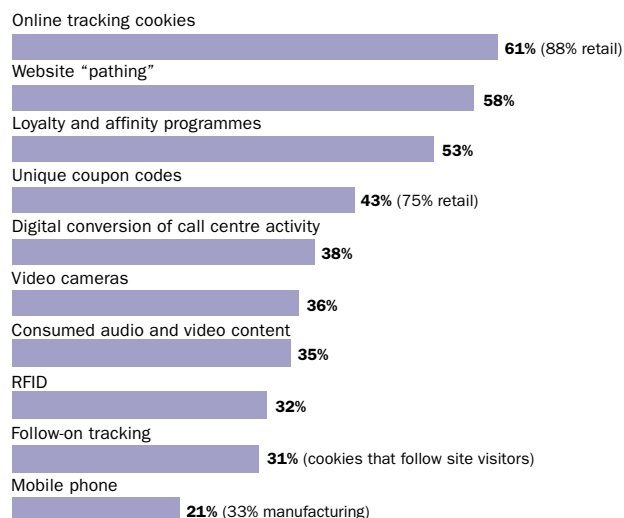
In % of responses



Source: Forbes Insights, "The Promise of Privacy: Respecting Consumers' Limits While Realizing the Marketing Benefits of Big Data," 2014  
© World Newsmedia Network 2014

## How companies collect consumer data

In % of responses from company executives



Note: 82% are also taking steps to correlate in-house data with third-party data.  
 Source: Forbes Insights, "The Promise of Privacy: Respecting Consumers' Limits While Realizing the Marketing Benefits of Big Data," 2014  
 © World Newsmedia Network 2014

What motivates companies to collect personal data? Thirty-six percent of the respondents say greater loyalty and stickiness on digital platforms, while 33 percent each say data collection makes for more effective marketing strategies and more optimal media spends. Meanwhile, 25 percent say better predictive capabilities, and 24 percent say premium pricing or optimal pricing.

Consumers reported measures they take to shield their online privacy. Almost one-third of the consumer respondents (30 percent) said they frequently decline to provide information about themselves, while 24 percent say they usually decline to provide information. Meanwhile, 29 percent said they sometimes "abort" a questionnaire when they feel they have to answer too many personal questions, while 22 percent said they always abort the questionnaire when there are too many questions.

Meanwhile, few take actions to block privacy invasions. One-third of the respondents said they rarely or never

## How B2C consumers shield online privacy

In % of responses

How often?	Always	Usually	Frequently	Sometimes	Rarely/Never
Decline to provide info	12%	24%	30%	21%	13%
"Abort" owing to too many questions	22%	16%	20%	29%	13%
Delete cookies to "hide"	11%	17%	21%	20%	31%
Use ad-blocker software	23%	20%	15%	11%	31%

Source: Forbes Insights, "The Promise of Privacy: Respecting Consumers' Limits While Realizing the Marketing Benefits of Big Data," 2014  
 © World Newsmedia Network 2014

delete cookies to "hide," while 21 percent said they delete cookies frequently and 20 percent said sometimes. Twenty-three percent said they always use ad-blocker software, which collects personal data, while 31 percent said they rarely or never do so.

Companies collect personal data of their customers in multiple ways. The most popular way is through online tracking cookies, with 61 percent of companies overall, and 88 percent of retail companies, saying they do so. Fifty-eight percent of the company respondents said website "pathing," or the monitoring of how consumers navigate through websites during each session, is their second most important way to collect personal data.

Loyalty and affinity programmes are utilised for data collection by 53 percent of the respondents, while 43 percent of the overall respondents and 75 percent of the retail survey respondents use unique coupon codes. For media companies, tracking content is an important practice. Thirty-five percent of the company executives said they collect data on consumed audio and video content.

Company websites are the most important places to collect data, according to the survey, with 65 percent of the executives saying so, followed by 60 percent for call centres, 53 percent in person, 50 percent on mobile, 49 percent from loyalty card activity, 46 percent on social media, and 44 percent each for direct mail and affinity programmes.

There is a disconnect between the percentage of company respondents who say they are observing customer interactions and how customers perceive they are being observed, according to the study. While 65 percent of business executives surveyed said they were likely or highly likely to observe website behaviours of consumers, 84 percent of the B2C and 53 percent of the B2B consumers perceived they were being observed on the websites.

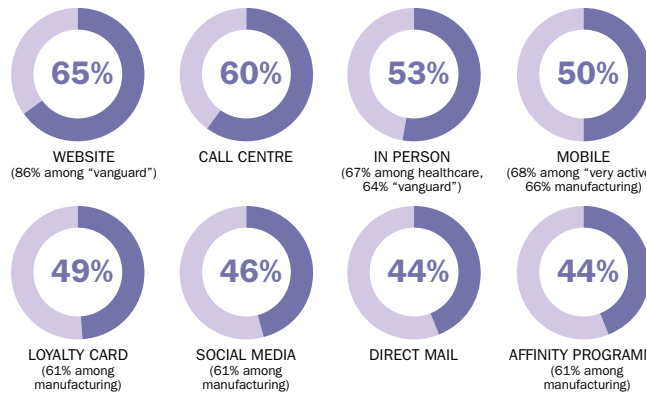
Fifty-three percent of business executives say consumers are being observed in person, while 47 percent of the B2C and 29 percent of the B2B consumers believe it is likely or very likely they are being observed in person. Perhaps the biggest disconnect is between consumers and companies regarding loyalty card observations. Seventy-three percent of B2C consumers perceive they

are likely or very likely to be observed, while 49 percent of executives say they are actually being observed. The study concludes that companies "need to do a better job of disclosing when, where and how they are tracking customers."



## Where consumer data is collected

In % of responses from business executives



Source: Forbes Insights, "The Promise of Privacy: Respecting Consumers' Limits While Realizing the Marketing Benefits of Big Data," 2014  
© World Newsmedia Network 2014

Company executive survey respondents report they are furthest along with using customer data to form a single view of the customer, but still have a long way to go to use the data to inform media spending and product and service marketing. According to the study, 88 percent of the respondents say they are achieving significant success with using customer data to form a single view of customers, but only 67 percent said they are successful in using customer data to inform

marketing and 64 percent said they are successful in using the data to inform media spending.

Meanwhile, 87 percent of the respondents say they are seeing business benefits both from informing marketing and from forming a single view of the customer, while only 60 percent say this is impacting media spending knowledge. Meanwhile, a significant percentage of the respondents say they have no plans in mind for forming a single view of the customer, 78 percent; informing marketing, 71 percent; and informing media spending, 48 percent.

## Information Week study

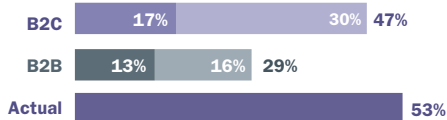
According to some of the most forward-thinking experts specialising in Big Data structure, such as Gartner and IBM, a data governance plan must be put into place in order for Big Data strategies to be successful. Data governance plans comprise the technical and security policies and procedures that become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

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. According to InformationWeek's 2013 survey on Big Data, budget constraints are the top

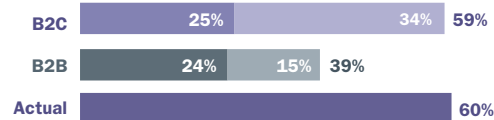
## Consumers' perceptions of being observed vs. reality

In % of responses

### IN PERSON



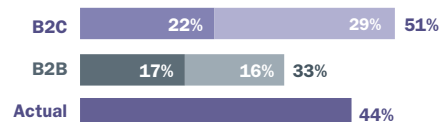
### CALL CENTRE



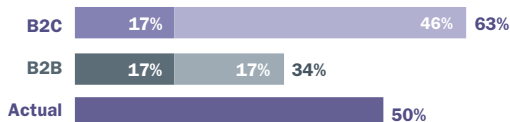
### WEBSITE



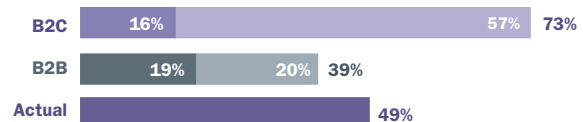
### DIRECT MAIL



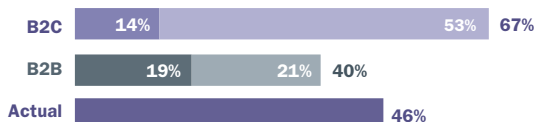
### MOBILE



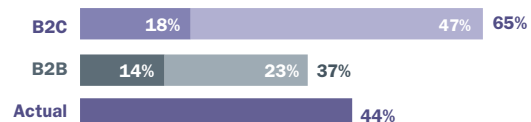
### LOYALTY CARD



### SOCIAL MEDIA



### AFFINITY CARD



Likely  
Highly likely

Source: Forbes Insights, "The Promise of Privacy: Respecting Consumers' Limits While Realizing the Marketing Benefits of Big Data," 2014  
© World Newsmedia Network 2014

## How far companies are in customer data journey

In % of responses

	Inform media spending	Inform marketing	Form a "single view" of customers
Achieving significant success	64%	67%	88%
Seeing business benefits	60%	87%	87%
Off to a good start	57%	68%	86%
See potential benefit but have yet to begin	53%	68%	N/A
Have no plans in mind	48%	71%	78%

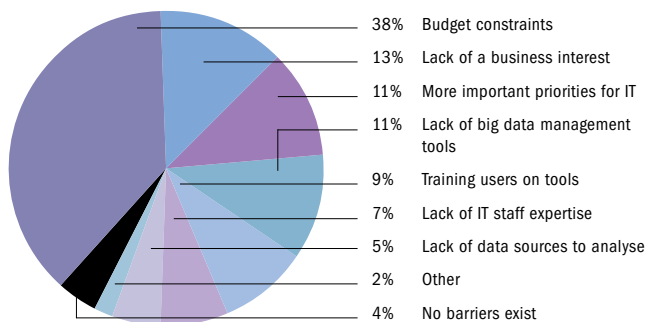
Source: ForbesInsights, "The Promise of Privacy: Respecting Consumers' Limits While Realising the Marketing Benefits of Big Data," 2014  
© World Newsmedia Network 2014

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). More than 40 percent of the respondents called their data analysis practices either "limited" or "abacus-like."

### Top barriers to successful use of Big Data

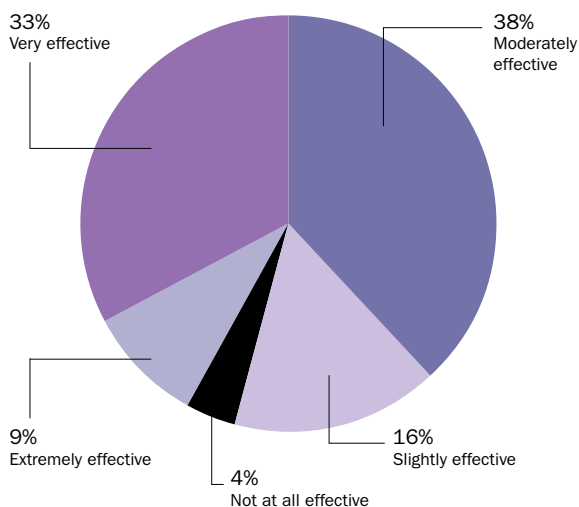
In percentage of respondents



Source: Information Week, "Blinded by Big Data," 2013  
© World Newsmedia Network 2014

### Critical data for decision-making

The degree to which organisations are effective at identifying critical data and using it for decisions



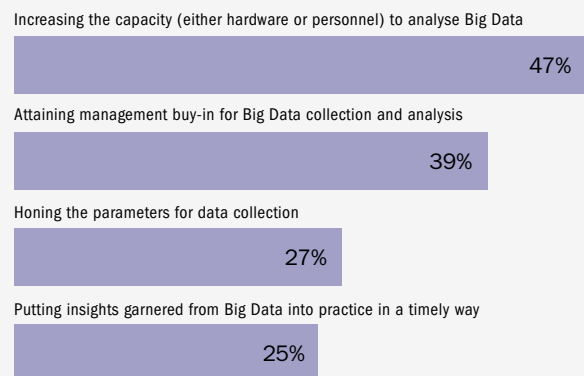
Source: Information Week, "Blinded by Big Data," 2013  
© World Newsmedia Network 2014

## KMPG STUDY

Developing "actionable" insights about the reams of Big Data generated at companies is a resounding theme in the many studies published in 2013 and 2014. It's the logical next step to gathering stockpiles of data across business enterprises. The No. 1 strategy made in response to Big Data collection is to increase the capacity of hardware or personnel in order to analyse Big Data (47 percent), according to KPMG's "Going Beyond the Data: Achieving Actionable Insights with Data and Analytics." The survey of CFOs and CIOs also found that attaining management buy-in for Big Data collection (39 percent) and honing the parameters of data collection (27 percent) were top Big Data strategies.

### Business strategies made in response to Big Data

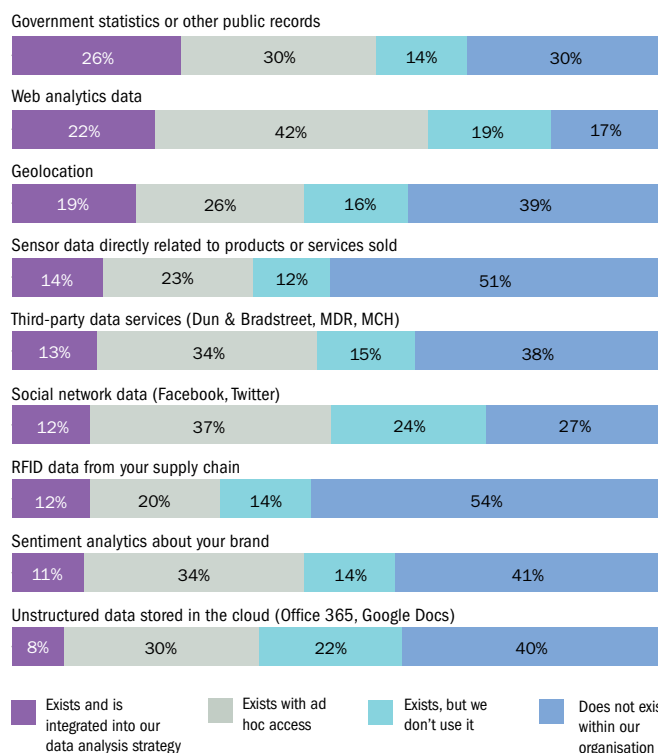
% of global CFO and CIO respondents



Source: KPMG, "Going Beyond the Data: Achieving Actionable Insights with Data and Analytics," January 2014, eMarketer  
© World Newsmedia Network 2014

## External Big Data sources

The degree to which external data sources are used as part of the respondents' data analysis and reporting programmes



Source: Information Week, "Blinded by Big Data," 2013  
© World Newsmedia Network 2014

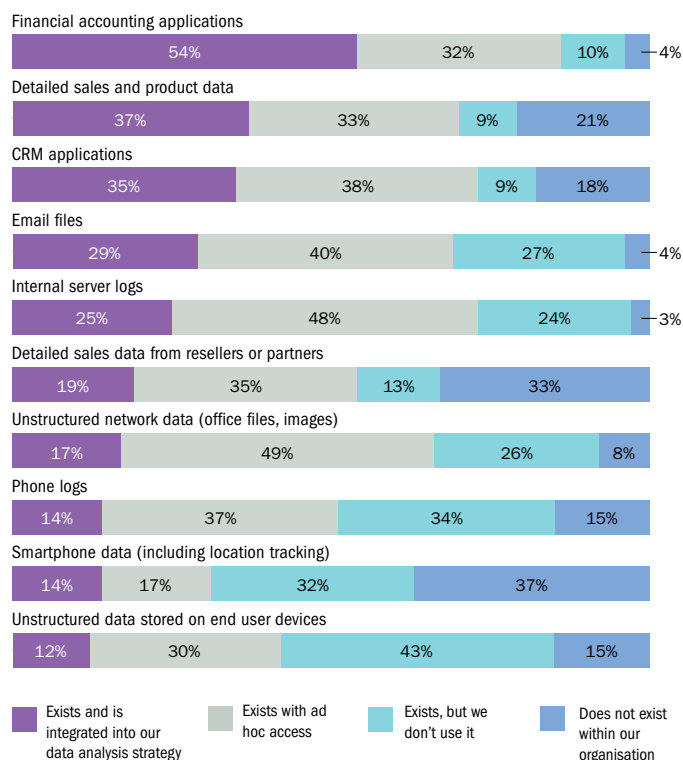
One of the key tools for assessing and leveraging Big Data is a data dashboard that could display data trends from a variety of company sectors, such as CRM, finance, online audience, sales and social media sentiment. The most popular application of a data dashboard is an audience usage dashboard. Ninety 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' 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. 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.

More companies are marrying their own data with that

## Internal Big Data sources

The degree to which internal data sources are used as part of the respondents' data analysis and reporting programmes



Source: Information Week, "Blinded by Big Data," 2013  
© World Newsmedia Network 2014

of external Big Data sources, according to the survey. In particular, governmental statistics and public records (26 percent); Web analytics data (22 percent) and geolocation data (19 percent) are some of the most popular data to be currently integrated into the respondents' data practices, according to the study.

Among the most applied internally available data used in respondents' Big Data strategies are financial accounting applications (54 percent); detailed sales and product data (37 percent); CRM applications (35 percent); email files (29 percent); and internal server logs (25 percent).

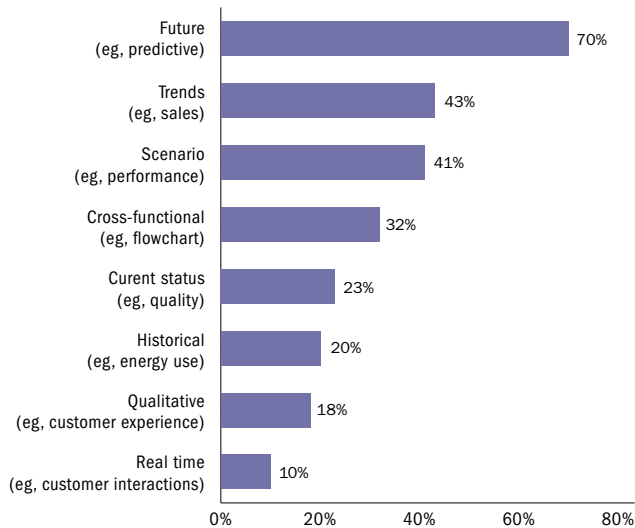
## The Economist Intelligence Unit study

The Economist Intelligence Unit conducted a global survey of 373 executives and in-depth interviews with thought leaders about Big Data strategies in 2013. The study found that the majority of these C-level respondents, 70 percent, look to Big Data analysis to be predictive in nature, in order to help them anticipate where their markets are headed and to inspire them to position themselves for growth.

Meanwhile, the second and third most important advantages to data strategies focused on trend data and

## Types of data insights critical to decision-making

% of C-level executive respondents in global survey



Source: The Economist Intelligence Unit, "Getting More Out of Big Data," 2013  
© World Newsmedia Network 2014

analysis, such as sales trends, 43 percent; and scenario-based data analysis, such as performance data, 41 percent.

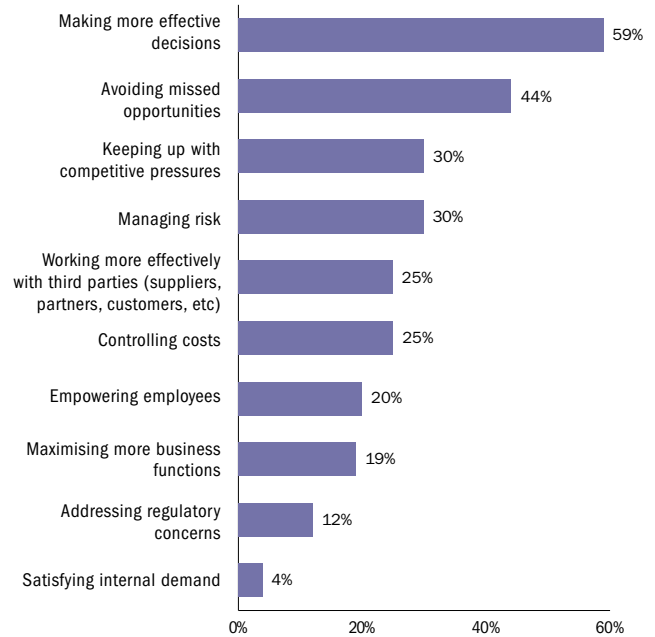
The study advises companies embarking on Big Data strategies that they must ask better questions about their objectives in order to achieve desirable outcomes. The most important perceived outcome is to make more effective decisions, 59 percent; avoid missed opportunities, 44 percent; keep up with competitive pressures and manage risk, at 30 percent each; work more effectively with third parties and control costs, at 25 percent each; and empower employees, 20 percent.

More than half of the respondents, 51 percent, say the quality, reliability or comprehensiveness of the data is the top data challenge in business today, followed by lack of effective systems to gather and analyze data, 43 percent; and lack of skills needed to interpret the data, 29 percent. Other key Big Data challenges for the business executives surveyed include lack of widespread understanding of what the data is used for, 28 percent; concern over disclosure of confidential or sensitive corporate information, 23 percent; and inability to track the impact of data, 21 percent.

The Economist's study broke down responses between strong performing companies and average/weak performing companies, and determined that strong performing companies are far more likely to have mapped out their Big Data strategies compared to weaker performers, at 55 percent compared to 35 percent; and that their data strategies are aligned with their overall business strategies, at 54 percent compared to 40 percent. Further, stronger performing companies

## Perceived business benefits of employing Big Data initiatives

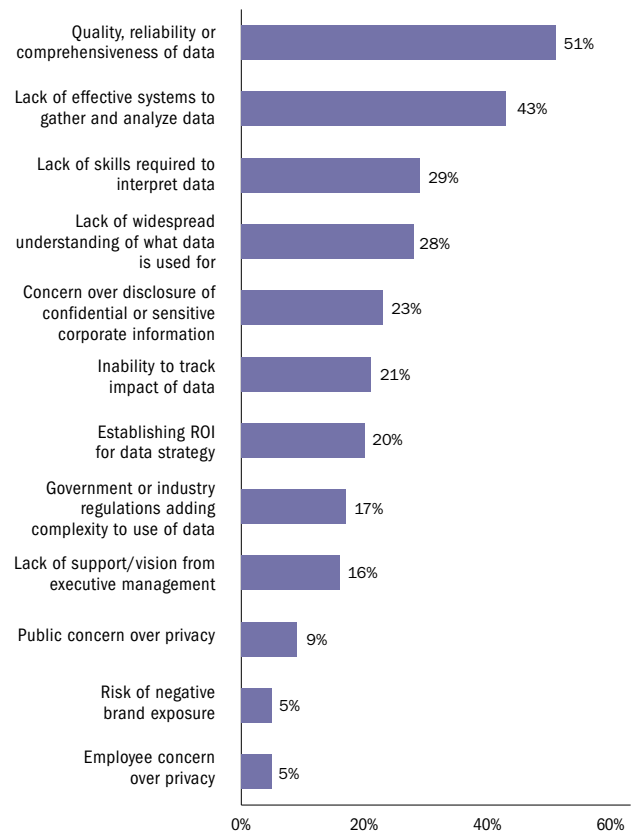
% of respondents in global survey



Source: The Economist Intelligence Unit, "Getting More Out of Big Data," 2013  
© World Newsmedia Network 2014

## Key data challenges in businesses globally

% of respondents in global survey

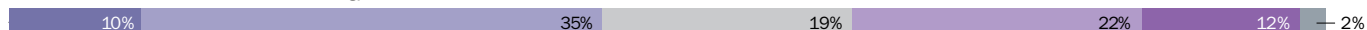


Source: The Economist Intelligence Unit, "Getting More Out of Big Data," 2013  
© World Newsmedia Network 2014

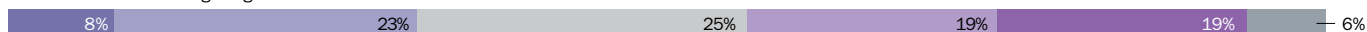
## Executives define business' data strategy

% respondents in global survey choosing range from 'strongly agree' to 'strongly disagree'

We have a defined, documented data strategy



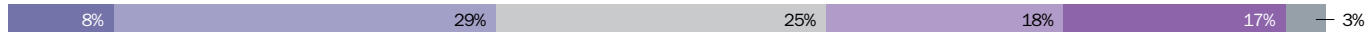
Our investment in deriving insight from data has increased more than 25% in 2012



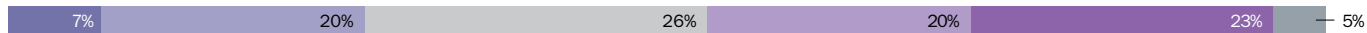
It is likely we will increase our data investment by 25% over the next three years



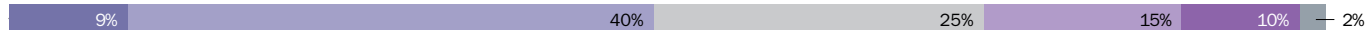
We have defined Key Performance Indicators (KPIs) for data efforts



We track the ROI of our data investments



We have incorporated insights from data business efforts into other business strategies



We are confident we understand the impact data will have on our organisation over the next three years



Our data strategy is aligned with our overall business strategy



We are prepared for the changes that will be needed to succeed in applying data to business

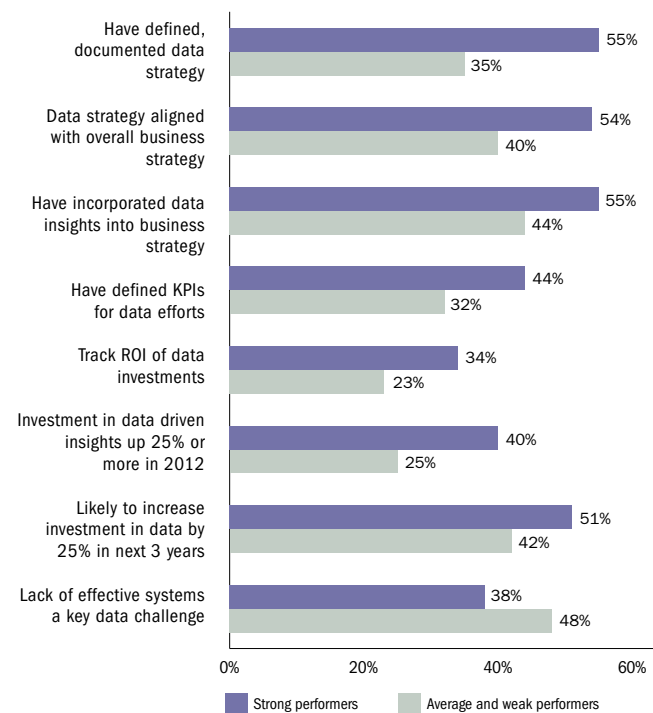


Strongly agree Somewhat agree Neither agree nor disagree Somewhat disagree Strongly disagree Not applicable

Source: The Economist Intelligence Unit  
© World Newsmedia Network 2014

## The state of businesses' data strategies

% of respondents in global survey agreeing



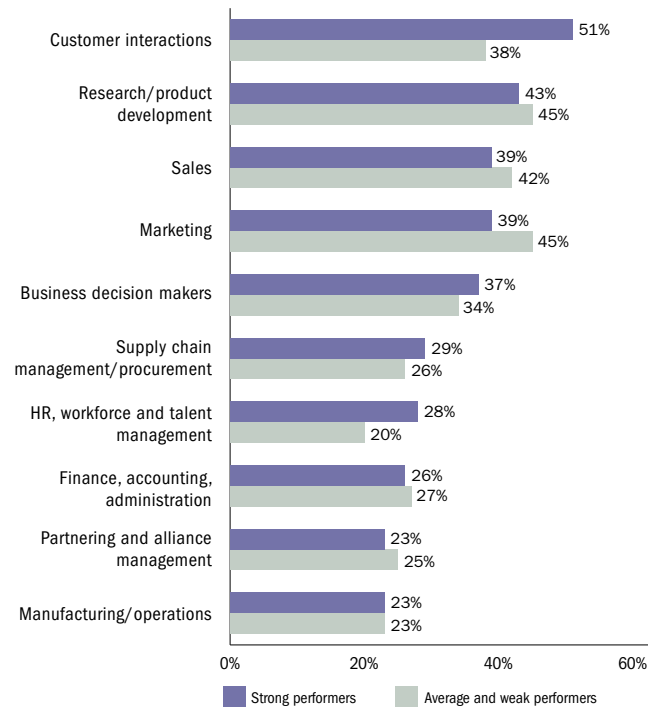
Source: The Economist Intelligence Unit, "Getting More Out of Big Data," 2013  
© World Newsmedia Network 2014

are more likely to have incorporated insights from their Big Data strategies into their business strategies, at 55 percent, compared to 44 percent for companies deemed average or weak performers.

The business areas most likely to receive investment to build Big Data capabilities also differ between strong-

## Business departments to receive data capabilities

% respondents in global survey selecting



Source: The Economist Intelligence Unit, "Getting More Out of Big Data," 2013  
© World Newsmedia Network 2014

performing companies and average or weak performers, according to the survey. Fifty-one percent of the strong performers surveyed said that those specialising in customer interactions receive data capabilities, and while only 38 percent of average or underperforming companies said so.



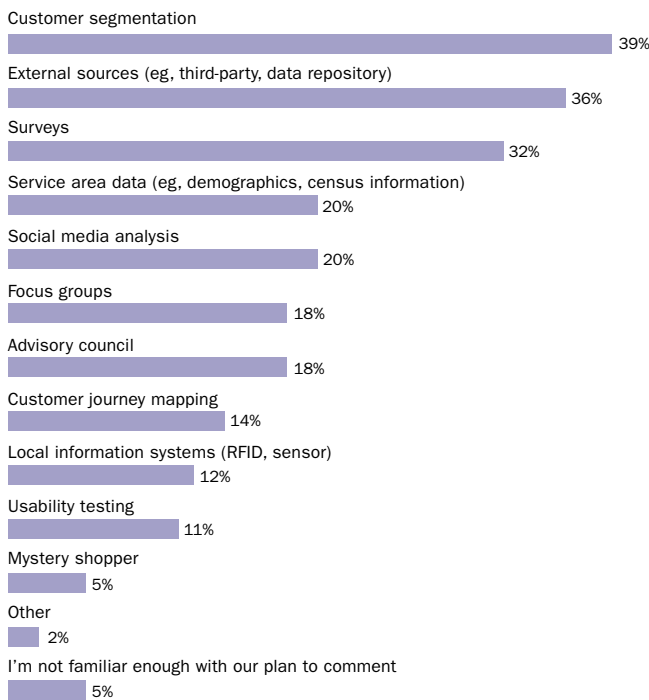
Research and product development was the second most important recipient of innovative data capabilities, with 43 percent of the strong performers and 45 percent of the weaker performers saying so. Sales and marketing were the next most popular recipients of Big Data focus, with 39 percent of the respondents saying so for each category, and with 42 percent for sales and 45 percent for marketing for the weaker-performing companies saying so.

Only 10 percent of the respondents strongly agree they have a defined, documented Big Data strategy, while 35 percent say they somewhat agree and 12 percent say they strongly disagree. Meanwhile, almost one-third strongly agree or somewhat agree that their company invested 25 percent more in 2012, while 38 percent either somewhat disagree or strongly disagree. However, 44 percent project that their Big Data strategy investment will increase 25 percent over the next three years, while 26 percent either somewhat disagree or strongly disagree.

According to the Economist survey, the most popular plans for deriving insights from data include customer segmentation, 39 percent; external sources such as third-party data providers, 36 percent; surveys, 32 percent; census data, 20 percent; social media analysis, 20 percent; and focus groups or advisory councils, 18 percent each.

### Methods for developing insight from Big Data

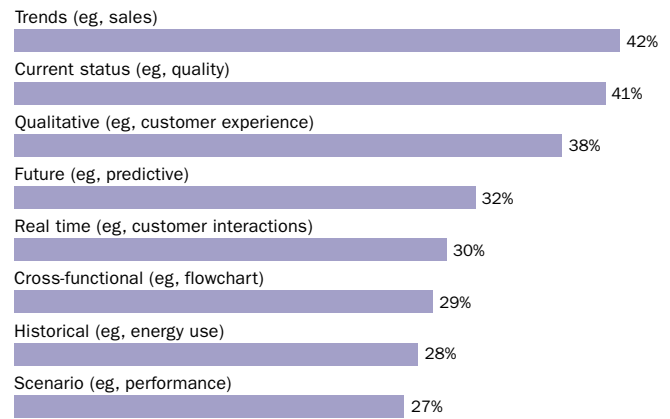
% respondents in global survey selecting up to three methods



Source: The Economist Intelligence Unit, "Getting More Out of Big Data," 2013 © World Newsmedia Network 2014

### Top data insights critical to decision-making

% respondents in global survey

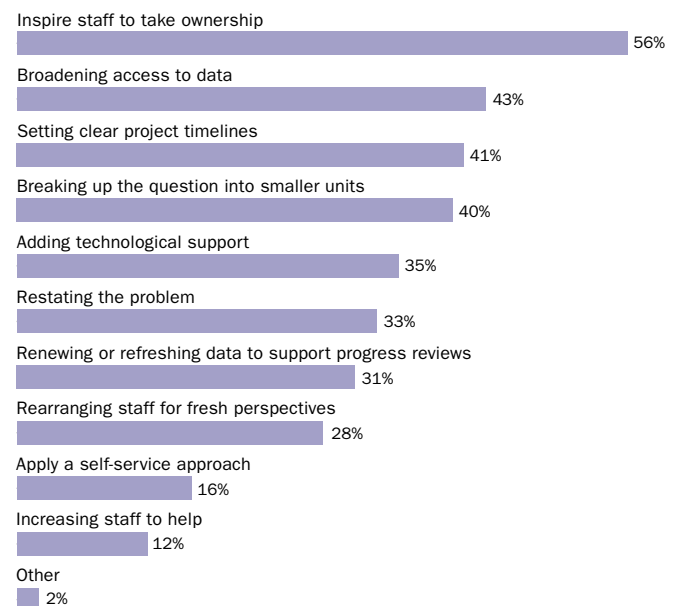


Source: The Economist Intelligence Unit, "Getting More Out of Big Data," 2013 © World Newsmedia Network 2014

The top data insights critical to decision-making, according to the survey, are: trends, such as sales, 42 percent; current status, such as quality, 41 percent; qualitative, such as customer experience, 38 percent; future, such as predictive, 32 percent; and real-time, such as customer interactions, 30 percent.

### Processes needed to convert data strategies to reality

% respondents in global survey selecting all that apply



Source: The Economist Intelligence Unit, "Getting More Out of Big Data," 2013 © World Newsmedia Network 2014

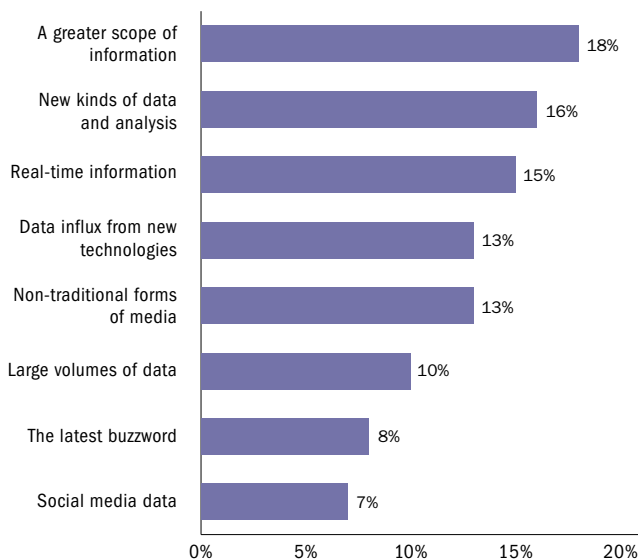
The processes that need to be implemented in order to convert Big Data strategies to reality, according to respondents, include inspiring staff to take ownership, 56 percent; broadening access to data, 43 percent; setting clear project timelines, 41 percent; breaking up the question into smaller units, 40 percent; and adding technological support, 35 percent.

## IBM, Said Business School study

IBM and the University of Oxford's Said Business School conducted a survey of 1,144 business people from 95 countries across 26 industries to learn about their strategies and opinions about the potential for Big Data in their businesses.

As with all major strategies on the ascendancy, the concept of Big Data may have as many definitions as there are those who attempt to define it. The survey makers narrowed down the descriptions to determine general definitions of Big Data and found that respondents, who were asked to choose two descriptors that best defined the concept, said Big Data was "a greater scope of information" (18 percent), "new kinds of data and analysis" (16 percent), "real-time information" (15 percent), and "data influx from new technologies" (13 percent) and "non-traditional forms of media" (13 percent).

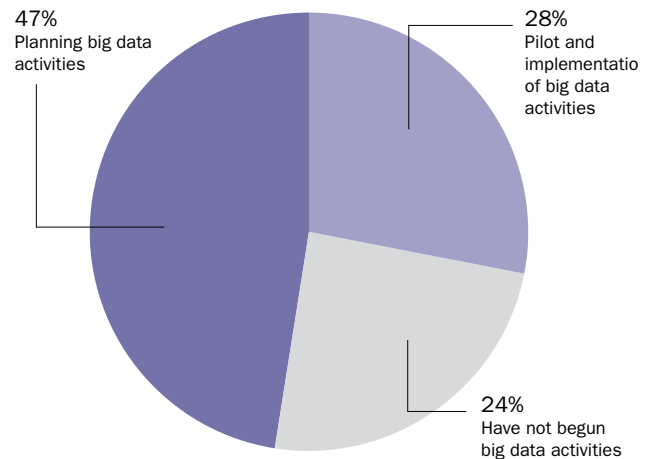
### Company executives define Big Data



Source: "Analytics: The real-world use of big data," a collaborative research study by the IBM Institute for Business Value and the Said Business School at the University of Oxford. © IBM 2012 © World Newsmedia Network 2014

The study corroborates other Big Data studies, showing that the majority of companies surveyed are still in the beginning stages of Big Data development, focused on understanding basic concepts or defining a roadmap for the future. Across the industries studied, the business case is strongly focused on addressing customer-centric objectives, and beginning pilot programmes for Big Data project implementation. The state of Big Data activity among respondents can be broken down into three areas, led by planning Big Data activities, 47 percent; pilot and implementation of Big Data activities, 28 percent; and those that have not begun Big Data activities, 24 percent, according to the survey.

### Company executives describe Big Data activity

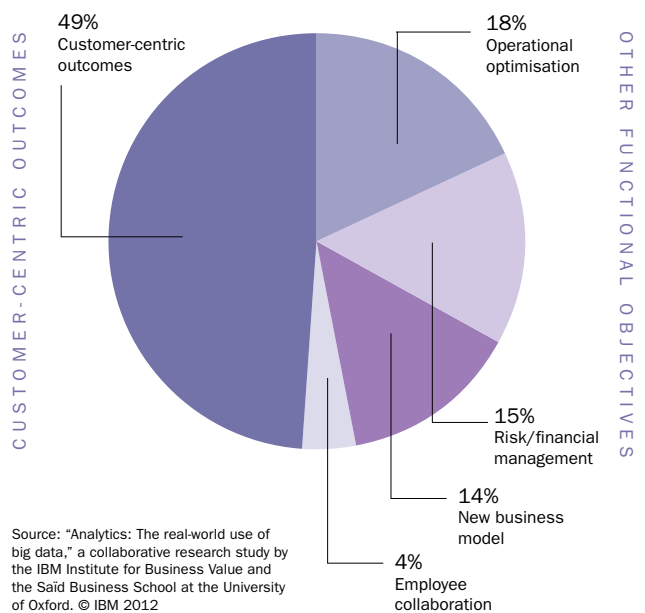


Source: "Analytics: The real-world use of big data," a collaborative research study by the IBM Institute for Business Value and the Said Business School at the University of Oxford. © IBM 2012 © World Newsmedia Network 2014

Respondents were asked about their expectations for Big Data's business-driven outcomes. Customer-centric outcomes (49 percent) are the most popular among survey responses, including the huge category of customer analytics. For example, customer analytics is a popular application of Big Data across industries, and particularly relevant and powerful for media companies' digital media properties.

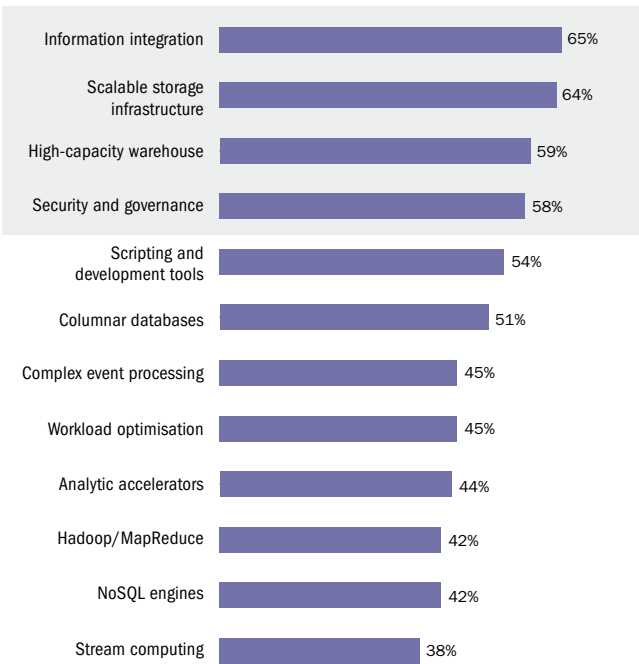
After customer-centric outcomes, respondents were divided over other Big Data outcomes. Eighteen percent said operational optimisation; 15 percent said risk/financial management; and 14 percent said new business model. Just 4 percent said employee collaboration.

### Company executives describe Big Data business-driven outcomes



Source: "Analytics: The real-world use of big data," a collaborative research study by the IBM Institute for Business Value and the Said Business School at the University of Oxford. © IBM 2012 © World Newsmedia Network 2014

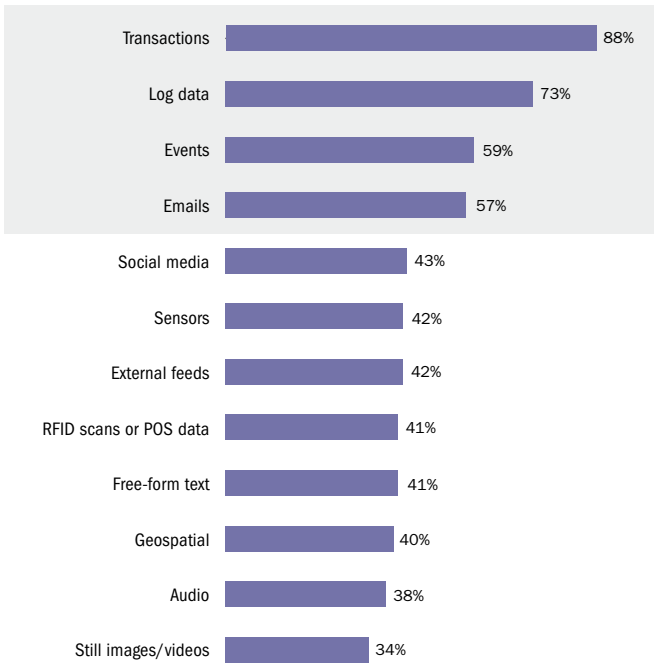
## Company executives describe Big Data infrastructure



Source: "Analytics: The real-world use of big data," a collaborative research study by the IBM Institute for Business Value and the Saïd Business School at the University of Oxford. © IBM 2012 © World Newsmedia Network 2014

Big Data architecture is the bedrock of a Big Data strategy. "The promise of achieving significant, measurable business value from Big Data can only be realised if organisations put into place an information foundation that supports the rapidly growing volume, variety and velocity of data," the report states.

## Company executives describe Big Data sources



Source: "Analytics: The real-world use of big data," a collaborative research study by the IBM Institute for Business Value and the Saïd Business School at the University of Oxford. © IBM 2012 © World Newsmedia Network 2014

Respondents rated integrated information as the most important part of Big Data architecture, at 65 percent; followed by scalable storage infrastructure, 64 percent; high-capacity warehouse, 59 percent; security and governance, 58 percent; and scripting and development tools, 54 percent.

## Approximate multimedia messaging file sizes, by type

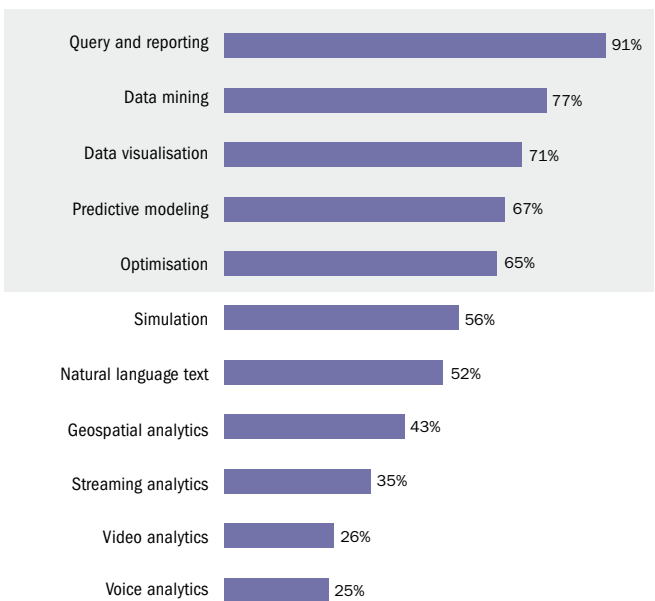
Type of MIM message	Approximate size (in KB)*
Text-only MIM (approximately 150 characters)	10
Photo	100
Audio file (one minute long)	150
Video file (one minute long)	12,000

\* File sizes are considerably compressed when sent via an MIM application and will not reflect its actual size.

Source: Deloitte, Technology, Media & Telecommunications Predictions 2014 © World Newsmedia Network 2014

File sizes for each type of multimedia varies considerably, and impacts many phone and smartphone users around the world who pay by file size rather than an "all you can consume" data package. These types of plans will severely impact the growth of multimedia smartphone usage around the world. However, messaging services like the Facebook-owned WhatsApp make access to multimedia more seamless and less expensive, at just US\$1 per year in some places, such as India.

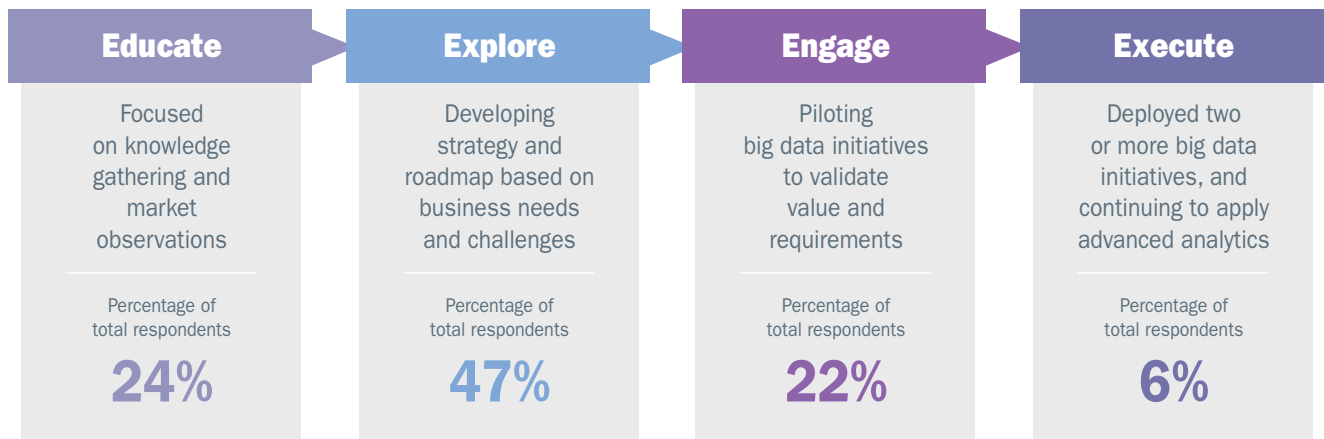
## Company executives describe Big Data capabilities



Source: "Analytics: The real-world use of big data," a collaborative research study by the IBM Institute for Business Value and the Saïd Business School at the University of Oxford. © IBM 2012 © World Newsmedia Network 2014

## Big Data adoption stages

Based on global survey of company executives



Source: "Analytics: The real-world use of big data," a collaborative research study by the IBM Institute for Business Value and the Said Business School at the University of Oxford. © IBM 2012 © World Newsmedia Network 2014

Most early data efforts are targeted at sourcing and analysing internal data, according to the study. The top Big Data projects executed by the respondents were transactions, 88 percent; log data, 73 percent; events, 59 percent; emails, 57 percent; social media, 43 percent; sensors, 42 percent; and external feeds, 42 percent. Most of the data collected, however, goes unanalysed, according to the report.

Respondents with active Big Data strategies were asked about the analytics capabilities available at their companies. Most companies surveyed were directing their efforts toward analysing structured data in organised databases. Analysing unstructured data such as streaming Twitter data, geo-location data, voice and video, continues to be a challenge for most companies, and especially media companies, since much of the data they produce is unstructured. Ninety-one percent of the respondents said query and reporting; 91 percent, data mining; 77 percent, data visualisation; 71 percent, predictive modeling, 67

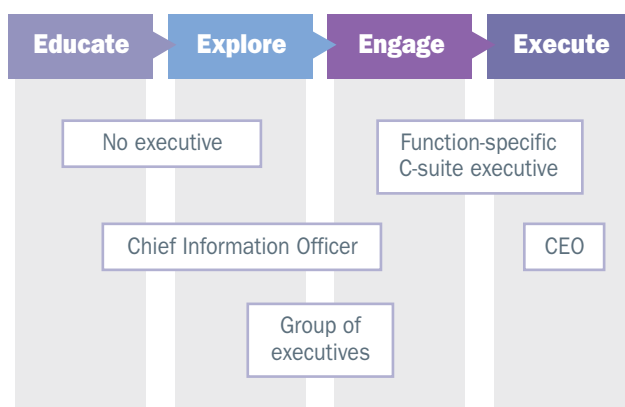
percent; and optimisation, 65 percent.

The IBM/Saïd Business School survey analysis of responses inspired researchers to break down the adoption stages for Big Data strategies. According to the study, 24 percent of the respondents are in the education stage, building a body of knowledge about Big Data and its capabilities for their organisations. Almost half, or 47 percent, are in the exploratory phase, developing their strategies and identifying business needs and challenges that could be addressed by a Big Data implementation. Meanwhile, 22 percent of the respondents are in the engagement stage, piloting Big Data initiatives in order to identify the value of various strategies, and their business-related requirements. Only 6 percent of the respondents were in the execution phase, implementing various Big Data projects, particularly customer analytics projects.

As the process of Big Data strategy implementation morphs from education, to exploring, to engagement, to execution,

## Big Data adoption sponsorship

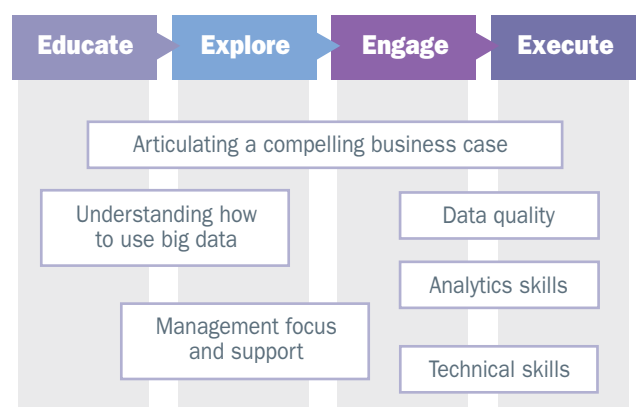
Based on global survey of company executives



Source: "Analytics: The real-world use of big data," a collaborative research study by the IBM Institute for Business Value and the Said Business School at the University of Oxford. © IBM 2012 © World Newsmedia Network 2014

## Big Data primary obstacles

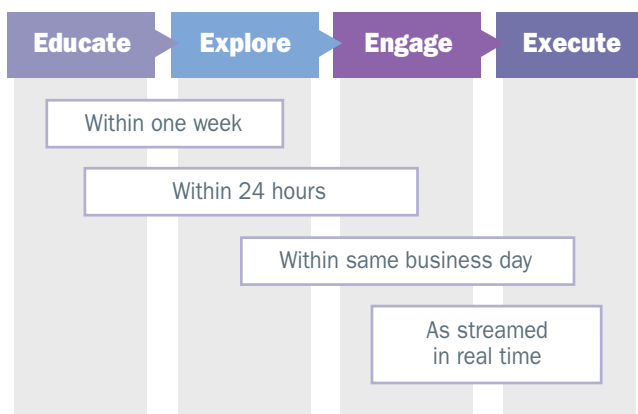
Based on global survey of company executives



Source: "Analytics: The real-world use of big data," a collaborative research study by the IBM Institute for Business Value and the Said Business School at the University of Oxford. © IBM 2012 © World Newsmedia Network 2014

## Big Data availability

Based on global survey of company executives



Source: "Analytics: The real-world use of big data," a collaborative research study by the IBM Institute for Business Value and the Saïd Business School at the University of Oxford. © IBM 2014 © World Newsmedia Network 2014

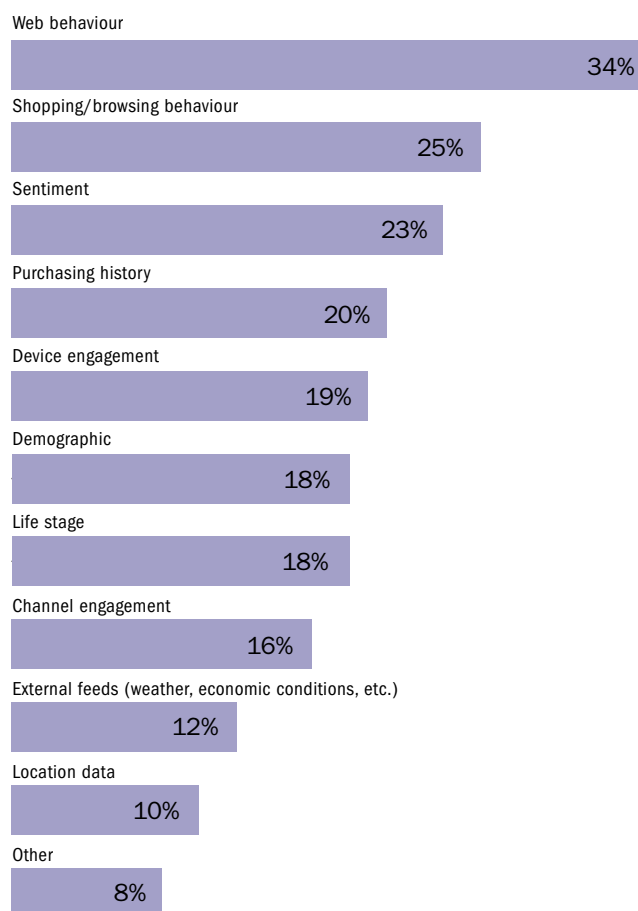
so too does the level of authority that approves it and shepherds it. Lower-level employees often can explore and educate themselves about Big Data opportunities, but any investment oftentimes comes from the chief information officer. Once a formal strategy is adopted, most organisations would begin an engagement phase, authorised and supervised by function-specific C-level executives, while the execution of Big Data initiatives is most often approved and sponsored by the CEO.

Like other strategies, Big Data initiatives encounter obstacles on their journey from the education to execution phases. According to the survey, throughout the stages, the ability to articulate a compelling business case is critical, including clear financial benefits, substantial consumer insights and measurable outcomes.

Respondents were asked how quickly data is required to be available for business users or processes. As the Big Data strategy progresses from education to execution, the imperative for quicker results is apparent, starting from "within one week" during the education phase, to "streamed in real time" during the execution phase.

## Inability to leverage Big Data insights in email marketing

% of respondents reporting disconnect between available data and inability to make it actionable in an email campaign



Source: StrongView, "2014 Marketing Trends Survey," December 2013, eMarketer © World Newsmedia Network 2014

Leveraging Big Data for actionable insights isn't always as straightforward as it sounds. According to StrongView's "2014 Marketing Trends Survey," a significant number of marketers surveyed found leveraging Big Data in their email campaigns to be challenging. One-third of the respondents said they reported a disconnect between the collection of Web behaviour data and making it actionable in an email campaign, while one-fourth found it a challenge to leverage shopping and buying behaviour in a campaign. Data collected about sentiment (23 percent), purchasing history (20 percent) and device engagement (19 percent) also proved to be challenging to leverage in the marketers' email campaigns.



## **About the author**

Martha L. Stone is a research associate at the Reuters Institute for the Study of Journalism at the University of Oxford from 2012 to 2014, specialising in the study of Big Data for Media topics, such as media practises, data journalism, Big Data strategy development and media organisations' cultural change issues. Stone has led two Big Data for Media conferences in London in 2013 and 2014; the first conferences in the world developed solely on the subject of Big Data.

## SELECTED RISJ PUBLICATIONS

Wendy N. Wyatt (ed.)

**The Ethics of Journalism: Individual, Institutional and Cultural Influences**

*(published jointly with I.B.Tauris)*

Raymond Kuhn and Rasmus Kleis Nielsen (eds)

**Political Journalism in Transition: Western Europe in a Comparative Perspective**

*(published jointly with I.B.Tauris)*

Nigel Bowles, James T. Hamilton, David A. L. Levy (eds)

**Transparency in Politics and the Media: Accountability and Open Government**

*(published jointly with I.B.Tauris)*

Julian Petley (ed.)

**Media and Public Shaming: Drawing the Boundaries of Disclosure**

*(published jointly with I.B.Tauris)*

James Painter

**Poles Apart: The International Reporting of Climate Scepticism**

Lara Fielden

**Regulating for Trust in Journalism: Standards Regulation in the Age of Blended Media**

David A. L. Levy and Robert G. Picard (eds)

**Is there a Better Structure for News Providers? The Potential in Charitable and Trust Ownership**

David A. L. Levy and Rasmus Kleis Nielsen (eds)

**The Changing Business of Journalism and its Implications for Democracy**

Tim Gardam and David A. L. Levy (eds)

**The Price of Plurality: Choice, Diversity, and Broadcasting Institutions in the Digital Age**

*published in association with Ofcom*

## CHALLENGES

John Lloyd and Cristina Marconi

**Reporting the EU: News, Media and the European Institutions**

*(published jointly with I.B.Tauris)*

James Painter

**Climate Change in the Media: Reporting Risk and Uncertainty**

*(published jointly with I.B.Tauris)*

Suzanne Franks

**Women and Journalism**

*(published jointly with I.B.Tauris)*

Naomi Sakr

**Transformations in Egyptian Journalism**

*(published jointly with I.B.Tauris)*

Nick Fraser

**Why Documentaries Matter**

Nicola Bruno and Rasmus Kleis Nielsen

**Survival is Success: Journalistic Online Start-ups in Western Europe**

John Lloyd

**Scandal! News International and the Rights of Journalism**

Richard Sambrook

**Are Foreign Correspondents Redundant? The Changing Face of International News**

James Painter

**Summoned by Science: Reporting Climate Change at Copenhagen and Beyond**

John Kelly

**Red Kayaks and Hidden Gold: The Rise, Challenges and Value of Citizen Journalism**

Stephen Whittle and Glenda Cooper

**Privacy, Probity, and Public Interest**

Stephen Coleman, Scott Anthony, and David E Morrison

**Public Trust in the News: A Constructivist Study of the Social Life of the News**

Nik Gowing

**'Skyful of Lies' and Black Swans: The New Tyranny of Shifting Information Power in Crises**

Andrew Currah

**What's Happening to Our News: An Investigation into the Likely Impact of the Digital Revolution on the Economics of News Publishing in the UK**