cre
council for research excellenc

A Primer for Defining and Implementing Big Data in the Marketing and Advertising Industry

Prepared for the Council for Research Excellence by
Gerard Broussard, Principal, Pre-Meditated Media
October 23, 2014

TABLE OF CONTENTS

FOREWORD

The practice of audience measurement and, more broadly, media research, is transforming rapidly. Not too many years ago we worked in a world where data were scarce and our capacity to extract useful information was limited. Today, data are growing ever more plentiful and our technology to mine those data is mind numbing. We are limited now mostly by our skills and imagination. At every media and marketing company, data resources are becoming a critical strategic asset. Where do marketers, media owners and advertising service firms begin in transitioning their businesses for the "big data" world? What critical skills do traditional researchers contribute to this data-rich landscape?

The Big Data Committee of the Council for Research Excellence commissioned Premeditated Media to produce this primer. It is intended to help orient advertising and media executives around the opportunities and challenges that present themselves as we go from megabytes of data on PC's to petabytes of data in the cloud.

It is always the challenge of a primer as to how much reader expertise should be presumed. This primer is not written for budding data scientists, but for practitioners who are competent in current standard research practices and who want to upgrade their knowledge and skills. Readers will engage a new vocabulary, learn about new applications, and discover resources for continued learning.
It is the intention of the CRE that this document will grow in wiki-fashion to keep the industry updated. If you are interested in joining the learning community that will take on this task, please let us know at info@researchexcellence.com

Stacey Lynn Schulman
Chair, Big Data Committee
Council for Research Excellence

Richard Zackon
Facilitator
Council for Research Excellence

CURRENT MEMBERS OF CRE BIG DATA COMMITTEE

Stacey Schulman, Chair	Katz Media Group
Michele Buslik	Assembly
Jon Cogan	Annalect
Laura Cowan	Bridge Member
Sam Garfield	Discovery
Janice Finkel-Greene	Magna Global
Paul Hockenbury	Comcast
Annette Malave	RAB
Michele Meyer	Univision
Dan Murphy	Univision
Pete Doe	Nielsen
Rosemary Scott	R Communications
Ceril Shagrin	Univision
Howard Shimmel	Turner
Richard Zackon	Facilitator

EXECUTIVE SUMMARY

Big Data Primer for Industry Clarity. Big Data has rapidly emerged as a promising way to generate new and exciting consumer and business insights that lead to superior marketing and advertising investment decisions. But the Big Data concept is often amorphous, lacking clarity and shape regarding its application in organizations.

This document serves as a Big Data primer for the marketing and advertising industry so that readers can get up to speed on key concepts including:

- A Big Data definition
- Why the approach is unique vs. existing practices
- Coexistence with market research methods
- Data science talent requirements
- Data quality
- Data privacy issues

Readers of this Executive Summary can elect to drill down further into any of the topics within the body of the report for greater detail and illustrative examples.

Structuring the Streams of Big Data. Advances in data-gathering technology—for example, sensors, audio and video watermarking, and coding—have enabled the recording of traffic patterns, solar power levels, and digital media assets. In the world of marketing and advertising, analyzing data from continuous streams of social media, photo sharing, mobile app usage, etc., requires transformation to a more structured format. For example, social media conversations can be distilled into positive, negative, or neutral. These "unstructured" data are often raw in their original state and require shaping so they can be analyzed alongside "structured" data like retail sales records or TV ratings that are organized in conventional row-and-column file formats. (p. 9)

Big Data Defined: Not Just About Size. Big Data in the Marketing and Advertising (BDMA) industry refers to the processing of data too big to handle on a single file server, most likely including unstructured data; it is also apt to utilize multiple data sources that reflect marketing and advertising activities and consumers' reactions to them. The number of data sources depends on the scope of the analysis. BDMA can include as many consumer touch points as are available or focus on a limited number of them to address specific marketing and advertising questions. (p. 10)

Why Big Data? Superior Decision Making. Big Data surfaces insights that would not have been possible through most currently available market and syndicated research. For example, large sources of electronically gathered, continuously measured data provide a platform for rapid test-and-learn feedback from consumers, unlike self-reported surveys, which require more time for processing and reporting. Other examples of Big Data implementation include: 1) helping a

marketer to better identify and focus on key geographic sales areas or 2) helping a media agency to more selectively pinpoint TV programs that deliver richly described target audiences with information beyond the standard age/gender metrics used today. (p. 14)

Market Research Will Continue to Answer the "Why?" Questions. Big Data is causing disruption in the world of traditional market research. Analysis of large-sample-size behavioral data is being viewed as a way to trump the labor-intensiveness, expense, and self-reporting vagaries of survey and consumer insight research. The market research discipline won't vanish anytime soon, however, since its role is to provide critical insights on the "Why?" of consumer behavior rather than the "What?" that Big Data specializes in. For example, Big Data can tell us what creative executions in a digital campaign were most successful in driving response, but not much about the emotions behind why consumers favored the creative. Some feel that market research will also be there to answer new questions about consumer actions that may arise in Big Data analyses. (p. 17–19)

Data Quality: Clean, Fit, and Valid. The "Big" in Big Data doesn't guarantee better. In fact, the ever-increasing volume and appearance of different types of data amplify data scientists' challenges in ensuring that data are fit for use. Gartner reported that poor data quality is the primary reason for 40% of all business initiatives failing to achieve their targeted benefits. The first step in addressing data quality issues is through rudimentary cleansing processes that involve attending to missing fields, spelling conventions, and typos, as well as conducting value and logic checks. Unfortunately, data cleansing is often a hot potato because at many organizations data ownership is not clear. George Ivie, CEO and Executive Director of the Media Rating Council, outlines the next quality assurance steps, "careful inspection of the underlying representativeness, ensuring consistency of reported metrics over time, and understanding how the data collection process might impact accuracy of reported data. There also needs to be full disclosure of the methods used by data suppliers to integrate/fuse data sets together to better understand the impact on business decision making." (p. 22)

Scouting Data Science Talent. Installing hardware and software to enable Big Data processing and analysis is relatively easy. Finding data scientists can be challenging, since they are in high demand and short supply. Data scientists are rare birds that possess 1) the ability to organize and work with large, unstructured data sets, combined with 2) an advanced statistical background for extracting insights from the data. Many data scientists are recruited directly from academia or from organizations outside the industry (e.g., NASA) and then learn the marketing and advertising business. Select universities now offer one- or two-year graduate degree programs to prepare candidates for business (see page 22 for a list of 20 top schools). (p. 20)

Privacy: Industry Addresses Consumer Unease via Self-Regulation. Consumers enjoy the enriched experiences made possible by advanced TVs, the PC Internet,

smartphones, and tablets. But they also fear that their privacy will be compromised if data collected from these channels are not properly managed, according to a study conducted by the White House during May 2014. Data privacy concerns are nothing new, and a few years ago the Digital Advertising Alliance (DAA), an industry group comprising seven advertiser, agency, and media trade associations, was formed to address critical privacy issues. The DAA developed key principles that speak to consumer privacy concerns, including transparency of data collection and use, choice of consumer data collected, and security and retention length of data. All members comply on a voluntary basis. (p. 25)

Getting Started: Strategic Plan and Technology Needed. Launching a Big Data initiative requires one part strategy, one part technology. Tangible goal and strategy statements will help improve odds for a successful and productive Big Data program. Examples:

- **TV Network—"**Reduce data storage costs of TV program social media posts by a minimum of 25% during next 18 months."
- **Marketer, Media Agency—**"Uncover new target segments within the Most Valuable Customer base and high-potential media touch points to reach them."

Once Big Data goals and strategies have been established, IT is then better able to determine the types of technology required for implementation. Hadoop and MapReduce are two core technological elements that are popular for deploying Big Data efforts. Hadoop is an open-source software platform that makes multi-server processing possible to handle large volumes of data. MapReduce is the algorithmic framework within Hadoop that serves as an air traffic controller of sorts for processing and generating data across computer clusters (groups of computers). Scripting language, machine language, and natural language processing (NLP) are instrumental in helping to provide structure for raw data. (p. 27)

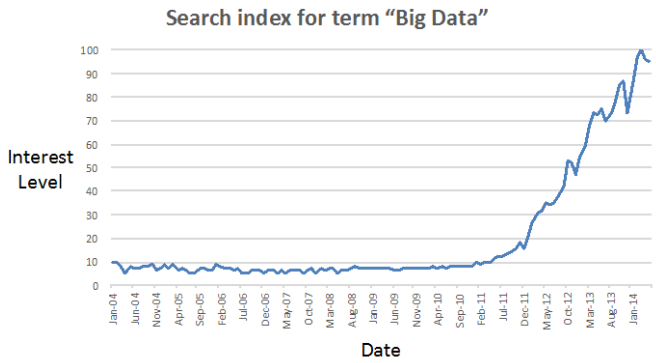
Marketplace Interviews. This section offers Big Data insights from ten companies and provides a richer marketing and advertising perspective and more color to the main Big Data document. Some highlights below (p. 30):

- **Definition—**Focus is on new insights and transactional metrics, not data size.
- **Successful Deployment—**Requires a strategy, management support, data talent, and internal education.
- **Hottest Areas in Marketing and Advertising—**Targeting and addressability, creating consumer multi-touch-point profiles, and speedy decision making.
- **Biggest Implementation Challenges—**Attracting data science talent and integrating data from disparate sources.

INTRODUCTION: BIG DATA NEEDS BIG DEFINING

In recent years, the words "Big Data" have relentlessly appeared in newspapers, magazines, on digital screens of every persuasion, and out of the mouths of business execs, academicians, and even the average consumer. And the crest of this Big Data tidal wave seems to be rising as rapidly as the quantity of data itself. According to Google Trends, interest in "Big Data" registered as a faint but consistent pulse for many years until starting its parabolic assent in late 2011, then growing nearly tenfold in magnitude by 2014.

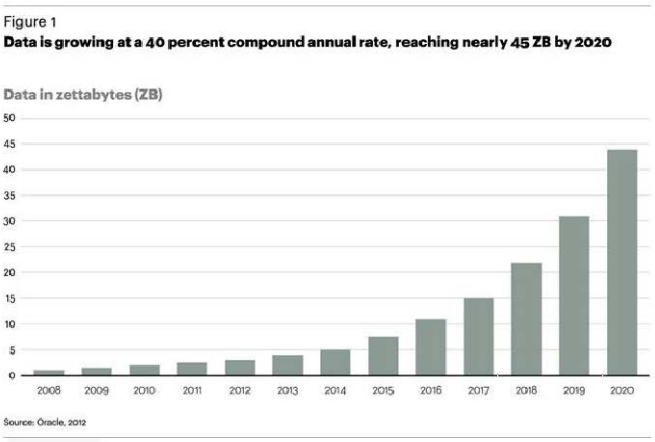
Chart A: Big Data Interest Over Time



Source: Google Trends, normalize search volume according to the ratio of term(s) to the entire volume of search

The expected volume growth of Big Data during the next six years makes the Google search queries to date seem like a trickle of water in a pond. In 2014, Oracle estimated that the combined size of all the data on Earth totals up to five zettabytes and will grow at the rate of 40% per year until 2020. Now for some perspective: one zettabyte of data is equal to a stack of books spanning the distance from Earth to Pluto 20 times, or 72 billion miles. In 2020, the data tally will be up to 45 zettabytes!

Chart B: Future Growth of Big Data Volume



Giant numbers aside, Big Data fixation goes hand in hand with Big Data vexation, as there appears to be some confusion about how to define the term, as well has how and when Big Data can and/or should be deployed in organizations. From the broadest view, Big Data is sometimes looked to as a panacea for improving the quality and effectiveness of practice across many different fields. According to *The Economist,* Big Data implementation translates to the ability to "spot business trends, determine quality of research, prevent diseases, link legal citations, combat crime, and determine real-time roadway traffic conditions."[1]

Not surprisingly, the marketing/media industry holds keen fascination with Big Data as a silver bullet for distilling vast mountains of consumer and media exposure data into meaningful insights that inform superior marketing strategies and investments. Consumer markets and media audiences have become so diversely fragmented that the task of securing and analyzing data from myriad touch points has morphed into an undertaking of monstrous proportions. These data can include retail scanner receipts, social media tweets, photos and conversations, online search activity, click-throughs, ad downloads, e-commerce transactions, survey research, TV audience ratings, and coupon redemptions, to name some core metrics. It seems that the marketing and advertising world looks to Big Data as the means for transforming disconnected industry metrics into statistically rigorous, speedy, and accurate systems for investment decision optimization. But the ABCs of how to get to this end state eludes many, which underscores the need for a Big Data definition and practical guidance for implementation.

The goal of this document is to offer the marketing/media industry informative clarity regarding the meaning, scope, and implementation of Big Data, which include the following:

- **Definition—**A Big Data framework of relevance and context for the marketing/ media industry so that practitioners can apply it to myriad business scenarios. This will include the concept of Big Data strategy.
- **Why Big Data?—**Elaborates on the Big Data value proposition through practical examples.
- **Impact on Market Research Practices—**Does the emergence of Big Data correlations that are predictive of <u>what</u> happens in the future eclipse today's more structured testing as to <u>why</u> thing happen?
- **Talent—**What makes good data scientists and where to find them.
- **Data Quality—**Examines whether "Big" equals quality, looking at the integration.
- **Getting started—**Setting a Big Data strategy and securing technology.

Defining Big Data: Technical Feat Evolves to Business Solution

[1] From "Data, Data Everywhere: A Special Report on Managing Information," February 25, 2010, *The Economist.*

The term "Big Data" evolved among scientists and business data subject matter experts as a way to describe the "challenge of processing data sets that are too big to fit on a single server, too unstructured to fit a row-and-column database, or too continuously flowing to fit into a static data warehouse."[2] How big is big? Perhaps the answer is in the eye of the beholder: "For some organizations, facing hundreds of gigabytes of data for the first time may trigger a need to reconsider data management options. For others, it may take tens or hundreds of terabytes before data size becomes a significant consideration."[3] While the perception of size is relative to each organization, the dividing line for what is considered big is more than just about amassing storage size to accommodate rapid data growth; it also encompasses the differences in data format and structure. The Gartner Group added data access speed as another Big Data element and offered the following three Big Data attributes, all conveniently starting with the letter "V":

- **Volume—**Sheer size of data and storage options to accommodate it.
- **Velocity—**The rate at which users can access data.
- **Variety—**Types of data, including structured (record, row, column) and unstructured (video streams, social media, traffic monitoring, etc.).

Structuring Data. Building on the last point above, unstructured data deserves a callout, given its prevalence across Big Data analyses within the marketing and advertising industry. "Unstructured" means that the data are not organized in a pre-defined manner when they are collected; some examples of unstructured data include conversations, photo postings, or search keyword entries. Structured data are categorized at the time of collection, such as during a multiple-choice survey that asks for the age or education level of a person by bracket. Analyzing data from continuous streams of social media, photo sharing, search, etc., requires transformation to a more structured format. Social media conversations about brands, for example, can be distilled into positive, negative, or neutral, while brand influencers are identified by how many people are reading their posts. These "unstructured" data are raw in their original state and reshaped so they can be analyzed alongside "structured" data like retail sales records or TV ratings that are organized in conventional row-and-column file formats.

The sheer technical feat of storing, restructuring, and managing Big Data is recognized as a major accomplishment among technical experts; however, new insights emerging from the data have mesmerized end users in business, science, and education. Now it seems there is a shift in the meaning of Big Data more closely reflect the analytic and insight aspects, as opposed to the advances in data technology that make them possible. Perhaps this explains the tremendous surge in

---

[2] From *Big Data at Work: Dispelling the Myths, Uncovering the Opportunities,* by T. H. Davenport, 2014, Boston, MA: Harvard Business Review Press.

[3] From "Introduction to Big Data," by R. Magoulas and B. Lorica, February, 2009, *Release 2.0, 2.0*(11).

interest revealed by the Google Trends in Chart A, as more end-user decision makers stood up and took notice.

As Thomas Davenport, academic and advisor on information and analytics, states, "the point is not to be dazzled by the volume of the data, but rather to analyze it—to convert it into insights, innovations, and business value."[4] Mayer-Shonenberger and Cukier make a further distinction that recognizes the empowerment of data size with the unique value that can be extracted, by referring to Big Data as "things that can be done at a large scale that can't be done at a smaller one, to extract new insight and create new forms of value."[5] Finally, Gartner recognizes the importance of analytics and insights in its most recent Big Data definition by tacking these attributes onto the three Vs: "Big data is high-volume, high-velocity, and high-variety information assets that demand cost-effective, innovative forms of information processing for ***enhanced insight and decision making***."[6]

DEFINING BIG DATA FOR THE MARKETING AND ADVERTISING INDUSTRY

With the proliferation of consumer digital media, the data emanating from smartphones, iPads, gaming consoles, TVs, laptops, and desktops are amassing at lightning speed and in high-volume fashion. This data stream bounds forward, each data source creating its own rivulet that sometimes joins with others, and sometimes not. The idea of creating a ubiquitous Big Data definition for an entire industry is daunting, considering the myriad data components and business interests that form its ecosystem. But perhaps if we think about the consumer as the center of the industry, we can simplify the task. And if we seek to achieve a more advanced understanding of consumer behavior, then integration of consumer touch points and their data equivalents would be the context in which Big Data could be viewed. The reality, however, is that each type of business within the marketing and advertising industry will likely deploy Big Data to meet specific goals of the organization that are in some way tied to individual or multiple aspects of consumer marketing.

Following is a hypothetical sequence of examples to demonstrate how a Big Data target segmentation analysis could be implemented by a marketer, translated by the marketer's media agency for media investment decisions, and further refined by the TV network holding company where the advertising is actually run.

First, let's start with Marketer A, whose goal is to identify the types of consumers expected to drive sales volume of Brand B during the next eight to twelve months.

---

[4] From *Big Data at Work: Dispelling the Myths, Uncovering the Opportunities,* by T. H. Davenport, 2014, Boston, MA: Harvard Business Review Press.

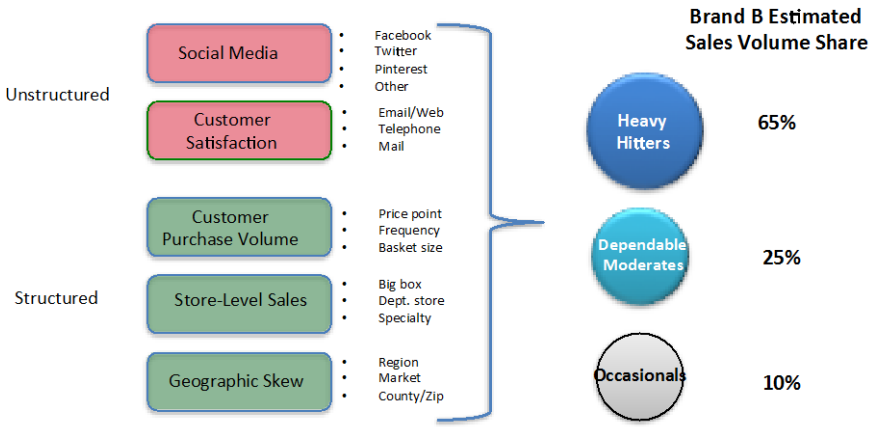
[5] From *Big Data: A Revolution That Will Transform How We Live, Work, and Think,* by V. Mayer-Schonberger and K. Cukier, 2014, Boston, MA: Mariner Books, Houghton Mifflin Harcourt.

[6] From http://www.gartner.com/it-glossary/big-data/2014

Marketer A's data scientists will use a combination of unstructured data sourced from social media activity and customer satisfaction feedback to develop a sense of what is being said about Brand B in the marketplace and who is saying it (see Chart C). Marketer A will then deploy modeling techniques to map this dialogue to the structured customer records that contain data on volume of purchase, type of retail outlet, and geography. The desired outcome would be to surface high-sales-volume customers who have a positive disposition towards Brand B; let's call them the "Heavy Hitters." For Marketer A, Big Data value is the incremental intelligence gained by combining multiple data sources to surface the Brand B Heavy Hitter target. Now Marketer A can provide this rich description of the Heavy Hitter to industry partners who will help maximize return on marketing and advertising investment.

Chart C: Marketer A, Brand B

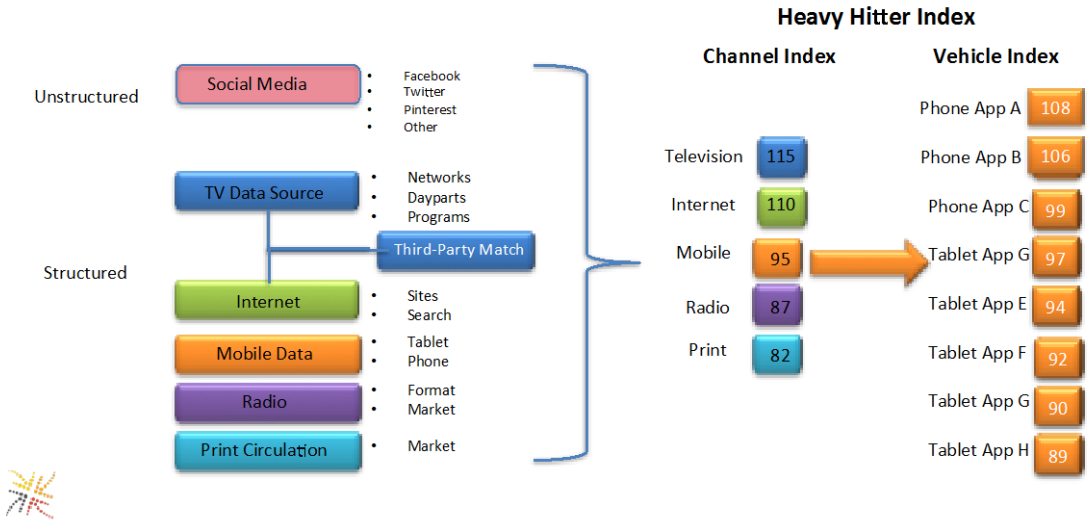
Pooling data together to identify high-propensity consumers, aka the Heavy Hitters



Marketer A now approaches its Media Agency with the Heavy Hitter target definition. Chart D shows how the Media Agency, in turn, translates the Heavy Hitter target across the various media investment channels, including social, TV, Internet, mobile, radio, and print, using algorithms. A Heavy Hitter index is assigned to each channel, followed by a vehicle index. For example, Mobile delivers the Heavy Hitter target at a rate slightly below the norm (95 index); however, Phone Apps A and B are above average. Notice also the data match between TV and Internet based on the Heavy Hitter target; this will help minimize over-saturation or under-delivery of advertising across the two media. For example, a portion of Heavy Hitters could be light TV viewers but could be targeted on the Internet, where they spend more of their time. Big Data produces value for the Media Agency by providing a means to sort through myriad channel and vehicle options that deliver the highest concentration of Heavy Hitters.

Chart D: Media Agency

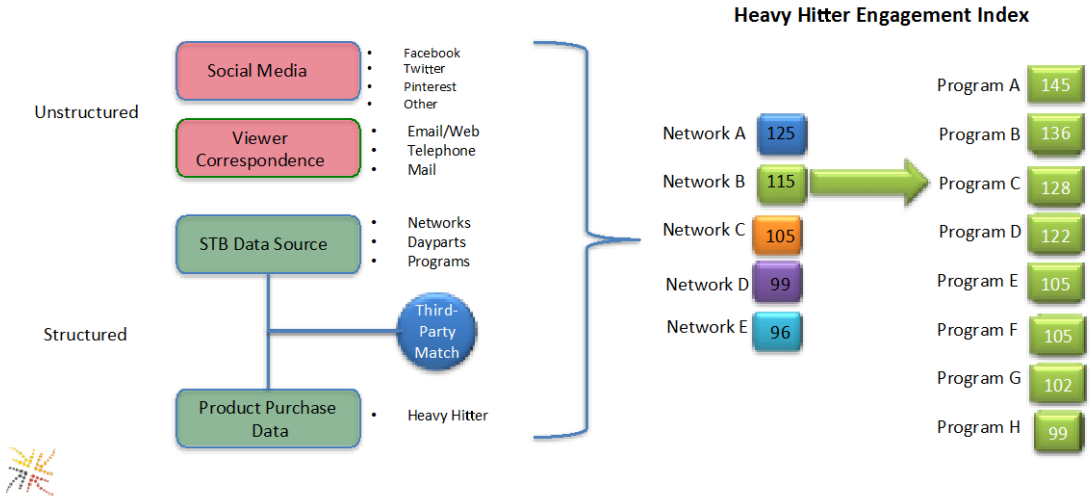
Identifying effective media that deliver Heavy Hitters . . .



With the permission of Marketer A, the Media Agency passes the Heavy Hitter definition to the TV Network holding company, which has a history of providing customized targeting solutions for Brand B. The Media Agency already has some ability to identify the TV networks that deliver high concentrations of Heavy Hitters but lacks three Network TV Group A data sources. Two unstructured sources, social media and viewer-to-network correspondence, provide indications of program engagement. The third directly links viewing and product purchasing, enabling TV Network TV Group A and programs to more precisely identify the Heavy Hitters within its own set of networks (see Chart E). Network B, for example, delivers engaged Heavy Hitters at 15% above the norm and several programs exceed that level. So Network TV Group A finds value by using Big Data techniques to further refine the Heavy Hitter target and identify those who are likely engaged with specific networks and TV programs.

Chart E: Network TV Group A

Finding networks and programs where the Heavy Hitters are . . .



So now we've seen that for any given marketing and advertising organization, the deployment of Big Data will vary according to specific needs and role that the organization plays in the industry ecosystem. In this particular example, one thread of commonality was the integration of unstructured data with the row-and-column data files that have been the cornerstone of business intelligence and analytics for decades. Classifying the size of the data as "Big" in terms of gigabytes, terabytes, or petabytes is uniquely relevant to each organization, as each company has different storage capacity, hardware, and qualified personnel to handle and extract insights from the data.

Following is a definition of Big Data, drawn from a collection of Big Data definitions that have emerged over the last couple of years and filtered through the lens of the marketing and advertising industry:

> Big Data in the Marketing and Advertising (BDMA) industry refers to the processing of data too big to handle on a single server, most likely including unstructured data, and likely involving integration of multiple data sources that reflect marketing/media activities and consumers' reactions to them. At its ideal level, BDMA can include as many consumer touch points as are available or focus on a limited number of them to address specific marketing and advertising questions. The BDMA approach must yield insight that was previously unachievable through more conventional business analysis.

WHY BIG DATA?
FOR BIGGER AND BETTER DECISIONS!

The abundance, granularity and richness of Big Data hold the promise of leading to more effective strategic and tactical decisions that result in superior marketing and advertising performance. There are five key aspects of Big Data that furnish the rigorous grounding for making better marketing decisions: 1) Precision 2) Prediction 3) Data Extension 4) Speed & Continuity and 5) Accuracy. Following are some examples of how each one bolsters the case for deploying a Big Data approach to marketing and advertising:

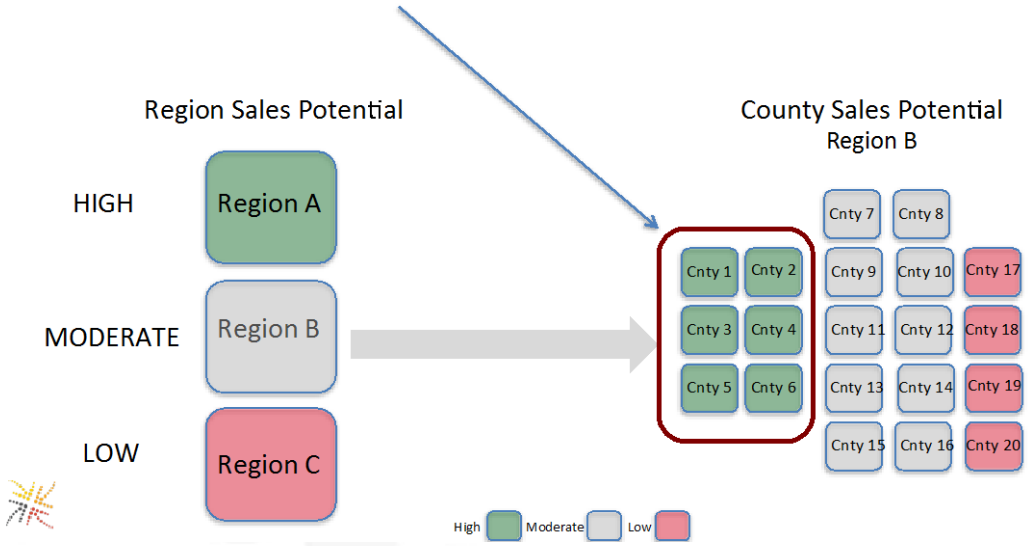
1) Precision

- More granular and descriptive data now make it increasingly possible to identify multiple target consumer segments for a single product or service; media investments and messaging can then be tailored to these sub-segments to more effectively drive sales and branding.

- Chart F demonstrates how a marketer can uncover areas of high sales potential within a business region with moderate sales activity. By focusing more on the counties with higher sales potential within Region B, the marketer can more effectively deploy resources such as sales force time and promotions to drive business.

Chart F: Precision

Focus on high sales potential by creating micro targets



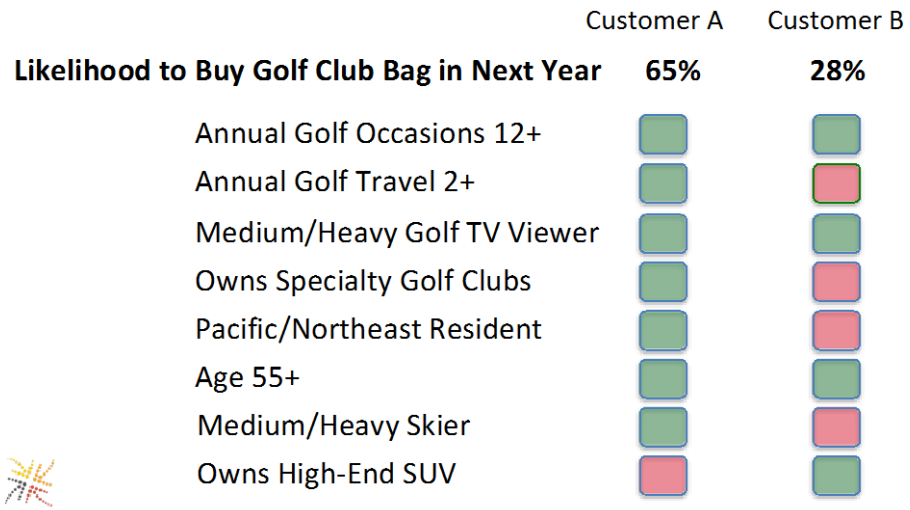
2) Prediction

- Big Data helps marketers develop valuable predictive tools to uncover the actions their customers are likely to take based on the tracking of their behavior over time; advances in data storage, processing speed, and data science applications have enhanced predictive modeling.

- Chart G shows how the combination of many descriptive and transactional data points can help when building a model for predicting future purchases—in this case, people who are likely to buy a golf club bag within the next year. Notice how Customer A's probability of purchase (65%) is more than double Customer B's (28%). This information is very helpful in allocating a marketing budget for either customer type in anticipation of their likelihood to purchase.

Chart G: Prediction

More data points = stronger predictive power



3) Data Extension

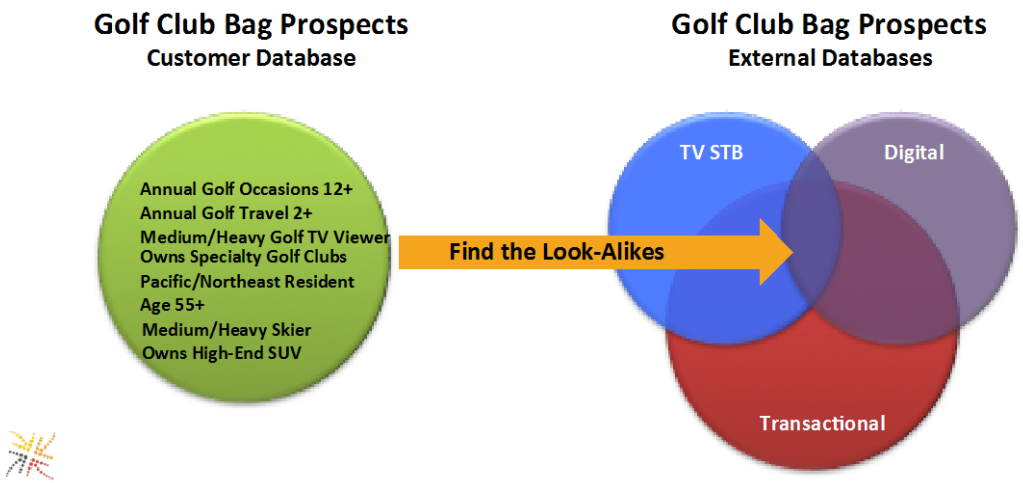
- Beyond existing-customer data, there are hundreds of external data sources available for integration to enable a fuller view of consumers' transactions and media touch points. This extended plane of data can translate to a deeper understanding of how to cultivate more business and/or loyalty from current customers, as well as how to attract new ones. Big Data processing speed and storage now make these data unions more feasible than ever before.

- Thinking about our golf example, Chart H demonstrates how Data Extension provides a means of finding new-customer prospects that meet the

transactional and descriptive profile of potential golf club bag purchasers. Plugging into these external databases also reveals a rich trail of electronic breadcrumbs in the digital and TV universe for reaching this target.

Chart H: Data Extension

Beyond customer database = finding more business opportunity



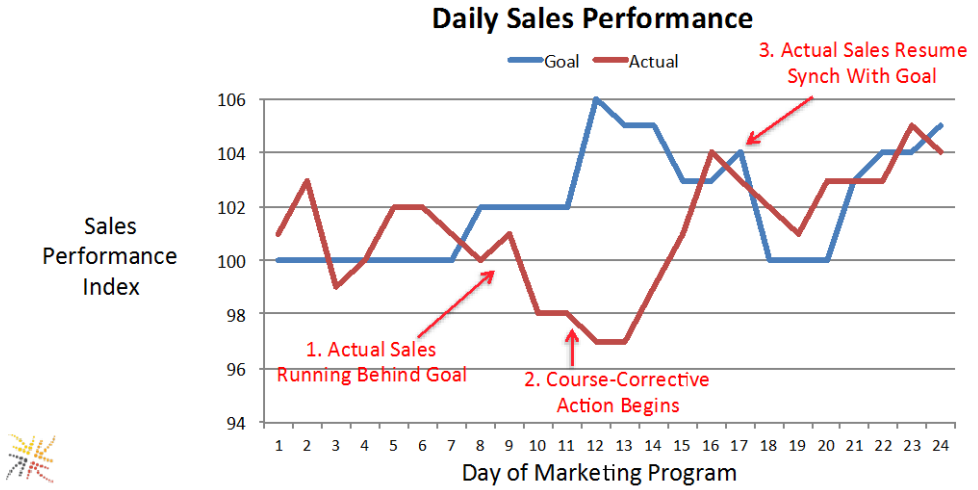
4) Speed & Continuity

- Electronic, continuous data collection provides constant and rapid feedback that offers the potential to adjust marketplace strategies and tactics within smaller time increments.

- Chart I provides a view of actual sales performance vs. goal during a 24-day period. In this particular example, a retailer may have been planning for a sales lift around a special promotion time (Days 11–14) that did not materialize and countered the downdraft with additional advertising or promotional incentives for consumers to visit the stores. By tracking daily sales to goals, the retailer can identify (a) when sales are falling short of goal, and (b) the start of course-corrective action, followed by (c) resumption of actual sales catching up to planned sales goals.

Chart I: Speed & Continuity

Always on, electronic access = opportunity for adjusting course



5) Accuracy

- Electronically gathered purchase and media consumption data provide a stronger foundation than survey-based data collection for determining the relationship between marketing and advertising and sales. Veteran researcher Larry Friedman offers this thought: "But now that we are able to link such memory-dependent survey metrics to data on actual behavior we can see how shockingly weak their relationship to reality often is."[7]

BIG DATA DISRUPTION: COEXISTENCE WITH MARKET RESEARCH

Big Data is steadily causing disruptive tremors in the world of traditional market research. Analysis of large-sample-size behavioral data is being viewed as a way to trump the labor-intensiveness, expense, and self-reporting vagaries of survey and consumer insight research. Setting up and executing classic research studies to isolate cause and effect is now giving way to faster, more immediate series of small experiments that fuel insights and feed decision making in a more expedient way. Harvard Business School Professor John Deighton points to the benefits: "When data is cheap, abundant and comprehensive, feedback is vast, the business objective is clear, and small failures are tolerable. Action beats analysis."[8] This extreme position

[7] From "Seeing With New Data," by L. Friedman, November, 2013, *Admap,* p.10–12.

[8] From "Big Changes Will Deliver a Big Future: What Marketing Decision-Makers Expect Their Customer Insight Teams to Deliver," by A. Riley and D. Smith, September, 2013, *ESOMAR, Best Paper Award, Congress.*

is characterized by the claims of some marketing technology companies that "segmentation . . . is now dead as we can simply use data analytics to uncover the predictable way in which consumers behave and target them individually rather than as a group."[9] While the signs of change are certainly present, the migration of traditional market research survey work to Big Data techniques will likely not occur overnight.

At the core of the Big Data vs. Market Research debate is whether it's critical to know simply "What" actions consumers are likely to take, as opposed to the motivational "Why" behind consumer decisions. For example, a Big Data approach to choosing which digital advertising creative to serve may be based purely on predictive algorithms rooted in previous purchase behavior, recent site visitation, historical creative performance, etc. The creative strategy, however, is more likely based on qualitative consumer research that includes survey-based creative testing and/or one-on-one in-depth interviews. So during the campaign, creative-selection algorithms will gravitate toward the best-performing ads based on consumer actions taken and optimize accordingly for maximum return. But the automated programs will not likely write the advertising copy anytime soon. . . . Well, not exactly! Modifications to wording, design, and color can be tested and served on the fly; however, the actual development of the original brand communication strategy and advertising development is still in the hands and minds of humans.

Digging deeper into statistical parlance, the "What" and the "Why" can also be thought of as correlation versus causation. It's here where some feel that Big Data could lead us astray, or at least leave us without well-grounded explanations for making decisions other than that two events are highly correlated. For example, Advertiser A knows that every Thursday, sales at its retail stores are five percent above the norm in one key market with no notable difference in per capita ad spending relative to other locales. The only unusual event that surfaces in this market is the fact that about 60% of the time, it is either overcast or raining on Thursdays. In fact, Thursday sales are even higher on rainy days. Does this mean that the sales lift is attributable to inclement weather? Should Advertiser A increase the media budget on Thursdays when the weather forecast calls for rain? What about other days of the week? What if the market is Seattle, where it tends to rain more than other parts of the country? There is likely some reason other than gloomy weather that explains why Thursday sales are up 5% in this market, and that factor might be better exploited by Advertiser A for future Thursdays. Of course, a capable data scientist would question the decision-making value of the high correlation between sales and weather on Thursdays in Seattle and, instead, request and test additional data that might provide further insight into this particular marketplace.

[9] From "Big Data: The Marketing Opportunity," by C. Strong, September, 2013, *Admap,* p. 25–27.

What will the ongoing function of survey research look like in a Big Data world? Some in the industry see Big Data and Market Research coming together to drive incremental insights for decision making. Colin Strong goes as far to say: "there is surely a strong role for market research to seize the Big Data agenda by operating as 'Consumer strategy consultants.'" Strong further contends that given their extensive marketing and advertising backgrounds, market researchers could enrich the Big Data process in a number of areas—for example, by conducting hypothesis testing for consumer strategy or providing subject matter expertise on data sources—while retaining the ongoing responsibility for survey tools. Survey research would still play an important but less primary role.[10]

But how, specifically, will survey research be deployed in concert with Big Data? It's very likely that surveys will be used to answer many of the new "Why?" questions that surface from Big Data insights. Bill Pink of Millward Brown asserts, "findings from our new data assets generate more questions, and those questions tend to be best addressed by traditional survey research. In this way, as big data increases, we see parallel growth in the presence and need for 'small data' to explore and answer the questions it raises."[11] Pink goes on to depict a scenario where a large advertiser would have access to retail store traffic and sales transactions data that could be combined with surveys to better understand motivations and point-of-sale behaviors for certain shopping segments. In this case, primary research would not need to focus on the "What?" that is happening (big data will do that), but instead concentrate on the "Why?"

Finally, it's important to note that Market Research will continue to play a major role in core ideation and strategy development for new products, creative development, and general consumer messaging. Rapid in-market concept testing, where a brand's future strategy is based on a sample of clicks or light sales volume, may produce less than optimal guidance. Wharton professor Pete Fader states, "people say, 'Let's just try stuff and see what works.' That can help you determine a winner, but it doesn't help you design what would have been the best. By doing careful research and determining the underlying drivers that cause people to take action, we can develop better products and services."[12]

---

[10] From "Beyond Big Data: How Big Data Needs Consumer Insights to Realise Its Full Potential," by C. Strong, June, 2013, *ESOMAR, 3D Digital Dimensions.*

[11] From "How Big Data Liberates Research," by W. C. Pink, 2013, *Millward Brown Point of View.*

[12] From "Finding a Place for Market Research in a Big Data, Tech-Enabled World," by P. Fader, January 29, 2014, http://knowledge.wharton.upenn.edu [Web log post].

CALLING ALL DATA SCIENTISTS! FINDING TALENT

Big Data holds the promise of uncovering valuable insights that will lead to more strategic and profitable decision making. Setting up the technical infrastructure, including software like Hadoop and scripting languages like Python, is relatively straightforward and inexpensive; the greater challenge is acquiring and orchestrating the data science talent to preside over the system.

Data science is the practice of extracting insights and knowledge from primarily large data sets that are likely to include unstructured data. Data scientists are the uniquely specialized practitioners of this process. They possess a rare skill set, are highly paid, and may be difficult to hold on to since they are constantly sought after to fill demand for a capability that is in limited supply.

The term "data scientist" may lack clear definition for marketing and advertising professionals who do not hold positions in the discipline of research/analytics. Non-practitioners tend to roll up the entire spectrum of quantitative skills and label them as "research" or "analytics." The two aspects that separate a data scientist from other quantitative professionals in the marketing and advertising field are 1) the ability to organize and work with large, unstructured data sets, combined with 2) an advanced statistical background for extracting insights from the data. In contrast, traditional market researchers work with much smaller samples of data and surveys designed to provide insights to answer specific questions—e.g., how much did advertising drive brand lift?

Data scientists first appeared in the marketing and advertising industry with the emergence of digital media and the electronic capture of clicks, page downloads, video streams, social media tweets, gaming activity, and applications. Since a high portion of digital data is unstructured in format, the data scientists' station in life was to organize these data and extract insights to provide marketing value. For example, classifying social media comments about a brand into positive, negative, and neutral sentiment over time provided one indication of a brand's health, to be considered with other performance metrics such as sales and brand awareness/favorability.

The ultimate data scientist is gifted with a wide range of skills that hit on nearly every aspect of data processing, such as setting up experiments and translating business problems into data-driven solutions and insights. This ideal person would have the ability to write code or programs that prep the data for analysis and have the scientific acumen and relentless curiosity to cycle through different experimental cuts of the data. The candidate would also possess quantitative skills that include statistical analysis, visualization analytics, machine learning, and evaluation of unstructured data, such as text, video, or images. From an organizational perspective, the ultimate data scientist would also play an advisory and business expert role by effectively solving recognized business challenges

through data analysis and communicating actionable results that everyone can understand.

The data scientist who excels across all ideal traits is truly a rare bird. Rather than search for this "super scientist," most organizations are apt to create teams comprising people with complementary talent sets. Perhaps the most obvious division in skill sets is the separation between technical and communicative/relationship competencies. For example, one data scientist may be highly suited for translating the business problem into a social media data solution design and communicating results in storytelling fashion, while another may have generated a brilliant solution for modeling the social media data that met the design requirements. In this case, both scientists may have technical and communicative ability; however, one holds an articulation advantage while the other candidate's strength is in the details of statistical approach.

Finding Data Scientists

The development of Big Data tools and analysis is relatively new in the scheme of business, and so is formal education and training for those seeking careers in the field of data science. Colleges and universities are now recognizing the need to include data science on their menu of offerings to fill the strong demand for the skill set. Coursework such as machine learning, scripting languages, and programming can be found at institutions that offer science/math degrees, as well as at business-oriented schools under the auspices of business analytics. IBM, through its Global University Program, actively recruits from a number of schools. Jim Sporer, IBM's director of this initiative, notes that 40 percent of participating schools fall into the engineering category, while another 40 percent are business schools. Sporer recognizes the need for balance: "The fact of the matter is we need marketing people who know big data analytics. We need health care people who know big data analytics."[13]

Table A is an alphabetized list, compiled by *InformationWeek*, of twenty universities[14] offering data science and analytics programs. These universities could serve as starting points for recruiting data scientists.

[13] From "As Demand for Big Data Analysts Grows, Schools Rush to Graduate Students With Necessary Skills," by S. Overly, September 15, 2013, *Washington Post.*

[14] From "Big Data Analytics Master's Degrees: 20 Top Programs," by D. Henschen, January 7, 2013, *InformationWeek.*

Table A: Big Data Analytics Master's Degrees

University	Major/Coursework
Arizona State University	Business Analytics
Bently University	Business/Marketing Analytics
Carnegie Mellon University	Business Intelligence and Data Analytics
Columbia University	Computational Biology, Computer Security, Machine Learning, Natural Language Processing
DePaul University	Computational Methods, Health Care, Hospitality and Marketing
Drexel University	Business Analytics
Indiana University, Bloomington	Business Analytics
Louisiana State University	Analytics
Massachusetts Institute of Technology	Business Administration
Michigan State University	Business Analytics
New York University	Business Analytics
North Carolina State University	Analytics
Northwestern University	Predictive Analytics
Purdue University	Business Analytics
Rutgers University	Analytics & Discovery Informatics and Data Sciences
Stanford University	Information Management and Analytics
University of California, Berkeley	Master of Information and Data Science
University of Cincinnati	Business Analytics
University of Connecticut	Business Analytics and Project Management
University of Maryland	Business for Marketing Analytics
University of San Francisco	Analytics
University of Tennessee	Business Administration, Statistics, Operations and Management Science

Source: D. Henschen, "Big Data Analytics Master's Degrees," *InformationWeek*, January 7, 2013

DATA QUALITY: CLEAN, FIT, AND VALID

As we've already seen, the "Big" in Big Data opens many possibilities for uncovering new insights at a faster rate to enable more frequent and informed decisions. But simply piling on terabytes of social media or transactional data, for example, does not guarantee actionable insights. Embarking on a Big Data analysis is like prepping your car for a long trip. First you plot out the destination with a GPS, which is akin to setting up the best analysis design to answer business questions. Next, you check the car's oil level, tire pressure, and engine coolant to ensure that you get to your destination safely; this requires a close examination of the structure and completeness of the data.

The concept of data quality can be viewed as a combination of practicing good data preparation hygiene as well as developing an understanding of the quality of the underlying data properties. These three stages encompass the data quality/prep process:

1. **Rudimentary data cleansing—**To ensure the correction and/or removal of inaccuracies, misspellings, inconsistencies, contradictions, disparities, data entry mistakes, missing fields, etc.
2. **Underlying quality check (one source)—**Examination of consistency, representativeness, and time period.
3. **Underlying quality check (multiple sources)—**Quality of data output resulting from Big Data integrations using disparate data sets.

Following is more detail about each of these three stages and why they're critical for generating insights for decision making.

Rudimentary data cleansing. The presence of dirty data can have serious consequences for decision making and corporate resources. Gartner reported that poor data quality is the primary reason for 40% of all business initiatives failing to achieve their targeted benefits.[15] From a resource perspective, data cleansing requires time and labor, and formal responsibility for the task is sometimes difficult to pin down within an organization. A study conducted by TekSystems reported that 60% of IT leaders (IT CIOs, VPs, and Directors) agreed that data owners aren't held accountable for ensuring data quality.[16] Not surprisingly, in the same body of research, 57% of IT leaders pointed out that it's not always clear who owns the data in the organization. Data cleansing may have become a corporate hot potato in the sense that there's a hesitancy to take on a task that is so clerical in nature. And sometimes the value of the data cleansing effort is in question. Mary Schacklett of Tech Republic described it this way: "At the end of the day, cleaning data can be hard to justify for ROI, because you have yet to see what clean data is going to deliver to your analytics and what the analytics will deliver to your business."[17]

The emergence of Big Data should elevate the practice of data cleansing to a new level of importance as the sheer volume and unstructured nature of many of the data sets are likely to compound the dirty data syndrome. Those who embark on a Big Data analysis should therefore ensure that the proper checks have been performed before they use the data to make business decisions.

Underlying quality check (one source). In addition to giving the data set a good scrubbing, it makes sense to go even deeper and perform a check on the soundness of the data for decision making. No data set will be perfect; however, it is critical that end users become acquainted with any holes or shortcomings for consideration when making decisions: George Ivie, CEO and Executive Director of the Media Rating Council, offers these four points for checking single data sources[18]:

1. **Underlying Data Values—**The valid values for each field and business rules surrounding each field within the data sets should be known. These field definitions should be tested for validity prior to using the data set. It helps to understand the processing used to derive the data set.

[15] From "Measuring the Business Value of Data Quality," *Gartner,* October 10, 2011.

[16] From TekSystems IT Industry Trends Survey, 2nd Quarter, 2013.

[17] From "Data quality: The ugly duckling of big data?" by Mary Shacklett, *TechRepublic,* February 11, 2014.

[18] From George Ivie, Media Rating Council, June 2014.

2. **Time Period—**The exact period of time that the data reflect should be known**—**e.g., November 1, 2013–May 31, 2014.

3. **Representation—**Assess what members of the population are included or absent**—** e.g., credit card holders versus cash customers.

4. **Consistency—**Data set should be consistent over time, and monitored for anomalies to ensure that the relationships remain true. Changes to the underlying fields or business rules for the data set should be captured and noted on a timely basis.

Multiple Data Sets. While data from a single source can be fraught with holes, Big Data integrations, using multiple sources, can compound the issues of inconsistency and incompatibility. The MRC's Ivie characterizes the challenges of comingling disparate data sources: "Big Data is almost always incomplete, having holes, missing data aspects, etc., which requires combining or matching with other data sets. All of these big-data- driven data integrations raise concerns about privacy, quality control, and employing best practices." Ivie points out some important procedural aspects to consider when integrating Big Data sets, especially where modeling is deployed to address deficiencies across sources:

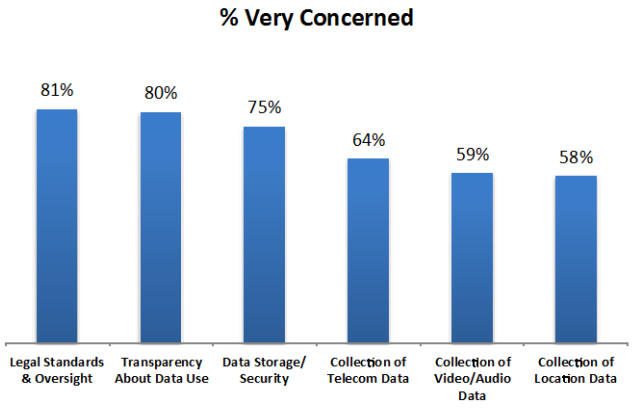
- **Strength of Integration Links—**The supplier of the integration product must demonstrate the statistical power and relevance of variables used for modeling or matching.
- **Link Comparability—**The variables used for matching or integration must be comparable in definition.
- **Model Priorities—**The supplier must have conducted appropriate research that establishes support for how variables and their respective weights are prioritized within the matching and integration process.
- **Distortion of Source Data—**The supplier must have conducted appropriate research on the extent to which the integration process itself causes distortion of the source data, on both an average and a worst-case basis.
- **Accuracy Assessment—**The accuracy of the modeling algorithm should have been assessed during the development process. This issue is most commonly studied through "split sample" or "fold-over" tests in which certain known characteristics of a sample are contrasted with the model-based predicted characteristics of that same sample. The assessment should include evaluation of both bias and distortion.
- **Sufficiency of Donors/Recipients—**The "matching success" of the method should be considered (if appropriate). On a related issue, the supplier must be able to justify the extent to which donor respondents (if applicable) are used multiple times during integration.

- **Choice of Static vs. Dynamic Process—**Data integration can involve single, all-purpose models, or distinct models for different tabulations (encompassing "ensemble" model groups or on-the-fly modeling). The supplier should provide rationale for this choice including evaluation of the technical pros and cons of each approach that are relevant to the supplier's application.
- **Reliability—**The statistical reliability of the integrated database must be quantified to a reasonable degree. At a minimum, broad averages must be provided that help the user understand the extent to which a typical integrated database measure may fluctuate from product release to product release.
- **Minimal Underlying Ascription—**If the databases being integrated are themselves dependent upon material amounts of data adjustment, the supplier must make a reasonable effort to minimize the dependence of the model on already-imputed data.
- **Ongoing Research—**A supplier is expected to have an ongoing program of methodological research in search of quality improvements.
- **Quality Control—**Data processing quality control procedures must be reasonable, as summarized by existing Standard A2: "Appropriate quality control procedures shall be maintained with respect to all external and internal operations which may reasonably be assumed to exert significant effects on the final results."

DATA PRIVACY: INDUSTRY SELF- REGULATION TO ADDRESS CONSUMER CONCERNS

There's been no shortage of excitement regarding Big Data's potential to help the marketing and advertising industry make crisper, quicker, and more informed investment decisions. But most consumers have a dimmer perception of Big Data, at least when it comes to data privacy. Consumers enjoy the enriched experiences made possible by technological advances in devices such as TV, smartphones, and tablets; however, they also fear that their privacy will be compromised if the data collected from these channels are not properly managed. According to a U.S. Government survey, their concerns are highest for legal standards and oversight, transparency, and the security of data storage (see Chart J). Secondarily, but not too far behind, is the worry about the way data are collected, particularly through telecom devices and physical location.

Chart J
Public concern spans virtually every aspect of Big Data



Source: White House, Office of the President, May 2014, 24,092 respondents

So the welcome advancements in technology are somewhat tempered by events that have become more commonplace, including identity theft, phishing, electronic embezzlement, personal espionage, etc. A core focus of consumer privacy concerns is the protection and integrity of Personally Identifiable Information (PII) information that can lead to the identification and contact of individuals. Reidentification may be achieved by piecing together bits of information about consumers that makes them unique.

Tackling data privacy issues is key to continuing the inroads Big Data is making in the marketing and advertising industry. The industry has been addressing data privacy in a self-regulatory manner, cognizant of the potential disruption that could befall businesses should the government need to intervene. Digital media that carry content and advertising on the PC Internet, smartphones, and tablets have been the most proactive, self-policing segment of the industry, given their heavy use of tracking tools for segmenting and targeting audiences. To that end, the Digital Advertising Alliance (DAA) emerged as a galvanizing driver of industry self-regulation and comprises the following industry associations:

- American Association of Advertising Agencies (4A's)
- American Advertising Federation (AAF)
- Association of National Advertisers (ANA)
- Better Business Bureau (BBB)
- Direct Marketing Association (DMA)
- Internet Advertising Bureau (IAB)
- Network Advertising Initiative (NAI)

The DAA first created the document, "Self-Regulatory Principles for Online Behavioral Advertising," to address the privacy concerns in online behavioral advertising space by establishing the following principles:

- **Education—:** iInform consumers about what Behavioral Advertising is.
- **Transparency—:** dDisclosure of data collection and data use.
- **Consumer Control—:** cChoice of whether consumer data is collected.
- **Data Security—:** rReasonable security and limited retention of behavioral data.
- **Material Change—:** oObtaining consumer consent before a policy change.
- **Sensitive Data—:** pParental consent for children's data, as well as consent for financial account numbers, social security numbers, pharmaceutical Rx, or medical records.
- **Accountability—:** iImplementation to enforce compliance with principles.

The DAA then carried these principles forward by publishing two other documents—one that covered all other non-behavioral data collection, and the other describing how to apply the principles to mobile media. All three publications can be found here: http://www.aboutads.info/principles.

From a government perspective, the FTC presides over the area of consumer privacy and has created its own guideline documents inspired by consumer and business input. The latest major FTC publication in the privacy arena is "Protecting Consumer Privacy in an Era of Rapid Change, March, 2012." The core spirit of this publication aligns with the DAA principles in that it covers security and retention of consumer data and disclosure of details about collection and use of consumer data. But FTC advocacy for an easy-to-use Do-Not-Track mechanism may be considered a bit too aggressive for some DAA members. Please find this document here: http://www.ftc.gov/news-events/press-releases/2012/03/ftc-issues-final-commission-report-protecting-consumer-privacy. Additional Privacy resources can be found in the Appendix.

GETTING STARTED: STRATEGIC PLAN AND TECHNOLOGY NEEDED

Launching a Big Data initiative requires data science talent, high volumes of data, and, of course, a collection of technology components that make algorithmic number crunching possible. The proper orchestration of these technological elements will lead to faster processing times, more efficient use of analytic staff, and hopefully, new insights for superior decision making. But technology alone won't guarantee that a Big Data initiative will produce value for an organization. Setting goals and strategy will help improve the odds for successful and productive Big Data undertakings. McKinsey's Barton and Court offer this viewpoint: "A clear vision of the desired business impact must shape the integrated approach to data sourcing, model building, and organizational transformation."[19]

[19] From Dominic Barton & David Court, "Three Keys to Building a Data-Driven Strategy," by Dominic Barton and David Court, Insights & Publications, McKinsey & Company, March 2013.

Setting a Strategy. Creating a Big Data strategy requires consideration of a wide variety of factors that can include a company's overall management objectives, technological maturity, richness of data sources, and intensity of data and insight usage within the corporate culture. Below are four Big Data benefit themes accompanied by example strategy statements relevant for the marketing and advertising industry. The first two themes, speed and cost savings, are tied to pure economic and efficiency gains made by introducing Big Data technology to an organization. The last two, new insights and new product development, focus on the use of Big Data for innovation:

1. Improved speed
   - ***Media Agency—****Reduce reporting time of TV audience ratings for high-sales markets from seven to two days after report month.*
   - ***Digital Network—****Reduce targeted ad serving decision time by a minimum of 20% to surpass competitor performance.*
2. Cost savings
   - ***TV Network—****Reduce data storage costs of TV program social media posts by a minimum of 25% during next 18 months.*
   - ***Marketer—****Maintain current sales reporting staff hires by improving efficiency and speed of reporting.*
3. New insights
   - ***Marketer, Media Agency—****Uncover new target segments within the Most Valuable Customer base and high-potential media touch points to reach them.*
   - ***Marketer—****Improve effectiveness of in-store advertising by combining in-store sales monitoring with follow-up shopper surveys for enhanced understanding of consumer actions.*
4. New product development
   - ***Mobile Media Firm—****Improve audience engagement by 25% over next 12 months by developing apps that allow users to query our site postings for hot music and cultural topics.*
   - ***Digital Search Engine—****Increase international audience by creating series of "Switzerland" apps that quickly translate music and film languages.*

Once Big Data strategies have been established, it becomes an easier task for IT to determine the types of technology required for implementation.

Putting Big Data Technology in Place. As discussed earlier, Big Data technology offers a way to process large volumes of structured and unstructured data that can't be handled by traditional databases and on a single file server. Traditional relational databases ingest data in the form of orderly rows and columns of numbers, while big data formats can run the gamut and include text conversations, audio signals, video, and graphics. A new generation of data software has emerged to handle Big Data—one that deploys the processing power of multiple servers for maximum scalability when analyzing large complex data sets. This document provides a high-level introduction to that technology so that nontechnical readers can, at a minimum, be conversant about basic requirements for Big Data implementation.

If you have been active in a business environment during the past couple of years, you may have seen or heard references to a trendy-sounding software by the name of Hadoop. Hadoop is the core Big Data ingredient that enables processing across multiple servers to handle large volumes of data. MapReduce is the algorithmic framework within Hadoop that serves as an air traffic controller of sorts for processing and generating data across computer clusters (groups of computers). MapReduce monitors available processing capacity across the clusters and allocates computing task resources accordingly. Scripting language, machine language, and natural language processing (NLP) are instrumental in helping to provide structure for raw data. These technologies are summarized in Table B.

Table B: Technologies for Big Data

Technology	Definition
Hadoop	Open-source software for processing Big Data across multiple parallel servers
MapReduce	The architectural framework on which Hadoop is based
Scripting languages	Programming languages that work well with big data (e.g., Python, Pig, Hive)
Machine learning	Software for rapidly finding the model that best fits a data set
Visual analytics	Display of analytical results in visual or graphic formats
Natural language processing (NLP)	Software for analyzing text—frequencies, meanings, etc.
In-memory analytics	Processing big data in computer memory for greater speed

Source: Thomas Davenport, "big data @work," Harvard Business Press, 2014

MARKETPLACE FEEDBACK
April 28, 2014–June 30, 2014

Ten interviews were conducted among an industry cross section of companies to provide a richer marketing and advertising perspective and more color to the main Big Data document. These interviewees included representation from advertiser, media agency, media company, and Big Data research firms. Discussions focused on the following Big Data topics:

- Definition within company
- How it is being used in the organization
- Structuring a plan
- Implementation requirements
- Implementation challenges
- Hottest areas of marketplace implementation
- Key requirements for accelerating Big Data in the marketing and advertising industry

Defining Big Data. When asked to provide a definition of Big Data, two key observations emerged:

1. **New insights over size—**Respondents recognized that size or volume is a core consideration in classifying data as big, but most emphasized new insights as the embodiment of the Big Data concept. Not surprisingly, those with greater technology expertise included data volume and required processing software in their Big Data descriptions.
2. **Transactional insights are key—**Leveraging actual consumer transactions such as retail and digital sales surfaced as an essential element of the Big Data definitions provided. The power of creating insights from actual consumer behavior was seen as a core Big Data value proposition. Mentions of unstructured data, including social media, photos, and bulletin boards, were far less prominent than transactional metrics when defining Big Data.

"'The word big is knee-jerk reaction to define by volume." ***VP Sales/Analytics, MVPD***

"Big Data is about people or devices not previously exploited to render insights." ***SVP Media Research, Multimedia Network***

"Big Data is blending of volume, technology, and quantitative approaches and the organizational requirements to get good at the entire process." ***COO, Digital DSP***

"Big data is about leveraging transactional data towards insights." ***EVP Media Research, Media Agency***

How Big Data Are Being Used in the Organization. The nature of Big Data implementation varied across companies according to specific need and company size and type. For example, a telecommunications marketer cited consumer touch-point data and social media conversations as critical adjuncts for surfacing consumer insights and informing their market mix model. A demand supply platform (DSP) COO mentioned targeting and predictive modeling based on digital response feedback. Across most of the discussions, integration of data sets to learn more about consumer touch points arose as a common Big Data deployment theme.

"Uncovering valuable ad exposure and sales conversion data that was not possible before." ***VP Sales/Analytics, MVPD***

"Integrating data sets that will bring TV/digital/syndicated and proprietary resource together for holistic consumer view for more effective targeting." ***SVP Media Research, Multimedia Network***

"Rebuilding predictive models of campaign performance every day for thousands of advertising campaigns based on billions of ad impressions." ***COO, Digital DSP***

"Understanding customer experiences, what customer touch points are most optimal, building on marketing mix model work." ***Marketing Research Director, Telecommunications Company***

Structuring a Big Data Plan. According to respondents, their organizations were in various stages of developing a formalized data strategy, although most do not use the term "Big Data" to describe their efforts. Instead, the initiatives tend to fall under corporate strategic umbrellas to provide intelligence and answer important business questions through data. The telecommunications marketer, however, established a Big Data Center of Excellence (COE) comprised of a large team of data scientists that use specialized processing skills to prepare data for a number of functional areas across the organization. Meanwhile, the DSP's handling of data is considered core to its business so the entire company can be considered one Big Data strategy and execution, although not formally recognized as such.

"We don't talk about strategy as Big Data but more as answering questions, helping clients." ***SVP, Director of Analytics, Media Agency***

"We were born as a technology company dealing with high-volume data transactions, required to process terabytes of data on a daily basis." ***COO, Digital DSP***

"Big Data is so new that the organization is prepping a plan that includes keeping pace with required skill sets, machinery, data visualization, etc." ***EVP, Media Research, Media Agency***

"*Only the most advanced companies have a truly structured plan, detailed by objectives and data sources. Companies still in early adopter stage, still trying to figure out what it (Big Data) means."* ***VP of Partnership Development, Big Data Syndicator***

"Data scientists in the Center of Excellence often do the heavy lifting of the data processing and hand over to research/analytic experts in functional departments to extract and write up the insights." ***Marketing Research Director, Telecommunications Company***

Implementing Big Data Across the Organization. After setting a strategy, Big Data needs to be incorporated into the organization. Successfully establishing a Big Data initiative within a company requires an all-hands-on-deck mentality that starts with top management and spreads across functional departments. Respondents identified three critical points of action that will increase the likelihood of Big Data success:

1. **Top Management Support—**Big Data implementation becomes more meaningful to the organization with leadership support and endorsement. Without upper-level backing, there's a tendency for nontechnical departments to minimize their commitment to familiarize themselves with and make use of data assets.
2. **Education—**Securing Big Data support and engagement from all functional areas within a company can be challenging when staff is unclear about what Big Data is and how it can benefit decision-making. The sophistication of Big Data education runs the gamut, from structured sessions mandated my management to informal efforts initiated by research/analytic experts.
3. **Centralized vs. Decentralized—**In larger organizations, management needs to decide how best to allocate Big Data staff**—** that is, central to the entire company or placed within functional departments. Establishing the exact proportional mix of this talent depends entirely on factors internal to the company. In general, however, merits of a centralized approach include economies of scale for processing data and cross-company access to data scientists with specialized skill sets. On the flip side, big data talent allocated solely to specific areas like product performance, marketing, advertising, etc., gain valuable subject matter expertise that can be applied directly to that specific aspect of the business.

"Difficult for unacquainted to understand what they can get from the data. What is the question? The push must come from the top." ***VP Sales/Analytics, MVPD***

"The problem is educating people who sometimes think of Big Data as a magic bullet." ***SVP Media Research, Multimedia Network***

"Everyone is expected to understand how our data systems work at the appropriate level of their functional area." ***COO, Digital DSP***

"Managers receive basic training on Big Data. Every major function has some Big Data initiative underway." ***Marketing Research Director, Telecommunications Company***

"Figuring out the right balancing points for centralized vs. decentralized organization can be a challenging issue." ***Marketing Research Director, Telecommunications Company***

Challenges of Deploying Big Data Initiatives. When asked what key issues faced them in deploying a Big Data initiative, respondents pointed to three areas of contention:

1. **Acquiring Talent—**Finding data scientists and analysts that can help deliver on Big Data goals and strategy can be difficult.
2. **Integrating Data From Disparate Sources—**The primary challenge is creating multiple consumer touch-point data via keys that link the same people across different sources of data. Secondarily, the task of data cleansing and formatting across multiple sources can be daunting.
3. **Data Ownership—**This is an ongoing tug-of-war in many organizations that is sometimes played between IT and business functional groups such as, for example,

"It's all about getting access to multiple data sets, full integrations. Match rates suffer when too many are put together." ***VP, Sales/Analytics, MVPD***

"Expertise, qualified people. Sometimes need to go outside and rent data scientists through strategic partnerships." ***SVP, Media Research, Multimedia Network***

"Connecting data from different sources, hiring top talent. Creating just-in-time activation based on most recent data." ***COO, Digital DSP***

"Biggest challenge is finding quality human resources, gathering and reporting integrated touch point data." ***SVP, Director of Analytics, Media Agency***

"Categorization and governance of data in organization—who owns it?" ***Partner, Product Strategy, Big Data Syndicator***

marketing or advertising. One key issue is identifying which groups hold core responsibility for acquiring, storing, processing, and analyzing data. In simple terms, business has the questions and IT has the data.

Hottest Applications of Big Data in the Marketing and Advertising Industry. Respondents provided their input on what they felt were the most promising applications of Big Data in the marketing and advertising industry. Following are the most significant areas for Big Data deployment:

1. **Targeting and Addressability—**The uses of Big Data with the most immediate traction of Big Data would be a) improvements in precision of target audience descriptions, and b) the ability to communicate more selectively with target prospects through addressable channels.
2. **Multi-Touch-Point Insights—**The notion of creating a consumer view that traverses multiple channels and/or touch points was felt to hold a great deal of promise and is driving the growth of third-party data integrations.
3. **Real-Time Decision Making—**Some respondents mentioned that enhanced targeting and addressable capabilities, when combined with data access speed, will drive growth in programmatic digital advertising and eventually television.

Key Elements for Acceleration of Big Data Deployment in the Marketing and Advertising Industry. Respondents were asked what kind of groundwork is a must to advance Big Data in the marketing and advertising industry. No participants offered actions specific to the marketing and advertising industry; however, they reiterated most of the critical requirements for success in their own companies:

1. **Strategy and Executive Support—**Virtually all were in agreement about the need to understand the steps, benefit, and outcome of a Big Data initiative that is supported by the company's leadership team.
2. **Securing Talent—**Ensuring that you have sufficient data science power
   - In-house: Staffing talent within the company
   - Partnering: Access to data scientists that are available at many third-party data companies
3. **Persons Authentication** - There is recognition of the need to validate that consumer information has been captured accurately across multiple data sources; this checkpoint is critical for ensuring that the right people are targeted for marketing and advertising programs.

"Need understanding at the management level to make it into the business plan." ***SVP Media Research, Multimedia Network***

"It's about the business and use cases—having a crisp vision for what you want to get out of Big Data." ***Partner, Product Strategy, Big Data Syndicator***

"Critical that it's woven into your business processes so the organization knows what to do with it." ***COO, Digital DSP***

"Not about doing it yourself; picking the right vendors is critical. Establish a network you can tap into." ***Marketing Research Director, Telecommunications Company***

"Most important aspect - authenticated consumer IDs at the center . . . currently lack ability to authenticate back to a person . . . need privacy by design?" ***President, Consumer Insights & Targeting, Big Data Syndicator***

APPENDIX

PRIVACY RESOURCES

WOMMA: http://www.womma.org/ethics/privacy-guide

Federal Trade Commission- FTC Issues Final Commission Report on Protecting Consumer Privacy: http://www.ftc.gov/news-events/press-releases/2012/03/ftc-issues-final-commission-report-protecting-consumer-privacy

Children's Online Privacy Protection: http://business.ftc.gov/documents/bus84-childrens-online-privacy-protection-rule-six-step-compliance-plan-your-business

Federal Trade Commission-Children's Online Privacy Protection Rule: http://www.ftc.gov/enforcement/rules/rulemaking-regulatory-reform-proceedings/childrens-online-privacy-protection-rule

SELECT BIG DATA COMPANIES (SYNDICATORS)
Following is a list of some of the top firms engaged in the practice of Big Data and their website contact information. Marketers, media agencies, media networks and digital publishers often use the services of these companies for 3rd-party matching to develop richer consumer target segments or tracking of transactional consumer purchase data. Because these firms provide data to a wide cross section of companies, they've been dubbed "Big Data Syndicators." While there is some overlap in company offerings they've been bucketed according to specialty.

GENERAL FIRMS

- Experian www.experian.com/marketing-services
- Acxiom www.acxiom.com
- Epsilon www.epsilon.com

ADVERTISING CAMPAIGN IMPACT TESTING

- Kantar-Shopcom www.kantarshopcom.com
- Dunn Humby www.dunnhumby.com
- DataLogix www.datalogix.com
- Symphony-IRI www.iriworldwide.com
- Nielsen Catalina www.ncsolutions.com
- Neustar www.neustar.com

DIGITAL PUBLISHER-CONSUMER DATA

- Exelate www.exelate.com
- Bluekai www.bluekai.com
- Quantcast www.quantcast.com