Blue Books

Experts' views for expert investors

Produced by CRÉDIT AGRICOLE SECURITIES

USA

Technology

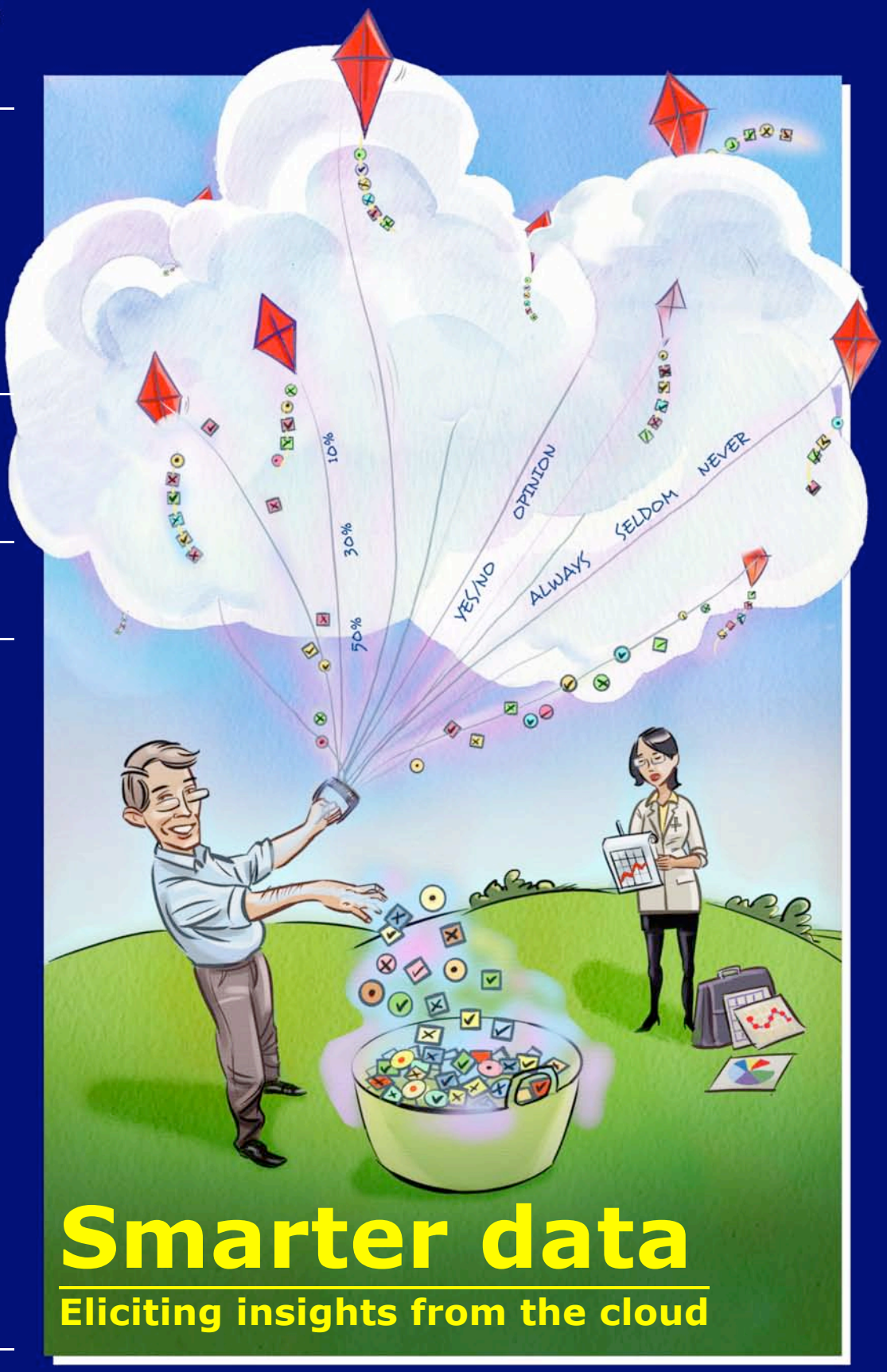
4 December 2012

Ed Maguire
Managing Director
Software Research
edward.maguire@clsa.com
(1) 212 549 8200

Guest authors

Neil Seeman
Founder and CEO
The RIWI Corporation

Eric Meerkamper
President
The RIWI Corporation



Smarter data

Eliciting insights from the cloud

www.clsau.com

Produced by CRÉDIT AGRICOLE SECURITIES

Straight to the source with CLSA

When industry innovations change as quickly as they are created, your ability to respond could mean the difference between success and failure. In this volatile environment, why rely entirely on broker research when you can tap into unfiltered, unbiased primary research?

CLSA U® is a value-added executive education programme created to allow you to gain firsthand information and draw your own conclusions and make better informed investment decisions.

CLSA U® offers tailored courses on a broad range of macro themes with a special focus on technology and telecoms. The format ensures you learn as we do and obtain firsthand information about prospects and trends in industries and sectors that underline the companies in your portfolio.

You will interact and learn from the trailblazers at the centre of today's fastest moving industries - experts, engineers and scientists who design, implement and shape the new technologies today, which impact the market tomorrow.

CLSA U® is not a one-off event. It is an ongoing education programme restricted to CLSA's top clients. The syllabus will constantly evolve to meet your needs and help you debunk the latest technologies, investment styles and industry trends that affect the markets and sectors you invest in.

For more details, please email clsau@clsa.com or log on to www.clsau.com

Neil Seeman

Neil is founder and CEO of The RIWI Corporation, a patented global data-capture technology company based on his invention, originally used for government-commissioned pandemic surveillance and drug-safety research. A former fellow of the free-market think tank, The Fraser Institute, Neil joined the University of Toronto Medical School and applied dynamic web-based performance and comparative reporting for over 150 hospitals. He joined IBM in October 2006, where he led research and consulting in social technologies in healthcare for clients around the world. Neil has authored over 800 essays, book chapters, editorials and over 30 peer-reviewed journal articles. He is co-author of four academic books, most recently, *XXL: Obesity and the Limits of Shame*, which was a finalist for the Donner Foundation's best book on public policy. His work has been profiled in *The Economist*, *The Washington Post*, *Business Week*, *CNBC*, *The Wall Street Journal* and other major media around the world.

Eric Meerkamper

Eric brings a wealth of experience in brokering data-capture partnerships with corporations, governments, thought leaders and non-governmental organizations (NGOs) around the world. He is former president and co-owner of DECODE Inc, a pioneering firm in using online and emerging research methodologies, which has conducted more than 500 consumer, employee and citizen insight projects for many of the world's largest companies, brands and organizations (1999-2012). Eric is co-founder and chair of the Centre for Social Innovation, which manages over 100,000 square feet of co-location space in Toronto and Manhattan. He is the former board chair of the Daily Bread Food Bank in Toronto and former VP of Bariston Inc (1995-99), an international market development firm, where he led Latin American market expansion.

Foreword

There's been a surge of interest in tools and techniques under the broad umbrella of "Big Data". With the increasing digitization of the world, the data generated by applications, interactions and transactions provide opportunities for businesses to gain deeper insights into how to improve sales, optimize operations and manage risk. Market research is traditionally a highly quantitative discipline. The explosive growth in terms of quantity, type and speed at which data are generated creates new opportunities and challenges for organizations seeking to glean insights that effectively guide their efforts.

Neil Seeman and Eric Meerkamper of RIWI discuss the crisis of credibility in the market research industry, the challenges of managing data collection, bias and interpretation, and the technology they have developed to address these issues. They provide a case study illustrating how their technology is applied to the energy-drink industry.

Market research traditionally relied on targeted surveys, focus groups and other approaches to collect proprietary information. While data were limited in scope, experienced analysts knew what to do with them. With the proliferation of online commerce and social media, the volume and variety of data available for analysis have surged, offering potential breakthroughs from the breadth of sources. In practice, there are significant practical problems - more data do not always mean better data. The analysis output is only as good as the underlying data set, and for market researchers, low-quality data result in "garbage in, garbage out."

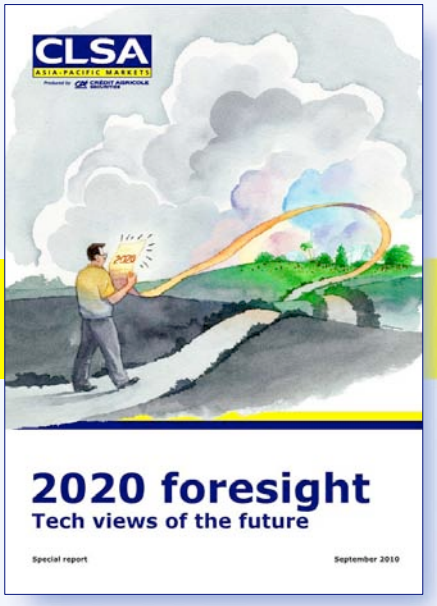
The RIWI approach involves lightweight "micro-surveys" of a handful of questions and seeks to avoid many of the pitfalls of traditional data gathering, which are explained in detail in this report. RIWI's surveys leverage the random nature of simply typing non-existent web address pages and can gather data from regions such as the Middle East and emerging markets where traditional data-gathering techniques do not have reach. Its approach is unique enough to have been awarded a patent, and in this report investors will get a sense of the potential power that can come from analyzing data collected in this way.

We are at the beginning of the Big Data era, and companies like RIWI are key to a new generation of insights, not just to predict consumer preference, but also to forecast important shifts in public sentiment and attitudes. RIWI's unique and proprietary approach paves the way for heretofore unattainable insights into customer preferences and behavior. For the considerable investments that hinge on these insights, potential ROI promises to be compelling.

Ed Maguire
Managing Director, Software Research

Contents

Looking to the future



Smarter data

Eliciting insights from the cloud

Market research is data dependent, and the industry is struggling to deliver valuable insights, overwhelmed by an explosion of data sources and a variety of analysis choices. RIWI Corporation's proprietary approach gathers unbiased data globally from targeted micro-surveys, enabling clients to derive insights, notably from countries where consumer data are difficult to access. This report surveys the landscape of Big Data and market research, showcasing RIWI's methodologies in an energy-drink case study.

Market research industry faces a "crisis of confidence" deluged with data and techniques

A bewildering array of new data sources includes traditional commercial panel surveys, computer-assisted telephone interviewing (CATI), focus groups, transactional databases, social-media mining and a surfeit of Big Data services in the marketplace. This CLSA U Blue Book discusses the important limitations of all traditional and emerging data-capture tools available today.

RIWI's patented data stream captures consumer insights across the world's micro-regions

The RIWI Corporation, which stands for "Real time, Interactive World Wide Intelligence", has developed a 24/7 data stream that taps into the pulse of thoughts, wants and purchasing intent of the everyday consumer in cities around the world with internet access. The approach is especially effective in capturing data in a privacy-compliant manner in emerging markets and hard-to-penetrate regions, such as nations in Africa, Asia and the Middle East.

We illustrate RIWI's approach through an energy-drink case study

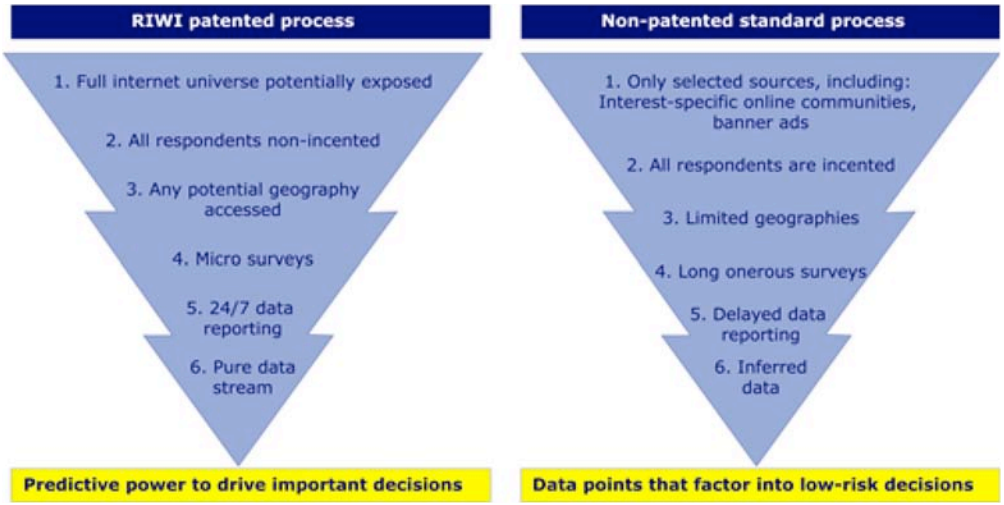
RIWI's novel approach to data capture is showcased in a case study across 11 countries. The study uses data to analyze the emerging power of energy-drink brands, comparing *Monster* to *Coke*. RIWI collects a "Passion Index", a "Preference Index" and a perceived safety reading across countries.

Data from fully representative survey respondents challenging to obtain

It is increasingly difficult to obtain data from fully representative survey respondents via the internet or more traditional routes. Though no data-capture scenario is perfect, the RIWI landscape imagines a reality where it *is* possible to predict scenarios hitherto considered impossible, such as long-tail negative or positive events in emerging markets.

RIWI gathers data from difficult to access consumers and geographies

Comparing the RIWI process with standard processes



Source: The RIWI Corporation

Too much data, too few insights

The volume and variety of data available for analysis have surged

With the proliferation of online commerce and social media, the volume and variety of data available for analysis have surged, offering potential breakthroughs from the breadth of sources. In practice, there are significant practical problems - more data do not always mean better data. The analysis output is only as good as the underlying data set, and for market researchers, low-quality information results in "garbage in, garbage out."

Credibility crisis in the global market research industry

Any market research of value must deliver demonstrable return on investment

The market research industry (MRI), US$31.2 billion in size as of 2010[1], is designed to better appreciate what consumers want, need or believe. Reliable market research is necessary to understand consumers and their preferences, identify opportunities for new brands to emerge and thereby increase profits, recognize and plan for industrial and economic shifts, monitor market competition and mitigate risk in business decisions.[2]

There are a growing number of primary data-collection tools, for example, phone surveys, in-person focus groups, online focus groups, crowd-sourcing, online panels, insights driven through games ("gamification") and surveys on social-media sites, River Sampling, social-media analysis and data harvested from mobile technology. Details on the opportunities and limitations of these methodologies, in the context of a world of ubiquitous data, will be explored throughout this Blue Book. Whatever the methodology, success hinges on its ability to produce results that predict or drive key outcomes.[3] In other words, the customer (the person who needs the information) is always right: the research must be able to answer the customer inquiry and deliver demonstrable return on investment. This is the metric that matters.

Basic issues, such as response rates, are declining in the market research industry

Recently, there has been a crisis of confidence within the market research industry. Serious questions have been raised about the quality of standard operating practices used by some of the leaders in global data capture, with a particularly acute concern about the reliability of online panels.[4] This is of note as the industry moves increasingly online for reasons of cost efficiency and to ensure they intercept people where they are most active.

With increased availability of ultra-broadband and fast network speeds, there is easier access to online resources, and people spend more and more of their time online.[5] This, in conjunction with increased mobile phone use and less landline use, has translated into a steady decline of contact, cooperation and response rates in phone-based surveys; since 1997, response rates have declined steadily in the USA. This is in part due to the fact that, as of the end of 3Q12, mobile use (generally inaccessible to surveys) had surpassed the one billion mark, specifically 1.038 billion. (Gartner predicts 1.2 billion smartphone devices and tablets to be purchased in 2013).[6,7]

In 2012, only 9% of US households that had been contacted for a survey had ever responded to a phone poll, with the definition of "responded" meaning simply answering the phone and beginning the interview process; it did not mean that the respondent actually completed the survey (see Figure 1)[8]. Meanwhile, some publicly reported industry response rates for internet intercept surveys are under 1%.[9]

Challenges in data integrity, data representativeness and innovation

Deriving insights from online tools therefore faces a three-pronged challenge: a notable decline in data integrity; increased difficulty in finding representative samples, or "true respondents" from around the world; and a lack of innovation in addressing the increased need to capture predictive data. Each of these challenges will be discussed in turn.

Declining integrity of data

> *'Clearly something's wrong. I am concerned about the impact on the credibility of the profession.'* [10]
>
> Bruce Cameron, Head of Return on Insight, a polling firm

Not all companies have the methodological rigor that can ensure data accuracy

Historically, the perception among lay people that market research is a complex and quant-heavy industry had reinforced the idea that the data produced by major branded polling companies were highly reliable. Their integrity appeared unassailable. On closer inspection, however, questions arise. Do all companies have the methodological rigor, or more importantly, the openness to revealing limitations that can ensure data accuracy and integrity?[11]

Response rates used by some companies to plan brand or counter-branding strategies, sometimes as low as 0.01%, are rarely reported. Even the oft-repeated phrase, 'this poll is accurate 19 times out of 20' - a reference to the 5% confidence interval for a traditional survey reported in any mainstream media outlet or MRI report - is increasingly subject to a major caveat, seldom noted, as in 'If this sample were reflective of the general [internet/telephone, etc] population, then the results would be correct 19 times out of 20, +/- x%.'

Five percent of professional, incented panelists participate in over 50% of all polls

At the May 2011 Market Research and Intelligence Association Conference[12] held in Kelowna, British Columbia, Canada, a leading MRI executive reported that, among major firms, approximately 5% of incented, enrolled panelists participate today in more than 50% of the industry's surveys. Attrition rates of these same "professional panelists," who do, in fact, respond to polls on anything from fabric softener to technology preferences, range from 80-100% annually. This information comes from the authors' interviews with global leading methodologists at major market research firms and former executives of some of the leading panel-supply companies that service market research firms.

It is getting increasingly difficult for researchers to reach cooperating survey participants

Figure 1

Surveys face growing difficulty reaching, persuading potential respondents (US data)

(%)	1997	2000	2003	2006	2009	2012
Contact rate (% of households in which an adult was reached)	90	77	79	73	72	62
Cooperation rate (% of households contacted that yielded an interview)	43	40	34	31	21	14
Response rate (% of households sampled that yielded an interview)	36	28	25	21	15	**9**

Note: Rates computed according to American Association for Public Opinion Research (AAPOR) standard definitions. Rates are typical for surveys conducted in each year. Source: PEW Research Centre 2012 Methodology Study

In its spring 2012 report on the global market research industry, the 11th edition of the GreenBook Research Industry Trends Report, which surveys the largest market research clients and suppliers in the world, found a growing disconnect between client needs and supplier capabilities to deliver quality data in a timely manner. The report[13] noted:

'The pattern that emerges is that clients/buyers are insisting that ever-more stringent deadlines be met, and expecting quality to remain high despite these pressures. Suppliers/providers are meeting the deadlines out of necessity, but are concerned that *quality must suffer as a result.* The key takeaway here seems to be that to clients/buyers, quality is akin to table stakes, and speed is a business necessity. It's up to suppliers/providers to meet both these expectations.' [Emphasis added]

The Achilles' heel in data capture: Coverage bias

Results that reflect the population sample targeted are harder and harder to reach

'The dirty little secret of the polling business is that our ability to yield results accurately from samples that reflect the total population has probably never been worse,' says Allan Gregg, who is chair of the respected market research firm, Harris-Decima.[14] In order to identify a representative random sample of the larger population by any public opinion poll or data-capture technique, whether conducted online, by print survey or by telephone or any other method, one needs to eliminate as much as possible the coverage bias of those targeted for polling.

For coverage bias to drop, respondents need to be randomly intercepted . . .

Coverage bias is eliminated when every potential respondent in the entire target population (eg, online population) has an equal probability of being randomly surveyed. The online population was 2.3 billion as of the end of 2011[15], and there will be nearly four billion internet users, more than 40% of the world's projected population, by 2015. [16] This internet population, albeit limited by access to computing devices in some regions of the world and, generally, more limited in the lower socio-economic strata, is increasingly the target of all major data-capture companies, except for in-person focus groups or targeted telephone interviews (eg, exit interviews of company executives).

. . . with equal probability of being exposed to the question(s) of interest

Unless the entire population (eg, national or regional) is approached, it is generally considered impossible to target potential respondents equally, with everyone having an equal and randomized probability of being asked to complete a given survey. For coverage bias to be eliminated, the potential respondents do not need to answer the question; however, they do need to be intercepted or exposed to the question with equal random probability, in (and this should go without saying) as unobtrusive and privacy-compliant a manner as possible. Robotic telephone calls disturbing the dinner hour, an increasingly common approach, does not have a representative reach nor is it unobtrusive in its manner of intercept.

Explosion in mobile telephony and call display means fewer people are eligible to be interviewed

In the context of telephone surveying, for example, substantial coverage bias creeps into the survey since people with mobile phones are less accessible to the surveyor than are other potential respondents; people who work outside the home are less accessible than are other potential respondents who stay at home during the work day; furthermore, the rising number of individuals who rely on call display to block unwanted companies from reaching them by telephone are also excluded as potential respondents.

Most major do-not-call laws in developed countries exempt polling companies; however, it is the authors' view that rising consumer activism and frustration with unsolicited calls may well result in more aggressive rules at national and sub-national levels removing such exemptions.

Arbitrary measures to produce a "random" sample: Traditional methods

Traditional probability-based approaches are very subjective and biased

The end goal for any surveyor is to obtain a representative random sample of the population of interest (eg, Canadians, Britons, or Australians who use the internet) in the final group of respondents. Other than coverage bias, there are a number of other biases in surveys that statistically differentiate the non-respondent pool from the respondent pool - notably, the fact that some individuals sharing certain characteristics (ie, sex, income or psychological profile) may be more willing to answer a certain type of survey than will others.

There are a number of approaches to reduce, but not totally eliminate, biases

A survey that seeks to be of good quality has historically engaged in a number of approaches to reduce such biases, but they can never be totally eliminated. One bias reduction method is to increase the sample size of those polled; this, in turn, reduces the margin of error of the final result, or the probability that the result observed is due to chance alone. Another approach commonly used is a type of multistage sampling or cluster sampling, where the surveyor assumes a number of variables a priori that can potentially affect the outcome, such as geographic area, sex or age; the surveyor then proceeds to survey a representative number of people based on these arbitrary and subjective a priori variables of interest, from one geographic area, or cluster, before moving to the next cluster (eg, based on area code, in the case of phone surveys).

It is important to note that the a priori variables, such as age and sex, are arbitrary and do not accommodate the full complexity of human diversity. Just because these variables have historically been used to segment data do not mean they are the most relevant to a particular industry. The more neutral the surveyor is in assessing the variables of interest, the more random the survey will be.

Adjusting respondent data after a survey is based on subjective post-processing of the data

An additional common method of reducing the margin of error is stratification. After the data have been collected, the surveyor seeks to correct for a number of variables that could be potentially skewing the final results. In the stratification approach or in the cluster sampling approach, the possible criteria for which the surveyor corrects, again, are inherently subjective; the most commonly used criteria in political polling are income, age and sex. Since the surveyor is gathering new information, it is, by definition, impossible for him or her to know all the possible variables that, in any particular enquiry, require adjusting.

Cluster sampling and stratified sampling are sometimes confused. In cluster sampling, the cluster is treated as the sampling unit so that the analysis is done on a population of clusters. In stratified sampling, the analysis is done on elements within strata. In stratified sampling, an algorithmically defined random sample is drawn from each of the strata, whereas in cluster sampling only the selected clusters are studied. An objective of cluster sampling is to reduce costs by increasing sampling efficiency; with stratified sampling, an objective is to increase precision.

Internet polling may be especially prone to bias

Because of these substantial challenges to obtaining representative random samples, all forms of polling - notably online polling - have been criticized. Online polling, despite being the dominant method by which companies now collect data from respondents, is especially prone to bias since there is often little randomization, if any, in identifying potential survey respondents. Online respondents who, for example, respond to a political survey on a partisan media website, are, by definition, overly representative of people interested in that particular political news issue on the news media site. Google

Consumer Surveys™, announced in March 2012[17], show bias since they only capture a small portion of the individuals who frequent the sites where the surveys are hosted, and moreover the potential respondents might be presented with a "paywall" prior to answering the survey (Google calls this a 'survey wall'), yielding an incentive such as a discount for, or access to, some of the site's services in compensation for answering the survey.

The internet is now at least a good a source as traditional methods

In its white paper, Google researchers, whose micro-surveys are limited to the USA (as of the time of writing), make the compelling argument that the internet population is comparable, in terms of geographic and other elements of diversity, to other traditional populations targeted by pollsters.[18]

An inherent English-language bias and skew towards incented, younger users

Imagine the future: Endless router of fresh, random daily respondents

No matter the standard MRI approach, it generally requires the recruitment of a large group of potential representative respondents, which is exceptionally challenging online. The challenge could be overcome if one could create a system whereby the universe of potential respondents surveyed each had an equal random probability of taking the survey. Inevitably, there would still be some bias in that those who actually responded would be those who had more time on their hands or were more inclined to respond to the particular survey for whatever reason (ie, "leisure bias"); there is also the inherent English-language bias and skew towards youth of the typical internet user. However, if one could equalize the probability of any internet user being exposed to a particular survey, one would substantially increase the likelihood of obtaining a representative global sample of internet users prior to adjusting, ex post, for any additional biases or non-random effects.

The number of people needing to be surveyed in order to achieve a representative random sample would drop, as would the number and complexity of the possible stratifications, or risk adjustments (eg, for age, sex, psychological profile, etc) to be done after the survey data have been collected. Such a process invention would dramatically reduce the post-processing time and labor that analysts would need to conduct and evaluate a scientifically valid online survey based on responses from a representative random internet sample.

Furthermore, one could apply this methodology, without using the inference-based limitations common to social-media analytics, to every country and city around the world, as RIWI does across 11 global markets.[19]

Predictive political polling is increasingly discredited around the world

Dewey trumps Truman (again and again)

'Dewey Defeats Truman' references the infamously inaccurate top headline on the front page of the *Chicago Tribune* on 3 November 1948, the day after US President Harry S Truman bested New York Governor Thomas E Dewey in the 1948 presidential election. Yet there are myriad modern examples that showcase the challenges of modern polling. Polls taken during recent political elections, such as those by two major firms in Ontario, Canada (population 13.4 million[20] people), were in vehement disagreement about which party leader led the popular vote, and announced that '*someone* leads in Ontario'[21] [emphasis added]. Some pollsters have been considerably off the mark when they claimed that particular parties, such as Alberta, Canada's fast-rising Wildrose Party, were set for 'sweeping majorities' and instead, after the fact, won only 17 of 87 seats.[22] Some 2010 UK election predictions were significantly distorted as they overstated the Liberal Democrats' vote.[23]

In the recent 2012 US presidential election, the most rigorous poll taken just prior to the election - done by Gallup - polled 2,700 likely US voters, and its findings were outside the margin of error as compared to the final results (Gallup had put Governor Romney up by one percent nationally). Relying on any one poll, as countless media reports had pointed out, would have been folly, since these were so varied, especially in the so-called "swing states". President Obama ended up winning by 2.4 percent (as of the time of writing), a statistically significant difference from the average of all the national polls aggregated by realclearpolitics.com, which had predicted a national popular vote differential favoring President Obama over Governor Romney by just 0.7 percent.

Predictive polling's accuracy began to be challenged as long as seven years ago

The fact that predictive polling is often inaccurate emerged in the academic literature as long as seven years ago. According to one 2005 journal summary of the situation in Europe, 'Large errors remain the norm in polls pertaining to European Parliament elections, suggesting a shared inability of Portuguese polling organizations in dealing appropriately with the problems caused by low turnout and "landslide" elections.'[24]

Major media and the public increasingly skeptical of market research methods

Given the recent media attention to polling errors, and particularly to the recurring defensive refrain within the MRI and the media that report on such surveys that 'polls are only a snapshot in time,'[25] many have begun to question the long-held belief that traditional market research is a reliable tool with which to gauge popular sentiment. The same challenge has confronted the capacity of social-media-sentiment analytics to predict elections (explored in more detail later). What is to be done? Is George Horace Gallup's early 20th century vision of predictive polling impossible? Is data integrity in a phase of inexorable decline?[26]

Inability to reach true respondents online

> *'Innovation and simplicity are often the best way to attack a complex problem.'*[27]
>
> Google White Paper (March 2012)

Many prominent approaches used to gain customer insights are highly questionable

The challenges of securing an even modestly representative group of the internet population have long plagued the online panel community. "River sampling" is one technique - surprisingly, now in common usage perhaps because of scant media attention to its extreme limitations - and it has proven to be an ineffective strategy. In brief, river sampling refers to gathering potential survey recruits using banner ads, pop-up ads, pop-under ads and similar instant "capture" promotions (eg, on shopping sites or on social-media sites such as Facebook™, itself a biased community in terms of user profile, and "power users" of Facebook - an even more biased sub-community).

Facebook implicitly acknowledged the growing limitations of its own micro-survey tool

Facebook, in fact, implicitly acknowledged the growing limitations of its own micro-survey tool, Questions™, which originally launched in July 2010 and was shuttered on 19 October 2012. On Facebook, the Questions Dashboard and option to share a poll to the news feed have been removed. Still, on Facebook walls, individuals, biased as they may be, who volunteer to participate in commercial panels, are thereafter screened for their reported demographic characteristics and then "randomly assigned", or "routed," to the appropriate survey. From here comes the metaphor of being captured from the "flowing river" of online persons. And yet, '[the] participants cannot be described as representing any larger defined population.' Adding a river

sample to an opt-in panel sample does not improve the forecasting of results, nor does it ensure that the two samples are different, as the same respondent could be in a panel and in a "river."[28]

False promise of data mining from social media and the open-source web

Social-media analytics do not deliver true consumer insights from "social chatter" online

Some analysts have chosen to incorporate data harvesting from social-media sites (eg, micro-blogs such as Twitter™ or Facebook). The business case for Big Data is as follows: significant amounts of "social chatter" online have provided vast opportunities for firms to gain insight into people's minds. Hence, by clever use of information extraction methods from Big Data, organizations can deliver answers to essential management questions such as, 'Where is the market heading for product x?'

Social chatter contains excessive data noise

The problem is that, by definition, these processes require extensive data management, the use of automated solutions, deep-dive analysis and external consulting from a new profession of "data scientists". More importantly, social chatter contains excessive data noise that makes it ineffective in capturing true user sentiment.[29] In fact, the potential data explosion concerns senior executives. IBM conducted its own survey of approximately 1,700 chief marketing officers, asking them what challenges they are facing today. Crunching data points and churning them into the promise of real insight ranked at the top.[30] As the political forecaster Nate Silver has noted, 'Big Data will never replace thinking or hypothesis-testing.'[31]

In order for a company to harness the value of Big Data, it needs three advantages: a) access to proprietary data; b) knowledge of what to do with it and how to process it; and c) the right relationship with the consumer in order to apply the data. Our view is that, even if Big Data offers insights of the kind its evangelists promote, very few firms possess point a, and, especially point b. For those companies that do invest, it will require significant costs in services to support Big Data efforts - about 20 times higher than software purchases. People with these consulting skill sets are rare and in high demand.[32]

On the openly accessible web, data are ubiquitous, and the rate of growth is breathtaking. By 2013, IBM estimates humanity will be generating five exabytes of data every 10 minutes; that compares to five exabytes of information created from the beginning of time to the year 2003.[33] This trend is accelerating as leading "open-data" cities such as San Francisco have opened up more than 200,000 once-closed data sets (eg, mapping urban revitalization projects, or directions to local parks and points of interest) to public view.[34]

Independent analysts question the abilities of Big Data to make algorithmic predictions

Despite the public benefits of this ocean of new data, independent analysts, such as Gregory Piatetsky-Shapiro, writing in *Harvard Business Review*, have questioned the abilities of Big Data to make algorithmic predictions about who will purchase certain goods or services, notably film purchases on Netflix™ (drawing on a data set of 480,000 anonymized users). In fact, Netflix introduced a competition to determine who could best use an algorithm to predict user trends based on the user data set (2007-09 data); the winning algorithm that matched the statistical demands of Netflix took three years (by the BellKor's Pragmatic Chaos team), yet the team's solution was so complex that Netflix never implemented it.[35] As early as 1999, a study exposed the challenges of open data to enable marketers to improve predictive accuracy using modeling activities to achieve the "'promised land" of purchase-intent insights.[36]

On Facebook alone, at least 30 billion pieces of content are "shared" among "friends" each month.[37] Facebook is pulling in billions of "likes" and 300 million photos per day.[38] On Twitter, the average number of Tweets sent per day in 2011 was 140 million.[39] A 21 January 2012 commentary in *CNN Online*, in an article, *Why social media couldn't predict the Iowa race*, stated that '[g]rowth in followers or high numbers of re-tweets are just an indication of notoriety or celebrity. Saying simple, stupid things that lots of people want to tell their peers about can get you tons of followers and re-tweets. But it doesn't mean anything definitive about grass-roots support.'

Unclear if current tools can effectively analyze complex linguistics

The analytical challenge is that language on such sites is far more complex than the media, some social-media-analytics companies or other observers would suggest. Analysis of how language is used would need to evolve dramatically and it is unclear if our current technological capture tools can create a picture of the complexity of the linguistic environment (eg, ever-changing notions of satire, short forms, nuance and cultural distinctions in language usage even in the same country). Just as important, in fields such as healthcare, mining open access application programming interfaces (APIs) or blogs or social media, shows that "power users" (ie, activated, literate "e-patients") are the people self-broadcasting thoughts that are later examined by data scientists.[40]

Thousands of tools to mine people's opinions and identify market "influencers"

There are thousands of tools in the current market to mine these people's opinions and identify market "influencers". An example is through the Appinions[41] platform (formerly known as Jodange), where 'organizations can identify, analyze, engage and measure relevant influencers.' According to the company, the system extracts and aggregates opinions from more than five million sources, including blogs, social networks, forums, and newspaper and magazine articles. Prominent paid tools include Sysomos™ and Radian6™.

On 19 October 2012, Toronto-based TrendSpottr™ announced a partnership with Salesforce.com™. The partnership strategy's goal is to mine social media to spot themes, "memes" and topics that might help predict emerging social-network gaffes. In August 2012, Facebook acquired Threadsy™, developer of the Swaylo[42] social-analytics tool that connects "influencers" to brands. Threadsy combines and aggregates e-mail, Facebook, Twitter and the broader social web into one interface. The Swaylo solution is used to measure the influence of users in their social network on a scale of zero to 10 by analyzing how their online activity impacts activities within their hub of "friends", or what is commonly now referred to as a "social graph". A new tool called Opinsy™, also open to text mining, is a platform which lets people build viral communities and followings around their opinions.

Power users, not necessarily everyday consumers, determine the popularity of products

Such data mining and "influencer" (notably Klout™) applications are being used in multiple sectors, including human-resource-recruitment activities and in the consumer-packaged-goods sector; yet disproportionate numbers of power users, not necessarily everyday consumers of brands, are the people online who promote or demote the popularity of people and products. Services such as Klout purport to measure influence based on anyone's ability to drive action across the social web. This necessarily excludes people who may wish to deliberately avoid the social web; as such, influencers such as CEOs of global companies or hedge fund managers would likely be excluded.

More problematic, the proliferation of "spam-bots" and paid marketers who explicitly (and robotically) promote company brands on sites such as Twitter means that mining antecedent data using such tools can yield the

Sarcasm is a huge obstacle for researchers and marketers

very opposite of information one may wish to infer. For example, interpreting the large amounts of sarcasm online is guesswork. Since sarcasm means stating the opposite of what is truly meant, sarcasm is a huge obstacle for researchers and marketers who create computer programs and natural language algorithms to analyze massive swaths of online chatter to gauge public opinions about products. Many people use emoticons (eg,:)) sarcastically; an exclamation mark may, in fact, convey a lack of enthusiasm.

Further, even if you can sort through this noise, all you will know is that your power users may love your brand or the brand you are analyzing, but the medium is a mess of information. It's like gathering insights about a popular beer at a college keg party by listening to the loudest and groggiest. Nevertheless, sites such as Twitter and Facebook, given their large size (despite the fact that some 9% of Facebook accounts are fakes) can be used effectively for exploratory work of niche populations (such as gamers).[43]

Limits of sentiment analysis have been well demonstrated . . .

The limits of sentiment analysis have been well demonstrated in many applications and tools. *Reuters'* new Social Dashboards™ creates a fascinating social stock index that overlays sentiment analysis on top of stock performance of the top-10 most discussed publicly traded companies to access regular people's perceptions about different brands and businesses. However, there is an admitted limit in the scope of such tools, especially given the capricious fluctuations and complexity of equities: 'Our audience,' reports the company, 'is made up of sophisticated news users, who are not just interested in what's trending on Twitter but how social can inform their business lives.'[44]

. . . because social-media sentiment is not randomized

With such tools, the consumer sentiment expressed is not representative of the internet population; it is not randomized. It would be like evaluating text on Twitter for the words "housing" and "bubble" and then trying to divine from this word combination - using some type of linguistic algorithm that ensured the adjacency of these two words - whether a rising incidence of Twitter users feel that housing prices are a "bubble" ready to pop.[45] Social-media sentiment is not randomized and thus limits the widespread applicability of sentiment analysis.[46]

Inability to capture predictive data; the promise of Big Data

> *'The accuracy of predicting future events is not reliant on Big Data, but the right data and understanding what data points to extract and use that provides the highest confidence indicators to predicting the customer's behaviour. I've found most companies may have "'Big Data", but either it's garbage data, decentralized or collectively unusable in its current format (due to no forethought).'* [47]
>
> Kael Kanczuzewski, Director of business development, GoLIMS

Five percent of professional, incented panelists participate in over 50% of all polls

Historically, the perception among lay people that market research is a complex and quant-heavy industry had reinforced the idea that the data produced by major branded polling companies were highly reliable. Their integrity appeared unassailable. On closer inspection, however, questions arise. Do all companies have the methodological rigor, or more importantly, the openness to revealing limitations that can ensure data accuracy and integrity?[48]

Intense investor interest in Big Data

As *Computerworld* reported in May 2012[49], corporate demand for tools that can manipulate and analyze massive volumes of structured and unstructured data has caught investor attention. Major US venture-capital (VC) firms in 2012 have invested hundreds of millions of dollars into companies selling a suite of Big Data technologies. VC firm Accel Partners recently established a US$100 million fund to finance the early stages and growth of Big Data companies. "Crystal Ball" software, Palantir.com (tag line: 'Palantir connects people who care to the data that matters') is just one hot VC-funded Big Data company. Why is there such intense investor interest in Big Data?

Accel Partners is not alone. VC and angel investors are increasingly investing in Big Data because of the emergence of the next generation of the internet, a profound disruption demonstrated by the manner in which the biggest internet companies, such as Google, Amazon, Facebook and Twitter, are improving how to target consumers. Google is increasingly focusing on "social search", notably its Knowledge Graph™, which is amenable to data mining (and contains more than 3.5 billion facts about relationships among public data).

The Wild West of new startups seeks to harness and make sense of unquantifiable databases

As data centers with massive cost efficiencies emerge, it is the "Wild West" of new startups that seek to make the case to investors that they can harness and then make sense of unquantifiable databases through new, highly scalable software architectural models such as Hadoop™ and Cassandra™.

As noted, the predictive promise heralded by promoters of Big Data, if decoded correctly, has encouraged a strong push to develop tools and strategies to tackle the unprecedented amount of data companies have at their disposal. However, as co-founder and CEO of Tellagence Matt Hixson notes, when analyzing data that have a social component, it is insufficient to look at it with a traditional lens, typically characterized by fixed environmental variables.

Common challenges of Big Data include serious interpretive concerns

Social data provided by individuals (eg, interests, behavior, etc) are constantly changing, and predictive tools will need to account for such dynamic variables.[50] One question, therefore, is how to tap into changing relationships and behavior while navigating the common challenges of Big Data sets (including data noise).[51] In the context of previously collected data sets in healthcare, this has been a serious interpretive concern. For instance, the analysis of historical data sets of drug response in depressed patients would miss the fact that men and women respond differently.

Confounders that have unintentionally crept into the data set can result in false conclusions

Because most patients with clinical depression are women,[49] adequate attention would not have been paid to appropriate recruitment of men and women. Not understanding confounders that have unintentionally crept into the data set can result in false conclusions. Early HIV data sets, for example, did not take into account the influence of intravenous drug use.[52] To understand health conditions, prospective studies or forward-looking data sets are best.[53]

Big Data: The solution to the global market industry's woes?

Just because correlations exist does not mean they mean anything

Big Data has been rife with many definitions. For the purposes of this CLSA U Blue Book, we mean it to refer to very large unquantifiable - and unintelligible - data sets. Herein lies what we call the "John Nash Problem". Mr Nash was the 1994 Nobel Laureate in economics, specifically, game theory, differential geometry and partial differential equations. With Big Data - often called by casual observers "Big Data analytics",[54] which technically refers to the analysis of large terabytes and exabytes of information freely available on the

web or on transactional databases - 1,000 geniuses in a room would not agree on what was statistically significant, since there would be, with statistical certainty, myriad positive and negative correlations. With too much data come too many correlations of presumed significance. The challenge, therefore, lies in asking the right questions, the preserve of experts in any given narrow field, not in the much easier task of mining unquantifiable databases and presenting customers with correlations and linear regressions.

Cognitive bias can badly impair one's understanding

What correlates with what is often interesting but may not be commercially meaningful. Cognitive bias, specifically the natural human impetus to impose meaning and structure on unstructured data, can badly impair one's understanding of the ever-growing complexity of a world in constant flux, or chaos. As is implied in John Nash's observation cited below, we need to be humble in the face of the complexity of human decision-making patterns, and we need to ask the right questions, rather than to impose our own patterns onto a complex and adaptive background.

We need to ask the right questions

This is the difference between astronomy and astrology. Astronomy asked questions and obtained useful answers. Astrologers obtained answers as well: Aries the ram, Taurus the bull, Gemini the twins, etc. But these answers did not lead to true prediction except by random chance.

> *'One aspect of this is that rationality of thought imposes a limit on a person's concept of his relation to the cosmos.'*
>
> John Nash, 1994 Nobel Laureate in Economics

Solutions for where we are now

Seeking new approach that avoids many pitfalls of traditional data gathering

The RIWI approach involves lightweight "micro-surveys" of a handful of questions and seeks to avoid many of the pitfalls of traditional data gathering. Its surveys leverage the random nature of simply mistyping a common web page address and can gather data from regions such as the Middle East and emerging markets where traditional data-gathering techniques do not have reach. Its 24/7 data stream taps into the pulse of thoughts, wants and purchasing intent of the everyday consumer in cities around the world with internet access.

The internet is now the common interview interface tool

As noted, with the increasing availability of ultra-broadband, even in relatively poor nations such as Peru and India, digital networks have begun to grow with accelerating speed. People are spending more time online, and, as more people migrate online, third-party access to data becomes more accessible[55]. The rapid consolidation of panel-supply companies that service large market research firms is an attempt to capture economies of scale as the "cost-per-complete" per interview that these firms can charge clients continues to drop.[56]

Large market research firms are also on an acquisition spree for new mobile and other devices to support intelligence-gathering on what consumers might like based on do-it-yourself (DIY) surveys and audience-impression information gleaned from mobile-enabled platforms (eg, Kantar's July 2012 purchase of UK-based Lumi SURVEY[57]). Telephone surveys are in fast decline as landline use decreases, call display emerges as a household necessity and mobile telephony remains relatively immune to telephone-survey intercept. From a purely economic standpoint, it thus makes sense that the internet is now the common interview interface tool since it reduces the costs and increases the ease with which survey respondents can be recruited.

The importance of mining web data has risen in significance . . .

The importance of mining data on the web has risen in significance. Companies such as Gnip[58] and Data Shift Inc analyze and categorize tweets and sell data packages based on tweet feeds directly from Twitter, with a range of 30 days to as far back at two years (as of the time of writing). Twitter, despite representing a biased type of internet user, will share tweets with the US Library of Congress. Purchasers of these data packages are able to isolate tweets based on topic (algorithmically based) and geographic location; it amounts to, at least in its commercial promise, holding a huge number of simultaneous, sporadic focus groups on brands or products. Gnip has reported that the information collected can also be used during natural disasters to help rescuers, to monitor illnesses such as flu outbreaks, and to analyze stock-market sentiment.[59]

. . . increasing the appetite for better decision-making IT tools to drive analytics

Despite the emerging awareness of the limitations of the data-capture tools described thus far in this CLSA U Blue Book, there is increasing appetite for data for better decision-making using IT tools to drive analytics. Gartner projects Big Data to drive US$232 billion in spending through 2016. It will further drive US$96 billion of global IT spending in 2012 and US$120 billion of IT spending in 2013.[60] At the same time, global economic uncertainty has made it increasingly difficult for companies to provide detailed 24/7 tracking or forecasting.

Some businesses are trying to access the supposed value in machine data in order to get answers quickly and, as such, are encountering the "John Nash Problem" referenced earlier. As today's decisions in the business world need to become more real time, the core systems and emergent profession of data

scientists who support those processes need to keep pace. It is only natural that data warehousing, business intelligence, decision support and related data-crunching systems and in-house analytics services begin to incorporate real-time data into central operating systems.

Many businesses now able to invest in software

Many CIOs at cash-rich companies in the USA spent the last three to five years installing complex systems from big vendors that handle tasks such as balancing the general ledger and tracking shipments. Those systems are in good shape now, leaving many businesses free to invest in other areas, notably software that analyzes data for trends. Splunk[61] is one such popular company, first introduced to the market in April 2012.[62] The multinational carrier Telefónica's new business unit, Telefónica Digital, recently announced Dynamic Insights, marking its formal entry into the Big Data marketplace. Mobile operators have impressive repositories of data in their businesses - not only from human activity on cellular networks, but also from WiFi networks.[63]

Beyond Big Data and towards smarter data

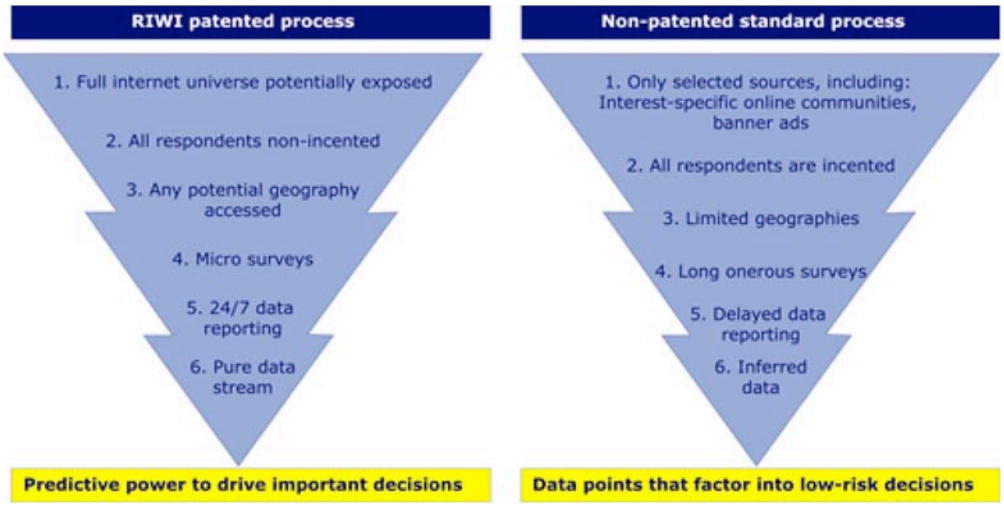
Leveraging the power of the URL bar

One approach to this new environment is RIWI, which stands for "Real Time Interactive World Wide Intelligence." The RIWI Corporation leverages the power of the URL bar (URL stands for "Uniform Resource Locator"), the fundamental architecture of the internet, with each country having a top-level domain (TLD), and with an evolving number of new TLDs in the process of being approved by the operational non-profit Internet Corporation for Assigned Names and Numbers (ICANN). The common nomenclature of the URL bar upon which domain names reside is 'http://' ('https://' in the case of secure sites). The vast majority of TLDs (approximately 98%) end in .com; as of October 2012, there are more than 142 million domains across the .com, .net, .org, .info, .biz and .us top-level domains.[64] These domains are generally rented from private domain registrars, including Godaddy™ or Register.com.

Comparing RIWI to traditional data-capture practices

Figure 2

RIWI approach to random global data capture vs standard quantitative practices



Source: The RIWI Corporation

The RIWI Corporation is thus far the only company to win a data-capture polling or survey-business-methods patent, including 21 related claims, since the US Patent and Trademark Office tightened its rules after the 2010 Supreme Court ruling (*Bilski* v. *Kappos*, 2010).[65] The company owns or rents

thousands of "nonsense" domain names, that is, real full-page URLs. RIWI refers to these domains as "broken" domains; this means they are not misspellings of trademarks.

Legal and privacy ethics compliance is assured

Legal and privacy ethics compliance is assured since the domains, being real domains, get used for a legitimate business purpose (micro-surveys or "snap-surveys" or "nano-surveys"). When someone enters one of these nonsense sites manually on the URL bar, a quick exit (for those not interested in the survey) is proffered. The site contains privacy disclosures assuring users that all data will be kept anonymous in perpetuity. There is no trickery. Users are not deceived into thinking (via images and/or words) that they have entered popular domains (such as Facebook, Twitter or Youtube™). Entry into these rotating domains is through typing errors via manual type-in made inadvertently by very large samples of random web users from around the world.

Two-thirds of web users have used direct manual navigation to a URL

In essence, the broken domains are strings of "nonsense" letters, or domains people think may exist but do not exist. Their usage is rising as manual type-in for a wide variety of reasons: notably privacy and security sensitivity; the fast rise of new TLDs (such as .hotel or .law, offered to companies that make bids for such URLs to ICANN); and the desire to avoid phishing from hidden links or malware embedded in emails, which increase on the internet every day. Foresight Research[66] estimates that approximately two-thirds of all web users at some moment or another have used direct manual navigation to a URL. This fraction is rising as new reports emerge about significant portions of redirects from major search engines, such as Bing, that result in exposure to malicious sites.[67]

Manual type-in is an essential measure to protect against malware and phishing attacks

To avoid malicious threats, users more frequently delete their "cookies" (leading the user, thereafter, to more frequently type in URLs of interest manually). When LinkedIn™ announced that 6.5 million of its users experienced potentially compromised LinkedIn passwords, it instructed users to re-register new passwords using manual type-in[68] within the URL bar. Similar concerns, together with the rise of the "https" designation (for more secure sites, notably online banking) are factors causing people to manually type their designated destination site into the URL bar. As "ethical hacking" expert, best-selling author and world-acclaimed computer security and digital intelligence consultant, Ankit Fadia, told the audience at the 19th annual CLSA Investors' Forum in September 2012, manual type-in is an essential measure to protect against potential malware and phishing attacks. Across all browsers, and all web-enabled devices, manual type-in is rising in tandem with web growth, according to RIWI statistics. The URL bar is like its own search bar, a fierce competitor to Google and Baidu™, and one that may potentially surpass them in usage frequency.

URL bar is increasing as a web browser of resort for manual type-in

The URL bar is increasing as a web browser of resort for manual type-in due to basic search needs and privacy assurance. As web usage mounts, so do random typographical errors on tablet and mobile that guide entry to online surveys. In RIWI's case, it acts as a patented statistical router that can capture a 24/7 thumbprint on what consumers want, think and need in every area of the world penetrated by the web. RIWI ramps its "nonsense" URL traffic up and down and targets new geographical regions depending on partner needs. Its patent secures it the right to use these nonsense domains for the purpose of asking questions and collecting geo-specific attitudinal data in every micro-region, down to a sub-city level, everywhere in the world.

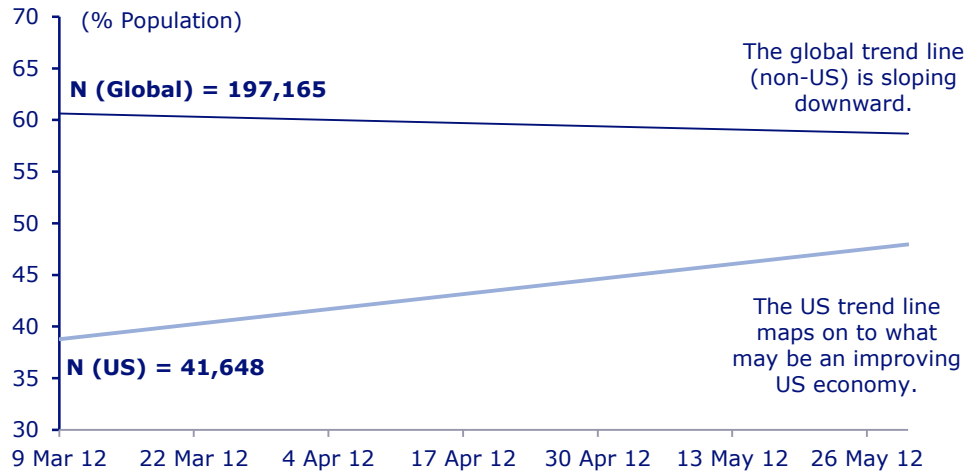
Fresh, new, non-incented respondents on a daily basis

One of the key advantages of the RIWI solution is to capture fresh, new, non-incented respondents (ie, those with no particular axe to grind) on a daily basis - randomized and impenetrable to internet coverage bias. RIWI captures the whole internet population, roughly from eight years of age to 85. As such, economic indicators known to be predictive of GDP, such as the US employment to population ratio (Emratio) - calculated by the US Bureau of Labor Statistics (BLS) monthly[69] for the 15-64 year old population, can be captured daily, and around the world, in every country and region with internet penetration. Many countries do not publish these economic labor-market indicators, or, if they do, only do so on an annual or bi-annual basis. Figure 3 illustrates RIWI Emratio data from 9 - 30 March 2012. The rising line reflects the US Emratio using RIWI.

RIWI collects daily labor-market data that can provide a unique macro perspective

Figure 3

RIWI Emratio trend line data from 9 March to 30 March 2012



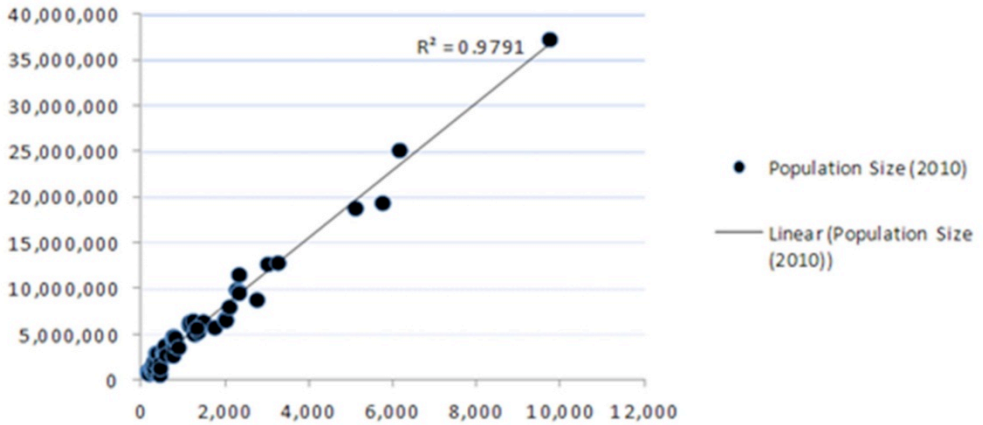
Source: The RIWI Corporation

Figure 4 demonstrates RIWI data are scientific in geographic precision, representative of the geography targeted.

RIWI can track responses by location to a great degree of specificity

Figure 4

Respondents to a peer-reviewed micro survey by state vs US population by state



For the Healthcare Confidence Index, N (US data) = 73,942, 47 states with samples >150 represented
The data here show an overwhelmingly significant linear regression (0.98 r-squared) plotting US state population sizes (Y-axis) vs. the respondent sample (N) captured from each state using *TimeTrender*.
No ex-post population re-weighting was done to achieve this correlation.

Source: The RIWI Corporation; Seeman, N ElectronicHealthcare, 10(4) 2012 e5-e11

Economic indicators never before tracked in real time are eligible for data capture

Other economic indicators, previously not capable of being tracked in real time, are eligible for data capture using the RIWI engine. This includes, for example, real-time labor-productivity rates or the changing number of Americans on food stamps, a fluctuating number now reported monthly and one that figured prominently in the recent US presidential debates. [70] Applications specific to finance are discussed on page 25.

These new daily and robust measures of already established indicators can be correlated with data that analysts may already be tracking for their industry vertical, such as retail. In Europe, RIWI's Emratio data can be correlated with established indicators during a period of downward slippage, such as the slide of Markit's Euro Zone composite purchasing managers' index (PMI) from July 2012 to below the 50 mark that separates expansion from contraction, and Germany's PMI, which hit a 37-month low.

RIWI Corporation's methodology: Q and A

Question: What precisely does the patent protect? How does the user experience work?

Patent protects usage of "broken" domains to obtain a statistically random sample

Answer: The RIWI patent protects the usage of "broken" domains to obtain a statistically random sample of internet users and the ability to use these samples to geo-locate to a specific region (ie, to know where respondents are coming from). Specifically, the patent covers a wide variety of approved claims (21 in number), including the ability to capture consumer sentiment and segment consumer responses based on how they respond.

RIWI's A/B testing on all domains owned, procured or leased ensures that rotating "nonsense" domains are statistically representative of the internet user population in each micro-region of the world (down to a city level). These domains are, collectively, referred to as "traffic". Every time RIWI allocates its own traffic and/or sources more micro-survey nonsense traffic from its partner providers, it runs A/B tests to ensure click-through responses and response rates are statistically similar to other micro-questions of a similar type that already have been run in the region of interest (eg, 1,000 cities in China).

Very unlikely a country can shut down all RIWI domains

Since RIWI's domains rotate, vary in number and are never popular domains (eg, Facebook), it's unlikely a country can shut down all RIWI domains or even a significant fraction of its domains (ie, issue a denial of service order), even in the theoretical possibility that it could find them or want to find them. To shut down all RIWI sites, a country would likely need to shut down the entire internet's top-level domains, which would run sharply against its commercial interests. In contrast, it may be very easy for a country, such as Iran or Syria, to shut down Facebook or Twitter or other virtual sandboxes from which Big Data companies mine open-access data.

Sites such as Facebook are now highly vulnerable to hacker attacks and server outages

Such sites also are now highly vulnerable to hacker attacks, or to server outages, so much so that Facebook was unavailable recently for many users in France, Germany, Norway, Italy and Greece. [71] Recently, Godaddy.com suffered an outage for three hours on 19 October 2012. In September, many Godaddy-hosted sites suffered an outage (for which one member of the global hacking group, Anonymous, claimed responsibility; Godaddy has denied this assertion, blaming server outage instead). On 18 October 2012, ahead of Google's earnings report, Google-owned YouTube™ suffered a brief outage. On 18 October 2012, popular software-development collaboration site GitHub™ endured for the second time in a month a distributed denial of service attack (DDoS), affecting more than two million users. In all, the

attack kept GitHub down for about 90 minutes. By contrast, RIWI is a "universal cross-platform" global data-capture tool, and therefore dramatically less vulnerable to hacker or server-outage shutdowns due to the large number of random and rotating sites hosted on different servers.

Whoever mistypes an intended URL has an equal probability of bumping into a RIWI question

When surfing the web, whoever mistypes his or her intended URL, or types in a domain that does not exist but thinks may exist (eg, hongkongsteak.com, which is not owned by anyone as of the time of writing), he or she has an equal probability of bumping into a RIWI question. The site the person encounters is a real web page, not a pop-up or "pop-under". The micro-survey stumbled upon (generally three to six questions, although RIWI can ask as many questions as it wishes) is visible on any web-enabled device/browser with a full-page visual and can be language-specific to target a specific region of interest.

Data is privacy compliant and anonymized

The potential RIWI respondent encounters a simple question on the page they enter (eg, 'Do you intend to buy a car in the next month?') with no advertising. He or she has the choice of answering the question or not. It is privacy compliant and anonymized data that get collected. The internet protocol (IP)-specific data are changed to privacy-compliant locators and chain linked to self-reported sex and age data if these are needed by the client. IP data are implicit data drawn directly from any visitor's browser. An IP address provides micro-geographical data. Whenever a visitor browses from any location, RIWI uses third-party and proprietary offline databases to confirm and convert the user domain and name into a usable unique identifier. As such, every random respondent is given an anonymized unique identifier. Various answers can be tied to one person in this manner. RIWI's bot-filter ensures the people answering are real humans, not "bots" trolling the web. RIWI collects responses in real time and segments them into "market intent" and other attitudinal data that are statistically representative of the internet population in every micro-region in the world.

Hundreds of thousands of people are being exposed to a random RIWI question every day

A person is only able to answer the question once (but may, by random change, be exposed to a different RIWI question in the future). Hundreds of thousands of people are being exposed to a random RIWI question every day, especially with RIWI ramping up its data-capture services in emerging markets, which traditional market research companies generally cannot reach and where social media have a much lesser presence. The number of exposed potential RIWI respondents is growing dramatically. And, as noted, because of the patented RIWI processes described here, there is no need for ex-post re-weighting of captured data by geography. RIWI has established this in peer-reviewed journal papers.[72]

RIWI has received a patent for its core technology

Figure 5

Summary of the RIWI invention

Title of invention: **Method of obtaining a representative online polling sample**

Approved until July 2030, the rare US business patent (11/676,540) covers, among many other claims:

Exclusive right to use RIWI's unique method to:

- **Ask questions** about what web users want, think and need about any product or idea or issues
- **Recruit respondents to a panel** of dedicated users who want to answer ongoing questions about specific products, ideas or issues
- **Link user opinions** to unique privacy-compliant, location-based identifiers - anywhere in the world, 24/7
- **Push a unique stream** of respondents to engagement for marketing campaigns

Source: The RIWI Corporation

Question: How do the patented processes assure proper weighting? In the developing world, weighting is always necessary because of uneven telecom/broadband access.

Proper weighting has been proven

Answer: The proper weighting has been proven in several government-commissioned peer-reviewed publications referred to earlier, notably in the areas of real-time public-health surveillance. Internet penetration is the main bias, as is leisure bias. It is possible to analyze the findings relative to the internet penetration in a country; however, this type of ex-post analysis would be an add-on analytical step. The RIWI engine delivers global random internet-user attitudinal data in as close to real time as possible. If, theoretically, someone is spamming any RIWI question site page, RIWI can identify this IP address immediately and weed out responses coming from it. RIWI can identify this theoretical possibility in seconds.

There is no cached history of the RIWI questions extant on the internet once the data-capture project or ongoing RIWI tracker ends. All of the data are captured by RIWI and kept in a history of its growing geo-segmented global data repository of tens of millions of global responses to date.

Proprietary algorithms focus on latency for "nano-surveys"

RIWI's approach is proprietary and has been developed and refined over the past three and a half years. The engine has been running 24/7 since April 2009. One major element of its approach to increase response rates relates to how quickly the user gets redirected to a new landing page (URL), how quickly questions get exposed to the respondents and how rapidly the respondents flip to follow-on questions. Speed (every one hundredth of a second) is critical to create "micro-snap" or "nano" questions that ensure a good response rate, rapid innate, emotional response and statistical purity.

Question: Why do people answer RIWI questions?

Answer: The answer to this question varies, and one might ask the same question of any survey intercept, such as "robo-calls", panel surveys or mall intercepts. Sometimes the first RIWI question is engaging (eg, a question, with male and female facial images, asking the respondent to select from among the most beautiful of the images) or relevant (eg, a "hot-button" political question). All questions render elegantly on any device. (Helpful in this respect to RIWI is the fact that, according to an October 2012 study from NPD DisplaySearch™, on average, the diagonal size of displays on electronic devices is expanding.)[73]

Questions are simple and snappy

Sometimes the question is emblazoned with the logo of the client organization, which may or may not increase a response rate. Questions are simple and snappy. RIWI makes it as easy as possible for people to exit, faster than one would be able to exit and land back on one's intended URL after a "page does not exist" error, or, in internet vernacular, a "404 error". RIWI response rates range from 4% to 40%. It averaged 36% in Egypt on the final two days of the recent Egyptian election and was able to predict the election outcome. It did so within the standard error, with more than 8,000 Egypt-based respondents.

RIWI continues to monitor volatility in the country, especially in light of post-election allegations in the western media that the new President, Mr Morsi, was instituting anti-democratic reforms, specifically, presidential limits on judicial powers.[74] In Figure 6, days 16 and 17 mark the election

period; another date of significance is day 21, when secularist support (ie, anti-Morsi sentiment) surpassed support for the President, who has been affiliated with the Muslim Brotherhood.

For the international intelligence community, best-practice market research in far-flung regions such as the Middle East involved stealth house-to-house qualitative interviews over several months, which resulted in very small, heavily biased samples. The situation is different today.

RIWI knows if a respondent is trying to pretend s/he is in another country

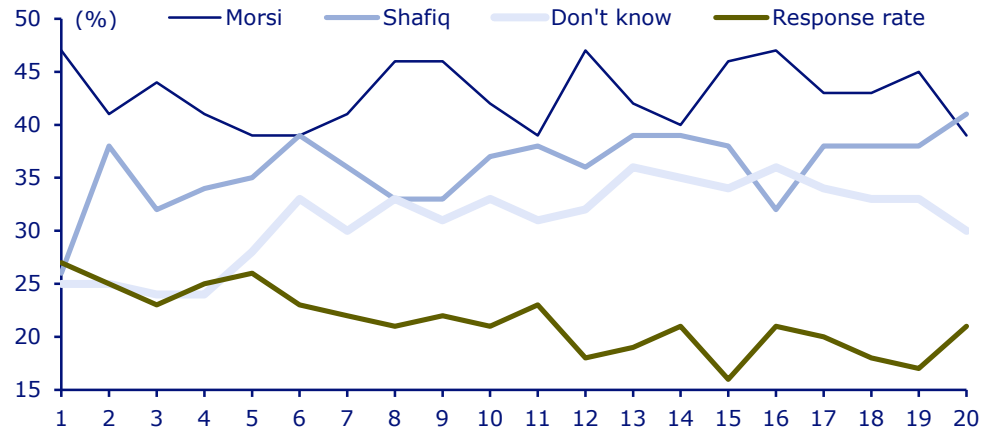
Question: What if web users in country X manage to secure IP addresses from country Y in order to fool the government representatives or others who monitor their web usage?

Answer: RIWI's technical processes know who these proxy server users are and can weed them out of the response sample. Since RIWI domains, rotating in real time, are not popular domains, the statistical odds of this happening is essentially nil. It has not happened once since the RIWI engine began running in April 2009. Also, RIWI surveys the average person using the internet (8-85 years in age). Only if a significant percentage of people in a country or region uses an IP address in a foreign jurisdiction via a proxy server and this group uses that foreign IP a significant percentage of time, would the results be statistically compromised.

RIWI was able to predict the election outcome in Egypt

Figure 6

RIWI was able to predict the election in Egypt and post-election volatility



Notes: RIWI Corp's technical-quality controls and inter-daily statistical checks completed. Running in Arabic and English. Average daily response rate: 30%. Response rate rose to peak of 39% on 22 June. New, non-incented internet-based respondents across Egypt answered this same question daily. RIWI's patented nano-survey tracker methodology is representative of the broad Egyptian (internet-enabled) population (age 8-85).
Official vote: Morsi wins by a ratio of 1.07 (Morsi):1 (Shafiq), RIWI data show Morsi wins by a ratio of 1.19 (Morsi):1 (Shafiq); RIWI Predictive Analysis: On two days of voting (16-17 June), RIWI data show Morsi wins by a ratio of 1.11 (Morsi):1 (Shafiq) +/- .06. Source: The RIWI Corporation

RIWI is also able to break down sub-country sentiment in hard-to-reach, poorer regions of the world

RIWI is also able to break down sub-country sentiment in hard-to-reach, poorer regions of the world, such as India. This can be valuable for an analyst or company forecasting whether a particular product (eg, US mineral water) will sell in India. Figure 7 examines popular support for the anti-corruption activist in India, Mr Anna Hazare.

Breaking down into sub-regions of India and capturing large samples

Figure 7

Tracking regional support in India for Anna Hazare

Region	% Yes	% No	% Don't know
Andhra Pradesh	81	11	8
Assam	74	20	7
Bihar	73	21	6
Chandigarh	80	12	8
Chhattisgarh	80	11	9
Daman and Diu	89	5	5
Delhi	79	13	8
Goa	75	18	7
Gujarat	77	13	10
Haryana	84	10	7
Himachal Pradesh	75	15	10
Jammu and Kashmir	85	15	0
Jharkhand	84	14	3
Karnataka	77	14	9
Kerala	64	20	16
Madhya Pradesh	80	13	7
Maharashtra	81	11	8
Meghalaya	50	50	0
Mizoram	63	25	13
Orissa	82	14	4
Puducherry	73	14	13
Punjab	78	13	9
Rajasthan	84	12	5
Tamil Nadu	75	14	11
Uttar Pradesh	77	15	8
Uttarakhand	79	12	9
West Bengal	68	18	14
Total	79	13	9

Notes: N = 26,005 India-based respondents; 18-26 August 2011. Source: The RIWI Corporation

Capturing macro-economic indicators

RIWI and global finance

RIWI data are able to track rising or falling consumer support for specific decisions, such as whether everyday Germans are willing to support back-stopping the euro. Investors can potentially leverage such consumer-driven intelligence when it is germane to market movement. In the case of Germany, RIWI data have suggested that a decline of popular German support for keeping the European Union intact *may* have predicted, at the 99th day (see left image in Figure 8), a signaling point downward in the value of the euro as against other currencies (see right side image).

There were indeed flashpoints in the currency markets that could have affected the euro's decline during this very period, notably labor market weakness in Spain and Greece. Still, macro-level consumer sentiment can be of potential significance to certain phenomena in the global finance sector. This is especially the case if the macro-sentiment is running 24/7. RIWI's downward sentiment projections in Germany towards the EU preceded the monthly release in late July 2012 of the European Commission's Economic Sentiment Indicator survey, an overall measure of confidence in the Euro Zone, which fell for the fourth straight month to 87.9 from 89.9.[75]

Real-time data capture of consumers' sentiment can provide valuable market-edge intelligence

To wit, macro-plays in the capital markets can be taken based on detecting shifts in consumer sentiment. For example, money managers can elevate their plays on a decline in platinum prices since platinum's primary use is to remove pollutants from the tailpipe emissions of diesel-fueled cars, and Europe is the world's largest market for such cars.[76] If, as a money manager,

you wish to play on a further weakening of Europe's economy and on rising uncertainty about member nations' commitment to the Euro Zone, real-time data capture of consumers' commitment to the Euro Zone could prove to be valuable market-edge intelligence.

Everyday Germans' support for the EU helps to predict relative currency value of euro

Figure 8

RIWI data and the slide in support for the euro

Source: The RIWI Corporation

Applications for finance sector, ranging from predicting real-estate bubbles to pandemics

Other applications of the RIWI methodology that may be useful for macro finance hinge on the engine's ability to: (a) penetrate emerging and all markets; (b) capture large samples of non-incented respondents through flash micro-surveys with high geographic precision; and (c) capture a sample that makes no a priori assumptions about what is the ideal breakdown of potential respondents.

The examination questions of interest could include:

- Is there a perception of a real-estate bubble in Miami or London or Barcelona? What is the comparison?
- What percentage of the population in each tier-1 and tier-2 city of China is planning on replacing their smartphone?
- What are back-to-school August/September sales looking like in key cities in the developed world?
- Early stage warning or data checks on existing indicators (eg, unemployment, job losses, home sales, mortgage trends).
- Consumer confidence around the world, country-to-country, in real time.
- Do consumers in over 1,000 Chinese cities prefer DIY home products (eg Home Depot)[77] as opposed to "do-it-for-me" stores and related services?
- Perceptions in leading super-power countries or high-GDP countries as to whether these countries are in decline or on the incline (eg, militarily, economically) relative to other nations.
- Using rotating images, how do individuals in different cities in the world perceive cosmetics?
- Multicity micro-surveys (eg, in China) on whether people in 1,000-plus Chinese cities are intent on buying, versus renting, a home.
- Multicountry trackers on perceived symptomatology of emerging pandemics.

Capturing data on an hourly/weekly or real-time basis

For the finance sector, RIWI data can be collected from any web-enabled device (ie, tablet, iPhone, etc) and any browser. The data are collected and secured anonymously. In addition to response data, "heat maps" of interest intensity in the specific issue/question (eg, employment, housing, food security) can be calculated, based on the hourly/daily changing response rate (eg, questions relating to whether people are supportive of a particular brand). The data can be reported hourly, daily, weekly or monthly. The masses of data can be "mashed up" to ensure that RIWI receives running tabulations on market intent by micro-regions around the world on a growing suite of issues.

Aiming to create 100-plus country trackers of consumer opinion in real time

Macro-indices of opinion (eg, macro real-estate indices, segmented by people's opinions in every city, as to the likelihood of a real estate crash) can be created in real time. The representative man-on-the-street data that RIWI collects can be compared to the non-representative text-based data or "social chatter" on the web (ie, gathered through scraping blogs, etc) described earlier. In many cases, RIWI data will show that the social-chatter noise is, in fact, unrepresentative of the population studied.

Case study - Will *Monster* overtake *Coke*?

Comparing the value of the *Monster* brand against competitors in 11 countries

We present RIWI's novel approach to data capture in a case study that not only highlights the challenges of data-collection management, bias and interpretation, but also the technology the company has developed to address these issues. Conducted from 27 July to 9 September 2012, the study analyzed the emerging power of energy-drink brands, comparing *Monster* to *Coke* across 11 key markets, specifically the USA, Canada, Mexico, UK, France, Spain, Germany, Poland, Russia, China and Japan.

RIWI's micro-survey technology was used to generate rapid answers from highly randomized and non-incented respondents. The goal was to minimize response bias and increase the quality and integrity of the data produced.

Micro-survey technology used to generate rapid answers from random and non-incented respondents

The study aimed to penetrate all markets of interest with new, fresh respondents every day; capture large samples of non-incented respondents through nano-surveys with high geographic precision; capture a sample that makes no a priori assumptions about what is the ideal breakdown of potential respondents; trend over time; produce very strong statistical confidence; and dramatically reduce, or altogether eliminate, coverage bias.

Each micro-questionnaire was presented to respondents in the national language of the country of access. The total tally of consumer respondents gathered was 90,209; the data below are broken down by country for total number of respondents, total internet users per country and corresponding measures of statistical confidence:

The total tally of consumer respondents gathered was 90,209

Figure 9

RIWI targeted over 90,000 users across 11 countries

Country	Respondents	Internet users	Statistical Confidence Margin of error: +/-1% Confidence interval (CI): 99%
USA	18,294	245,203,319	$p < 0.001$
Canada	8,183	27,757,540	$p < 0.001$
Mexico	7,559	42,000,000	$p < 0.001$
UK	4,620	52,731,209	$p < 0.001$
France	5,290	50,290,226	$p < 0.001$
Spain	5,951	30,654,678	$p < 0.001$
Germany	4,981	67,364,898	$p < 0.001$
Poland	5,170	23,852,486	$p < 0.001$
Russia	5,649	61,472,011	$p < 0.001$
China	17,984	513,100,000	$p < 0.001$
Japan	6,528	101,228,736	$p < 0.001$
Total	**90,209**		

Interpretation: For each country, the sample of respondents is representative of the internet user population, at least 99 times out of 100. Source for "Internet users" - Key ICT indicators for the ITU/BDT regions, International Telecommunications Unions (ITU), Geneva, 16 November 2011. Source: The RIWI Corporation

The large numbers of respondents that were able to be micro-surveyed in short periods of time by this technology generate high levels of statistical confidence.

Focusing on affinity of respondents for the *Monster* drink vs others

The primary focus of the research was to probe a highly randomized population of respondents on their affinity for the *Monster* brand compared to competitor brands. Additionally, respondents were asked about their perceptions of the safety of consuming energy beverages.

The Preference Index emulates the in-store selection experience

The energy-beverage market includes various direct and indirect competitors, with many similar products simultaneously available at the point of sale. In combination with a rapid on-shelf final purchase decision and relatively low brand-switching costs, this makes for a highly competitive and dynamic purchase environment. For this reason, we determined that creating a direct head-to-head and base-lined "Preference Index" would provide additional insights into the current relative brand strength and future prospects in all the markets surveyed.

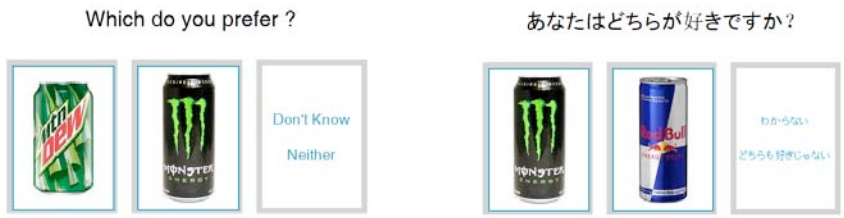
Because of the category's highly competitive and dynamic purchase environment, it is critical that the research measurement tool serve to emulate, as much as possible, an in-store experience and the associated rapid decision-making that needs to occur.

Respondents were exposed to randomized rotating product images, on random rotating URLs, of *Monster* **and** a competitive product, and are then asked to indicate which they prefer:

Surveys are simple, often binary choices

Figure 10

Mountain Dew or Monster?



Source: The RIWI Corporation

This visual, intuitive and rapid-response format captures a consumer's immediate brand preferences and measures the head-to-head comparative strength of *Monster* versus the identified competitive brands in every market.

Design interface and direct phrasing allow respondents to rapidly provide an answer

Similarly, the design interface and direct phrasing of all questions is specifically constructed to allow the respondent to rapidly, and truthfully, provide an answer through a one-click technology that cannot be penetrated by bots or fake panelists. The question renders elegantly on any browser interface and any tablet or web-enabled device such that it "takes over" the full screen.

Survey questions included:

- Have you ever seen [brand] before?
- What do you think of [brand]?
- Have you ever bought [brand]?
- Would you consider buying [brand]?
- How often do you drink [brand]?
- Which do you prefer? (head-to-head visual comparison)
- Do you drink energy drinks like Monster/Red Bull, etc?
- Do you think energy drinks such as Monster/Red Bull are safe?

Key findings are illustrated in the following analysis and accompanying graphs.

The study established relative popularity and consumption benchmarks in each market measured

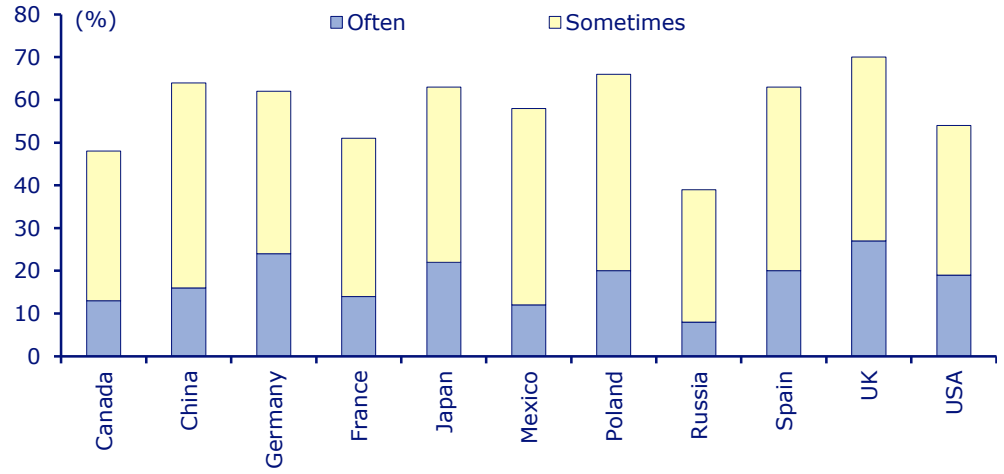
Overall consumption of energy beverages

The survey respondents were asked whether they drink energy drinks and could choose to interpret this as including local, traditional or other energy beverages, as well as globally branded products such as *Monster* or *Red Bull*. Respondents were allowed to opt out and not respond to any particular question, thus those who chose to respond are likely more aware or interested in the energy-beverage category and are the consumers of greatest interest for the purpose of this study. The levels of self-reported usage illustrated in Figure 11 are particularly valuable as a comparative measurement of the *relative* popularity and consumption rates of energy drinks in each market among engaged consumers. This research indicates that energy-drink consumption varies significantly among the 11 markets surveyed, with between 39% (Russia) and 70% (UK) of respondents indicating that they consume energy beverages at least occasionally, and a core group 10-20% of respondents who indicate they are frequent users.

Results can be compared easily across markets

Figure 11

Energy-beverage consumption across 11 markets



Source: The RIWI Corporation

Brand Affinity: The Passion Index

While market share and sales growth can be effectively measured through retail sales, it is more difficult to measure a brand's health, momentum and staying power. This is especially hard to do using traditional online intercept methodologies, since these approaches are not able to capture a representative sample of web users. In the case of brand categories, such as energy beverages that employ significant online marketing, the online RIWI intercept can be effective for assessing brand health.

It is hypothesized that understanding consumer "passion" for a brand is one way to measure brand health and is probably important in emerging and fast-growth categories, such as energy drinks. One objective of this 24/7 tracking research was, therefore, to generate a relative "Passion Index" for *Monster* and each of its competitors in each country surveyed.

RIWI can capture micro-emotional responses as well as rational responses

Brand affinity and passion are believed to be, to a large extent, an emotional, as opposed to a purely rational, response. Hence, it is important that the research measurement tool worked to emulate an in-store experience with its associated rapid, emotional decision-making, thus decreasing response bias. Respondents were shown a pack shot of either a can of *Monster* or a

competitive product and asked to indicate their affinity ('Love it', 'Like it', 'It's okay', 'Don't like it'), as follows:

Survey can include gradations of preference

Figure 12

Monster versus competitor pack shots

Source: The RIWI Corporation

It is important to note that strong passion scores reflect the intensity of support for a brand and may translate into higher momentum and customer loyalty, as opposed to being a measure of current market share.

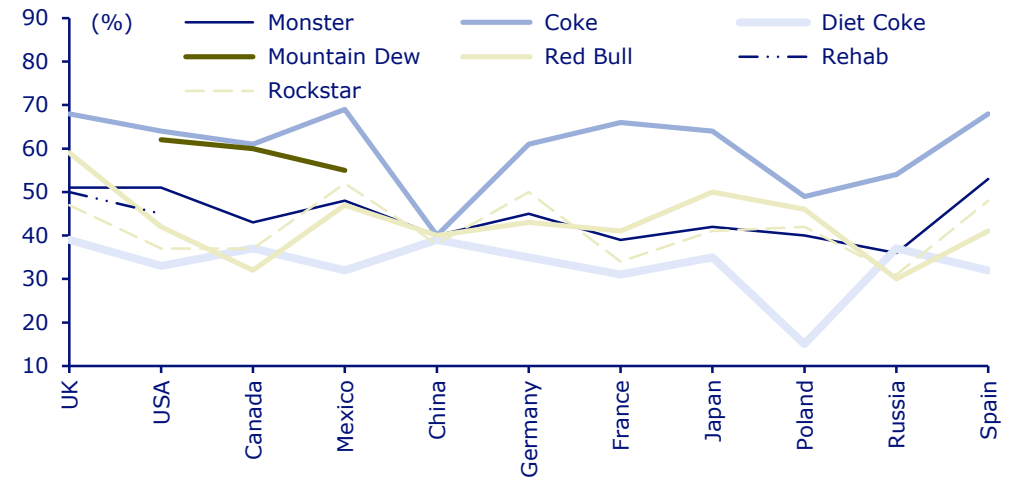
Consumers can have opinions and expectations of brands even if they have never consumed them

The passion question was asked only of those who said they had *seen* the product before but had not necessarily consumed it, as an individual does not have to have access to a particular product, or to have ever tried it, to have an opinion when asked. Even in markets where there is no or limited availability of the product, such as China, it is possible that brand opinions can be created through seeing online or other advertising, such as extreme sports sponsorships. Also, since respondents were shown a pack shot, it is also likely that expressed opinions were partially a measure of how appealing they found the packaging, and by extension, their neuro-associational expectations of the potential appeal of the products.

Passion Index measures intensity across markets

Figure 13

Passion Index for different beverages across 11 markets



Source: The RIWI Corporation

For example, in China, 2,483 people responded to the question of what they think about *Monster*, with 542 (22%) saying they 'like it', and 436 (18%) saying they 'love it', hence the combined (40%) support score illustrated in Figure 13, with the remainder being 'ok' (40%), 'don't like it' (11%) and 'don't know' (9%).

Including *Monster* in the China survey, knowing that it is not currently in the market or at least not widely available, was done with the intention of gaining some insight to the potential receptivity or competitive barriers to entry if/when *Monster* moves into that market, as opposed to a measure of current actual market penetration.

This analysis dispels the idea that *Monster* will become the Next Coke, at least for now

The analysis of the Passion Index indicates that *Coke* clearly remains the overall leader among consumers who 'love' or 'like' a brand, dispelling the idea that *Monster* will become "The Next Coke", at least for the near future. Of particular note is China, where *all* brands, including *Coke*, show almost identically low levels of consumer passion.

On this Passion Index, *Monster* and *Red Bull* clearly fare better than most of the other energy products (*Rockstar*, *Rehab*); however, *Monster* falls short of *Coke* and *Mountain Dew*. (Note: *Rehab* was only provided as a comparator option in the USA and UK).

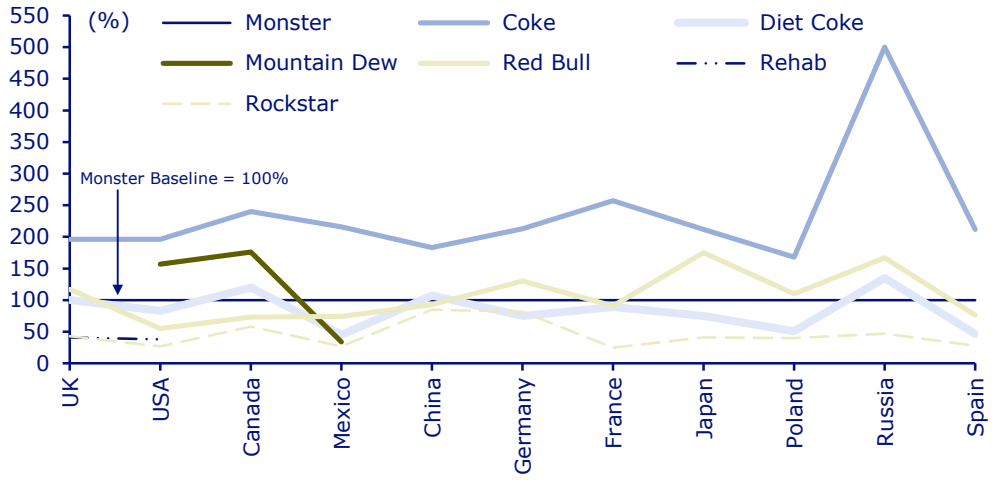
Head-to-head competitive brand findings: The Preference Index

Monster may be battling with *Red Bull* to become the "Coke" of the energy-beverage market, but neither looks likely to replace *Coke* as the overall preferred refreshment beverage in the markets surveyed. In the head-to-head product preference question, *Coke* clearly beats *Monster*. As Figure 14 illustrates, *Coke* is preferred over *Monster* at least by a factor of 2 (China) and sometimes 5 (Russia).

Comparing preferences head to head

Figure 14

Preference Index for different beverages across 11 markets



Source: The RIWI Corporation

***Monster* and *Red Bull* are battling for supremacy among energy drinks**

Mountain Dew also shows significant strength on both the Passion Index and head-to-head comparisons with *Monster* in the USA and Canada. China is currently the most undefined market. While *Coke* dominates on passion and preference, *Red Bull* and *Monster* are locked in second place and share very similar passion and preference scores. *Monster* is preferred in head-to-head comparisons to *Rockstar* and *Rehab*. However, *Monster* and *Red Bull* are strong in different markets, with *Monster* preferred in North America, and *Red Bull* strong in Europe, as follows:

- *Monster* is preferred over *Red Bull* in the USA, Canada, Mexico and Spain.
- *Red Bull* is preferred over *Monster* in Germany, Japan, Russia, France and in the UK.
- *Monster* and *Red Bull* are essentially equally preferred in China and in Poland.

Energy beverages have been in the safety spotlight

Perceived safety of energy drinks

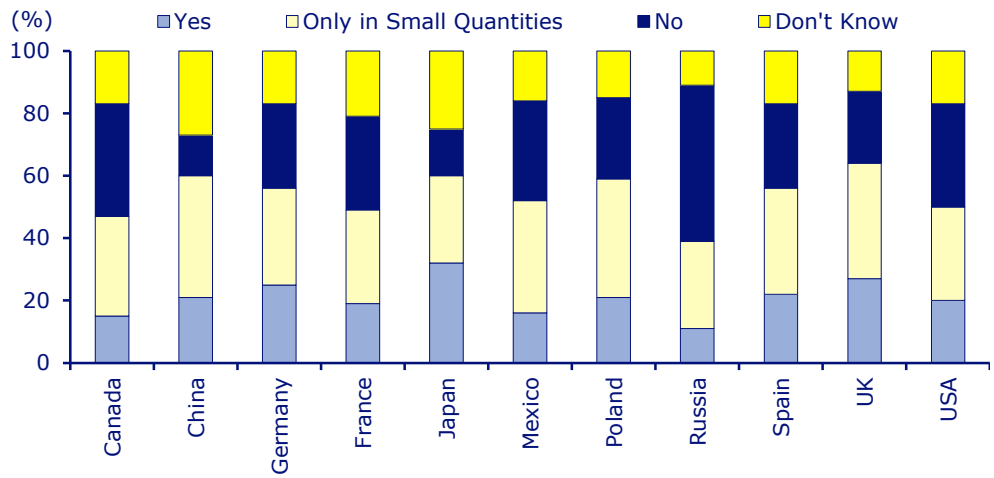
In North America particularly, energy beverages have been in the safety spotlight, with calls to limit their sales in certain jurisdictions. Over the past five years, criticism has centered around research linking select energy-beverage ingredients (eg, sugars and acids) in particular beverages to diabetes, obesity and dental decay.[78] Reports in mainstream media state that 'long-term exposure to the various components of energy beverages may result in significant alterations in the cardiovascular system . . . the safety of energy beverages has not been fully established.'[79]

In order to determine the market perception of the safety of energy beverages in all markets being surveyed, RIWI posed the question to respondents, 'Do you think energy drinks like *Monster*/*Red Bull* are safe?', with the answer options provided being 'Yes', 'No', 'Only in small quantities' and 'Don't know'.

Comparing perceptions across countries

Figure 15

Perceptions of energy beverage safety across 11 markets



Source: The RIWI Corporation

The safety perception of energy drinks remains a significant issue for the category globally

In the 11 markets examined, consumers have a broad variety of opinions and perceptions regarding the safety of energy drinks. On the whole, consumers are as likely to believe they are safe, not safe or safe in small quantities. In established markets, such as Europe and North America, consumers are more likely to think energy drinks are unsafe, with Russians the most likely to think that energy drinks are not safe, followed by Canadians and Americans. There are market opportunities in Asia: Chinese consumers are the least likely to consider energy drinks unsafe, followed by Japanese consumers.

Moving to smarter data

There are challenges in global data capture despite a universe of data tools and applications

As we have learned in this CLSA U Blue Book, it is increasingly difficult, some would say impossible, to obtain data from fully representative survey respondents whether via the internet or more traditional routes. Most data-capture observers will agree on the following:

- It is very difficult to gather robust sample sets from every region of the world.
- Opinions of respondents can change from day to day.
- Predictive algorithms are based on subjective assumptions, representing little more than guesswork.

Although differentiated in the marketplace, the RIWI patented solution is not perfect. No data-capture method is. RIWI's specific, acknowledged biases in peer-reviewed research are: 1) the use of the internet population, which means a skew towards a younger population and, inherently, a more technologically savvy population; 2) the English-language nature of the internet (rarely, if ever, acknowledged by many data-capture providers). RIWI does, however, pose survey questions in the language of the country from which the responder derives; and 3) social leisure bias - a bias that affects every data collection method known to humankind, including the focus group.

It is important to acknowledge the John Nash problem in Big Data

The "John Nash problem" needs to be acknowledged. Specifically:

- Social-media analytics are getting noisier, making Big Data analysis highly limited.
- Commercial panels are based on repeat, incented, primed respondents.
- Underlying probability samples are based on subjective bias.
- Coverage bias and non-response bias in polling methods are rising.
- Decision-makers need 24/7, global, actionable data of online populations.

It is possible to draw on macro consumer opinion to predict white and black swans

The RIWI landscape imagines a reality where it *is* possible to predict scenarios hitherto considered impossible, especially long-tail black-swan (ie, negative) or white-swan (positive) events in emerging markets. Such events include potential inflection points such as:

- Whether Germans in every city in Germany will allow the EU to survive (a white swan).
- Shenzhen-based Tencent™ upsetting user adoption of Facebook.
- Famine in sub-populations subjected to risk of a monoculture (a black swan)[80].
- Risks of over-dependence on one source of energy (a black swan)[81].
- Southern Europeans cutting back on coffee consumption.

RIWI hopes to re-invigorate George Gallup's classic vision

Reflecting, in this context, on the key messages in this report, RIWI hopes to re-invigorate George Gallup's classic vision of using consumer sentiment to monitor, predict and divine intelligence in a reliable manner from global consumer populations, whether these consumers reside in the 11 major markets used for the *Monster* case study presented here, or in any other of the more than 190 countries around the globe.

Endnotes

[1] Global Market Research 2011- An ESOMAR Industry Report in cooperation with KPMG Advisory. (2011). Retrieved October 11, 2012, from http://www.esomar.org/uploads/industry/reports/global-market-research-2011/ESOMAR-GMR-2011_Preview.pdf.

[2] Guide to Market Research. (2012). Retrieved on 11 October 2012, from http://canadabusiness.ca/eng/page/2691/.

[3] Crisis or Opportunity (n.d.).In *Sentis.* Retrieved on 11 October 2012, from http://www.sentisresearch.com/tags/esomar.

[4] More Dirty Little Secrets of Online Panel Research. (2012). In *GreyMatter Research & Consulting* Retrieved on 11 October 2012, from http://www.greymatterresearch.com/index_files/Online_Panels_2012.htm.

[5] Om, Malik. (4 September 2012). When it comes to startups, timing really is everything. In *GIGAOM*. Retrieved 11 October 2012, from http://gigaom.com/2012/09/04/onlives-troubles-and-why-timing-really-is-everything/.

[6] Lunden, Ingrid. (16 October 2012). Mobile Milestone: The Number Of Smartphones In Use Passed 1 Billion In Q3, Says Strategy Analytics. In *Tech Crunch*. Retrieved 19 October 2012, from http://techcrunch.com/2012/10/16/mobile-milestone-the-number-of-smartphones-in-use-passed-1-billion-in-q3-says-strategy-analytics/

[7] Lomas, Natasha. (6 November 2012). Gartner: 1.2 Billion Smartphones, Tablets To Be Bought Worldwide In 2013; 821 Million This Year: 70% Of Total Device Sales. In Techcrunch. Retrieved 6 November 2012, from http://techcrunch.com/2012/11/06/gartner-1-2-billion-smartphones-tablets-to-be-bought-worldwide-in-2013-821-million-this-year-70-of-total-device-sales/

[8] Evans, John. (25 August 2012). What Happens When Pollsters Are No Better Than Psychics? In *Techcrunch*. Retrieved 11 October 2012, from http://techcrunch.com/2012/08/25/what-happens-when-pollsters-are-no-better-than-psychics/.

[9] Lavrakas, Paul. (May 2010). An Evaluation of Methods Used to Assess the Effectiveness of Advertising on the Internet. Retrieved 11 October 2012, from http://www.iab.net/media/file/Evaluation_of_Internet_Ad_Effectiveness_Research_Methods.pdf.

[10] Alberta's election polls miss the mark. (24 April 2012). In *CBCNews Canada*. Retrieved on 11 October 2012, from http://www.cbc.ca/news/canada/albertavotes2012/story/2012/04/24/albertavotes2012-poll-prediction-failure.html.

[11] More Dirty Little Secrets of Online Panel Research. (2012). In *GreyMatter Research & Consulting* Retrieved on 11 October 2012, from http://www.greymatterresearch.com/index_files/Online_Panels_2012.htm.

[12] Tradeshow. (n.d.) Retrieved 12 October 2012, from http://www.mria-arim.ca/Conference2011/TRADESHOW/generalinfo.asp.

[13] GRIT Report (Spring 2012). Retrieved 6 November 2012, from

http://www.greenbook.org/PDFs/GRIT-S12-Full.pdf

[14] Bryden, Joan. (14 February 2012). Election polls are fun, but they don't mean a thing: pollsters. In *Thestar.com*. Retrieved 11 October 2012, from http://www.thestar.com/news/canada/article/938345--election-polls-are-fun-but-they-don-t-mean-a-thing-pollsters.

[15] Internet World Stats. (31 December 2011). Retrieved 11 October 2012, from http://www.internetworldstats.com/stats.htm.

[16] Global Internet Traffic Projected to Quadruple by 2015. (1 June 2011). In *Cisco*. Retrieved 11 October 2012, from http://newsroom.cisco.com/press-release-content?type=webcontent&articleId=324003.

[17] Wong, Martin. (4 April 2012). What You Need to Know About Google Consumer Surveys. In *Smartt*. Retrieved 11 October 2012, from http://www.smartt.com/what-you-need-know-about-google-consumer-surveys.

[18] MacDonald, P., Mohebbi, M., and Slatkin, B. (n.d.). Comparing Google Consumer Surveys to Existing Probability and Non-Probability Based Internet Surveys. Retrieved 11 October 2012, from http://www.google.com/insights/consumersurveys/static/357791852901773780/consumer_surveys_whitepaper.pdf.

[19] Seeman, Neil (29 November 2011). Method of obtaining a representative online polling sample. Retrieved on 11 October 2012, from http://www.google.com/patents/US8069078.

[20] Ontario Fact Sheet. (September 2012). In *Ontario Ministry of Finance*. Retrieved on 11 October 2012, from http://www.fin.gov.on.ca/en/economy/ecupdates/factsheet.html.

[21] Grenier, Eric. (21 September 2012). Ontario Polls Paint Confusing Picture Of Political Landscape. In *The Pulse*. Retrieved on 11 October 2012, from http://www.huffingtonpost.ca/2012/09/21/ontario-polls-nanos-forum_n_1903388.html#slide=1553021.

[22] Wingrove, Josh. (22 April 2012). Wildrose Party set for sweeping majority, latest poll shows. In *The Globe and Mail*. Retrieved on 11 October 2012, from http://www.theglobeandmail.com/news/politics/wildrose-party-set-for-sweeping-majority-latest-poll-shows/article2410297/.

[23] Poll Watch: How accurate were election opinion polls? (7 May 2010). In *BBC Election 2010*. Retrieved on 11 October 2012, from http://news.bbc.co.uk/2/hi/uk_news/politics/election_2010/8667801.stm.

[24] Pre-Election Polls in Portugal: Accuracy, Bias, and Sources of Error, 1991–2004 *Int J Public Opin Res* (Winter 2005) 17(4): 399-421 first published online 18 March 2005). Retrieved on 11 October 2012, from http://ijpor.oxfordjournals.org/content/17/4/399

[25] Alberta's election polls miss the mark. (24 April 2012). In *CBCNews Canada*. Retrieved on 11 October 2012, from http://www.cbc.ca/news/canada/albertavotes2012/story/2012/04/24/albertavotes2012-poll-prediction-failure.html.

[26] FMC Program Segments 1930-1960: George Gallup and the Scientific Opinion Poll. (n.d.). In *The First Measured Century*. Retrieved 13 October 2012, from http://www.pbs.org/fmc/segments/progseg7.htm.

[27] Simplicity and Enterprise Search: A New Model for Managing Your Enterprise Information. (2003). Retrieved on 11 October 2012, from http://www.google.ca/url?sa=t&rct=j&q=&esrc=s&source=web&cd=1&ved=0CCQQFjAA&url=http%3A%2F%2Fwww.google.com%2Fenterprise%2Fpdf%2Fgoogle_simplicity_enterprise_wp.pdf&ei=eDp5UNagH-y-0QH85YGIDw&usg=AFQjCNH3zJDR0rS03dv_y8ufk8E290Tg3A&sig2=vz7ymnC3NYAC2i6diILPpw.

[28] DiSogra, Charles. (2008). River Samples: A Good Catch for Researchers? In *Knowledge Networks*. Retrieved on 11 October 2012, from http://www.knowledgenetworks.com/accuracy/fall-winter2008/disogra.html.

[29] Garber, Megan. (31 January 2012). Be Better at Twitter: The Definitive, Data-Driven Guide. In *The Atlantic*. Retrieved on 11 October 2012, from http://www.theatlantic.com/technology/archive/2012/01/be-better-at-twitter-the-definitive-data-driven-guide/252273/.

[30] From Stretched to Strengthened. (n.d.). In *IBM*. Retrieved on 11 October 2012, from http://www-935.ibm.com/services/us/cmo/cmostudy2011/cmo-registration.html.

[31] Fox, Justin. (4 October 2012). Why Data Will Never Replace Thinking. In *HBR.org*. Retrieved on 11 October 2012, from http://blogs.hbr.org/fox/2012/10/why-data-will-never-replace-thinking.html.

[32] Williams, Alex. (17 October 2012). Big Data To Drive $232 Billion In IT Spending Through 2016. In *Tech Crunch*. Retrieved on 19 October 2012, from http://techcrunch.com/2012/10/17/big-data-to-drive-232-billion-in-it-spending-through-2016/

[33] Why Invest in Remarkable Content? (n.d.) In *Acquisition Engine*. Retrieved on 11 October 2012, from http://www.acquisitionengine.com/.

[34] On 18 October 2012, San Francisco announced plans to hire a Chief Data Officer. This trend is emerging in cities across North America. See Ha, Anthony. (15 October 2012). In *Tech Crunch*. Retrieved 19 October 2012, from http://techcrunch.com/2012/10/15/san-francisco-chief-data-officer/

[35] Piatetsky-Shapiro, Gregory. Big Data Hype and Reality - Mars Rover vs Mars Candy. 18 October 2012. In *HBR.org*. Retrieved 19 October 2012, from http://www.kdnuggets.com/2012/10/my-hbr-blog-big-data-hype-reality.html

[36] Piatetsky-Shapiro, Gregory and Brij Mason. Estimating Campaign Benefits and Modeling Lift. In *Proceedings of KDD-99* (AAAI Press). Pp. 185-193. San Diego, California. Retrieved 18 October 2012, from http://www.kdnuggets.com/gpspubs/kdd99-estimating-campaign-benefits-modeling-lift.pdf

[37] Pring, Cara. (11 January 2012). 100 social media statistics for 2012.In *The Social Skinny*. Retrieved on 11 October 2012, from http://thesocialskinny.com/100-social-media-statistics-for-2012/.

[38] Constine, Josh. (22 August 2012). How Big Is Facebook's Data? 2.5 Billion Pieces Of Content And 500+ Terabytes Ingested Every Day. In *Tech Crunch*. Retrieved on 11 October 2012, from http://techcrunch.com/2012/08/22/how-big-is-facebooks-data-2-5-billion-pieces-of-content-and-500-terabytes-ingested-every-day/.

[39] #numbers. (14 March 2011). In *blog.Twitter*. Retrieved on 11 October 2012, from http://blog.twitter.com/2011/03/numbers.html.

[40] Rizo C., Deshpande A., Ing A., and Seeman N. A rapid, Web-based method for obtaining patient views on effects and side-effects of antidepressants. J Affective Disorders 130(1-2):290-293 2011.

[41] Website can be found at www.appinions.com.

[42] Website can be found at https://pro.swaylo.com/.

[43] Sage, Adam. (9 August 2012). Facebook and Surveys – Thoughts on Going "Not by the Book". In *Survey Post*. Retrieved on 11 October 2012, from https://blogs.rti.org/surveypost/2012/08/09/facebook-and-surveys-thoughts-on-going-not-by-the-book/.

[44] Teicher, David. (2 February 2012). Reuters' New Social Dashboard Charts CEO Influence, Chatter About Companies.In *AdAge Digital*. Retrieved on 11 October 2012, from http://adage.com/article/digital/reuters-launches-social-dashboard-readers-journalists/232488/.

[45] This 'emoticon' is the popular manner in which to express sadness on Twitter.

[46] Teicher, David. (2 February 2012). Reuters' New Social Dashboard Charts CEO Influence, Chatter About Companies. In *AdAge Digital*. Retrieved on 11 October 2012, from http://adage.com/article/digital/reuters-launches-social-dashboard-readers-journalists/232488/.

[47] Rayport, Jeffrey. (5 September 2012). Use Big Data to Predict Your Customers' Behaviors. In HBR.org. Retrieved 14 October 2012, from http://blogs.hbr.org/cs/2012/09/use_big_data_to_predict_your_c.html?referral=00563&cm_mmc=email -_- newsletter- _- daily_ alert- _- alert_date&utm_source=newsletter_daily_alert&utm_medium=email&utm_campaign=alert_date#disqus_thread.

[48] More Dirty Little Secrets of Online Panel Research. (2012). In *GreyMatter Research & Consulting* Retrieved on 11 October 2012, from http://www.greymatterresearch.com/index_files/Online_Panels_2012.htm.

[49] Jaikumar, Vijayan. (4 May 2012). In *Computer World*. Retrieved on 11 October 2012, from http://www.computerworld.com/s/article/9226840/Investors_are_pouring_funds_into_big_data?taxonomyId=18.

[50] Rayport, Jeffrey. (5 September 2012). Use Big Data to Predict Your Customers' Behaviors. In *HBR Blog Network*. Retrieved on 11 October 2012, from http://blogs.hbr.org/cs/2012/09/use_big_data_to_predict_your_c.html?referral=00563&cm_mmc=email-_-newsletter-_-daily_alert-_-alert_date&utm_source=newsletter_daily_alert&utm_medium=email&utm_campaign=alert_date#disqus_thread.

[51] McCall, J., Stipp, H. (4 September 2012). Does the 18-to-49 Demographic Matter Anymore? An *HBR Management Puzzle*. In *HBR Blog Network*. Retrieved on 11 October 2012, from http://blogs.hbr.org/cs/2012/09/does_the_18-to-49_demographic.html?referral=00563&cm_mmc=email-_-newsletter-_-daily_alert-_-alert_date&utm_source=newsletter_daily_alert&utm_medium=email&utm_campaign=alert_date.

[52] Volkow, Nora. (n.d.). The Link Between HIV/AIDS and Drug Abuse. Retrieved 12 October 2012, from http://hiv.drugabuse.gov/english/about/message.html.

[53] This article illustrates how important and unknowable confounders are when looking at huge data sets historically. See: BaumeisterAA Hawkins MF , Lee Pow J , Cohen AS, "Prevalence and incidence of severe mental illness in the United States: an historical overview," *Harv Rev Psychiatry* Sep;20(5):247-58 2012. Historically, due to a lack of understanding of confounding variables in very large antecedent data sets, the increasing rates of mental illness during the 19th century were attributed to immigration rather than to more likely causes - alcoholism, syphilis and disorders related to senescence.

[54] Companies such as IBM, Accenture and Microsoft, among others, are prominent service providers in this fast-growing field. See, in the case of IBM: http://techcrunch.com/2012/10/08/puredata-meet-ibms-new-big-box-for-big-data/. Other companies, such as Alteryx, are also entering this 'data-crunching' market to service smaller firms. See: http://techcrunch.com/2012/10/07/alteryx-opens-big-data-analytics-apps-studio-for-the-rest-of-us/

[55] Om, Malik. (4 September 2012). When it comes to startups timing really is everything. In *GIGAOM*. Retrieved 11 October 2012, from http://gigaom.com/2012/09/04/onlives-troubles-and-why-timing-really-is-everything/.

[56] Warren, NJ. (7 September 2011). Lightspeed Research and GMI Form One of the Largest Global Online Sample Providers in the Market Research Industry. In *PRWeb*. Retrieved 12 October 2012, from http://www.prweb.com/releases/2011/9/prweb8772951.htm.

[57] Website can be found at www.lumimobile.com.

[58] Website can be found at http://gnip.com/twitter.

[59] Lipka, Mitch. (1 March 2012). Twitter sells your feed to Big Data. In *The Globe and Mail*. Retrieved 11 October 2012, from http://www.theglobeandmail.com/technology/digital-culture/social-web/twitter-sells-your-feed-to-big-data/article550140/.

[60] Williams, Alex. 17 October 2012. Big Data To Drive $232 Billion In IT Spending Through 2016. In Tech Crunch. Retrieved 19 October 2012, from http://techcrunch.com/2012/10/17/big-data-to-drive-232-billion-in-it-spending-through-2016/.

[61] Website can be found at http://www.splunk.com/.

[62] Savitz, Eric. (4 April 2012). Talk About Big Data: Splunk IPO Has 1999-Style Day One Spike. In *Forbes*. Retrieved 11 October 2012, from http://www.forbes.com/sites/ericsavitz/2012/04/19/talk-about-big-data-splunk-ipo-has-1999-style-day-one-spike/.

[63] Lunden, Ingrid. (8 October 2012). Telefonica Wants To Turn Its Mobile Data Into A Big Data Business, Launches Dynamic Insights Unit. In *Tech Crunch*. Retrieved 11 October 2012, from http://techcrunch.com/2012/10/08/telefonica-wants-to-turn-its-mobile-data-into-a-big-data-business-launches-dynamic-insights-unit/.

[64] Domain Counts & Internet Statistics. (October 2012). In *Whois Source*. Retrieved 11 October 2012, from http://www.whois.sc/internet-statistics/.

[65] *Bilski* v. *Kappos*, 130 S. Ct. 3218, 561 US __, 177 L. Ed. 2d 792 (2010).

[66] Getting Your Brand Right Online. (17 July 2003). In *NMK*. Retrieved 11 October 2012, from http://www.nmk.co.uk/article/2003/7/17/getting-your-brand-right-online.

[67] Lardinois, Frederic. (5 October 2012). Report: Bing Users Disproportionately Affected By Malware Redirects. In *Tech Crunch*. Retrieved 11 October 2012, from http://techcrunch.com/2012/10/05/report-bing-users-are-disproportionately-at-risk-of-facing-malware-redirects/.

[68] Reuven, Cohen. (6 June 2012). Linkedin Hacked: A Few App Suggestions For Protecting Your Online Passwords. In *Forbes*. Retrieved 11 October 2012, from http://www.forbes.com/sites/reuvencohen/2012/06/06/linkedin-hacked-a-few-apps-suggestions-for-protecting-your-online-passwords/.

[69] As prominent observers, such as Jack Welch (2012) and AustanGoolsbee (2003), former chairman of the Obama administration's Council of Economic Advisers, have noted, there is some subjectivity and bias in the BLS calculations. The labor market participation data reported each month are gathered over a one-week period by census workers, by phone in 70% of the cases, and the rest through home visits. In sum, they try to contact 60,000 households, asking a list of questions and recording the responses. Welch, Jack. (11 October 2012). Jack Welch: I Was Right About That Strange Jobs Report. In *The Wall Street Journal*. Retrieved 11 October 2012, from http://online.wsj.com/article/SB10000872396390444897304578046260406091012.html.

[70] Food Stamp Nation: Uncle Sam's free grocery rolls keep growing. (6 September 2012). In *The Wall Street Journal*. Retrieved 11 October 2012, from http://online.wsj.com/article/SB10000872396390443847404577631580385950886.html.

[71] Etherington, Darrell. (11 October 2012). Facebook Currently Down For Some European Users, Anon Hacker Claims Responsibility [Update: Not A Hack].In *Tech Crunch*. Retrieved on October 11, 2012, from http://techcrunch.com/2012/10/11/facebook-currently-down-for-some-european-users-anon-hacker-claims-responsibiilty/.

[72] Rizo C., Ing A., and Seeman N. Assessing and Responding in Real Time to Online Anti-vaccine Sentiment during a Flu Pandemic. Healthcare Quarterly 13(5):8-15 2010.; Seeman N. Results of a New Healthcare Confidence Index. Electronic Healthcare, 10(4) 2012: e5-e11.

[73] Etherington, Darrell. (16 October 2012). Study: Average Display Size Climbing On All Devices Except For Mobile PCs. In *Tech Crunch*. Retrieved 19 October 2012, from http://techcrunch.com/2012/10/16/study-average-display-size-climbing-on-all-devices-except-for-mobile-pcs/

[74] Trager, Eric. (15 August 2012). Egypt's New President Moves Against Democracy. In *The Wall Street Journal*. Retrieved 11 October 2012, from http://online.wsj.com/article/SB10000872396390444772404577587611785521668.html.

[75] Hannon, Paul. (30 July 2012). Crisis Saps Euro-Zone Confidence. In *The Wall Street Journal*. Retrieved 11 October 2012, from http://online.wsj.com/article/SB10000872396390444226904577558483329408526.html.

[76] Cui, Carolyn. (9 August 2012). Bet on Platinum's Fall Is Anti-Euro Wager. In *The Wall Street Journal*. Retrieved 11 October 2012, from http://online.wsj.com/article/SB10000872396390444900304577577270132138302.html.

[77] Burkitt, Laurie. (14 September 2012). Home Depot Learns Chinese Prefer 'Do-It-for-Me'. In *The Wall Street Journal*. Retrieved 11 October 2012, from http://online.wsj.com/article/SB10000872396390444433504577651072911154602.html.

[78] Future Flavor Trends in Soft Drinks: New opportunities for natural, healthy, and exotic flavors. (November 2008). In *Global Business Insights*. Retrieved 14 October 2012, from www.globalbusinessinsights.com/content/rbig0005m.pdf.

[79] Higgins JP, Tuttle TD, Higgins CL. Energy beverages: content and safety. *Mayo ClinProc* 85(11):1033–1041 2010.

[80] Taleb, Nassim Nicholas, Antifragility, Robustness, and Fragility, Inside the 'Black Swan' Domain (September 1, 2010). Available at SSRN: http://ssrn.com/abstract=1669317 or http://dx.doi.org/10.2139/ssrn.1669317

[81] See Note 111, *Supra*.

Analyst certification

The analyst(s) of this report hereby certify that the views expressed in this research report accurately reflect my/our own personal views about the securities and/or the issuers and that no part of my/our compensation was, is, or will be directly or indirectly related to the specific recommendation or views contained in this research report.

Important disclosures

CLSA (which for the purpose of this disclosure includes subsidiaries of CLSA B.V. and Credit Agricole Securities Asia B.V., Tokyo Branch)/Credit Agricole Securities (USA) Inc ("Credit Agricole Securities (USA)")'s policy is to only publish research that is impartial, independent, clear, fair, and not misleading. Analysts may not receive compensation from the companies they cover.

Regulations or market practice of some jurisdictions/markets prescribe certain disclosures to be made for certain actual, potential or perceived conflicts of interests relating to a research report as below. This research disclosure should be read in conjunction with the research disclaimer as set out at www.clsa.com/disclaimer.html and the applicable regulation of the concerned market where the analyst is stationed and hence subject to. This research disclosure is for your information only and does not constitute any recommendation, representation or warranty. Absence of a discloseable position should not be taken as endorsement on the validity or quality of the research report or recommendation.

Neither analysts nor their household members/associates may have a financial interest in, or be an officer, director or advisory board member of companies covered by the analyst unless disclosed herein. Unless specified otherwise, CLSA/Credit Agricole Securities (USA)'s did not receive investment banking/non-investment banking income from, and did not manage/co-manage public offering for, the listed company during the past 12 months, and it does not expect to receive investment banking relationship from the listed company within the coming three months. Unless mentioned otherwise, CLSA/Credit Agricole Securities (USA) does not own discloseable position, and does not make market, in the securities.

The analysts included herein hereby certify that the views expressed in this research report accurately reflect their own personal views about the securities and/or the issuers and that unless disclosure otherwise, no part of their compensation was, is, or will be directly or indirectly related to the specific recommendation or views contained in this research report or revenue from investment banking revenues. The analyst/s also states/s and confirm/s that he has/have not been placed under any undue influence, intervention or pressure by any person/s in compiling this research report. In addition, the analysts included herein attest that they were not in possession of any material, non-public information regarding the subject company at the time of publication of the report. Save from the disclosure below (if any), the analyst(s) is/are not aware of any material conflict of interest.

Key to CLSA/Credit Agricole Securities (USA) investment rankings: BUY: Total return expected to exceed market return AND provide 20% or greater absolute return; O-PF: Total return expected to be greater than market return but less than 20% absolute return; U-PF: Total return expected to be less than market return but expected to provide a positive absolute return; SELL: Total return expected to be less than market return AND to provide a negative absolute return. For relative performance, we benchmark the 12-month total return (including dividends) for the stock against the 12-month forecast

return (including dividends) for the local market where the stock is traded. For example, in the case of US stock, the recommendation is relative to the expected return for S&P of 10%. Exceptions may be made depending upon prevailing market condition.

Prior to 1 Jan 2012, our investment rankings were: BUY = Expected to outperform the local market by >10%; O-PF = Expected to outperform the local market by 0-10%; U-PF = Expected to underperform the local market by 0-10%; SELL = Expected to underperform the local market by >10%.

Overall rating distribution for Credit Agricole Securities (USA) Equity Universe: Buy / Outperform: 67%; Underperform / Sell: 33%; Restricted: 0%; data as of 30 September 2012. Investment banking clients as a % of rating category: Buy / Outperform: 77%; Underperform / Sell: 23%; Restricted: 0%; data for 12-month period ending 30 September 2012. For a history of the recommendations and price targets for companies mentioned in this report, as well as company specific disclosures, please write to: (a) Credit Agricole Securities (USA), Compliance Department, 1301 Avenue of the Americas, 15th Floor, New York, New York 10019-6022; and/or (b) CLSA, Group Compliance, 18/F, One Pacific Place, 88 Queensway, Hong Kong.

CLSA and/or CAS has/have produced this publication/communication for private circulation to professional, institutional and/or wholesale clients only. The information, opinions and estimates herein are not directed at, or intended for distribution to or use by, any person or entity in any jurisdiction where doing so would be contrary to law or regulation or which would subject CLSA and/or CAS to any additional registration or licensing requirement within such jurisdiction. The information and statistical data herein have been obtained from sources we believe to be reliable. Such information has not been independently verified and we make no representation or warranty as to its accuracy, completeness or correctness. Any opinions or estimates herein reflect the judgment of CLSA and/or CAS at the date of this publication/communication and are subject to change at any time without notice. Where any part of the information, opinions or estimates contained herein reflects the views and opinions of a sales person or a non-analyst, such views and opinions may not correspond to the published view of CLSA and/or CAS. This is not a solicitation or any offer to buy or sell. This publication/communication is for information purposes only and does not constitute any recommendation, representation, warranty or guarantee of performance. Any price target given in the report may be projected from 1 or more valuation models and hence any price target may be subject to the inherent risk of the selected model as well as other external risk factors. This is not intended to provide professional, investment or any other type of advice or recommendation and does not take into account the particular investment objectives, financial situation or needs of individual recipients. Before acting on any information in this publication/ communication, you should consider whether it is suitable for your particular circumstances and, if appropriate, seek professional advice, including tax advice. CLSA and/or CAS

do/does not accept any responsibility and cannot be held liable for any person's use of or reliance on the information and opinions contained herein. To the extent permitted by applicable securities laws and regulations, CLSA and/or CAS accept(s) no liability whatsoever for any direct or consequential loss arising from the use of this publication/communication or its contents. Where the publication does not contain rating, the material should not be construed as research but is offered as factual commentary. It is not intended to, nor should it be used to form an investment opinion about the not rated companies.

Subject to any applicable laws and regulations at any given time CLSA, CAS, their respective affiliates or companies or individuals connected with CLSA/CAS may have used the information contained herein before publication and may have positions in, may from time to time purchase or sell or have a material interest in any of the securities mentioned or related securities or may currently or in future have or have had a business or financial relationship with, or may provide or have provided investment banking, capital markets and/or other services to, the entities referred to herein, their advisors and/or any other connected parties. As a result, investors should be aware that CLSA, CAS and/or their respective affiliates or companies or such individuals may have one or more conflicts of interest.

Regulations or market practice of some jurisdictions/markets prescribe certain disclosures to be made for certain actual, potential or perceived conflicts of interests relating to research report. Details of the disclosable interest can be found in certain reports as required by the relevant rules and regulation and the full details are available at http://www.clsa.com/member/research_disclosures/. Disclosures therein include the position of the CLSA Group only and do not reflect those of Credit Agricole Corporate & Investment Bank and/or its affiliates. If investors have any difficulty accessing this website, please contact webadmin@clsa.com on (852) 2600 8111. If you require disclosure information on previous dates, please contact compliance_hk@clsa.com.

This publication/communication is distributed for and on behalf of CLSA Limited (for research compiled by non-US analyst(s)) and /or CAS (for research compiled by US analyst(s)) in Australia by CLSA Australia Pty Ltd; in Hong Kong by CLSA Research Ltd.; in India by CLSA India Ltd. (Address: 8/F, Dalamal House, Nariman Point, Mumbai 400021. Tel No: +91-22-66505050. SEBI Registration No: BSE Capital Market Segment: INB010826432; BSE F&O Segment: INF010826432; NSE Capital Market Segment: INB230826436; NSE F&O Segment: INF230826436); in Indonesia by PT CLSA Indonesia; in Japan by Credit Agricole Securities Asia B.V., Tokyo Branch, a member of the JSDA licensed to use the "CLSA" logo in Japan; in Korea by CLSA Securities Korea Ltd.; in Malaysia by CLSA Securities Malaysia Sdn Bhd; in the Philippines by CLSA Philippines Inc. (a member of Philippine Stock Exchange and Securities Investors Protection Fund); in Thailand by CLSA Securities (Thailand) Limited; and in Taiwan by CLSA Limited, Taipei Branch.

United States of America: Where any section of the research is compiled by US analyst(s), it is distributed by CAS. Where any section is compiled by non-US analyst(s), it is distributed into the United States by CLSA solely to persons who qualify as "Major U.S. Institutional Investors" as defined in Rule 15a-6 under the Securities and Exchange Act of 1934 and who deal with Credit Agricole Corporate & Investment Bank. However, the delivery of this research report to any person in the United States shall not be deemed a recommendation to effect any transactions in the securities discussed herein

or an endorsement of any opinion expressed herein. Any recipient of this research in the United States wishing to effect a transaction in any security mentioned herein should do so by contacting CAS.

United Kingdom: Notwithstanding anything to the contrary herein, the following applies where the publication/communication is distributed in and/or into the United Kingdom. This publication/communication is only for distribution and/or is only directed at persons ("permitted recipients") who are (i) persons falling within Article 19 of the Financial Services and Markets Act 2000 (Financial Promotion) Order 2001 (the "FPO") having professional experience in matters relating to investments or high net worth companies, unincorporated associations etc. falling within Article 49 of the FPO, and (ii) where an unregulated collective investment scheme (an "unregulated CIS") is the subject of the publication/communication, also persons of a kind to whom the unregulated CIS may lawfully be promoted by a person authorised under the Financial Services and Markets Act 2000 ("FSMA") by virtue of Section 238(5) of the FSMA. The investments or services to which this publication/communication relates are available only to permitted recipients and persons of any other description should not rely upon it. This publication/ communication may have been produced in circumstances such that it is not appropriate to categorise it as impartial in accordance with the FSA Rules.

Singapore: This publication/communication is distributed for and on behalf of CLSA Limited (for research compiled by non-US analyst(s)) and /or CAS (for research compiled by US analyst(s)) in Singapore through CLSA Singapore Pte Ltd solely to persons who qualify as Institutional, Accredited and Expert Investors only, as defined in s.4A(1) of the Securities and Futures Act. Pursuant to Paragraphs 33, 34, 35 and 36 of the Financial Advisers (Amendment) Regulations 2005 with regards to an Accredited Investor, Expert Investor or Overseas Investor, sections 25, 27 and 36 of the Financial Adviser Act shall not apply to CLSA Singapore Pte Ltd. Please contact CLSA Singapore Pte Ltd in connection with queries on the report. MICA (P) 162/12/2011.

The analysts/contributors to this publication/communication may be employed by a Credit Agricole or a CLSA company which is different from the entity that distributes the publication/communication in the respective jurisdictions.

MSCI-sourced information is the exclusive property of Morgan Stanley Capital International Inc. (MSCI). Without prior written permission of MSCI, this information and any other MSCI intellectual property may not be reproduced, redisseminated or used to create any financial products, including any indices. This information is provided on an "as is" basis. The user assumes the entire risk of any use made of this information. MSCI, its affiliates and any third party involved in, or related to, computing or compiling the information hereby expressly disclaim all warranties of originality, accuracy, completeness, merchantability or fitness for a particular purpose with respect to any of this information. Without limiting any of the foregoing, in no event shall MSCI, any of its affiliates or any third party involved in, or related to, computing or compiling the information have any liability for any damages of any kind. MSCI, Morgan Stanley Capital International and the MSCI indexes are services marks of MSCI and its affiliates. The Global Industry Classification Standard (GICS) was developed by and is the exclusive property of Morgan Stanley Capital International Inc. and Standard & Poor's. GICS is a service mark of MSCI and S&P and has been licensed for use by CLSA Asia-Pacific Markets.

EVA® is a registered trademark of Stern, Stewart & Co. "CL" in charts and tables stands for CAS estimates unless otherwise noted in the source.

CLSA ASIA-PACIFIC MARKETS
Produced by CRÉDIT AGRICOLE SECURITIES

Research & sales offices

www.clsa.com

Australia
CLSA Australia Pty Ltd
CLSA House
Level 15
20 Hunter Street
Sydney NSW 2000
Tel: (61) 2 8571 4200
Fax: (61) 2 9221 1188

China - Beijing
CLSA Limited - Beijing Rep Office
Unit 10-12, Level 25
China World Trade Centre Tower 2
1 Jian Guo Men Wai Ave
Beijing 100004
Tel: (86) 10 5965 2188
Fax: (86) 10 6505 2209

China - Shanghai
CLSA Limited - Shanghai Rep Office
Room 910, 9/F
100 Century Avenue
Pudong New Area
Shanghai 200120
Tel: (86) 21 2020 5888
Fax: (86) 21 2020 5666

China - Shenzhen
CLSA Limited - Shenzhen Rep Office
Room 3111, Shun Hing Square
Di Wang Commercial Centre
5002 Shennan Road East
Shenzhen 518008
Tel: (86) 755 8246 1755
Fax: (86) 755 8246 1754

Hong Kong
CLSA Limited
18/F, One Pacific Place
88 Queensway
Hong Kong
Tel: (852) 2600 8888
Fax: (852) 2868 0189

India
CLSA India Ltd
8/F, Dalamal House
Nariman Point
Mumbai 400021
Tel: (91) 22 6650 5050
Fax: (91) 22 2284 0271

Indonesia
PT CLSA Indonesia
WISMA GKBI Suite 901
Jl Jendral Sudirman No.28
Jakarta 10210
Tel: (62) 21 2554 8888
Fax: (62) 21 574 6920

Japan
Credit Agricole Securities Asia BV
Tokyo Branch
16/F, Shiodome Sumitomo Building
1-9-2, Higashi-Shimbashi
Minato-ku, Tokyo 105-0021
Tel: (81) 3 4580 5533 (General)
(81) 3 4580 5171 (Trading)
Fax: (81) 3 4580 5896

Korea
CLSA Securities Korea Ltd
30/F, One IFC
10 Gukjegeumyung-ro
Yeongdeungpo-gu,
Seoul, 150-712
Tel: (82) 2 397 8400
Fax: (82) 2 771 8583

Malaysia
CLSA Securities Malaysia Sdn Bhd
Suite 20-01, Level 20
Menara Dion
27 Jalan Sultan Ismail
50250 Kuala Lumpur
Tel: (60) 3 2056 7888
Fax: (60) 3 2056 7988

Philippines
CLSA Philippines, Inc
19/F, Tower Two
The Enterprise Center
6766 Ayala corner Paseo de Roxas
Makati City
Tel: (63) 2 860 4000
Fax: (63) 2 860 4051

Singapore
CLSA Singapore Pte Ltd
80 Raffles Place, No.18-01
UOB Plaza 1
Singapore 048624
Tel: (65) 6416 7888
Fax: (65) 6533 8922

Taiwan
CLSA Limited
Taiwan Branch
27/F, 95 Dunhua South Road
Section 2
Taipei 10682
Tel: (886) 2 2326 8188
Fax: (886) 2 2326 8166

Thailand
CLSA Securities (Thailand) Ltd
16/F, M Thai Tower
All Seasons Place
87 Wireless Road, Lumpini
Pathumwan, Bangkok 10330
Tel: (66) 2 257 4600
Fax: (66) 2 253 0532

United Kingdom
CLSA (UK)
12/F, Moor House
120 London Wall
London EC2Y 5ET
Tel: (44) 207 614 7000
Fax: (44) 207 614 7070

USA - Boston
Credit Agricole Securities (USA) Inc
99 Summer Street
Suite 220
Boston, MA 02110
Tel: (1) 617 295 0100
Fax: (1) 617 295 0140

USA - Chicago
Credit Agricole Securities (USA) Inc
227 W. Monroe Street
Suite 3800
Chicago, IL 60606
Tel: (1) 312 278 3604

USA - New York
Credit Agricole Securities (USA) Inc
15/F, Credit Agricole Building
1301 Avenue of The Americas
New York 10019
Tel: (1) 212 408 5888
Fax: (1) 212 261 2502

USA - San Francisco
Credit Agricole Securities (USA) Inc
Suite 850
50 California Street
San Francisco, CA 94111
Tel: (1) 415 544 6100
Fax: (1) 415 434 6140

CLEAN & GREEN™

At CLSA we support sustainable development. We print on paper sourced from environmentally conservative factories that only use fibres from plantation forests. Please recycle.

CLSA Sales Trading Team

Australia	(61) 2 8571 4201
China (Shanghai)	(86) 21 2020 5810
Hong Kong	(852) 2600 7003
India	(91) 22 6622 5000
Indonesia	(62) 21 573 9460
Japan	(81) 3 4580 5169
Korea	(82) 2 397 8512
Malaysia	(60) 3 2056 7852
Philippines	(63) 2 860 4030
Singapore	(65) 6416 7878
Taiwan	(886) 2 2326 8124
Thailand	(66) 2 257 4611
UK	(44) 207 614 7260
US	(1) 212 408 5800

SYSTEM CERTIFICATION ISO 14001:2004 SGS

CLSA is certified ISO14001:2004

Key to CLSA/Credit Agricole Securities investment rankings: **BUY**: Total return expected to exceed market return AND provide 20% or greater absolute return; **O-PF**: Total return expected to be greater than market return but less than 20% absolute return; **U-PF**: Total return expected to be less than market return but expected to provide a positive absolute return; **SELL**: Total return expected to be less than market return AND to provide a negative absolute return. For relative performance, we benchmark the 12-month total return (including dividends) for the stock against the 12-month forecast return (including dividends) for the local market where the stock is traded.