Big Data and the Future of R&D Management

Deliverable 1: A Primer on Big Data for Innovation

Co-Chairs: Jeffrey Alexander, SRI International
            Mike Blackburn, Cargill
            David Legan, Kraft Foods (previously)

The co-chairs acknowledge gratefully the advice and research support from Prof. Diego Klabjan, Northwestern University, and Andrew Windsor, independent consultant, in providing a review of relevant literature and initial organizational case studies to inform this analysis.
# Table of Contents

1. **Introduction** ............................................................................................................... 1
2. **What Is Big Data?** ....................................................................................................... 2
   2.1 Big Data: A Constellation of Buzzwords ............................................................... 2
   2.2 “Uncomfortable Data” ......................................................................................... 4
   2.3 Analytics: Answering Big Data Questions ............................................................. 5
3. **Exploiting the Value of Big Data: Key Dimensions** ...................................................... 8
   3.1 Strategy ................................................................................................................ 9
   3.2 People ................................................................................................................ 10
   3.3 Technology ........................................................................................................ 11
   3.4 Process Integration ............................................................................................. 12
4. **Implications for R&D Management** .......................................................................... 14
   4.1 How Big Data Can Change R&D Planning ......................................................... 14
   4.2 How Big Data Can Enable New Approaches to R&D ......................................... 15
   4.3 How Big Data Might Disrupt R&D ..................................................................... 15
5. **Conclusions and Future Work** ................................................................................... 17
   5.1 Research Framework: Toward a Big Data CMM? ............................................... 17
   5.2 Project Schedule ................................................................................................. 18
   5.3 Role for IRI Members ......................................................................................... 19
6. **Contributors** ............................................................................................................. 20
1 Introduction

The Big Data project is one focus area of the IRI study “Digitalization and its Implications for R&D Management. This project was kicked off in early 2015 and will be completed in October 2016. This document is the first deliverable in the project, providing an introduction to the topic and a framework for understanding its relevance to IRI members.

Big Data is a term that is widely used, often described but with no common agreed definition. In the early phases of the Digitization Project ideation phase there was broad input from the IRI members to elicit areas of interest. There was a wide range of questions uncovered and it was clear that the IRI membership had a huge spectrum of understanding of and involvement with Big Data in their operations These extend from large research programs and groups focused on Big Data to those trying to understand what Big Data is and whether they should be concerned.

There is a great deal of published information available on the technical developments in Big Data creation, acquisition, storage and analysis. Another set of literature looks at the operational impact of Big Data on companies, primarily in customer-facing functions such as marketing and customer service. Publications on the impact of Big Data on the R&D activities, to the extent that they exist, generally come from government research institutions and academia.

The objective of this document is to help IRI members understand the concepts encompassed in the term “Big Data,” to see how Big Data is already having an impact on organizations, and to establish the framework for the next phases of this project.
2 What Is Big Data?

“There’s no point in being precise if you don’t know what you’re talking about.”

John von Neumann, pioneer in computer science

One of the first questions addressed by the working group was, “What do we mean by Big Data?” This proved difficult to answer, as much has been written about “Big Data” but with many different meanings, interpretations, and implications. Early discussions made it clear that IRI members held diverse opinions about Big Data and its importance. So, our first question became, “Is it actually helpful to define Big Data?”

Our conclusion is that Big Data is going to mean different things to different organizations. For those organizations accustomed to working with massive datasets, Big Data implies a scale far beyond state-of-the-art data management technologies. For other organizations, Big Data may be any dataset that cannot be handled by Microsoft Excel. Or, as Dr. Bill Pike of the Pacific Northwest National Laboratory puts it, Big Data is data of sufficient size and complexity to challenge contemporary analytical techniques. So, the more useful approach is to look at the characteristics of what we call Big Data, and how those characteristics relate to the way that organizations are accustomed to using data.

2.1 Big Data: A Constellation of Buzzwords

Big Data as a concept is really a confluence of a whole set of trends in computing, information processing, computational methods, and analytical tools. We recognize the cynical view of Big Data that has emerged recently—one member defined Big Data as “a word consultants use to extract money from us.” Another prevailing view is that Big Data is just another iteration of technology trends from the past, such as data warehouses and data mining. So what is going on here, anyways?

The origins of the term “Big Data” date back to the early 2000s, from two parallel technology developments. The high-energy physics community, faced with a deluge of data being generated from experiments at various particle accelerators and colliders, pioneered the development of grid computing—highly-distributed infrastructures for data storage, processing, and access. Terms such as “eScience,” or “Fourth Paradigm Science” (coined by
Jim Gray of Microsoft Research) described the emerging approach of conducting scientific experimentation entirely with data previously generated from other experiments. The underlying technologies were picked up by Web search firms like Google and Yahoo!, and were implemented in their data centers as completely new approaches to data storage and management. Google developed MapReduce and the Google File System, and Yahoo! picked up MapReduce and introduced it to the open source community as Hadoop.\(^1\)

From discussions with thought leaders and our internal discussions, we see a number of technical buzzwords that are part of the conversation around Big Data:

- **Open Data:** data that is typically in the public domain and readily available, such as government (public domain) data
- **“Found” Data:** data generated for a specific purpose but analyzed for a different purpose—for example, analyzing credit card transactions to discern consumer purchasing patterns
- **Cloud Computing:** data management and data storage on virtual, remote arrays of servers, such as Amazon Web Services or Microsoft’s Windows Azure service
- **Machine Learning:** new modes of analysis, based on techniques and algorithms from computer science, where automated algorithms can process huge datasets efficiently and produce human-interpretable results
- **Predictive Analytics:** the use of machine learning and related techniques to generate very accurate predictions about future events
- **Ubiquitous Sensing:** the deployment of sensors across a wide array of environments to collect very detailed data with an enormous number of observations
- **Social Media:** the rise of new modes of intensive online human interaction involving very large sets of users (e.g., Facebook, Twitter, Reddit)
- **“Datafication”:** the trend towards capturing more and more aspects of social and physical phenomena as digital data

Big Data really is the product of the confluence of these trends. Our capacity to capture, store, process and analyze large volumes of diverse types of data has expanded tremendously in the past 10 years, leading to the characterization of Big Data as the combination of five common attributes:

1. Volume—Big Data as being “big”  
2. Variety—Big Data generated from many sources with different characteristics  
3. Velocity—Big Data generated continuously by sources in near real-time  
4. Variability—Big Data generated with differences in timeframes and “burstiness”  
5. Veracity—Big Data from sources that may not be entirely trustworthy (such as data on social media)

The first three “V’s” were coined by Doug Laney in 2001, when he served as an analyst at the market research firm Meta Group. The two additional V’s have been added to help organizations in assessing the trustworthiness of insights generated by Big Data. The issue of trustworthiness highlights an important issue raised by Big Data—the degree to which the scale, scope, and sophistication of Big Data analysis can overcome the potential pitfalls inherent in dealing with high-velocity, diverse data streams.

### 2.2 “Uncomfortable Data”

Discussions within the working group and at IRI meetings demonstrated that there is no single set of criteria that defines Big Data across all organizations. In fact, one problem with the term Big Data is that it does not describe a particular technology or approach. As noted by Stephen Hoover, CEO of the Palo Alto Research Center (PARC) in his remarks at the 2015 IRI Annual Meeting, “Big Data isn’t a solution—it’s a problem.” The challenge to organizations is that we are all going to need to deal with Big Data sooner or later. In fact, the market research firm GartnerGroup recently removed Big Data as an item on its annual “hype cycle” chart of emerging technologies, arguing that Big Data is so pervasive that it’s no longer emerging. The question isn’t whether or not organizations will exploit Big Data for competitive advantage—it’s what they will do with the Big Data that is going to become part of the operating environment.

Instead of talking about Big Data, participants at the 2015 Winter Meeting suggested that the term “Uncomfortable Data” might be more relevant. Uncomfortable Data refers to any set of data so large and unwieldy that it defies analysis using the tools and methods normally employed by an organization. Uncomfortable Data emphasizes the challenge

---


3 Sharwood, Simon. “Forget Big Data hype, says Gartner as it cans its hype cycle.” The Register, 21 August 2015. Available at http://www.theregister.co.uk/2015/08/21/forget_big_data_hype_says_gartner_as_it_cans_its_hype_cycle/
posed by Big Data—it requires new technologies and new approaches to enable organizations to use data effectively in improving decision-making and operations.

Uncomfortable Data also brings out how Big Data challenges the role of data in organizations. Due to their attributes of variability, (questionable) veracity, etc., high-velocity datastreams are inherently uncertain. Data is being collected and transferred at such high rates and in such volume that an analyst may not have the time, resources, or ability to understand all the structural properties of those datastreams. The data may be biased, for example, or it may be corrupted. The pace of business will not always allow analysts the luxury of cleansing and validating those data. In a world of Uncomfortable Data, decisions become probabilistic—the data tells us what we think we know to be true, but we can't be completely sure. So, organizations need to decide what level of uncertainty they are comfortable in managing, and calibrate their use of Big Data to match their preferences, if possible.

The other dimension of the Big Data challenge is complexity. In traditional data analysis, we’re accustomed to examining fairly well-defined phenomena using just a few attributes. As noted by Ananth Krishnan, Chief Technology Officer of Tata Consultancy Services, a company might define each customer simply with a name, address, phone number and account. Now, a company can tap Big Data to describe each customer in terms of product preferences, purchasing patterns, travel history, number of touchpoints—a huge number of variables that produce a much more nuanced, but also more complex view of each customer. This is a tremendous range of data that must be integrated and correlated to gain a deeper understanding of that customer. So, the data environment is becoming very complex, with so much information available that it exceeds human cognitive capacity. New tools like machine learning can reduce that complexity to a degree, but organizations must still become comfortable with dealing with large numbers of variables simultaneously, both in their internal data and in data about the external environment.

### 2.3 Analytics: Answering Big Data Questions

If Big Data is the defining challenge of a data-rich environment, what is the solution? The answer, in short, is advanced analytics—new techniques for ingesting, parsing, and processing Big Data datasets to then build new types of models that will identify latent relationships between variables leading to new insight. This interaction between data, models, and analysis are the core of the promise of Big Data for applications such as R&D.
INFORMS (the Institute for Operations Research and the Management Sciences), in 2010, commissioned a report by the consulting firm Capgemini to study the rise of analytics and its implications for the society. The study sought to confirm whether or not the field of analytics was a new successor to what had been labeled as ‘operations research,’ and to what extent it might differ from OR. In what is now a seminal work in understanding this new field, Capgemini argued that analytics in many ways superceded OR to become a broader phenomenon, in large part because advanced analytics is driven not just by methodologies and techniques but by business concerns.4

The report adopted the definition that Analytics facilitates realization of business objectives through reporting of data to analyze trends, creating predictive models for forecasting and optimizing business processes for enhanced performance. Capgemini drew further distinctions among a set of progressively more sophisticated forms of analytics:

- **Descriptive analytics:** using data to find out what has happened in the past.
- **Predictive analytics:** using data to find out what might happen in the future (i.e., forecasting and estimation).
- **Prescriptive analytics:** using data to identify the courses of action that are likely to produce the best outcomes under given conditions.

In one framework for comparing these approaches,5 descriptive analytics is comparable to traditional business intelligence solutions—the compilation of statistics and major findings about past activities and conditions in a given time period. Predictive analytics takes traditional forecasting but applies new techniques to create very sophisticated models of an environment. The work of quantitative hedge funds in modeling the stock market is one example of predictive analytics at work. Modern approaches use advanced statistical methods and machine learning algorithms to isolate and study thousands of variables simultaneously in a predictive model. This enables organizations to construct very complex models of an environment and observe the interactions of those many variables to determine which ones drive the emergence of a potential future result.

Prescriptive analytics applies techniques such as optimization, simulation, and heuristics-based decision-making to test the potential consequences of pursuing alternative strategies or courses of action. This type of analysis provides an organization with a means of

---


understanding trade-offs between decisions, and may improve the quality of decisions by integrating factors beyond the capabilities of human cognition. One area of prescriptive analytics, for example, is in computer-assisted diagnosis, where a physician can enter in observations about a patient and rapidly scan the entirety of the medical literature to identify diseases or disorders that might be generating those symptoms, rather than relying solely on memory to generate those diagnoses.6

3 Exploiting the Value of Big Data: Key Dimensions

“Data is like crude. It’s valuable, but if unrefined it cannot really be used.”

The above quote is commonly attributed to Anne Winblad of the venture capital firm Hummer Winblad. In reality, it was first written by Michael Palmer, Executive Vice President for Member Relations at the Association of National Advertisers. This isn’t so surprising, as the marketing and advertising functions in a company are among the most prolific users of data, and in particular of data generated from external sources. We found that many examples of how companies are exploiting Big Data for competitive advantage came from marketing—for example, mining Twitter feeds to understand customer sentiment about a brand, or to detect product complaints. Other functional areas are beginning to tap into the potential of Big Data, demonstrating how it can have a significant and transformative effect on how organizations and markets will behave in the future. In this section, we’ll describe some of the ways that organizations have generated real business value from Big Data.

The working group realized early in our deliberations that we needed a framework to understand how Big Data might affect R&D management, and especially how R&D organizations might need to change to operate effectively in a world of Big Data. We took inspiration from the work on Capability Maturity Models, a structured method of benchmarking and evaluating how well organizations are able to perform various types of functions (e.g., software engineering, supply chain management). We are examining the way that Big Data may change how organizations approach strategy, people, and technology, and finally, how they integrate Big Data with existing business processes.

---

7 http://ana.blogs.com/maestros/2006/11/data_is_the_new.html
3.1 Strategy

The strategic implications of Big Data should be obvious now, in light of the case studies in Section Two. Big Data promises to change the way that organizations make strategic choices and how they decide on a course of action. In some cases, the actual business strategy of an organization may shift due to the capabilities enabled by Big Data techniques. In the case of Caterpillar, for example, the company is developing an entirely new service-based business by taking data generated by its own products in the field, analyzing those datastreams, and selling the resulting insights back to its own customers.8

More commonly, an organization will need to find ways to ensure that it is fully exploiting Big Data analysis in its existing decision-making processes. For example, Amazon has created strategic advantage by allowing algorithms to play a role in setting product prices, rather than leaving that task to human analysts. In a marketplace the size of Amazon’s online presence, the dynamics of supply and demand can change so quickly that humans would be unable to respond in a useful timeframe.

As one expert told us, this issue of timing is a critical part of ensuring that Big Data produces business value. For those decisions that require human judgment, Big Data can provide extremely valuable insights, especially through predictive analytics. The organization needs to ensure that the collection, parsing, and processing of data takes place in time to produce an analytical output that arrives in time to influence the decision. For this reason, Big Data is an activity that is now being assigned to a Chief Data Officer or Chief Analytics Officer, not the Chief Information Officer. The CIO is tasked with running infrastructure, maintaining enterprise architecture, and then assisting in the development of business applications for generating new business value. In contrast, the CDO or CAO needs to be someone who is very familiar with strategy and business needs, and thus able to direct Big Data analysis towards those concerns.

Given that Big Data entails significant technical and process challenges, as described below, organizations need to invest time and effort in understanding how to deliver timely Big Data insights to leaders, in a manner that those leaders can understand and evaluate.

3.2 People

The skillsets required to deal with Big Data are different from those that most firms are accustomed to acquiring. In 2012, DJ Patil and Tom Davenport popularized the term “data scientist” in the Harvard Business Review. As they put it, “Data scientists are the people who understand how to fish answers to important business questions from today’s tsunami of unstructured information.” These are people who have two critical sets of skills—an understanding of the nature of problems facing an organization, and the ability to find, build and model datasets that can be analyzed to produce relevant findings. They are the embodiment of what some have called “pi-shaped” experts. Most experts are “T-shaped,” meaning that they have some degree of knowledge across a broad range of topics, but very deep knowledge in only one domain. Data scientists need to complement that domain knowledge with equally deep knowledge of the tools and methods of data science—statistics, coding, and data management.

While most firms recognize the importance of IT, and especially highly-skilled IT staff, to corporate competitiveness, data scientists different from most other IT workers. To a large extent, data scientists are trained as scientists—they performed scientific research in graduate school, and that research often required the analysis of large, messy datasets. With the rise of eScience, graduate students in science and engineering are likely to develop skills

---

in statistics and coding, either formally or informally, as part of their studies. Their domain expertise in their degree field—chemistry, physics, biology, etc.—may qualify them as scientists, their hands-on experience in using data as part of the research process makes them data scientists.

As Patil and Davenport point out, data scientists are analogous to those in the 1990s with Ph.D. degrees in physics and math who went to work doing quantitative modeling on Wall Street. For those “quants,” their skills in rigorous modeling and multi-variate analysis made them extremely valuable, and they could pick up the finance and economics on the job. In the data science world, the critical skills are machine learning, advanced statistics, data management, and predictive modeling. However, the better that data scientists can pick up business skills and awareness, the better they will be at ensuring that the organization can exploit data analytics more fully. As noted by Clay Heaton of RTI, “Big Data is optimal when the people who are using the technology are also familiar with the business space as well as the technology side.”

In their short book Analyzing the Analyzers, Big Data experts Harlan Harris, Sean Patrick Murphy, and Mark Vaisman asked working data scientists what skills they used most often in their jobs. In varying degrees, they mentioned some combination of business skills (e.g., product development), machine learning and analytics techniques, mathematical approaches such as operations research or Bayesian reasoning, programming (coding) skills, and advanced statistics. Data scientists are now in high demand. The number of position announcements recruiting for data scientists has skyrocketed since 2012. Many universities now offer specialized degree programs in data science or analytics. These programs might be found in engineering schools, business schools, or information schools (i-Schools), illustrating the interdisciplinary and changing nature of this type of expertise.

### 3.3 Technology

Experts interviewed by the working group tended to downplay the technical aspects of Big Data, although of course they are a major force in the emergence of this trend. Big Data as a buzzword would not exist but for innovations like Hadoop, MapReduce, and a host of database and analytical systems and tools. Most of the “tools of the trade” in Big Data are open source projects, such as Mallet (topic modeling using text analytics), R (advanced

---

statistical analysis) and D3 (data visualization). Most of these came out of university laboratories, and are still evolving.

The dynamic and experimental nature of Big Data technology is part of what sets it apart from previous generations of data technologies, such as data mining and data warehousing. Clay Heaton, a researcher in the Center for Data Science at RTI International, pointed out that most of the tools he uses are still experimental and require extensive customization and modification. That tends to run counter to the preferences of an organization’s central IT unit, which would prefer commercial off-the-shelf technology and stable platforms. It also explains why data scientists need to be programmers as well as analysts, because the tools and techniques used may depend heavily on the datasets available, the nature of the problems being investigated, and the degree of certainty (or uncertainty) that is acceptable. Identifying Big Data by its associated technologies would be difficult, as the toolset for Big Data today is different from what it was just a few years ago, and it will continue to change. What an organization needs is a tolerance for experimentation, tinkering, and creative approaches driven by what its data scientists feel might work.

3.4 Process Integration

The final domain we are investigating is how organizations are integrating Big Data into their existing operations, and the degree to which existing processes may need to adapt to a Big Data environment. Marketing and market research are early adopters of Big Data, and we are already seeing major shifts in those functions. Instead of relying primarily on customer surveys and direct customer feedback, organizations now seek to understand customers by mining social media data, analyzing transaction records, and delving into a variety of new and existing pools of data. One result is a dramatic decrease in the cycle time from product release to market assessment and feedback. In a digital marketplace like Amazon.com, the success of a product launch can be judged in a few days. Even for more traditional “brick-and-mortar” retail environments, manufacturers are receiving information on customer response within a few weeks.

Process integration is likely to be one of the most challenging aspects of Big Data. As noted earlier, Big Data provides us with new analytical capabilities and new resources for understanding the environment, but all with some degree of uncertainty and complexity. Organizations will need to adjust to a probabilistic world, where decisions must be made very quickly but perhaps based on analysis that is not fully verifiable. As Bernie Meyerson of IBM put it, organizations will need to speed up their management and control processes
to keep pace with their information processes. Annual planning cycles may become irrelevant if feedback on the success of a strategy or campaign is available within weeks, or even days. We will continue to explore this aspect of Big Data as this project unfolds.
4 Implications for R&D Management

Based on internal discussions, consultations with IRI members, and our thought leader interviews and research, we see three major ways that Big Data might transform the practice of R&D management in the future—in the planning, execution, and possible disruption of the R&D function.

4.1 How Big Data Can Change R&D Planning

The essence of R&D strategy is the process by which the organization selects which projects to undertake and continue, and which to ignore or abandon. Big Data may offer very new insights that will change the management of the R&D pipeline. Already, companies like Caterpillar, Rolls Royce Aerospace, and Tesla are using Big Data generated from sensors to plan their next product improvements and to determine what features and components need to be enhanced or created. Maintenance data, for example, provides an extremely rich source of opportunities for increasing Mean Time Between Failures.

The rise of predictive analytics may enable even more substantial change. Experts point to the health care and pharmaceutical sectors as the vanguard in this trend. To a large degree, drug prescriptions are made using a trial-and-error process—the doctor observes the patient’s symptoms, and prescribes a drug that seems likely to address the underlying cause, and then waits to observe the effects on the patient. Bernie Meyerson, Chief Innovation Officer at IBM, discussed how both the genomics revolution and the implementation of electronic health records is revolutionizing the development of new therapies. Combining those two massive datasets, data scientists can observe patterns in the drugs prescribed, the outcomes of those therapies, and the genetic profile of the patients. In the future, doctors may be able to know in advance that the genetic markers that indicate the presence of a disease, along with the patient’s specific genetic profile, necessitate the use of a very specific drug that would be the most effective treatment. The entire domain of drug development is moving towards this genomic-based approach.

One objective pursued by a number of initiatives is the application of predictive analytics to technology and market forecasting. If data scientists can build models that predict the success or failure of a product, or that can foresee the emergence of a new disruptive technology, then organizations will be able to plan their R&D project by anticipating where the market will be, not based on retrospective analysis of what worked in the past.
4.2 How Big Data Can Enable New Approaches to R&D

The world of Big Data may change the entire scientific research process, as the rise of eScience already indicates. One aspect of this is the rise of “virtual experimentation,” where researchers and engineers can test hypotheses about new products using digital datasets before a single prototype is produced. In the chemical and materials sectors, for example, databases of chemical reactions and molecules might enable firms to understand the properties of a new compound without physical testing.

One potential shift in R&D is the rise of “machine-generated hypotheses.” In their article on “Metaknowledge” in *Science*, researchers James A. Evans and Jacob G. Foster described how the machine analysis of the entire corpus of scientific literature in a particular domain, such as biomedical research, could uncover hidden biases in how researchers select, design, and conduct experiments. An algorithm could, for example, determine that a “gap” exists in the literature concerning research into the application of a relevant therapy as treatment for a particular condition, inspiring a new line of research. To a certain extent, advanced analytics and artificial intelligence may become an integral part of how research is designed and conducted.

4.3 How Big Data Might Disrupt R&D

A third area for exploration is the extent to which traditional R&D organizations and approaches could be disrupted by new players using the capabilities of Big Data. One example in the headlines today is Hampton Creek Foods, a start-up in the food sector. While traditional competitors in that market have been developing products using food chemistry, Hampton Creek has been analyzing data on the properties of plant proteins through genomic analysis to invent entirely new foods. One such effort, for example, uncovered a protein that mimics the emulsifying property of eggs, leading to the development of a vegan version of mayonnaise.

Hampton Creek has garnered much attention, both by taking on a key product category of companies like Unilever, and by challenging traditional frameworks of food regulation. The Food and Drug Administration recently ordered Hampton Creek to stop selling its version of mayonnaise under the name Just Mayo, claiming that the term mayonnaise only applies to a type of product that contains eggs. Regardless of the outcome of the controversy,
Hampton Creek exemplifies how the application of Big Data in an unconventional area could enable a small team of data scientists could take on large incumbents in the market by developing new products with completely novel tools and techniques.
5 Conclusions and Future Work

The Big Data Working Group is scheduled to continue its research for the next year. Our plan is to pursue further investigation of this phenomenon using a common research framework, designed to help IRI members in understanding the potential impact of Big Data and how they may need to be positioned to deal with that impact.

5.1 Research Framework: Toward a Big Data CMM?

Our working group has developed an agenda of major research questions and subquestions by matrixing the three perspectives on the impact of Big Data on R&D management against the four organizational dimensions, as shown below:

<table>
<thead>
<tr>
<th>RQs</th>
<th>Strategy</th>
<th>People</th>
<th>Technology</th>
<th>Process Integration</th>
</tr>
</thead>
<tbody>
<tr>
<td>How will Big Data inform R&amp;D activities?</td>
<td>How could overall R&amp;D management improve from the use of Big Data?</td>
<td>Who will be using Big Data to inform R&amp;D management, and what will they need to know?</td>
<td>What Big Data technologies and systems will R&amp;D management use to improve decision-making?</td>
<td>How will R&amp;D management practices &amp; processes change as Big Data becomes pervasive?</td>
</tr>
<tr>
<td>How will Big Data enable new R&amp;D activities?</td>
<td>What new capabilities and approaches to R&amp;D become possible thanks to Big Data?</td>
<td>How will the skills and knowledge of R&amp;D teams change to make use of Big Data?</td>
<td>What Big Data technology and systems will become part of the R&amp;D function?</td>
<td>How will R&amp;D activities change as Big Data becomes pervasive?</td>
</tr>
<tr>
<td>How will Big Data disrupt traditional R&amp;D?</td>
<td>How can Big Data create/identify opportunities to disrupt markets and industries? How might competitors use Big Data against incumbents?</td>
<td>Who will use Big Data as a tool for disruption, and what will they look like?</td>
<td>What technologies on the horizon enable future disruptive opportunities?</td>
<td>What should companies do to predict/exploit disruptive opportunities using Big Data?</td>
</tr>
</tbody>
</table>

Our hypothesis is that an organization’s ability to leverage more fully the potential gains of Big Data in its R&D function will depend on how well it addresses each of the cells in this
One could envision, perhaps, a method of rating how an organization is able to manage the impact of Big Data on R&D planning, execution, and disruption by modifying its strategy, staffing, technology, and process design. The end result may be a type of “capability maturity model” that suggests a roadmap from organizations that have yet to prepare for the Big Data world, to those that are masters at harnessing Big Data for R&D management. For this project, we propose to build at least the framework on which a future CMM might be developed.

### 5.2 Project Schedule

To execute our research agenda, we have a rough roadmap of activities and deliverables planned for the coming year. The first set of activities, consisting of thought leader interviews and organizational case studies, was launched over the summer and will continue with the development of this document. We anticipate another interim report to be completed in time for the 2016 IRI Annual Meeting, with the final report presented at next year’s Member Summit.
5.3 Role for IRI Members

As with any other IRI research-on-research effort, the success of this project will depend on contributions from all IRI members, not just those involved in the working group. By participating in this effort, we believe that IRI members will be able to gain a more holistic view of the implications of Big Data for their R&D organizations, and be able to understand how Big Data trends and techniques are evolving in the broader context of data science and “datafication.” At the end, members should:

- Have a toolkit for discussing big data with senior management
- Appreciate how to leverage current and emerging trends in Big Data and data science to improve competitive positioning, create new value, and help their teams

In the next and final phase of the project, we plan to profile a wide range of organizations that are leveraging Big Data and advanced analytics to obtain competitive advantage or other types of value. We suggest that members nominate potential case study organizations to the Big Data group, whether those are your own organizations or ones you have encountered in other contexts. We also continue to solicit ideas of “thought leaders” to interview for this project. We intend to use these inputs to derive some proposed best practices in the use of Big Data for R&D management. In turn, those best practices will provide the basis for an amended research framework that can, in turn, become the framework for a future Capabilities Maturity Model. The value of this project will be enhanced if members can provide us with a broad, diverse set of case study organizations across multiple industries and sectors, so that we can observe how Big Data is affecting R&D management in many different manifestations.
6 Contributors

A partial list of active contributors to the development of this report:

Jeffrey Alexander, SRI International
Kevin Ambrose, Crown Holdings
Manjula Ambur, NASA Langley Research Center
Jennifer Bennett, Phillips 66
Mike Blackburn, Cargill
Ivan Cornejo, Georgia-Pacific
Marta Fernandez, Arup
James Foster, Archer Daniel Midlands
Rob Fuchsteiner, Ingersoll-Rand
Richard Golding, Crown Holdings
Louis Gritzo, FM Global
Sandra Hopkins, Georgia-Pacific
Diana Hoyt, NASA
Semyon Kogan, Gen3 Partners
Darin Latimer, Dahaner
David Legan, Kraft
Pushpa Manukonda, Deere & Co.
Mark Matlock, Archer Daniel Midlands
Nancy McMillan, Battelle
A.J. Rao, USG
John Reid, Deere & Co.
Lou Reyes, J.M. Smucker
Ramanathan Sugumaran, Deere & Co.
Mark Zhuravel, TATA Consultancy Services
Thought Leader Interviews

Evan Burness & Merle Giles, National Center for Supercomputing Applications
Jon Ellenthal, Walker Innovations
Clay Heaton, RTI International
Ananth Krishnan, TATA Consultancy Services
Bernie Meyerson, IBM
Bill Pike, Pacific Northwest National Laboratory