



ExecBlueprints™

in partnership with Aspatore Books

Action Points

I. What Challenges Does “Big Data” Present?

In the process of establishing a single, managed environment for your organization’s data, you should be prepared to address at least two levels of challenges: those relating to data quality and security, and those generated by your user community who may either criticize the system for its inflexibility or praise it unduly for its “unlimited capacity.”

II. The Bottom Line

Determining the ROI of centralizing your information into a big data store could prove tricky, unless you have solid metrics that demonstrate a clear link between effective data management and more revenue (or greater efficiency). You can measure the system’s effectiveness by asking users to rate its ability to solve problems or by conducting audits.

III. Must-Have Practices for Managing Big Data

To effectively handle this key IT responsibility, you and your team need to be both operationally diligent in monitoring performance and logging activity as well as sufficiently flexible to easily accommodate new technologies and regulations. To avoid disruption, however, perform incremental adjustments rather than massive roll-outs.

IV. The Golden Rules for Working with Non-Technical Users

A properly trained staff is your best tool for big data analysis. To truly engage your users, you must not only expose them to in-person and online trainings, but also seek their opinions on surveys and in user groups. Finally, take it a step further with your key decision makers and power users: invite them to participate in your system reviews.

V. Essential Take-Aways

Although it is true that big data stores will enable your organization to streamline data management while also discovering deeper insights, many challenges in the areas of architecture, tools, governance, analytics, and the cloud await the naïve explorer. What is necessary for success? Robust support mechanisms, able leadership, and usable data.

Contents

About the Authors	p.2
W. Greg Price	p.3
Perry Rotella	p.8
Sam Schoelen	p.10
Dan Hoogterp	p.12
Ideas to Build Upon & Action Points	p.15

The CTOs/CIOs from Troy University, Verisk Analytics, Continental Resources, and Bankrate on:

Addressing the “Big Data” Issue: What You Need to Know

W. Greg Price

Chief Technology Officer, Troy University

Perry Rotella

Senior Vice President and Chief Information Officer, Verisk Analytics

Sam Schoelen

Chief Information Officer, Continental Resources Inc.

Dan Hoogterp

Senior Vice President and Chief Technology Officer, Bankrate Inc.

These days, you’re probably hearing a lot of hype about “big data.” Vendors are currently hawking a wealth of new tools, all of which promise to help your organization unlock previously inaccessible insights from your proprietary information. According to the authors, there is no doubt that big data, i.e., organization-wide data that’s being managed in a centralized repository, can yield valuable discoveries that will result in improved products and performance — if properly analyzed. Nonetheless, you must look before you leap. First, is your company culture ready for such a move? How will data managers be affected when scores of discrete data silos are gathered and reviewed as a whole? How will you involve leadership and others in ongoing decision-making processes? How will you choose your architecture and tools from the dizzying array of options that are currently available? How will you stay up-to-date in this rapidly evolving field? Finally, how will you train your company’s users so that they can actually leverage the new capabilities? This ExecBlueprint explores these and other key concerns. ■

About the Authors



W. Greg Price

Chief Technology Officer, Troy University

W. Greg Price is presently chief security and technology officer for Troy University. Additionally, Mr. Price is the founder and director of the Alabama Computer Forensics Institute (a U.S. Department of Justice-funded research and training facility) and the elected president of the Pike County Board of Education. In his position at Troy, Mr. Price is responsible for all administrative computing systems and telecommunications at Troy University, which has over 50,000 end users across the globe.

With the Alabama Attorney General, Mr. Price is also a co-founder of the Alabama Cyberkids Awareness program. Since 2006, Mr. Price has spoken to thousands of Alabama youth on the benefit of Internet safety. His latest book, *Thieves of Innocence*, was released in the winter of 2011. The book details the Internet exploits used by child predators. In 2012, he received the International Laureate for his work for promoting child welfare through technology.

Mr. Price has worked as a researcher for two Fortune 500 companies, has

three provisional technology patents, and has co-authored a security planning book and numerous technology articles. His information security team at Troy University was recognized by *SC Magazine* as one of the Top 3 Security Teams in the U.S. for 2009, 2010, and 2011. Mr. Price holds over 30 IT certifications.

[Read Greg's insights on Page 3](#)



Perry F. Rotella

Senior Vice President and Chief Information Officer, Verisk Analytics

Perry F. Rotella is senior vice president and chief information officer of Verisk Analytics. He is responsible for the development and execution of information technology initiatives in support of the company's business operations. Before joining Verisk, Mr. Rotella was senior vice president and chief information officer of Moody's Corporation, where he served as the first CIO for the corporation and was responsible for all

aspects of global information technology strategy, support, and operations.

Before that, he served as CIO for AIG's Domestic Brokerage Group and as AIG's global chief technology officer. Previously, he spent 15 years at American Management Systems (AMS), where he filled a number of roles in the consulting organization and became chief technology officer for the Insurance Technology Group.

Mr. Rotella also serves as the president of the New York chapter of the Society for Information Management (SIM), an association of senior IT executives and prominent academicians and consultants who come together to share their intellectual capital for the benefit of members and their organizations.

[Read Perry's insights on Page 8](#)



Sam Schoelen

Chief Information Officer, Continental Resources Inc.

Sam Schoelen is the CIO of Continental Resources, a provider of advanced technology solutions and support to businesses, governments, and academia.

He has served Continental Resources for more than one half of his 15-year tech career, running corporate IT and managed services for the past two years.

[Read Sam's insights on Page 10](#)



Dan Hoogterp

Senior Vice President and Chief Technology Officer, Bankrate Inc.

Dan Hoogterp has served as senior vice president and chief technology officer of Bankrate Inc. since May 2005. From 2002 to 2005, he was chief executive officer of TQuist LLC, a boutique technology consulting company focused on content and document processing. From 2001 to 2002, he served as executive vice president and

chief technology officer of Enamics Inc., a category leader in business technology management. From 1999 to 2001, he served as SVP and CTO of Sagemaker Inc., a provider of enterprise information portals to energy, telecom, insurance, and other industries. From 1991 to 1999, he served as CEO and chief architect of Retrieval Technologies Inc.

In each position, Mr. Hoogterp focused on linking business operations and drivers with innovative, timely technology and product solutions. He is an occasional speaker at events in search engine and online service areas.

[Read Dan's insights on Page 12](#)

W. Greg Price

Chief Technology Officer, Troy University

Questions Leading to the Development of a Core Data Repository

A number of years ago, while attending a C-level meeting, I witnessed a series of activities that continue to fuel my fascination with what we now call “big data.” A discussion arose regarding an

By measuring and tracking big data, we can take deeper dives and discover greater insights and value.

elemental component of our university’s strategic plan. A participant presented a report, another participant presented a competing report, and so on. Once the dust settled, the room turned to the integrated technology expert (me) and inquired, “Why”? How was it that a single issue could have multiple, competing reports prompting different conclusions? Was the data source corrupt? Was the reporting software functioning improperly? A litany of questions arose.

In a prior engagement, I had worked within a large corporate environment, where data was our core asset. All data management exercises were conducted with pristine accuracy, a single methodology was employed to assess and benchmark effectiveness, and it was all based on the same set of large, yet, consistent data stores.

Fast forward to the large, public university environment and one encounters something drastically different. My employer is such a large institution. We operate

globally, with online and in-class sessions at any combination of venues one can imagine. As a result, data was being stored locally, semi-regionally, disaggregated, and it was available for consumption by those who were aware of its ubiquitous, yet, somehow elusive presence. We suffered from a horrible case of data sprawl.

W. Greg Price
Chief Technology Officer
Troy University

I began asking questions. I began to seek the systems within which the data existed. I sought clarity to a cloudy and confusing topic: “How were we making informed decisions?”

Over the course of a few years, we not only answered questions raised by competing responses to unique inquiries, we remedied the matter decisively. The ad-hoc data sources were collected, assimilated, and forged into a core data repository. Those aware of university culture can appreciate that this occurred somewhat begrudgingly — yet, with senior management endorsement, we shed the visceral decision-making tactics and began to rely solely on clean, big data. Informed decisions are the norm now.

The Mission of Data Management

Often, in the big data discussions, skeptics of centralized data management will suggest that



W. Greg Price
Chief Technology Officer
Troy University

“We manage all of the data, everywhere. That is our mission for data management – we endeavor to centralize the storage and management of all data elements.”

- *Founded Alabama Computer Forensics Institute in 2007*
- *Founding member and president, Alabama River Region Chapter of Infragard, a member of the American College of Forensics Examiners and the Institute of Computer Forensic Professionals*
- *Named among top 100 IT leaders in the world by ComputerWorld in 2010*
- *Trained at Troy University, Harvard University, the SANS Institute, and IBM education facilities*

Mr. Price can be e-mailed at greg.price@execblueprints.com

centralization reduces creative use of data, creates bottlenecks, and builds one monolithic group that is the one true source of data.

I agree. Decentralized data and decentralized data management do present the opportunity for leaner operations, for faster access to data. However, if one only peers at the cover of a book, how can one appreciate its contents?

Proper centralization creates opportunity and fosters completeness. Decentralization often prompts partial-picture analysis and misinterpretation of the whole.

Importance of Properly Trained Staff

A properly trained staff is your best tool for big data analysis. We conduct scores of live, in-person training events and offer continuous online training. We measure staff knowledge and require updating of skills to access the large data stores. In so doing we do not trivialize the importance of information that the masses of data we maintain can yield. We would love to employ data scientists; however, that is not practical for our diverse and

widespread workforce. Instead, we seek employees who possess critical thinking skills, statistical training, and the ability to wonder.

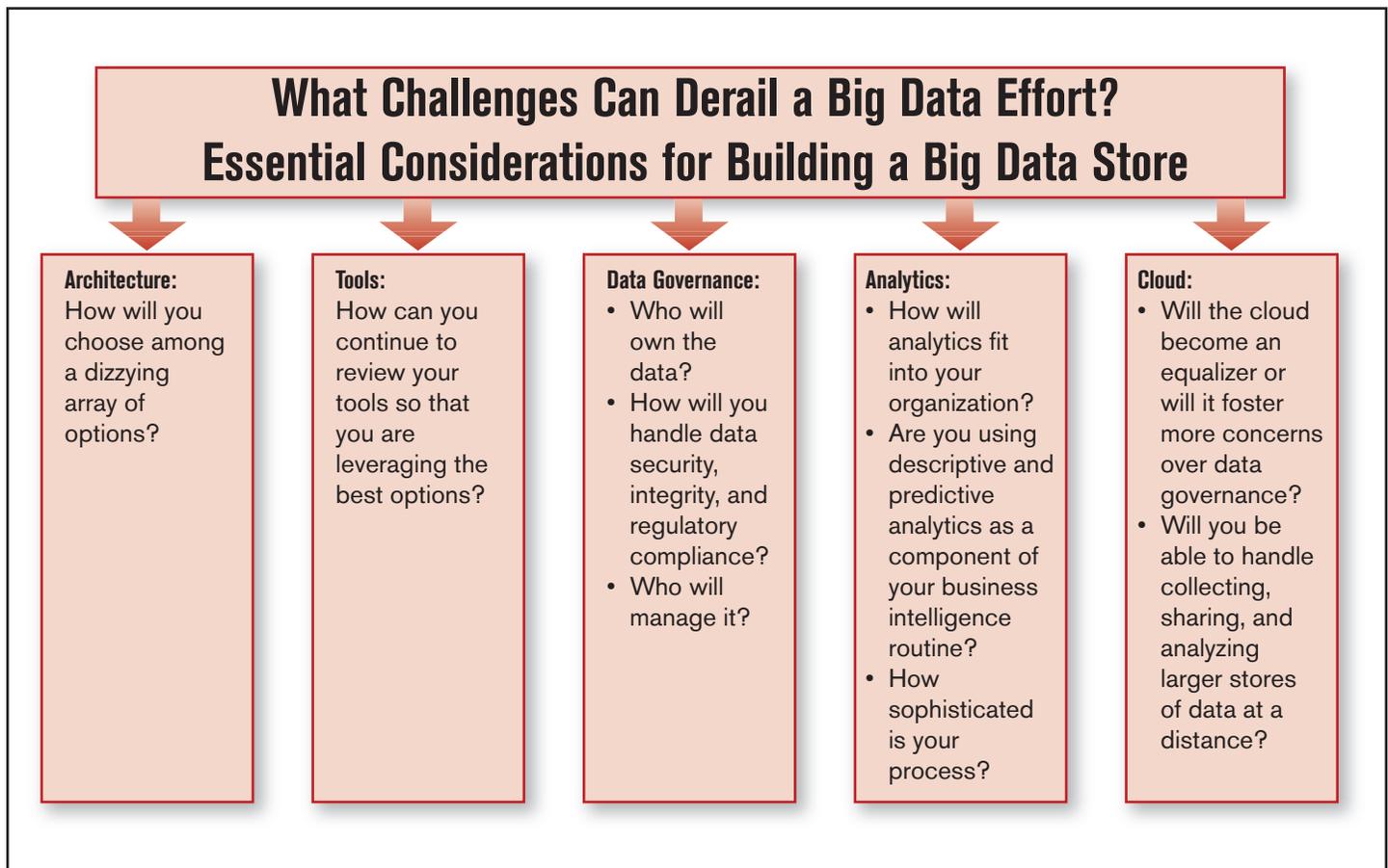
Types of Data Managed

A popular saying goes, “You can’t manage what you don’t measure.” We generate and store enormous amounts of data. In its natural and disparate form, common data is scattered. In the aggregate, big data is formed. By measuring and tracking big data, we can take deeper dives and discover greater insights and value. The derived knowledge results in improved decision making and a quality byproduct: improved performance.

We manage massive stores of student data on over 40,000 active

students in 17 time zones. All interactions with our systems — such as site statistics, event data, and transaction data — are logged. Additionally, we maintain large systems for government agencies in our data center. As part of our hosting contract, we manage reporting systems for those agencies.

The system metrics are nearly synonymous for all of our hosted data. However, based on the constituent base, the data types differ. The major functions of the university are instruction, research, and administration. Within those verticals, student data is varied. Within instruction, the data is academic progress. If the student participates in research, those elements are stored. A student’s business and non-instructional needs are



addressed through the administrative function. Woven among these different silos are common denominators. The common denominators are the keys to managing seamless data stores across large data sets.

Our data storage increases significantly every year. For example, from 2005 through 2010, our data storage needs doubled approximately every 18 months. Currently, our data storage capacity doubles every 10 months. Our operations growth and added features to our Web presences have substantially contributed to this increase.

Best Practices for Staying Up-to-Date and Warehousing Data

Our development cycle is constant. Frequently, we refer to our continuous development cycle as a “continuous deployment process.” We not only continuously refine our deliverables but we are always debuting minor adjustments, which, based on our analytics, will result in easier adoption among our user base. By addressing problems incrementally, we avoid massive roll-outs of wholesale changes and thereby minimize disruption from potential product failure and user base shock. In this way, we maintain business continuity.

When warehousing data, we follow an IT best practices guide that we have created through a collaborative effort. The document is part of this continuous development program, where IT assets meet regularly with key decision makers and power users. Collectively, the group reviews the best practices guide quarterly — a schedule that is necessary, owing to rapid changes in deliverables, our

Expert Advice

Paperless Initiatives Beginning in 1997, Troy University began an effort to digitize all paper files. The continuous deployment of new paperless initiatives is a strategic initiative for the organization. The importance of a fully digitized environment is multifaceted. In a global organization, the need to present data quickly despite location is pivotal to success. With a properly deployed paperless environment and thoroughly trained employees, the delay effects of time and distance are erased.

frequent instruction cycles, and technology updates. We employ an information technology infrastructure library (ITIL)-based effort that documents the plans, effort, and changes associated with data storage/management.

Although we are cognizant of what we store, an outsider’s opinion or service often identifies a better way of using our data. The world is flush with rich, unexpected pools of talent and we have discovered that data are being managed to form new sources of knowledge that no single source can address. We consider new solutions, looking for gaps in our deliverables. Effective big data management also entails appreciating its impact on a customer. A vendor who does not understand that big data has the potential to prompt considerable change in management style and cause cultural disruptions within the entity does not meet with my group. When reviewing a new product, if I see no testimony to the potential managerial changes that evidence-based decision tools will effect among a vendor’s literature, I disregard the solicitation.

Big Data Tools

Despite the size of the data, tools that manage data are relatively similar and have the ability to scale. Piles of data are thrown at the big

data tools and they must be able to handle large volume analysis within a tolerable amount of time. Big data analysis reveals meaningful patterns within large sets of data. In addition, discrete analysis of small data will continue to be valuable for agile situations, but this will likely still be performed using big data tools.

Many tools, including open-source tools, are available for big data analysis. Hadoop is the most common framework, and we deploy it within our environment. The largest issue with current software platforms, however, is that the tools were not designed for the typical business user. Clearly, change is underway as solution providers race to provide easier-to-use tools.

Data Management Responsibilities

At Troy University, the IT department is responsible for all aspects of data management across physical locations, departments, and functions. We are participants in a group that reviews usage of data; additionally, we participate in all decisions based on data management, a role that has changed significantly in the past eight years.

Before that time, IT managed core services only. However, after observing significant deficiencies in

data reporting, mismanaged data repositories, and an unsustainable data sprawl, IT petitioned to collect the solutions and data into a single managed environment. Beyond the traditional meetings and collaborative efforts, Troy IT offers numerous training efforts, an annual IT Summit and quarterly data report surveys. In these activities we gather end-user comments and suggestions and work to develop enhancements to current offerings.

By participating in a wide variety of industry-sponsored training and conferences on data management, we are able to not only maintain current awareness of new developments, but we have a voice in product design. We encourage power users from different support units to attend the events and participate in post-conference discussion groups. With a unified voice as a large institution, we approach vendors, sponsors, and organizers with our issues. By avoiding a fragmented and unorganized approach, we safeguard against being lost in the shuffle.

Best Practices for Centralization of Data

Even before the advent of big data management, issues with managing data did exist, especially in the areas of security and quality. Threats to data occur from many vectors. Whether one is concerned about the potential loss of data through nefarious efforts, system failure, or simple data entry errors, the issue of data sanctity is always paramount.

Our best practices are leveraged on centralization of data. Through the use of homogenous

infrastructure and properly trained staff, one can mitigate threats. Contaminated data elements or rogue use of incomplete data reporting mechanisms present significant issues to big data.

Using big data tools, we are able to proactively monitor system performance and identify bottlenecks, breakdowns, and anomalies. System analysis and end-user activity provide insight into effectiveness of our solutions. By cross-utilizing data analysis tools to conduct resource reviews, one can effectively assess areas of inefficiency. Using such metrics, we can remedy the solution based on factual information, rather than scores of disconnected logs and the sometimes whimsical musings of overworked systems administrators. By removing intuition from the troubleshooting equation, salient resources can be appropriated.

Onboarding of New Employees

Many organizations fail to thoroughly address the routine of onboarding new employees. Groups can expend significant effort in identifying and recruiting solid employees but then fail to execute a rigorous orientation to the business. Often, based on stellar interviews, applications, and references, new employees are dropped into a job. The new employee often flounders and seeks assistance from others while wading through cumbersome orientation materials.

Within our university, IT is recognized as a strategic enabler for onboarding employees, nearly all of whom will engage in some service delivered by IT during the course of performing their jobs.

As such, we participate in the orientation of all new employees. In fact, IT's role in the orientation has changed from an occasional event to a key supporter and advocate for thorough training. The tools that present the orientation sessions are managed by IT.

Human resources officials manage a rich orientation to policies, expectations, and culture. Depending on the new employees' projected involvement with IT systems, they will be exposed to a variety of online learning tools and in-person sessions. These all include proficiency evaluations that seek to validate the person's comprehension of the materials and the effectiveness of the learning tool.

Big Data Challenges and Future Considerations

Many challenges face big data management. Beware. Much hype abounds about big data. Vendors have harvested a wealth of new tools, all promising to illuminate deep insight into hidden unknown recesses and produce value. But, the correct support mechanisms must exist, corporate leadership must appreciate the effort, and you must have usable data.

The major challenges that could derail any big data effort include: architecture, tools, governance, analytics, and cloud.

Architecture: What architecture will you deploy? As the vendors offer more products, a dizzying array of options will arise. Competing products will inevitably create confusion, especially in an area where so much promise exists. Open source will continue to play a pivotal role in big

data architecture. Look at what you have in place now — that is a great place to start. What platforms do you employ, where are your staff efficiencies? Use these answers to mold your architecture platform decisions.

Tools: Tools are maturing, so your review of tools should be instrumental and ongoing. Although custom-fit solutions are difficult to find now, as specialization grows, one will be able to find tailored solutions for each industry, avoiding the expense of resources associated with unique skills. And, given the success of key firms, the tools and approaches alike will only become more user-friendly.

Data Governance: Data governance is a constant concern. In the realm of big data, the concern crosses every strata of the organization. How will you handle data security, integrity, and regulatory compliance? The ongoing dance of data ownership will inevitably rise to the forefront. How will data managers be affected when scores of discrete data silos are gathered and reviewed as a whole? How does big data fit within my company culture? Who will manage it?

Analytics: How do analytics fit into your organization? Are you using descriptive and predictive analytics as a component of your business intelligence routine? Who directs your business intelligence effort? Is the process sophisticated or is it a collection of queries and reporting mechanisms that litter the landscape? As you become more sophisticated and gain a

deeper understanding of the relationship among your data stores, who will manage, who will review, and who will benefit from the discovery of the non-obvious?

Cloud: Will Big Data as a Service give rise to less expensive, rapidly deployed tools for smaller organizations? Will the cloud become an equalizer or will the cloud foster more concerns over data governance? I do not doubt that the cloud will present a vast array of tools to the big data enthusiast; however, as many organizations continue to struggle with basic cloud operations, will IT resources be drained in this effort as well — collecting, sharing, and analyzing larger stores of data at a distance? Can you handle such an exercise and is your organization primed for such activity?

Return on Investment of Data Management Initiatives

Typical for educational organizations, accurate assessments of our return on investment for data management initiatives have been elusive, so far. However, big data offers the potential to produce such reviews and metrics. By gathering all available data and analyzing the effectiveness of marketing programs, academic programs, and recruiting efforts, one could measure the overall effectiveness of spent resources. However, this is a slippery slope in academia. Our purpose goes beyond the structure of the corporate world — while we need financial resources to operate,

our mission is not rooted in financial objectives. How will big data tools help us evaluate our effect on the communities we serve, on research effectiveness, and on the knowledge proposition that we engage with students? These are compelling questions; perhaps big data will be able to assist us in addressing them.

Measuring the Effectiveness of Tools and Procedures

Measure, measure, and measure. How can one gauge the effectiveness of tools and procedures? IT change does not always need to be large and expensive. Identify a business unit to evaluate the usefulness of your big data proposition. Keep the evaluation group lean, but include the key decision makers and solid IT resources. Further, the group should endeavor to build good rapport; they need to function as a team and understand that they are studying how to use big data to remedy a problem that directly affects them. Despite the fact they are focused on the term, “big,” encourage the group to be agile. Can they use the results of big data analysis to implement a new process, enhance an existing process? If so, can the group’s results be reproduced, perhaps within a similar unit?

Through a structured, targeted approach, one can measure success and use such momentum as a catalyst for smaller, incremental changes as well as continuous, less-intense deployment of big data tools throughout the organization. ■

Perry Rotella

Senior Vice President and Chief Information Officer, Verisk Analytics

Data Management

Because we provide data and analytics solutions to property/casualty insurers, health care companies, and financial services, Verisk's business is all about data. Our data helps manage supply chain risk across multiple industries. We're trusted to manage significant amounts of private data, including automotive violations, auto claims, and policy information; health claim

building characteristics has provided us with tremendous amounts of detailed information about millions of commercial properties.

The size of the data sets we're analyzing is exploding. Therefore, we evolve our policies to respond to new technologies and implement data management solutions for infrastructure services such as columnar databases, text mining, and cloud computing offerings.

Data management requires rigor, a deep understanding of best practices, and the ability to determine whether new advancements will truly support business objectives — or whether they're merely “buzz.”

Perry Rotella

Senior Vice President and Chief Information Officer
Verisk Analytics

information; and mortgage information. Each vertical area has different regulations as well as different privacy needs.

Currently, we house more than 3,000 terabytes. We manage data contributed by customers, data that we purchase, publicly available data, and data that we “manufacture” ourselves. As an example of how we manufacture data, a Verisk team that surveys

In some of our markets, the data sets contain detailed transactional information from our clients, which essentially means the data we manage is on an order of magnitude larger than any one of our clients' data.

Warehousing Data

The current focus of our standard operating procedures revolves



Perry Rotella

Senior Vice President and
Chief Information Officer
Verisk Analytics

“In a field that's evolving quickly, technology business leaders need to stay informed and engaged with new solutions.”

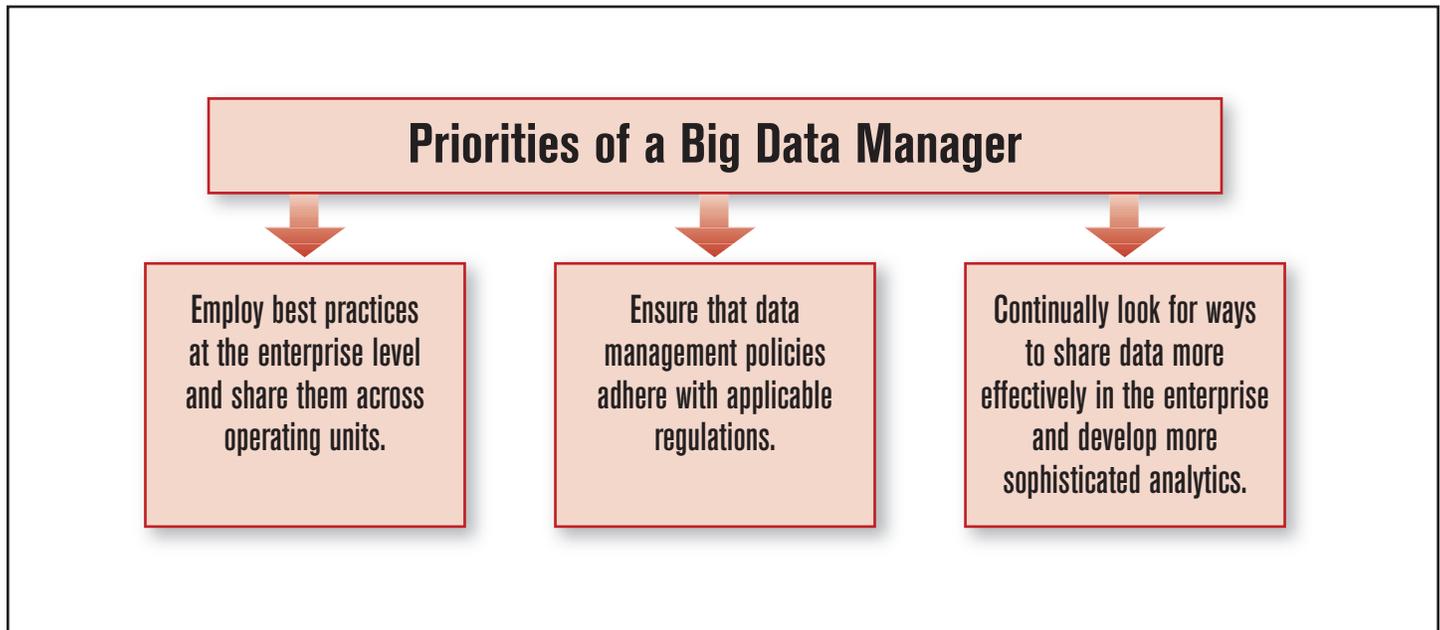
- Previously SVP and CIO, Moody's Corporation
- President, New York chapter, Society for Information Management (SIM)
- Lecturer on information technology issues, Executive Master of Science in Technology Management program, Columbia University
- B.A., Mathematics and Economics, University of Pennsylvania

Mr. Rotella can be e-mailed at
perry.rotella@execblueprints.com

around security and privacy. We warehouse data within each of our operating units. To ensure data is protected, we employ best practices at the enterprise level and strictly share them across operating units while continually looking for ways to better share data in our enterprise and develop more sophisticated analytics. Concurrently, we're tracking cloud computing trends very closely, as these solutions are evolving rapidly but still do not meet all of our security and privacy needs.

Expert Advice

Working with Vendors Verisk works with vendors in a variety of ways, and we consider certain technology tools to be the engine behind our analytic offerings. Sometimes we use a shared marketing approach rather than licensing products by data volume, which gives the vendor credit for the underlying technology and allows us to match vendor pricing to our own value proposition. Specifically, this type of arrangement informs our customer which technology is being used in case the customer wants to adopt that vendor's product in-house. For example, if we use a specific technology as part of a fraud scoring solution, our clients can adopt the same method in order to investigate fraud.



Data Regulations

As a data analytics company, we have to be mindful of the regulatory environment as we look to leverage our proprietary data across business lines. The regulatory environment on data is vast and changing. One of the most prominent regulations is the Health Insurance Portability and Accountability Act (HIPAA). Everyone in our organization who touches that data is HIPAA-certified, and we make sure our infrastructure teams and data centers have the appropriate certifications that our clients require. With different regulations around financial service data, such as the Fair Credit Reporting Act (FCRA), we need to ensure we physically separate data. In these

ways, we physically and operationally ascertain we're in line with required practices.

Leveraging Knowledge

As the data management and analytics space continues to change rapidly, keeping up with advancements and understanding best practices are critical. There are many opportunities for technology leaders to stay informed. I belong to the Society for Information Management (SIM), which is a membership organization for IT leaders. I participate in SIM's Advanced Practices Council, which comprises a peer group of CIOs. We commission research from leading university researchers at universities

such as Harvard, the University of Virginia, and MIT; recently, our agenda has been very much related to big data. At our next meeting, we'll be reviewing some of the latest research around managing data streams and real-time data flow. I hope we'll gain insight into best practices for warehousing data and real-time uses of data. We leverage the researchers, but then we also leverage each other's ongoing experience.

Data management requires rigor, a deep understanding of best practices, and the ability to determine whether new advancements will truly support business objectives — or whether they're merely “buzz.” ■

Sam Schoelen

Chief Information Officer, Continental Resources Inc.

Current Data Management Policies

Because every department is different, we do not have any strict policies for data management. Everybody has different needs, and they evolve over time. For example, marketing could have a policy regarding the amount of data that can be stored at any one time, but if they implement a new system that requires 10 times more data, I need to be able to adjust and look into the cost of managing this extra data. To accommodate these types of changes, we have a flexible way of managing our data policies.

Change and Documentation

Changes to our data storage procedures occur on a day-to-day basis. I would love to publish a policy and just have everything work within that but it is not realistic. For example, our engineering team might come up to us one day and tell us they need 100GB just to dump images in a week for a large client and we have to make it happen. This has happened to us in the past and we have to accommodate.

We have decent documentation for our policies and procedures, but because we change on a day-to-day basis, we do not document every single change. While we do have standard security and software use policies and specific documents

The top challenge of data management is getting people to understand the limitations of data and what they can do with it.

Sam Schoelen

Chief Information Officer
Continental Resources Inc.

for our service offerings, we do not have a specific documentation process for every change we make. Internally we do not need this type of change management system.

Data Management Decision Making

In regards to data management, we are a support function, which many people do not understand. They believe that we are the ones making decisions; however, sometimes we may only be the ones flipping the switch. In the end, it is the business that makes the decisions and IT provides support. IT decides how to most effectively put in place the company's decisions concerning data management.

When IT was first introduced and throughout much of its history, technology was not a business enabler. The role of IT has evolved slowly, but we are starting to see IT departments obtain more freedom



Sam Schoelen

Chief Information Officer
Continental Resources Inc.

"Our company is so dynamic that we could make a change every single day depending on what is going on and what the customers' needs are."

- With company over 7 years
- Currently runs corporate IT and Managed Services
- Company is provider of advanced technology solutions

Mr. Schoelen can be e-mailed at sam.schoelen@execblueprints.com

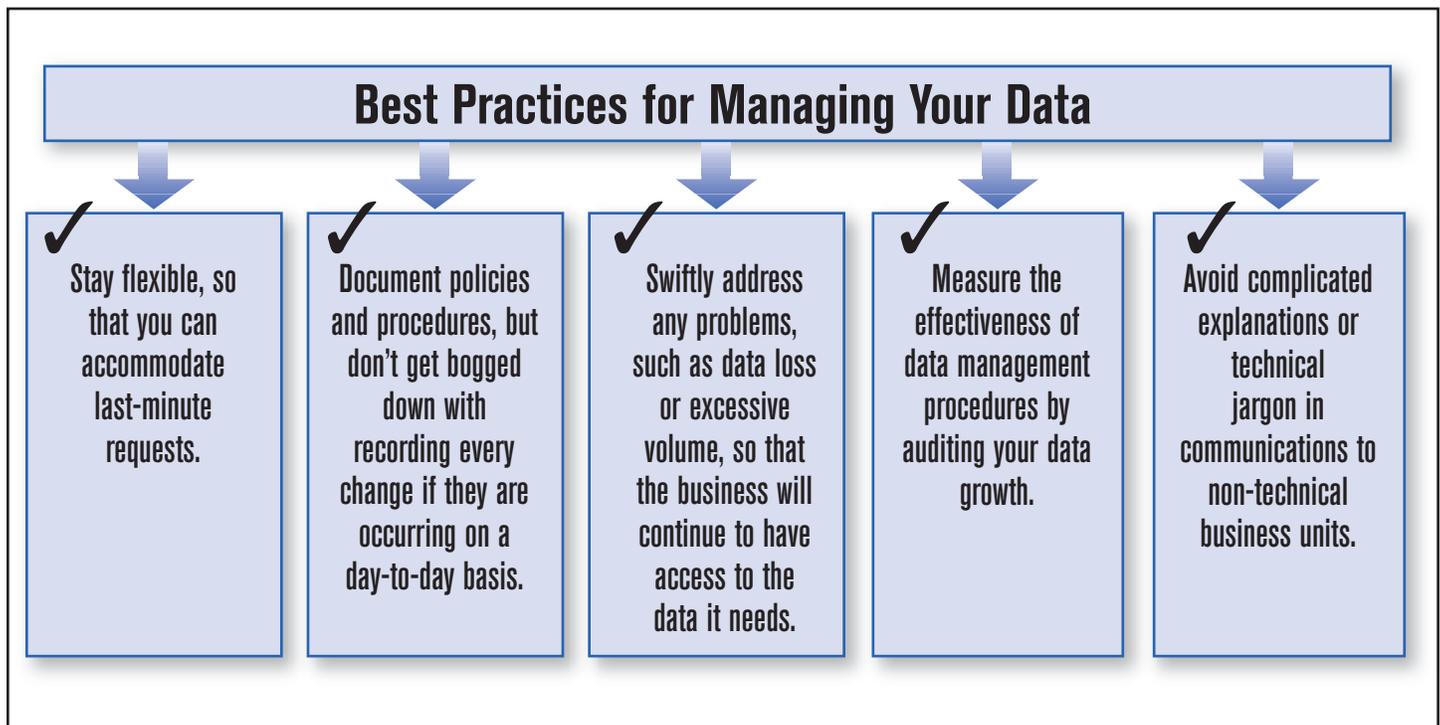
to decide the best approaches for implementing technology decisions. IT used to just keep the wheels turning when the business asked for a certain function but now we see that IT can come to the business with a function and help them out.

Developing Responsive Data Management Policies

The best practice for dealing with data management issues depends entirely upon the problem. However, if you are losing data, then obviously you need to get your backups, and if you are looking at problems with data being too large, you need to audit your data. Basically you need to always do

Expert Advice

Every once in a while you will find users who store all of their family photos on the print server and you need to inform them of the policies prohibiting such practices. We use the results of our audits to determine the effectiveness of our policies and procedures. From this simple audit it is easy to see who is conforming to policy and who isn't. I wouldn't say this shows exactly whom to flag but it does show you the first places to check for infractions.



whatever it takes to ensure business will always have the data it needs.

There is an easy metric for measuring the effectiveness of data management procedures — look at your data growth. When you do your audits, you can find out how well you educated your users.

Communicating Data Limitations

The top challenge of data management is getting people to understand the limitations of data and what they can do with it. Everybody

thinks the system should be unlimited and capable of doing anything they want. Most people want what they want and do not care about feasibility. One employee, for example, asked for an unlimited e-mail inbox. He did not understand that Microsoft places limits on the mailbox with the particular software we were using. Most people will react negatively when you tell them that they need to use less data or store the data differently because they want to be able to store anything without any issues whatsoever. Company employees

are generally not able to comprehend this unless the company invests a great deal of time explaining data limitations.

Moreover, when working with other business units, IT cannot have a policy that nobody understands, so simplification is important. You cannot give overly complicated explanations or communicate in confusing technical terminology. With a simpler approach, people are more likely to understand and follow the rules. ■

Dan Hoogterp

Senior Vice President and Chief Technology Officer, Bankrate Inc.

Data Collection and Management

Our non-financial operating and traffic data is managed in systems that aggregate traffic, behavior, and demographic information across our network. Our revenue tracking data is stored in specific financial systems and is generally

As we explore new tools and new ways to add value for both consumers and our advertisers, we are finding synergies and latent value among data that can now be feasibly extracted with current technologies.

Dan Hoogterp

Senior Vice President and Chief Technology Officer

Bankrate Inc.

not part of the larger traffic and interaction data. We also have systems for different sites or interactions that collect specific usage data from individual systems. We have data sets that range from millions to billions of events per year, with some data created at thousands of data points per second. We generally maintain separation between data that directly drives revenue processing and other data that is generated from systems around

our company. The revenue-related data has distinct requirements for data handling.

We have seen that both the core function, or primary data, and the related data associated with interactions have increased in the past few years. The amount of ancillary data that we collect has been

growing, as well as the number of systems and channels through which that data is collected or generated. Five years ago, we might have recorded a visit to a page with a few added points of data; today, we capture 15 events that occurred on that page, and we may be capturing it from two or three complementary systems. This has driven a substantial volume increase and opens the opportunity for substantial optimizations as this data is fully exploited.

Warehousing Data

Our systems have a strict path for revenue data, as well as systems into which we aggregate the ancillary data, such as usage data, visitor behavior, log data, and other information from the back-end systems associated with particular sites or services. Various views or excerpts of that data are aggregated for specific site and corporate purposes.



Dan Hoogterp

Senior Vice President and
Chief Technology Officer
Bankrate Inc.

"We are ensuring that 'big data' is a solution to real challenges and not a science project or fad that we chase with unclear benefits."

- Has served as SVP and CTO for Bankrate since 2005
- Previously CEO of TQuist LLC
- Focuses on linking business operations and drivers with innovative technology and solutions

Mr. Hoogterp can be e-mailed at dan.hoogterp@execblueprints.com

We have not yet undertaken the aggregation of user interaction data from around the company into a single system or data warehouse.

These practices have been effective in the past, as our individual properties have distinct audiences and offerings. However, as we explore new tools and new ways to add value for both consumers and our advertisers, we are finding synergies and latent value among data that can now be feasibly extracted with current technologies. There is a trend to aggregate more and more of our low-level data from our network, which spans personal finance in multiple

Expert Advice

When we incrementally improve engagement through optimization, making calculated improvements that provide our customers and advertisers added value, we improve our ability to operate our network and gain leverage through the data. Our core application of big data tools will support this direction. This is clearly the best use case for big data in improving our business value.

categories, business models, and over a dozen unique sites.

Over time we will see the back-end data converging to a greater extent and we will gain more leverage across the network as cumulative effects of this data can be leveraged, along with the efficiencies of more commonality in the application tools across the network.

Use of Big Data

We are on the lower end of the scale in terms of big data. We are lower volume than many high-traffic consumer sites and business operations; however, we are high volume in terms of our categories of focused financial sites. Essentially, general news, social, and consumer sites operate at much higher traffic levels but typically have lower per-visit economics.

While we have many of the challenges and opportunities of big data, we are on the smaller end of the scale because we are not processing trillions of page views or generating enormous scale data streams in comparison to the largest enterprises or web sites. Instead, we are serving billions over a longer period of time at a much higher monetization rate. This changes the mix and the value of the data, but the basic path and opportunity is the same. So, we benefit from the tools and techniques, but we won't need the scale of infrastructure to extract the value that firms with a larger data volume challenge will face.

Generally, we are approaching big data on a case-by-case basis at our firm. The value of aggregating and processing previously inaccessible volumes of data is applied

at specific points where the value case can be clearly made. Additionally, we are moving from physical systems to highly virtualized environments. The projects we are planning will be executed in an infrastructure that is much more scalable on a horizontal basis and has a much more scalable physical data layer underneath it. Our software and operating architectures are evolving from classic monolithic databases and architectures designed for a small number of system images to horizontally scalable many-image and distributed data/service approaches. As tools and services prove value in areas of the firm, we will find opportunities to standardize and leverage the success and knowledge more broadly across the firm.

It is important to have working sessions with distinct teams across the enterprise. We are working on projects with specific vision and drivers to accomplish product and site objectives in each area. We are not building a technology infrastructure and then plugging it into a business need. Instead, we are working from the business trends and trajectories that we need to enable with technology. To date, there is a synergistic relationship between big data and our web optimization, analytics, engagement, and satisfaction needs. However, we must be clear that the business need is driving the technology and not the other way around.

Following Trends

We all wish there was an easier way to stay on top of everything, particularly if the noise could be filtered, but we know our challenges and we

constantly look for the technologies that will address our specific situation.

We have architecture teams that look at various technologies, prototypes, and different approaches. A broad team continues to read publications, technical news media, and attends conferences. The goal is to focus on elements that connect to our business needs. In our business, we want to be some distance behind the bleeding edge, but still relatively early adopters when technology drives business outcomes. There is a substantial economic tradeoff in trying to use an early product or technology. We adopt things when they are emerging and useful, but with evidence of operational success.

Best Practices for Data Management

We have a data-hungry organization. As a result, there are good systems in place for capturing information, and we have extensive internal dashboards from the system management and other perspectives. We run hundreds of tests on our own data to validate internal systems that run constantly, so we have a pretty good alerting system and pulse on the operation. If a problem were to arise, it would set off dozens of alarms, and our employees would find it before it became a serious issue. Although our overall data management is still mainly traditional monolithic SQL of one form or another, the baseline of data management is very strong.

Effective Storage Practices

Generally, the data systems and the database administration in the

firm are responsible for the long-term storage of data. The end users are only granted permission for the data they need for their business role.

The actual lifecycle management of the data falls on the data management team. In many cases, we save data indefinitely, because there have been instances where data from 15 years ago has proven to be useful today.

In addition to shared file areas, we use various SharePoint portals that do a good job of keeping versions and making storage efficient for documents. We have moved many routine documents into SharePoint or other workflow systems. There are also common systems for presentations and sales materials around the company that help reduce redundancy in storage.

A combination of automated tape libraries and de-duplicating

disk storage appliances are used for long-term archiving of inactive data.

The Year Ahead

There are new infrastructures going into place, and with them, related procedures to control who can get access to our data and for what purpose. There will be scheduled routine inquiries run against that data as well.

We will also work both with internal groups and external firms to leverage more of our data across the firm. We will also be evaluating various tools and we are bringing in outside vendors to aggregate data to provide enhanced optimization or operating dashboards.

Challenges of Big Data

The key challenge is to make sure that you are tying your data to

the business. More isn't always better. You want to ensure that the first and second best use of the data is 100 percent covered in the roadmap that you establish. Otherwise, could potentially build a huge warehouse that is really not used for any ongoing value generation purpose.

Secure and reliable data management does not have a direct ROI, but it is an absolutely necessary cost of business in our operation. Understanding how our firm is running depends on having our data systems intact and secure. Many ongoing competitive and operational decisions require that we make the best use of our data feasible. These needs may not be revenue producing, per se, but the benefits are strong and the cost compared to those benefits and the size and scale of our company, the cost of that opportunity and security is actually quite low. ■

Ideas to Build Upon & Action Points

I. What Challenges Do “Big Data” Present?

While most institutions develop big data stores to solve problems relating to deficiencies in data reporting, mismanaged data repositories, and unsustainable data sprawl, these data sets can also present their own difficulties that need to be proactively addressed, including:

- Slower access to data
- Increased vulnerability to damage from attacks (because all data has been collected into one repository)
- Potential losses from catastrophic system failures or data-entry errors
- Incomplete data-reporting mechanisms
- Users’ assumptions that capacity is now infinite
- Users’ perception that data centralization reduces creative uses of data

II. The Bottom Line

It’s true that many tools exist to measure the effectiveness of your data management tools and procedures, but you may be harder pressed to calculate a return on investment, especially in non-profit or academic settings. Proven methods for assessing both areas include:

ROI:

- Analyzing the cost savings of efficiencies that big data stores have enabled, especially as they relate to marketing and sales processes
- Comparing sales revenues before and after the consolidation into a big data environment
- Determining the extent to which centralized data management has contributed to overall product and service quality

Effectiveness of tools and procedures:

- Asking user groups to assess big data’s capacity to implement a new process, or enhance existing processes
- Conducting audits to determine how your business units are using — and adding to — the big data store

III. Must-Have Practices for Managing Big Data

Generally, IT is responsible for all aspects of data management and, accordingly, should participate in decisions relating to this critical function — even while not losing sight of the fact that its primary objective for performing this responsibility is to serve the data needs of the overall business. In other words, even though IT manages the data, it does *not* own all of it. Accordingly, it must do whatever is necessary to ensure that the organization always has the data it needs. To accomplish this (especially in a big-data environment), IT leaders must ensure that:

- Using big data tools, system performance is monitored frequently to locate bottlenecks, breakdowns, and anomalies.
- All interactions within the system, including site statistics, event data, and transaction data, are logged.
- Physically and operationally, the big data store adheres to all applicable regulations.
- Common denominators are used to manage data stores across large data sets.
- A best practices guide is developed and reviewed regularly so that it remains current and accurate — and then shared across business units.
- An IT infrastructure library (ITIL) documents plans, effort, and changes associated with the big data store.
- Policies and procedures are continuously updated to respond to new technologies and regulations.
- Problems are addressed incrementally, through minor adjustments, to avoid massive, potentially disruptive, roll-outs.

IV. The Golden Rules for Working with Non-Technical Users

If people are going to use the big data store effectively to advance the objectives of the organization, they need both basic knowledge and input into the system’s development and maintenance. In other words, your processes and procedures cannot be a total “black box” to your users. Proven approaches for educating and engaging this audience include:

- Ensuring that new employees receive a position-appropriate IT orientation that includes a proficiency evaluation

- Communicating policies and procedures in simple language that avoids complicated explanations and technical jargon
- Conducting periodic surveys to gather end-user suggestions
- Holding live, in-person training events along with continuous online training opportunities
- Encouraging power users from different business units to attend IT conferences and participate in post-conference discussion groups
- Involving key decision makers and power users in your regular system reviews

V. Essential Take-Aways

Proper centralization of your organization’s data assets creates opportunities for rich exploration as well as diligent record keeping. Yet the field of big data management is still evolving, and requires IT leaders to continuously educate themselves on the advancements so that they will be able to ascertain which ones truly support the organization’s objectives, and which others just create “buzz.” In other words, while opportunities abound, there are still minefields to navigate.

Potential concerns include:

- How does big data fit within the organization’s culture?
- Who will manage it?
- Which architecture should you choose?
- How will you develop the right tools for your environment?
- How will you handle data security, integrity, and regulatory compliance?
- What role (if any) will cloud-based services play in your big data configuration?
- As analytics become more sophisticated, who will benefit from the discovery of the non-obvious?

Potential new features include:

- The ability to extract greater insights and value from the data
- The apprehension of appreciable patterns within large sets of data
- Fully-digitized libraries that are accessible from anywhere ■



10 KEY QUESTIONS AND DISCUSSION POINTS

- 1 What are your company's current data management policies? What types of data are being managed?
- 2 What standard operating procedures does IT presently follow when warehousing data? How often do you make changes to these procedures?
- 3 How do you define "big data" at your company? How does this definition compare to industry standards?
- 4 What are your best practices for working with vendors to address the unique challenges posed by big data? How is big data management different?
- 5 How do you stay abreast of big data trends in your industry?
- 6 How are company employees oriented to appropriate and effective data management and warehousing practices? What is IT's role in this orientation process?
- 7 In the next 12 months, do you plan to introduce new data management systems or procedures to help you address concerns with big data? How were they chosen? What do you hope they will accomplish?
- 8 What are the top five challenges IT executives face when it comes to managing big data? How can you address them?
- 9 In the next 12 months, what data management needs do you anticipate your company will have? Will these needs involve new roles or responsibilities for IT?
- 10 How do you benchmark big data management practices at your company? What is measured? How often?