



VENTANA RESEARCH



# Data Preparation: An Enterprise Imperative

A Modern Approach to Data and Analytics

White Paper

Sponsored by





# Table of Contents

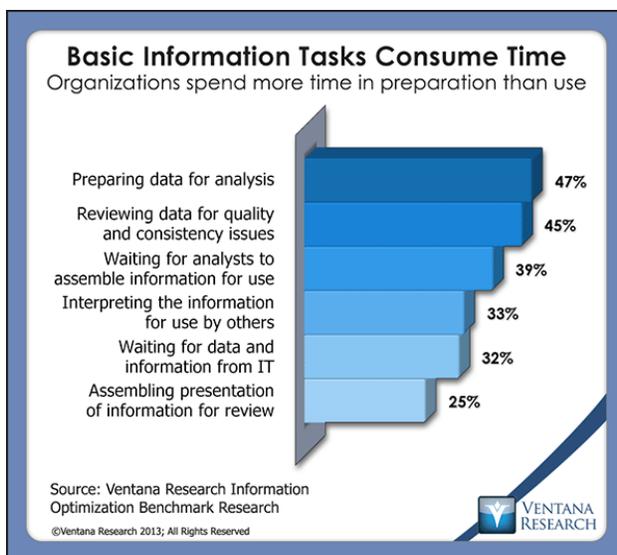
<b>An Essential Step in Using Data</b>	<b>3</b>
<b>Elements of Data Preparation</b>	<b>5</b>
Access and Explore	5
Data Analytics (Data about Data)	6
Manipulation and Quality	7
Enhancing Value	7
Use and Reuse of Data	8
<b>Technology Considerations</b>	<b>8</b>
<b>Aligning Business and IT</b>	<b>9</b>
<b>Benefits of Data Preparation</b>	<b>10</b>
<b>About Ventana Research</b>	<b>11</b>



## An Essential Step in Using Data

Data is fuel for today's businesses; they cannot run without it. Like other types of fuel, it must be processed and made ready to use. Many users, however, are unfamiliar with what is required to ensure that the data available to them meets their requirements for use. In our view, for the critical process of preparation to run smoothly, they need to understand why it is important and how it works.

In an organization, data is created and stored in many applications and systems. In its native form it often is not ready for use in business



activities, chief among which today is analytics, an increasingly critical aspect of today's information usage. Our research repeatedly shows that the manual approaches traditionally used to get data ready for use in analytics require far too much time – according to our benchmark research on information optimization, analysts spend nearly half of their time in preparing data for analysis or reviewing its quality and consistency. Such time-consuming methods prevent analysts from focusing on analysis and delay them from

providing guidance to decision-makers.

The data preparation process includes the steps of accessing, searching, aggregating, enriching, transforming and cleaning data from different sources to create one uniform data set. It's a process that should be done in a consistent manner but also needs to be flexible enough to be able to provide data as needed for a diverse array of business uses. Many organizations view being able to do this as a challenge: Almost two-thirds (65%) of the participants in our benchmark research on information optimization said it is very important to simplify making information available; organizations that have the largest number of data sources asserted that most often.

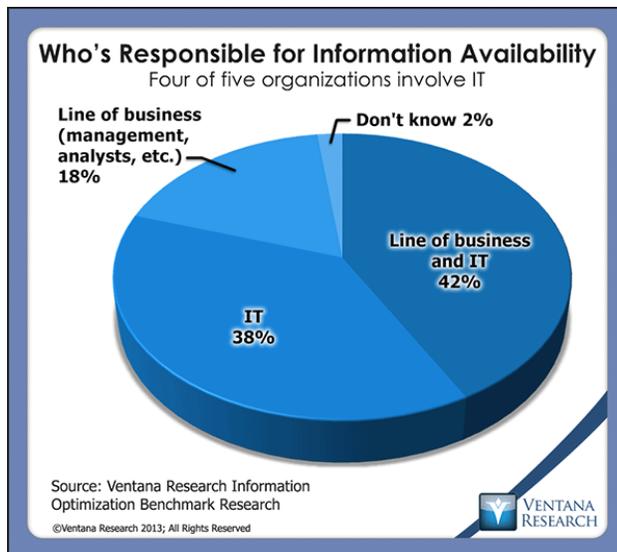
Unfortunately, most organizations lack the skills required to improve the availability of information. Only one-fourth (24%) said they are very confident that they have the people and resources needed to



improve availability of information within the organization. Even fewer have that level of confidence that they can provide it to partners and customers.

Every line of business brings its own needs to the company's data store. Customer service, for example, wants to optimize customer interactions and ensure satisfaction. Marketing needs to understand customer preferences and adjust strategies and campaigns to meet them. Sales needs to understand what affects its progress toward goals and achieving targeted outcomes. Operations needs to determine how efficiently departments are functioning and to identify processes that must be improved for them to be most effective. Yet

the data they and other departments all use must come from a single trusted source so that information is consistent across the enterprise.



In many organizations business users may require help from IT to get the data they need. But that doesn't mean that making data available is primarily IT's responsibility. Smart organizations engage both business and IT in this effort. And our research shows that more organizations (42%) make the

line of business and IT jointly responsible for improving information availability than leave it to IT alone (38%).

Systematic data preparation is essential to free analysts to spend more time actually working to derive business insights through analytics. As noted, our research finds that preparing data for analysis (47%) and reviewing data for quality and consistency (45%) are the tasks in which most analysts spend the largest portion of their time. It also is important to enable IT professionals to easily review data so they can address a range of ad hoc information-related needs.

Our research finds much room for improvement. Only 10 percent of organizations are satisfied with their existing approach to the design and deployment of information; four times as many (43%) are making improvements. In efforts to optimize information, IT is the process



area that the most research participants (42%) said needs improvement to provide information more readily.

Improvement will benefit an organization's various uses of data. Preparing analytics requires data in the right form for calculation. Big data involves assembling large volumes of data from many disparate sources, so the data must be in forms capable of combination.

Business intelligence and reporting require data adaptable into forms that are easy to understand. Transactional data in business applications, ranked first or second in importance by most research participants, must be consistent regardless of what application it comes from.

Automated data integration does not address the need to quickly combine data from various sources to meet business needs that arise during the course of a day.

Automating data integration can ease the burden of data preparation as well as streamlining the execution of well-designed processes in IT. But automation alone does not go far enough. In particular, automated data integration does not address the need that can arise during the course of a day to quickly combine data from various sources to meet business needs.

## Elements of Data Preparation

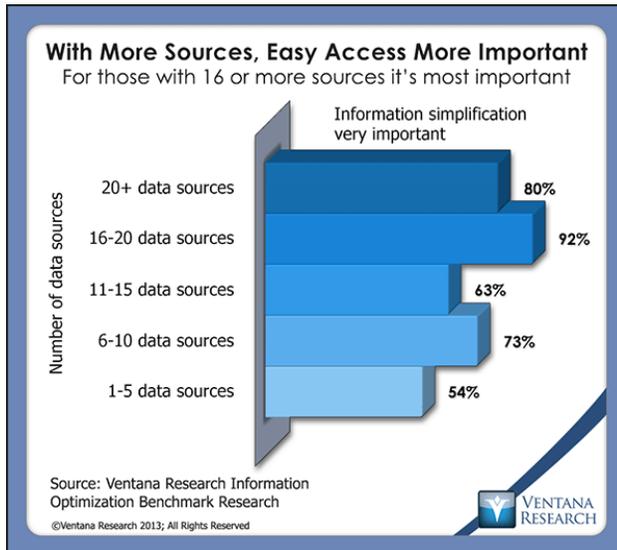
Our research confirms repeatedly that it is essential for effective performance to offer efficient access to information across the enterprise. For example, our information optimization research showed that extracting information from sources across the enterprise is very important in two of five (39%) organizations. Data preparation is an essential part of enabling this access. Ventana Research has constructed a framework for data preparation which consists of the following five elements.

### Access and Explore

The ability to import and export data from all applications and systems is essential to enabling access to all the information needed for business. Almost two-thirds (61%) of organizations have to integrate



five or more sources to create complete information for use, our research finds; 18 percent have to integrate more than 20.



From their many sources organizations must create one complete data set. Joining disparate data sources during transformation and processing of data is a critical capability for more than half (56%) of organizations. Moreover, effective data acquisition must be dynamic; additional data may be introduced at any step of the analytic process – preparation, discovery and analysis – so businesses must be able to combine data

dynamically to support their as-needed use in analytics. They also must be able to explore the sources to determine which contain data relevant to particular tasks.

Using big data technology is growing increasingly popular as organizations struggle to make sense of today's huge and rapidly increasing amounts of data and derive insights from it. New technologies such as data warehouse appliances, in-memory databases, specialized databases and Hadoop are being used by between one-fifth and almost half of organizations. Big data is increasingly central to these technologies. Exploiting big data effectively requires first being able to access and explore it, then being able to prepare it for use in analytics.

In addition, cloud computing applications are now generating data organizations need. Two in five (41%) research participants said it is very important to integrate data in cloud computing applications with other enterprise data. However, cloud data sources are not always under the organization's control; for example, more than one-third of research participants use Internet information, economic data and social media sources. The need to be able to blend data from cloud-based sources with that from enterprise systems intensifies pressure to be able to assemble useful data sets.

### Data Analytics (Data about Data)

Three particular techniques are relevant in determining whether data being prepared is relevant for the given purpose.



- Profiling analyzes parts of data sets (such as rows or columns) to find issues in the data and determine where further manipulation may be needed.
- Discovery finds new patterns in data that can help guide proper data preparation or determine issues that need to be addressed in operational systems.
- Exploration provides methods to navigate through data to better understand it and determine which is useful.

### **Manipulation and Quality**

Ensuring the quality of data is critical to be able to create reliable information that users trust. Yet only one in nine (11%) organizations in our benchmark research on big data integration said they are very confident in the quality of information being generated – the same percentage that said they are not confident. Several techniques facilitate assurance of data quality.

- Interactive manipulation of data (sorting, filtering and ranking) is part of establishing data quality.
- Segmentation splits groups and data for business uses or testing.
- Editing makes bulk changes to field labels or descriptions, data formats or data itself to help align data to proper formats.
- Reconciling compares and reviews data side by side and makes changes to ensure consistency.
- Masking obscures sensitive data to comply with privacy rules and regulations. This is the data task least well addressed by today's data integration approaches; it is completely adequate in only 9 percent of organizations.

### **Enhancing Value**

Gaining full value from data is critical for businesses, and effective methods to do so should be readily available for responsible individuals. Two capabilities are key to effective data preparation.

Enriching data supplements data with other fields or attributes to provide more robust views by enabling ad hoc and dynamic interaction with sets of data. It also provides the ability to apply governance in accordance with standard conventions.

Matching uses semantics to link common fields across critical categories of data such as product, customer, employee and others often used.



## Use and Reuse of Data

For data to be useful it must be available to all qualified users, and they must have confidence in its accuracy.

Governance reviews data to ensure that it meets prescribed quality standards and is suitable for use by any authorized department or individual.

The ability to share data across an organization is increasingly important. Publishing sets of data allows them to be used by technologies other than the ones in which they were created, from analytics and business intelligence to operational systems.

New approaches to collaboration enable users to discuss where information can be used or reused for business. More than three in five organizations (62%) use business and social collaboration technology with business information.

## Technology Considerations

Effective data preparation requires the use of technology that supports a range of administrative and information management capabilities. The ability to define, model and lay out information for use in applications is important or very important to four out of five (79%) organizations.

Effective data preparation requires the use of technology that supports a range of administrative and information management capabilities.



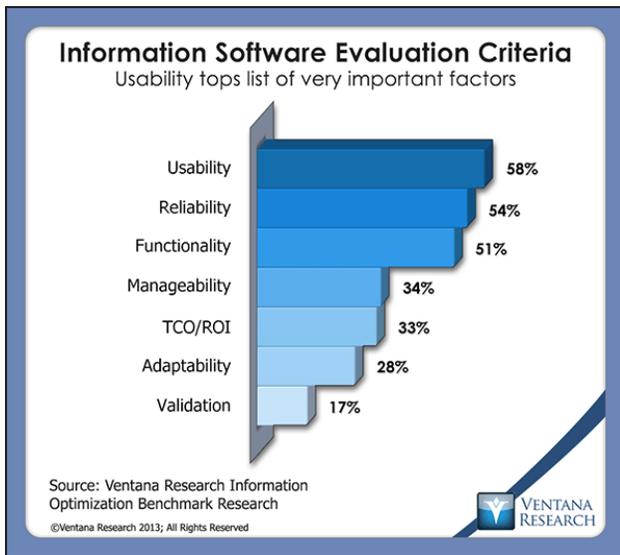
have enough capabilities to integrate and normalize data from disparate sources.

We find, however, that in many cases technologies currently in use are not able to meet the range of needs in the lines of business and IT. Only one in four (26%) organizations are satisfied with the current technology used to provide information. And almost half (45%) said they do not

Spreadsheets are widely used but ineffective for data preparation. Our benchmark research on spreadsheets in the enterprise finds that seven out of 10 people in participating organizations download information



into a spreadsheet all the time or frequently. Yet almost one-third (31%) of organizations find conflicting versions of spreadsheets, and fewer than half said spreadsheets are accurate and timely. In addition, half acknowledged that there are better alternatives to spreadsheets for important processes.



Having technology that is usable by various roles within business as well as IT is critical to ensure effective and accessible data preparation. The expanded use of business intelligence and analytics beyond technical specialists has created an emphasis on tools that business people feel comfortable with, as our research projects consistently show. Regarding information optimization, for example, usability is the most important software evaluation criterion, cited by more than half (58%) of organizations.

## Aligning Business and IT

Business and IT should understand and agree on each other's roles in preparing data for use. However, several factors may work against this cooperation. Entrenchment of budgets and priorities (for 42%) and different charters or directions set by the organizations (37%) are the most common issues between business and IT in integration of big data. Resistance to allowing business users to integrate big data that is not prepared by IT is found in 23 percent of organizations. Management should take steps to get business and IT talking about their differences and the streamlining of data preparation by sharing common software.

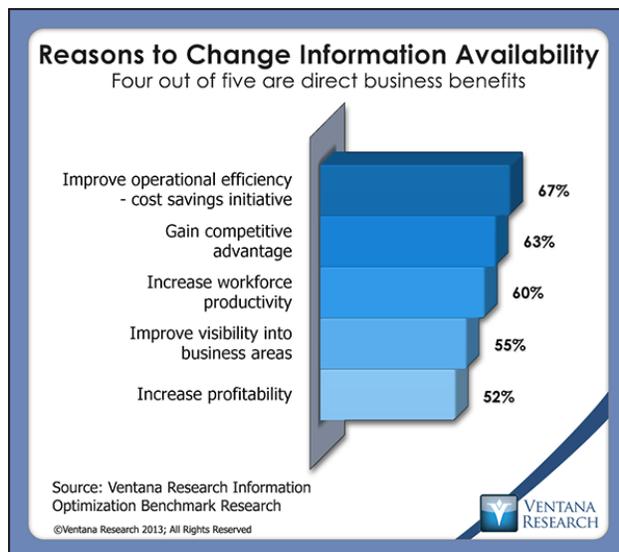
Adopting a collaborative approach can save time and resources, and we find evidence of this occurring. In making new information available, the largest percentage (52%) of organizations said that business analysts will work with IT to design and deploy systems.



In both business units and IT support it is important to focus on skills as well as tools. To work with big data, data skills were the category of skills most often named as needed (by 60 percent of organizations). In our information optimization research almost two out of three (63%) participants said SQL is a critical technology standard for making information available. We conclude that a flexible yet consistent approach, like that found in data preparation software, will prove optimal in making data ready for any use.

## Benefits of Data Preparation

As is often the case with technology deployment, those proposing new investments in data preparation must identify tangible benefits that can be realized in a defined time period and at reasonable cost. Being



able to show a combination of business and operational benefits is likely to serve well. Our research shows that improving operational efficiency is the most often cited reason (in two out of three organizations) for changing how information is provided, but gaining a competitive advantage is a close second, cited by 63 percent.

Data preparation software also can reduce backlogs of requests made to IT and the time it previously took to address them. Any system selected thus should ensure timeliness in providing adequately prepared data.

An appropriate system also should be able to handle a variety of evolving data needs of the business. Technology that is not adaptable or flexible to change is one of the top reasons (cited by 56%) that organizations are not satisfied with their current data integration technology.

Effective data preparation can save time and increase efficiency across the enterprise. It enables those whose responsibilities include data management to make it ready for particular uses. It reduces reliance on manual and custom coding and scripting processes and frees



technical resources to work on other high-value tasks. It provides IT staff and business users a single common work environment in which to provide data, track its use and establish safe, reliable processes.

## About Ventana Research

Ventana Research is the most authoritative and respected benchmark business technology research and advisory services firm. We provide insight and expert guidance on mainstream and disruptive technologies through a unique set of research-based offerings including benchmark research and technology evaluation assessments, education workshops and our research and advisory services, Ventana On-Demand. Our unparalleled understanding of the role of technology in optimizing business processes and performance and our best practices guidance are rooted in our rigorous research-based benchmarking of people, processes, information and technology across business and IT functions in every industry. This benchmark research plus our market coverage and in-depth knowledge of hundreds of technology providers means we can deliver education and expertise to our clients to increase the value they derive from technology investments while reducing time, cost and risk.

Ventana Research provides the most comprehensive analyst and research coverage in the industry; business and IT professionals worldwide are members of our community and benefit from Ventana Research's insights, as do highly regarded media and association partners around the globe. Our views and analyses are distributed daily through blogs and social media channels including [Twitter](#), [Facebook](#), [LinkedIn](#) and [Google+](#).

To learn how Ventana Research advances the maturity of organizations' use of information and technology through benchmark research, education and advisory services, visit [www.ventanaresearch.com](http://www.ventanaresearch.com).