The state of data quality

An Experian Data Quality white paper

Experian®
Data Quality

The state of data quality

Contents

Executive Summary

The perception of data across organizations is changing. No longer is data just viewed as a secondary component of business.

Today information contained within a database is viewed by senior management and many departments as a critical factor in decision making, customer interaction and service delivery. In fact, 93 percent of companies believe data is essential to their marketing success. However, data quality inaccuracy leaves organizations at risk.

Unfortunately, the level of global inaccurate contact data has increased from 17 percent to 22 percent, up five percent in just 12 months.

With the increasing volume of information collected through a variety of channels, there is more room for human error. This combined with the prevalence of segmented, departmental approaches to data accuracy is preventing stakeholders from analyzing, improving and controlling data problems.

This year's study revealed that 66 percent of companies lack a coherent, centralized approach to data quality. But, ad-hoc approaches are dividing resources and further segmenting information.

Poor data quality is having a negative effect on budgets, marketing efforts and most importantly, customer satisfaction. Organizations that are not able to control the quality of their data are unable to effectively communicate with their customer base.

Data quality is the foundation for any data-driven effort and in order to succeed in the year ahead, organizations will need to look at prioritizing data accuracy and accessibility.

Thomas Schutz, SVP, General Manager of North American Operations
Experian Data Quality
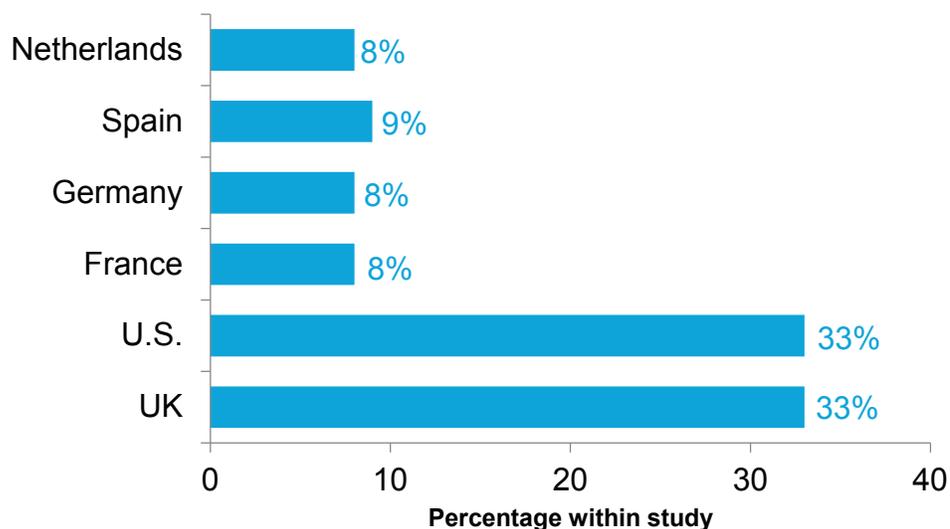
The state of data quality

Introduction

Research overview

In December 2013, Experian Data Quality commissioned a research study to look at current approaches to data quality. This report, 'The state of data quality,' reviews the evolution of data quality and consumer interaction while providing best practices for data management.

Research methodology

Over 1,200 respondents globally took part in the research, produced by Dynamic Markets for Experian Data Quality. Individuals from the U.S., UK, France, Germany, Spain and the Netherlands completed the survey. Industry sectors included in the sample were finance, public sector, retail, manufacturing, utilities and education. Respondents consisted of C-level executives, vice presidents, directors, managers and administrative staff connected to data management, across a variety of functions.

Countries in sample

Country	Percentage within study
Netherlands	8%
Spain	9%
Germany	8%
France	8%
U.S.	33%
UK	33%

Percentage within study

Key findings

The state of data quality

An abundance of channels

Organizations are interacting with consumers in countless ways. On average, companies use 3.4 channels to collect customer or prospect contact data. Multinational companies operate through more channels than those who operate in a single country. The most common channel for interacting with customers is the organization's website, followed by a sales team and the call center.

The number of channels has remained consistent year-over-year. And while websites are dependably the most popular channel over the past few years, mobile is gaining in prominence. Today, half of organizations are capturing customer contact data through mobile applications.

Organizations are not just collecting information through select channels, they are also sending marketing messages to consumers to create brand awareness and drive purchases. These marketing communications are sent through a number of channels, the most popular of which is email, followed by social media and then mobile telephone. Social media prevalence is increasing year over year, up five percent from 2012.

With email being the most popular marketing communication channel, it is not surprising that 83 percent of companies acquire customer or prospect email addresses for their email marketing efforts. These addresses are collected through an average of three channels, the most popular being the company's website and the call center. U.S. companies actually collect customer and prospect email addresses in a wider variety of ways, compared to other countries.

While this general diversification of channels is not new, companies are starting to increase their focus on cross-channel marketing. Cross-channel marketing is the coordination of different channels to provide the customer with a consistent experience rather than a more segmented multi-channel approach. While 87 percent of companies now engage in cross-channel marketing, 83 percent of them face challenges in this area of their operation.
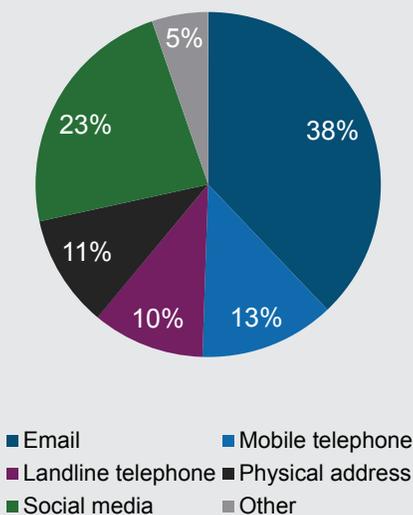
The principal challenges associated with cross-channel marketing relate to data. Having accurate information on the consumer and having enough information on the consumer are the two primary challenges when providing consistent communication across channels. Respondents in the U.S., Germany and Spain actually face a wider variety of challenges when they engage in cross-channel marketing, compared to respondents in the UK, France and the Netherlands. Multinational companies also relate to a greater number of cross-channel marketing challenges, compared to companies that only have national offices.

These cross-channel marketing challenges directly correlate the current state of data quality across the globe.

Channels to collect consumer data

Legend: Physical store, Face-to-face, Website, Mobile website, Mobile application, Catalogue, Call center, Other

Values: 42%, 60%, 73%, 36%, 35%, 24%, 54%, 14%

Marketing communication channels

Legend: Email, Mobile telephone, Landline telephone, Physical address, Social media, Other

Values: 38%, 13%, 10%, 11%, 23%, 5%

The state of data quality

Data quality continues to be a challenge for many organizations as they look to improve efficiency and customer interaction through data insight. 91 percent of companies suffer from common data errors. The most common data errors are incomplete or missing data, outdated information and inaccurate data.

Because of the prevalence of these common errors, the vast majority of companies suspect their contact data might be inaccurate in some way. Globally, the average amount of inaccurate data has risen to 22 percent from 17 percent just 12 months ago. U.S. organizations actually believe they have the highest percentage of inaccurate data at 25 percent.

The level of inaccurate data is staggering when one considers how much businesses are relying on information for business intelligence and improved customer interaction.

The main cause of inaccurate data remains to be human error, which has consistently been the main cause of errors over the past three years. While all other causes clearly lagged behind the front runner, they include a lack of communication between departments and technical limitations.
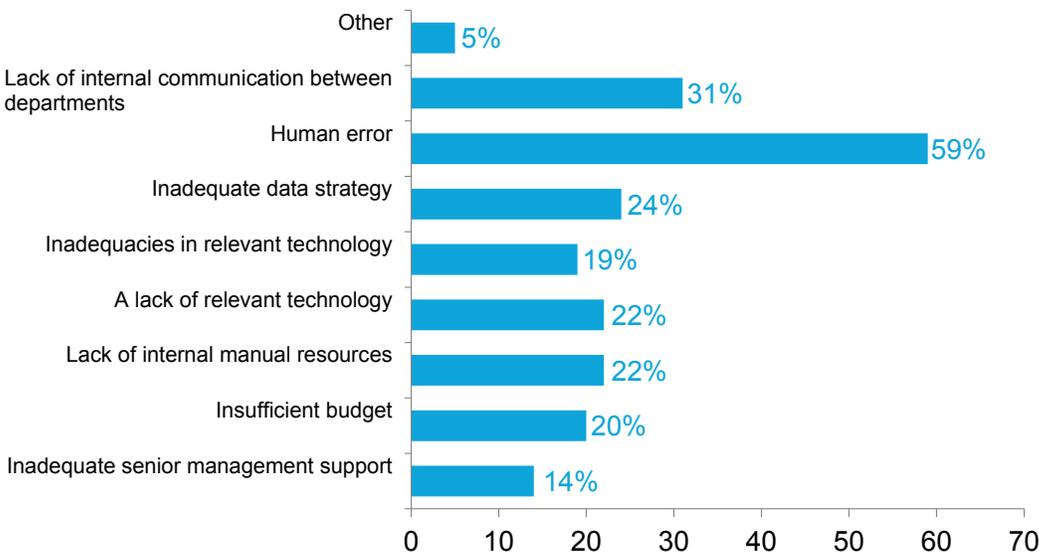
Information collected across various channels is frequently exposed to human error as consumers and individual employees enter information manually. Collectively, 78 percent of companies have problems with the quality of data they collect from various channels. Globally, call centers produce the poorest data quality, followed by websites. However, in the U.S. that order was reversed, with company websites causing the most challenges.

The level of inaccurate data relates to a lack of a sophisticated data management strategy, which many organizations are struggling to centralize.

On average, U.S. companies believe
25 percent of their data
is inaccurate.

Common data errors plague
91 percent
of organizations.

Reason for data inaccuracy

Reason	Percent
Other	5%
Lack of internal communication between departments	31%
Human error	59%
Inadequate data strategy	24%
Inadequacies in relevant technology	19%
A lack of relevant technology	22%
Lack of internal manual resources	22%
Insufficient budget	20%
Inadequate senior management support	14%

The state of data quality

Strategy development

Organizations conceptually see the benefit of having accurate data. The main drivers for having a data quality strategy include increased efficiency, enhancement of customer satisfaction, and more informed decision making. Respondents in the U.S. say more of these factors account for why their organization has a strategy to maintain data quality. On the industry side, respondents in manufacturing, financial services and utilities relate to more of these issues as drivers towards having a strategy to maintain high-quality contact data, compared to those in education and the public sector.

Interestingly, the benefit of cost savings continues to drop as a motivator for a data quality strategy, which reflects the current desire for data to inform overall business strategy rather than just to serve an operational purpose.

There are several key areas that make up a given strategy. They include:

1. Management of data quality

2. Utilization of third parties

3. Adoption of strategies

4. Selection of implementation method

Today, only 30 percent of companies manage their data quality strategy centrally, through a single director. That means 66 percent of companies lack a coherent, centralized approach to data quality. Given the number of channels and departments that interact with data, it is challenging to possess quality information when each department has different standards and methods for data management.

Third parties are frequently used for data management strategy suggestions. 64 percent of companies have used or still use third parties for their data quality strategy. Utilizing third parties for data management is most common in retail and manufacturing, compared to other industry sectors. When reviewing variances across company size, among these generally large companies, more of the smaller ones use third parties for data management.

However, given the decentralization of data management strategies in general, it is most likely that these third party engagements are for one-off campaigns or departmental management practices.

The main drivers for having a data quality strategy include increase efficiency, enhancement of customer satisfaction, and more informed decision making.

In general, senior managers selected a greater volume of drivers when it comes to maintaining high-quality contact records.

Data management strategies are often segmented between departments.

30 percent
manage data quality centrally.

66 percent
lack a coherent, centralized approach to data quality.

Third parties are frequently used for data management strategies.

64 percent
of companies have used or still use third parties for their data quality strategy.

Some companies take advantage of automated software techniques. One in three companies use dedicated point-of-capture software to verify information as it is entered. In addition, another one in three companies use dedicated back-office software to clean data after it is submitted. Automated processes are an indicator of the sophistication of data management methods. Companies who use automated data management methods are more likely to have their data strategy managed centrally, by a single director.

However, many companies still rely on manual data cleansing methods. 53 percent of companies perform manual data cleansing tasks. These include processes like manually reviewing data in excel or one-off manual corrections for seasonal campaigns. While overall the prevalence of manual methods is down from previous years, organizations do need to look at utilizing more automated methods to prevent human error.

Finally, there are different deployment methods for data management strategies. Some organizations choose to deploy on premise software to manage data quality. However, software-as-a-service (SaaS) deployments continue to gain in popularity. Ignorance about SaaS solutions for data management has fallen quiet considerably during a 12 month period from 15 percent last year to nine percent today.

Today, over half of companies are using SaaS to manage data quality. Only nine percent of organizations have no plans to implement a SaaS solution for data quality. The U.S. and France stand out in that more companies are already using SaaS technology to manage data quality. Manufacturing and retail are leading adopters with roughly one in five managing all of their contact data through SaaS technology.

However, knowledge about SaaS solutions seems to vary by department and seniority level. In general, the more senior staff and those in IT and data management roles are better informed of their organization's position with respect to the use of the cloud.

The lack of a centralized strategy and consistent automated methods fuel a large percentage of data inaccuracies. This leads to a number of consequences.

Data management methods vary greatly by company

55 percent use
automated methods of data management

and manual methods are used by

53 percent
of organizations.

Ignorance in SaaS solutions for data management has fallen quite considerably during a 12 month period

53 percent
of organizations are using SaaS to manage data quality.

The state of data quality

Consequences of poor information

With a quarter of information believed to be inaccurate in U.S. organizations, companies are facing many consequences.

First, inaccurate data is affecting the company bottom line. 77 percent of companies believe their bottom line is affected by inaccurate and incomplete contact data and on average, respondents believe 12 percent of revenue is wasted. Despite increased knowledge about data quality and the benefits of utilizing data-driven techniques, the average percentage of wasted revenue has not changed in this survey since 2007.

However, changes in business practices have brought about new consequences. Some relate to customer engagement and loyalty programs that have made a strong surge in the past few years. 84 percent of companies have a loyalty or customer engagement program. Unfortunately, 74 percent of respondents have encountered problems with these programs. The main causes are inaccurate information, not enough information on the consumer, and an inability to analyze customer information. All of these issues relate to data accuracy and accessibility.

Another trend is business intelligence and analytics, frequently referred to today as big data. 89 percent of companies now use their data in a strategic way for business intelligence and analytics. In fact, the U.S. stands out with more companies conducting business intelligence and analytics on their data, compared to the UK and France.

These programs also encounter problems due to poor data quality. 81 percent of organizations encounter problems when trying to generate meaningful business intelligence, mainly due to data inaccuracies. Other problems include a lack of information, a lack of flexible data and systems, followed by an inability to consolidate data across channels.

Finally, marketers are continuing to communicate through email, however, 67 percent have experienced email deliverability problems in the last 12 months. These problems result in poor customer service, an inability to communicate with subscribers and unnecessary costs.

- ■ Having inaccurate information on the consumer
- ■ Having enough information about the consumer
- ■ Inability to analyze customer information
- ■ Inability to access disparate customer information
- ■ Inability to create compelling offers
- ■ Lack of customer participation
- ■ Other

Problems with loyalty campaigns

- 5%
- 22%
- 15%
- 16%
- 13%
- 11%
- 18%
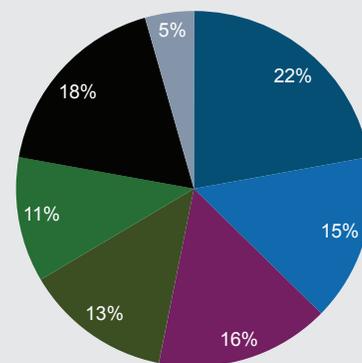
- ■ Cannot consolidate data across channels
- ■ Inaccurate data
- ■ Not enough information available
- ■ Too much information available
- ■ No analytics resources
- ■ Lack of training
- ■ Lack of flexible data
- ■ Other

Problems with business intelligence

- 3%
- 12%
- 22%
- 16%
- 9%
- 10%
- 13%
- 15%
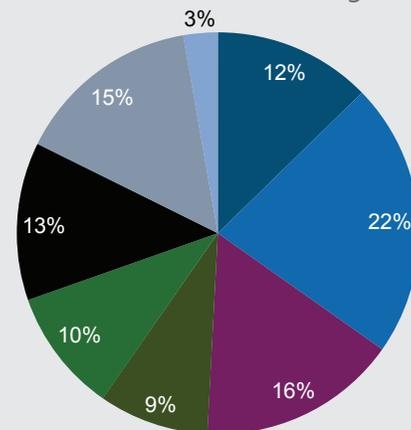
Value of data sets

As organizations look to gain value from there data, certain data sets have emerged as being more important that others, particularly for marketing. In fact, 93 percent of respondents think some form of data is essential to their marketing success. Contact data tops the list of data and information deemed to be essential to marketing success, followed by sales data and demographic data.
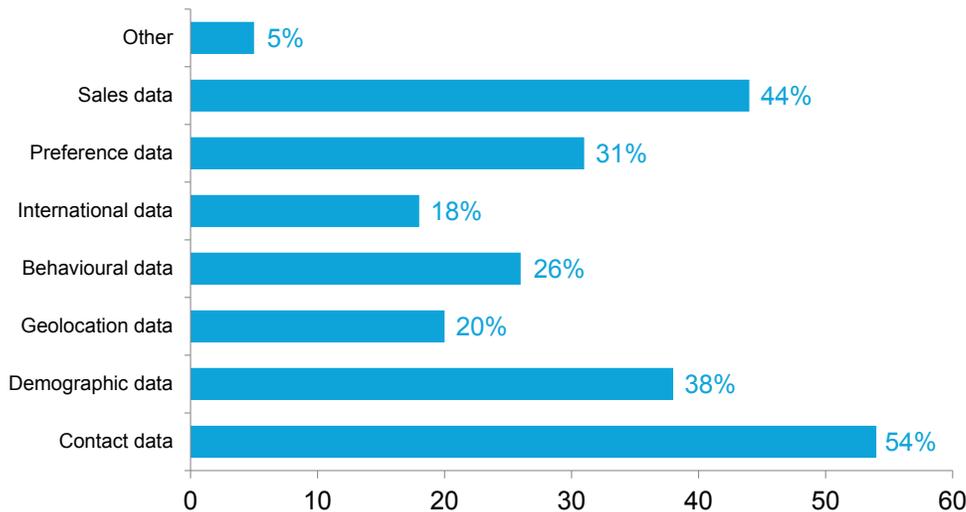
To gain additional insight that may not be contained within an existing database, many organizations are looking to third parties to append insight for marketing and business intelligence purposes. 94 percent of companies append enrichment data to their contact information.

On average, companies append three different types. The top three data sets are business data, geolocation data and demographic data. U.S. companies seem more switched on to data enhancement and append a wider variety of enrichment data. One data set in particular is preference data, which is especially popular as a route to marketing success for the U.S., compared to other countries.

Enrichment data is used by

94 percent

of organizations.

On average, companies append

three different

data sets to their customer contact information.

Information essential to marketing success

Category	Percentage
Other	5%
Sales data	44%
Preference data	31%
International data	18%
Behavioural data	26%
Geolocation data	20%
Demographic data	38%
Contact data	54%

The state of data quality

A five year evolution – what has changed?

Looking back over the past five years of the Experian Data Quality study, what has changed about our perception, use and motivations for data quality?

First, the perception of inaccurate data is increasing. Globally, the average amount of inaccurate data has risen to 22 percent from 17 percent just 12 months ago. This increase is due to the increasing volumes of data from multiple sources. Organizations generally acknowledge that the proliferation of digital channels and mobile technology has brought more information than ever before. While some of this is unstructured data that is difficult to mine, there is a general increase in overall customer and prospect information.

In addition to increased volumes, data quality has moved from a primarily operational function that was associated with efficiency and cost savings, to a strategic function aligned with consumer insight and overarching business intelligence.

The increase in data has also led to the concept of big data, which came about within the past few years and is generally a buzz word most of us have read about. While many of us have read about the term, there still is no single, consistent definition that has emerged in the market. Respondents in this year's survey were asked to review several definitions for big data, and no single answer emerged as the front runner.

45 percent of respondents believe the term refers to a large, unified single source database. Although this took the highest percentage, multiple data sources and predictive analytics definitions were not far behind. More senior level respondents selected more of these possible interpretations of what the term big data means to them, whereas 40 percent of administrative level staff admitted they did not know what the term meant.

Finally, we have seen the explosion of SaaS. Five years ago, few organizations implemented cloud or hosted solutions. Today with the explosion of software-as-a-service platforms, many organizations are looking to utilize this deployment method for their data quality management software to decrease implementation time and ensure consistently updated technology. However, as we have seen in previous years, security still remains a concern around SaaS that large organizations need to research further.

Data quality has moved from a primarily operational function that was assoicated with efficiency and cost savings, to a strategic function aligned with consumer insight and overarching business intelligence.

The average amount of inaccurate data has risen to

22 percent
for global organizations from

17 percent
only just 12 months ago.

Creating a centralized
data strategy

The state of data quality

Creating a centralized data strategy

Given the importance of data-driven efforts, the level of inaccurate data needs to decrease for organizations to gain actionable insight. Improved data quality leads to better consumer interactions and informed business decisions.

Based on the research findings, organizations need to look at reviewing a centralized approach to data management. One-off, ad-hoc projects of the past no longer suffice given the volume of data and the speed at which it needs to be accessed.

Given the number of channels through which data is being entered, organizations need to look at creating a centralized strategy for data management. A central approach ensures consistency across departments, access to many data sources for customer information and improved best practices related to data management.

In order to create a centralized approach, organizations need to focus on three key areas.

Create a central task force	Data management is not the responsibility of one department alone.

Consolidate data	In order to create a centralized approach, organizations need to consolidate different sources of information.

Implement data best practices	Strong tools and processes will prevent the most common errors within a database.

These areas allow organizations to create a centralized source for data with consistent management practices and easy access to valuable consumer information.

Create a central task force

Data management is not the responsibility of one department alone. Many departments input and utilize data for daily operations. Information is entered through websites, call centers, sales representatives, and many more. Then departments such as billing, customer service, fulfillment and marketing utilize information to communicate with customers or supply them with basic goods and services.

In order to manage data quality as it relates to all of these areas, a centralized task force needs to be created. This group should consist of stakeholders and individuals to execute plans from across the organization.

Stakeholders should consist of members from various departments that have a stake in the quality of information. These individuals can provide detail around how information is collected and utilized. From there, create a data map to showcase all workflows within the organization. This will help understand what tools or processes should be implemented and a prioritization structure for those projects.

IT should also be involved to implement priorities and source technology that will need to be utilized to enforce and maintain data management. IT can also provide insight into technical resources available, given other business priorities.

Benchmarks can be taken and regular progress can be checked as organizations move through the process of implementing new solutions. The group can meet on a regular basis to review new statistics and processes, and identify if the quality of information is improving for all departments.

Companies that have a centralized approach to how they review and manage their data quality strategy are more likely to utilize SaaS technology for data quality management compared to those who are decentralized.

The majority of organizations manage their data quality in a decentralized way.

66 percent
lack a coherent, centralized approach.

Those who use more sophisticated, automated methods for data management are more likely to have their data strategy managed centrally.

The state of data quality
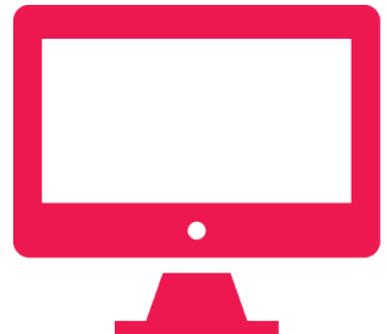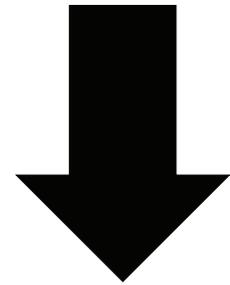
Consolidate data

One of the first tasks in any central strategy is consolidating information. According to the Experian Data Quality study, the average large organization has eight different databases. That statistic most likely does not include other spreadsheets or sources of data that may exist outside of a database, which can be numerous in large organizations.

In order to create a centralized approach to data management, organizations need to consolidate different sources of information. This will allow data to more easily be accessed, but also allows for consistency in management and standardization processes.

In order to consolidate data, organizations should take several steps.

1. Identify all sources of information that should be consolidated and who the owners are for each source.

2. Identify the data infrastructure. Is there an existing database where all information can be stored or is a new system required that can better accommodate multiple departments? The data quality task force can help determine the need.

3. Clean and standardize as much existing information as possible. Most often, contact data will be an easy source of information that is contained within each source. These details can be utilized to identify duplicate records across sources in order to consolidate information for each client into a single central source.

4. Utilize software to identify duplicates and remove the possibly of human error. Once potential duplicates are found, a golden record can be identified and all information can fall within that record.

Creating a centralized source can take some time; however, it will be invaluable for quickly accessing customer information and improving business intelligence. Centralization should be paired with data best practices in order to ensure the quality of information.

Implement data best practices

Once information is consolidated, organizations should look to implement best practices around data management within their newly formed system. In order for departments to adopt a new central source, it needs to be easy to utilize and have quality information. Otherwise individuals will revert back to old data sources that better fit their business need.

It is important to note that data quality is not meant to be a one-off cleaning engagement; data is constantly entering a database, expiring or changing format.

Just five years ago, the term big data did not exist and the concept of unstructured data from social media and mobile phones was just starting to come to the forefront. As organizations become more sophisticated with their data management practices, it is important that they have a best practices strategy and then adjust it over time based on market need.

Implementing data management best practices will help combat incomplete or missing data, outdated information and inaccurate data, which are the most common errors within a database. By preventing these errors, information can be maintained over time and be fit for its desired purpose.

There are six key best practices.

1. Create benchmarks around data accuracy

As with any business initiative, it is important to track progress in order to demonstrate that time and budget associated with a project are bringing a return on investment. The data quality task force should look to create benchmarks around data accuracy to help show improvement, but determine what processes or investment should be continued and what did not work as well as planned.

Benchmarks can easily be taken around package delivery, returned mail, email deliverability, or customer service calls. Businesses can also look to leverage third party consultants to benchmark segments within the database. Over time, these same benchmarks can be reviewed once processes and tools have been implemented.

2. Verify data upon entry

Today, most businesses utilize information as soon as it is entered for loyalty offers, marketing efforts, fulfillment, or billing. While information has always needed to be accurate from an operational standpoint, consumer communication has become more of a factor given the speed at which organizations need to follow-up with relevant marketing offers. Inaccurate data affects customer interaction almost immediately.

Implementing data management best practices will help combat incomplete or missing data, outdated information and inaccurate data, which are the most common errors within a database.

Organizations vary in the way they manage their data quality.

99 percent state they have a data management strategy in place.

38 percent perform regular manual analysis in Excel.

55 percent use some sort of automated method to manage data accuracy.

While there is less confessed ignorance among IT and data management professionals around the term big data, they do demonstrate the widest array of opinion on what the term means.

The state of data quality

Therefore, it is important to check the validity of information as it is being entered. Software tools can be put in place to verify structured customer information, such as email address, mailing address and telephone number.

This standardized and validated information allows organizations to more easily find existing accounts and accurately append third party data sets that rely on basic customer information. With cross-channel marketing efforts, point-of-capture verification processes help to ensure the accuracy of customer information, but also the prevention of duplicate accounts, which keep marketers from valuable consumer insight.

By validating customer information, organizations can prevent inaccurate data and ensure that communications not only reach the consumer, but that any customization techniques are more accurate across channels.

3. Validate information with consumers when possible

Data expires within a database quickly, in fact, it is estimated that two percent of contact data goes bad each month, which is almost a quarter of the database annually. To keep consumer information up to date, it is important to validate information with the consumer as often as possible.

Not all of these methods involve directly reaching out to the consumer. Marketers can watch outbound communication efforts for signs customer information may not be accurate.

For example, marketers can watch delivery rates and open rates from email campaigns. If an email bounces or customers go for long stretches without opening messages, the email address may not be active. Marketers can flag email addresses for update at the next consumer interaction, or reach out to the customer for changes in communication preferences.

However, when customers call into a call center or go into a branch or store location, associates can verify existing information to make sure it is still the best method of communication for that individual.

4. Improve searching functionality

Duplicate records cause problems for organizations by spreading out account history and creating incomplete customer records. Duplicates are often created when recent information is entered and it cannot be reconciled with an existing record.

Most often this is because the record cannot be found due to a slight variation, such as a name abbreviation or a mis-keyed email address.

Basic searching functionality within a database is often poor, requiring an exact match to find an existing record. More sophisticated searching can be put in place to find potential matches and identify more possibilities for the account than just a one-for-one match.

5. Check the database for duplicate entries on a regular basis

Even with validated data and improved searching, duplicates will inevitably be created due to the nature of human error. Stakeholders should be sure to check the database on a regular interval to ensure no duplicate accounts have been created and to consolidate information whenever possible.

6. Review data management processes annually

Data management and information requirements change constantly across an organization. Looking back on the past five years, data techniques have changed dramatically as evidence by the research.

To ensure that data is fit for purpose and can be used in the desired way by the organization, the data quality task force should look to review data management practices annually and identify new ways that information is being utilized or processes that may not be fulfilling their purpose based on benchmark data.

By reviewing management processes on a regular basis, organizations can make certain they are able to use their valuable data asset to its maximum potential.

Among those with contact data accuracy issues, the main cause of such problems is believed to be human error, followed by an inadequate data strategy and a lack of internal manual resources.

Duplicate data is among the top three data quality errors for

30 percent
of organizations.

Stakeholders should be sure to check the database on a regular interval to ensure no duplicate accounts have been created and consolidate information whenever possible.

The state of data quality

The utilization of data across organizations has shifted. Businesses see the potential for this valuable asset to provide an avenue for better consumer interaction and business decision making with the evolution of big data.

To ensure data is fit for purpose, organizations need to take steps to ensure its accuracy, accessibility and completeness. Data management best practices should be implemented to standardize data, better consolidate it into a single record for each client, and append additional data sets when required.

These practices need to be managed centrally across an organization to consolidate resources and ensure all information receives similar validation and standardization.

Data quality is the foundation for any data-driven effort. As the data proliferation continues, organizations need to prioritize data quality to ensure the success of these initiatives.

About Experian Data Quality

Experian Data Quality is a global leader in providing data quality software and services to organizations of all sizes. We help our clients to proactively manage the quality of their data through world-class validation, matching, enrichment and profiling capabilities. With flexible software-as-a-service and on-premise deployment models, Experian Data Quality software allows organizations around the world to truly connect with their customers by delivering intelligent interactions, every time.

Established in 1990 with offices throughout the United States, Europe and Asia Pacific, Experian Data Quality has more than 13,500 clients worldwide in retail, finance, education, insurance, government, healthcare and other sectors. For more information, visit http://www.qas.com.

For more information about how you can improve your data quality management strategy, please call 1 888 727 8330 or visit us online at www.qas.com.

Experian®
Data Quality

Intelligent interactions.

Every time.

09/2013