

The Real Cost of Bad Data  
**Six Simple Steps To Address Data Quality Issues**

A White Paper

---

WebFOCUS   iWay Software   Omni

# Table of Contents

---

<b>1</b>	<b>Introduction</b>
<b>3</b>	<b>Getting Started – Building a Business Case</b>
<b>5</b>	<b>Fixing Your Bad Data in Six Steps</b>
5	Step 1: Gain an Understanding of the Problem
5	Step 2: Create the Data Steward Responsibility
6	Step 3: Agree on the Impact of Bad Data
7	Step 4: Decide What to Do With Bad Data
7	Step 5: Start the Cleansing Process
8	Step 6: Implement and Maintain Broad-Reaching Data Quality Processes
<b>9</b>	<b>Best Practices in Data Quality</b>
<b>11</b>	<b>Beating Bad Data: Real-World Case Studies</b>
<b>12</b>	<b>Conclusion</b>

# Introduction

Inadequate data governance can cost millions. From regulatory fines for systems and control failures, to the labor expenses incurred from the extra effort required to produce accurate information for critical business decisions, to the legal fees spent cleaning up a publicity nightmare from a data-driven error. Moreover, when it comes to patient or consumer safety, poor data quality can also cost lives.

In addition, most organizations tend to handle data quality in a very tactical fashion. Either the IT organization rectifies any issues, or the individual business units must create business processes to fix it. When IT is in charge, the problem goes into the queue and, more often than not, gets relegated to a low priority. So business units, who face the problem on a day-to-day basis, must step in to deal with it.

Key findings from a Gartner<sup>1</sup> study revealed:

- Poor data quality is a primary reason for 40 percent of all business initiatives failing to achieve their targeted benefits
- Data quality effects overall labor productivity by as much as a 20 percent

A survey performed by The Data Warehouse Institute demonstrates similar results, with respondents claiming that poor data quality has led to lost revenue (54 percent), extra costs (72 percent), and decreases in customer satisfaction (67 percent).<sup>2</sup>

So why aren't data quality problems being handled at an organizational level?

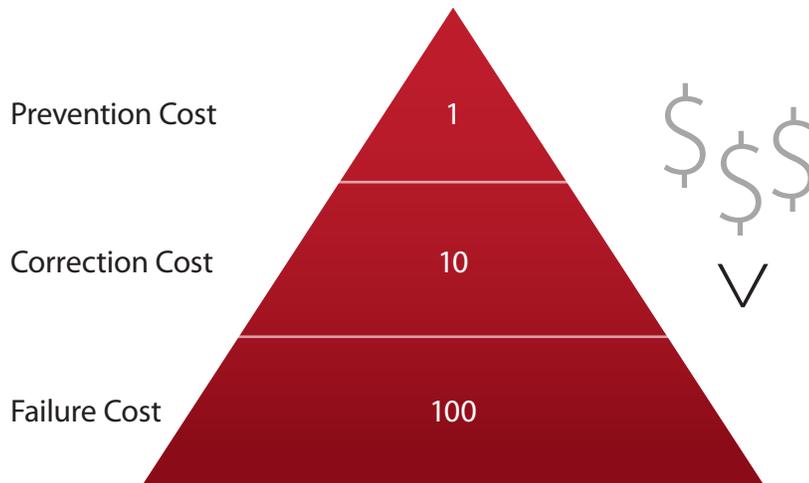
Typically, the numbers cannot be adequately digested when business processes and workarounds to these issues are constantly being developed. Oftentimes, changing the status quo means changing the organizational culture, and being able to treat data quality as an ongoing strategic program, not a series of one-off projects. Cultural change is difficult to foster, and if there are perceived business processes to deal with the problem anyway, it just seems easier to leave things be.

Joel E. Ross' 1-10-100 rule<sup>3</sup> gives us something important to consider. The rule explains how failure to take notice of one cost escalates the loss in terms of dollars. There are many costs of non-quality, such as: prevention, appraisal, internal failure, and external failure. Of these types of costs, prevention cost should probably take priority because it is much less costly to prevent a defect than to correct one. The principle is not unlike the traditional medical axiom: "An ounce of prevention is worth a pound of cure." The relationship between these costs is reflected in 1-10-100 rule as depicted in the following illustration:

<sup>1</sup> Friedman, Ted; Smith, Michael. "Measuring the Business Value of Data Quality," Gartner, October 2011.

<sup>2</sup> Eckerson, Wayne. "Data Quality and the Bottom Line," The Data Warehouse Institute.

<sup>3</sup> Ross, E. Joel; Perry, Susan. "Total Quality Management: Text, Cases and Readings," CRC Press, June 1999.



**The above illustration shows that \$1 spent on prevention will save \$10 on correction and \$100 on failure costs. As one moves along the streams of events from design to delivery the cost of errors escalates as failure costs becomes greater.**

The earlier you detect and prevent a defect, the more you can save. For example, if you catch a two-cent chip problem before you use it, and throw it away, you lose two cents. If you don't find it until it has been soldered into a computer component, it may cost 10 dollars to repair the part. If you don't catch the components until it is in the motherboard and in the customer's hand, the repair will cost hundreds of dollars. Indeed if a \$2,000 computer has to be repaired in the field, the expense may exceed the manufacturing cost.

In this white paper, we'll outline six simple, yet highly effective steps for preventing and fixing your bad data. You'll learn how to garner executive support for a data quality management program, and what processes you'll need to follow to ensure optimum integrity in your data. We'll also highlight some best practices in data quality, and share some real-world success stories.

## Getting Started – Building a Business Case

In a recent article in *Harvard Business Review*, Thomas C. Redman, president of Navesink Consulting said, “Most senior executives are well aware that their data are not up to snuff. But absent a business case, data quality doesn’t make the cut among competing priorities.”<sup>4</sup>

Before you embark on creating and implementing a formal data quality program, you’ll need to build your business case and round up support (and funding) from the right executive stakeholders. Gaining the needed backing, however, is easier said than done. You can best validate your arguments in the following ways:

### **Showcase High-Profile Problems**

Identify a few high-profile functional areas where bad data is causing problems. This will help you demonstrate the most pressing data quality needs, and indicate where the most significant near-term results can be achieved.

CRM systems are often a likely culprit. Dun & Bradstreet data shows that close to 96 percent of CRM records have some level of inaccuracy, while CRM degradation is approaching two percent per month.<sup>5</sup> These systems can be directly tied to revenue generation, customer satisfaction and loyalty, or other critical goals – and will therefore grab the attention of your executive team.

### **Define “Quality”**

It is important for all those who interact with or oversee data to have a common understanding of what “quality” means. Therefore, you’ll need to define a few metrics or benchmarks that your information can be measured against. For example, customer numbers must fall within certain ranges, or product numbers must follow a certain format.

### **Estimate the Costs**

Tally up the money you’re losing due to the bad information contained in the systems you identified earlier. For the average company, this may be as high as 15 to 20 percent of operating revenue, according to estimates from the U.S. Insurance Data Management Association.<sup>6</sup>

Be sure to include direct costs, such as non-value-added work performed when people must locate, validate, or cleanse the data they need to do their jobs, as well as the costs associated with re-work when mistakes are made due to poor information. For example, a healthcare firm may need to re-run expensive tests because the patient ID number was incorrect, or a retailer may need to re-process an order because erroneous part/product numbers caused the customer to receive the wrong item.

You’ll also need to consider indirect costs, such as lower-than-expected campaign response rates or poor up-sell/cross-sell penetration due to inaccurate customer mailing or e-mail addresses. Don’t forget the costs associated with the poor decisions that result from invalid or corrupt information.

<sup>4</sup> Redman, C. Thomas. “Making the Case for Better Quality Data,” HBR Blog Network, August 2012.

<sup>5</sup> “The Big Payback on Quality Data,” Dun & Bradstreet, May 2012.

<sup>6</sup> Lizard, Sophie. “The Costs of Data Quality Failure,” Match2Lists.com, December 2011.

### **Profile Your Information**

Using the quality metrics and rules defined in prior steps, run data profiles against the systems and sources you identified as being the most problematic. This will clearly demonstrate exactly where your bad data lives, and just how severe those issues may be. Although your executive team is probably already aware that there is a problem, they may be quite surprised at the magnitude of it.

Data profiling is explained in more detail in the next section of this whitepaper.

### **Compile and Present Your Findings**

Create an executive briefing that outlines the specific data discovery processes you followed, the results of those processes, and the costs associated with any bad data you identified. You'll also need to include recommendations for resolving any problems – suggested technology solutions, a roadmap for implementing those solutions, etc.

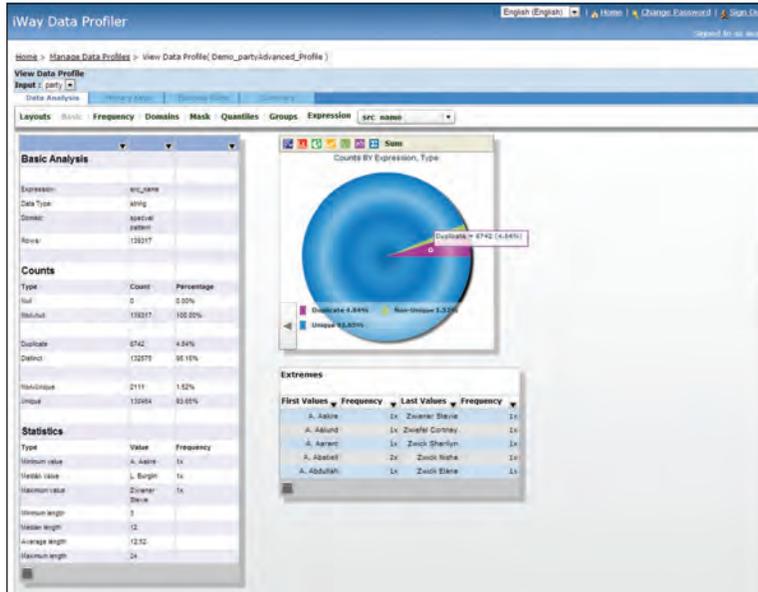
Once your findings have been compiled and documented, schedule an executive briefing to present your case, ask for support, and secure the needed budget funds. Be sure that the executives responsible for those processes that are most impacted by bad data are present at the meeting, since they will be the ones who find your proposal most compelling, and will have the most to gain by supporting it.

# Fixing Your Bad Data in Six Steps

## Step 1: Gain an Understanding of the Problem

Your organization must come to some consensus about where the bad data is, and what impact it is having on your business. The best way to accomplish this is through data profiling. Data profiling is a learning activity, focused on discovering what you have and how bad it is.

Several types of gaps may be uncovered during the profiling process. Data may be incomplete. For example, customer records in a customer relationship management (CRM) system may be missing zip codes or e-mail addresses, or part numbers may be absent in entries within a materials management database. Or, a human resources application may be missing employee records. In this second case, you must determine the full definition of “employee” to perform your assessment (i.e. will contractors and/or firms you outsource to be included?).



**Through comprehensive data profiling, your company can perform a full assessment of its most troublesome data quality problems.**

You might also discover inaccuracies or inconsistencies within your data. Are values contained within your database fields correct? Some common problems could include zip code entries that contain letters, or customer e-mail addresses without the “@” symbol. Lastly, duplications and redundancies are likely to be found. For example, products may have been entered into an inventory database more than once, or there may be multiple records for a single customer in a CRM system.

## Step 2: Create the Data Steward Responsibility

Your data stewards may be the most important people in your data quality initiative. These stakeholders will be responsible for creating the rules for data generation, handling, maintenance, and sharing, and for outlining the processes that will help ensure quality across the organization.

Your data stewards will also serve to enforce these guidelines, to monitor and measure information integrity on an ongoing basis, and to modify quality practices as needs, data sources, and other factors change.

Keep in mind that you probably already have people performing data stewardship functions. Formalizing the role with responsibilities around the data will serve two purposes. First, it will demonstrate to the organization the importance of data in your business processes. Second, it will convey the importance of people in those processes.

### Step 3: Agree on the Impact of Bad Data

To further your exercise in building the case for correcting data quality problems, you must gain full insight into the impact of the bad data. Through the data profiling performed in the first step, you've come to an understanding of what is wrong. In this step, you'll determine why it is wrong, and how it is affecting your business.

Perhaps the best way to begin this step is to create a lifecycle chart of the data you looked at during the profiling step. Where does the data originate? What are the downstream applications that use this data? What is the impact to those downstream applications if the data is not right? Identify touch points during the lifecycle where data can be manipulated, and highlight places where the data is output, so you can clearly see what happens when quality is compromised. At every step, associate a value to right data and wrong data. When the lifecycle is complete, simply add up the numbers.



**Once you know which data is bad, you'll need to determine why and understand how it impacts your business.**

It is important to note that, to achieve certain data quality goals, you may need to modify certain business processes. In particular, any manual activities or workflows that may be significantly contributing to information integrity issues must be closely evaluated, and altered as needed to minimize errors going forward.

#### **Step 4: Decide What to Do With Bad Data**

Once issues are uncovered during profiling, and you have some insight into the extent of the problems, what will you do to correct them? All opportunities, challenges, and risks must be prioritized, and a methodology for handling them must be outlined, implemented, and enforced.

The plan must answer these questions:

- Which errors will be accepted (the ones that fall within acceptable limits), and which ones will be rejected (the most problematic mistakes)?
- What will happen to rejected errors? Will the data be excluded, or will it be sent to someone for further review?
- How will correctable errors be fixed? Will they be handled manually, or will default values be applied?

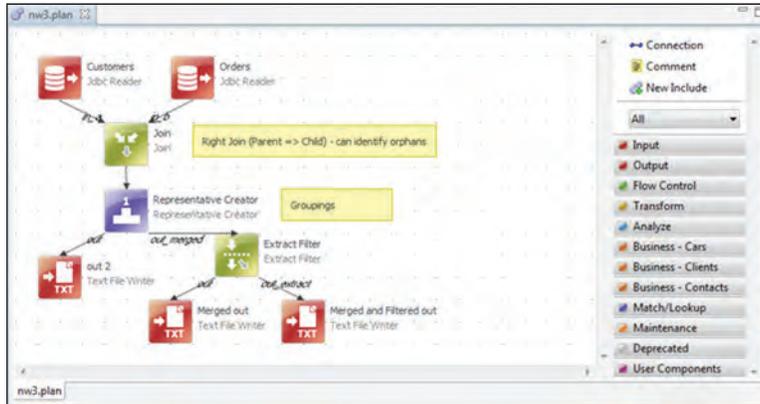
Note that these plans can vary for certain types of data errors. For example, the processes for handling formatting mistakes may differ from the steps taken to address redundancies, missing information, or invalid data.

Successful data quality management strategies don't necessarily require you to make your information completely problem-free. In fact, such an unattainable goal would prevent your plans from getting off the ground in the first place. Additionally, keep in mind that 'right' can be 'good enough', depending on the impact analysis performed during step three.

#### **Step 5: Start the Cleansing Process**

You've identified the data that is problematic, come to a determination as to what the corrected data should look like, and assigned people to be responsible for the data going forward. So, what's next?

It's time to put the processes, procedures, and plans into place to begin cleansing the data. This may seem like a technical exercise, and should involve your IT organization to work with the actual tools themselves. The work you've put into the previous steps, however, will make the application of the plans quite easy. You and your data stewards have created the blueprint, just as an architect would do. Now you have to turn the blueprint over to the contractor to begin the construction, using the tools at their disposal.



**Comprehensive cleansing tools can help efficiently make the needed corrections to existing dirty data.**

## Step 6: Implement and Maintain Broad-Reaching Data Quality Processes

You've got a process in place to resolve your initial business problem, and your business unit is now operating more efficiently. Is it operating at peak efficiency? Are there other data-related problems your business unit should be looking to correct? If so, you've got the people, processes, and technology in place to deal with them.

Before you start talking about it, however, you need to make sure you have ways to measure and monitor your progress towards your data quality goals. You have already developed metrics, and it is important to continually keep them refreshed and updated. An objective measure needs to be created to demonstrate your success, and that measure must be continually monitored.

Once these measures are put into place and are being regularly tracked, you're still far from done.

Remember step three, where you created the data lifecycle? You successfully fixed the data problem in your business unit, but now it's time to talk to other business units. Although downstream consumers of your improved data are happy that you've fixed their problems, you too are a consumer of data, and need to share your story with the producers of the data you are having problems with.

Now that problem of creating organizational change is a groundswell that is being dealt with one problem at a time. Slowly, your entire organization will benefit from your pioneering efforts and the cultural change will start to occur.

# Best Practices in Data Quality

In addition to these steps, there are other actions you can take to ensure the integrity of information. Below are some techniques that have been implemented by organizations that demonstrate a data quality mindset.

## **Treat Data as an Asset**

Long ago, companies learned to treat their workforce – those who make decisions and carry out mission-critical tasks – as an asset. They have since invested in training and other valuable programs to get the most out of their employees and maximize their contribution to the organization.

The time has come to treat the information used to support the activities of those employees as an asset as well. Organizations must develop a culture that makes data quality a high priority. Employees at all levels – not only those who oversee data, but those who consume and use it – must realize the importance of trusted information, and foster its maintenance and enrichment whenever possible. Consistent, accurate, and timely information is a foundation for better performance, higher outcomes, and, for publically traded companies, higher valuations.

## **Take a Proactive Approach**

Many companies rely on a reactive approach to data quality, taking steps to locate and cleanse invalid or corrupt information on a regular basis. What they don't realize is that once bad data has entered the environment, some irreversible damage may have already been done. Additionally, because data moves throughout a business quickly, other systems may have been compromised. The dirty data will quickly permeate other sources.

Think of data flowing like a river. Stopping contaminated water at its source would be less costly and require less effort than cleaning a large body of water – the lake that the dirty river runs into – after it has been polluted. Additionally, countless people may drink the water and become ill before that cleansing is complete.

That's why a more proactive approach to data quality is needed. A real-time data quality firewall can catch bad records as soon as, or even before, they enter a database. This will eliminate the need to devote human and financial resources to locating and rectifying corrupt information that has been scattered across countless systems and sources.

## **Enforce, Enforce, Enforce: Data Governance is Key**

Many companies that implement data quality programs often fail to follow up and ensure related policies and procedures are being followed. There are many possible reasons for this, but one of the most common, according to the BeyeNETWORK blog, is "the approach that data quality and data governance are one-time projects, instead of ongoing programs requiring their own continuous monitoring and improvement. Without an effective data governance plan, data quality improvement is reduced to little more than data cleansing – this may provide quality data temporarily, but nothing more."<sup>8</sup>

<sup>8</sup> Medved, Joy. "An Easier Way to Enforce Data Governance Compliance," BeyeNETWORK, February 2009.



**An effective data governance plan, complete with supporting solutions, is crucial to the achievement of continuous data quality.**

In his "Obsessive-Compulsive Data Quality" blog, Jim Harris writes, "The difference between success and failure for your data governance program is the ability to enforce your policies."<sup>9</sup> In other words, your data quality plans are useless if they aren't properly implemented and followed.

Harris suggests that the best way to do this is to:

- Document your policies in a clear and easy-to-understand way
- Communicate the procedures, but allow for open discussion and debate
- Define and measure quality metrics, so the impact can be monitored
- Correct any troublesome business processes, technologies, data, or people that may be contributing to quality problems
- Refine your policies as your organization, its goals, and its data environment evolve

<sup>9</sup> Harris, Jim. "Jack Bauer and Enforcing Data Governance Policies," Obsessive-Compulsive Data Quality Blog, June 2010.

# Beating Bad Data: Real-World Case Studies

How do these principles apply in the real world? Let's look at how some organizations are leveraging innovative tools and best practices to enhance data quality across their business.

## **CenterStone Technologies**

This leading developer of sales order management solutions is driving revenue and improving customer service with a versatile, scalable, high-performance environment that automates and accelerates data transfer and cleansing, while improving data quality. Due to increased efficiency and reduced costs, CenterStone will gain a projected, cumulative three-year net benefit of more than \$1 million for the project, with a return on investment of 371 percent and a payback period of three months.

## **Love's Travel Stops & Country Stores**

Love's is the seventh largest privately held company in America, according to Forbes, operating more than 280 stores and 130 tire care centers across 39 states. To match point-of-sale values with remittance data, in order to spot anomalies and determine the source of recurring problems, Love's implemented a unique environment built on advanced integration and data quality management solutions. Designed to ensure the quality of financial data by validating that transactions are routed correctly at each stage of a multi-faceted communications process, the environment is helping the company to better detect, analyze, and resolve problems with financial issues, which in turn is improving cash flow and accelerating payments.

## **State of New Hampshire Department of Revenue Administration**

This agency, which manages tax collection for 253 state municipalities, sets the tax rates, and oversees the statewide property tax system, was having trouble capturing, comparing, and analyzing taxpayer data, and struggled to make projections. It deployed a cost-effective platform for retrieving, analyzing, and enhancing the quality of its data, which resides in a vast array of accounting and billing systems. New Hampshire can now provide better services to taxpayers, and deliver more accurate forecasts to the legislature, helping civic leaders judge the economic vitality of the state and the practicality of the taxes.

## **Senior Health Insurance Company of Pennsylvania**

The Senior Health Insurance Company of Pennsylvania (SHIP) incorporated data quality into its plans to create an advanced business analysis environment designed to support long-range planning, enhanced communication, and industry-wide information sharing, while increasing efficiency and reducing costs. Information about tens of thousands of payments and thousands of claims, as well as details about policyholder locations, claim types, ages, and costs are automatically cleansed while being loaded into a warehouse, where it is made available for analysis. Thanks to information that is timely and trusted, the environment is helping SHIP managers to reduce fraud and maintain the level of reserves they need to pay clients for the next 50 years or more.

# Conclusion

Promoting data quality across your organization doesn't have to be an overwhelming challenge. By following six simple steps, you can achieve success by creating a culture that fosters information integrity and implementing the proactive processes and controls needed to ensure that your data is trusted, timely, and complete at all times.

What do companies like CenterStone, and Love's Travel Stops and Country Stores, and SHIP rely on to support their mission-critical data quality initiatives? Comprehensive data integration and integrity solutions from Information Builders:

### **iWay Data Quality Suite**

Real-time data quality management and profiling capabilities help to optimize the consistency, completeness, and correctness of data across all information assets.

### **Master Data Suite**

Leading-edge master data management (MDM) technology consolidates millions of records, and makes unified and validated master data instantly available to a wide range of enterprise applications and systems.

### **iWay Integration Suite**

Regardless of where data resides, whether it's in a structured or unstructured format, integrity solutions from Information Builders can seamlessly cleanse and enrich it, ensuring a consistent and accurate view of data from every interaction point.

These powerful data and application integration technologies, combined with innovative data quality and master data management capabilities, make up the iWay information asset management platform.

### **About Information Builders**

Information Builders helps organizations transform data into business value. Our software solutions for business intelligence and analytics, integration, and data integrity empower people to make smarter decisions, strengthen customer relationships, and drive growth. Our dedication to customer success is unmatched in the industry. That's why tens of thousands of leading organizations rely on Information Builders to be their trusted partner. Founded in 1975, Information Builders is headquartered in New York, NY, with offices around the world, and remains one of the largest independent, privately held companies in the industry. Visit us at [informationbuilders.com](http://informationbuilders.com), follow us on Twitter at [@infobldr](https://twitter.com/infobldr), like us on Facebook, and visit our LinkedIn page.

# Worldwide Offices

## Corporate Headquarters

Two Penn Plaza  
New York, NY 10121-2898  
(212) 736-4433  
(800) 969-4636

## United States

**Atlanta, GA\*** (770) 395-9913  
**Baltimore, MD** (703) 247-5565  
**Boston, MA\*** (781) 224-7660  
**Channels** (770) 677-9923  
**Chicago, IL\*** (630) 971-6700  
**Cincinnati, OH\*** (513) 891-2338  
**Dallas, TX\*** (972) 398-4100  
**Denver, CO\*** (303) 770-4440  
**Detroit, MI\*** (248) 641-8820  
**Federal Systems, DC\*** (703) 276-9006  
**Florham Park, NJ** (973) 593-0022  
**Gulf Area** (972) 490-1300  
**Hartford, CT** (781) 272-8600  
**Houston, TX\*** (713) 952-4800  
**Kansas City, MO** (816) 471-3320  
**Los Angeles, CA\*** (310) 615-0735  
**Milwaukee, WI** (414) 827-4685  
**Minneapolis, MN\*** (651) 602-9100  
**New York, NY\*** (212) 736-4433  
**Orlando, FL** (407) 804-8000  
**Philadelphia, PA\*** (610) 940-0790  
**Phoenix, AZ** (480) 346-1095  
**Pittsburgh, PA** (412) 494-9699  
**Sacramento, CA** (916) 973-9511  
**San Jose, CA\*** (408) 453-7600  
**Seattle, WA** (206) 624-9055  
**St. Louis, MO\*** (636) 519-1411, ext. 321  
**Washington DC\*** (703) 276-9006

## International

**Australia\***  
Melbourne 61-3-9631-7900  
Sydney 61-2-8223-0600  
**Austria** Raffaisen Informatik Consulting GmbH  
Wien 43-1-211-36-3344  
**Brazil** InfoBuild Brazil Ltda.  
São Paulo 55-11-3285-1050  
**Canada**  
Calgary (403) 718-9828  
Montreal\* (514) 421-1555  
Ottawa (613) 233-7647  
Toronto\* (416) 364-2760  
Vancouver (604) 688-2499  
**China** Information Builders China  
Beijing 86-10-5128-9680  
**Estonia** InfoBuild Estonia OÜ  
Tallinn 372-618-1585  
**Finland** InfoBuild Oy  
Espoo 358-0-207-580-840  
**France\***  
Puteaux +33 (0)1-49-00-66-00  
**Germany**  
Eschborn\* 49-6196-775-76-0  
**Greece** Applied Science Ltd.  
Athens 30-210-699-8225  
**Guatemala** IDS de Centroamerica  
Guatemala City (502) 2412-4212  
**India\*** InfoBuild India  
Chennai 91-44-42177082  
**Israel** SRL Software Products Ltd.  
Petah-Tikva 972-3-9787273  
**Italy** Information Builders Italia S.r.l.  
Agrate Brianza 39-039-596620  
**Japan** KK Ashisuto  
Tokyo 81-3-5276-5863  
**Latvia** InfoBuild Lithuania, UAB  
Vilnius 371-67039637  
**Lithuania** InfoBuild Lithuania, UAB  
Vilnius 370-5-268-3327  
**Mexico**  
Mexico City 52-55-5062-0660

## Middle East

Innovative Corner Est.  
Riyadh 966-1-2939007  
■ Iraq ■ Lebanon ■ Oman ■ Saudi Arabia  
■ United Arab Emirates (UAE)

**Netherlands\*** Information Builders (Benelux) B.V.  
Amstelveen 31 (0)20-4563333

**Nigeria** InfoBuild Nigeria  
Garki-Abuja 234-9-290-2621

**Norway** InfoBuild Norge AS c/o Okonor  
Tynset 358-0-207-580-840

## Portugal

Lisboa 351-217-217-400

**Singapore** Automatic Identification Technology Ltd.  
Singapore 65-69080191/92

## South Africa

Fujitsu (Pty) Ltd.  
Cape Town 27-21-937-6100  
Sandton 27-11-233-5432  
InfoBuild (Pty) Ltd.  
Johannesburg 27-11-510-0070

**South Korea** UVANSYS, Inc.  
Seoul 82-2-832-0705

**Southeast Asia** Information Builders SEAsia Pte. Ltd.  
Singapore 60-172980912

■ Bangladesh ■ Brunei ■ Burma ■ Cambodia  
■ Indonesia ■ Malaysia ■ Papua New Guinea  
■ Thailand ■ The Philippines ■ Vietnam

## Spain

Barcelona 34-93-452-63-85  
Bilbao 34-94-452-50-15  
Madrid\* 34-91-710-22-75

**Sweden** InfoBuild AB  
Solna 46-7-024-656-50

## Switzerland

Dietlikon 41-44-839-49-49

**Taiwan** Galaxy Software Services, Inc.  
Taipei (866) 2-2586-7890, ext. 114

## United Kingdom\*

Uxbridge Middlesex 0845-658-8484

**Venezuela** InfoServices Consulting  
Caracas 58212-763-1653

\* Training facilities are located at these offices.