Data Quality: The Next Disaster Waiting to Happen?

Managing utility assets relies more and more on managing data quality. Mismanagement can lead to catastrophe.

By Robert Thurlby and Mark Brulé

Consider the following situation: A U.K. electricity distributor, under pressure to reduce the number of outages, redesigns the maintenance process for a type of substation of which it has approximately 100 in its network. The new process is tested for rigor and safety by simulation against the asset database and then introduced. Soon, there is an incident involving a technician using the new process. At the subsequent inquiry it is established that, although the technician was using the process correctly, the substation layout differed from the standard layout, and its differences were not recorded in the asset database. The utility is found guilty of negligence and heavily fined.

Consider a second situation:

An engineering maintenance organization wins a large contract to refurbish a set of capital assets. To refurbish each asset is a six-month task, and there are financial penalties for delays. The price quoted was based on the engineering drawings of the asset provided by the operating company. When the first refurbishment starts, however, the drawings are found to bear little relationship to the actual asset. Ten years of operation and maintenance have resulted in significant changes to

the asset's structure. Consequently, the refurbishment program's first month is spent producing accurate drawings, and the project's completion is delayed. The operating company invokes the penalty clause; the contractor countersues for provision of inaccurate information. The case is still unresolved, but both organizations have lost critical time and money.

Apocryphal stories? No, both are real instances. Exceptional? Again no. Evidence indicates there are far more instances of this type of problem than there were even five years ago. Each illustrates the impact data quality has on our ability to manage assets and meet performance standards effectively.

The Problem: Bad Data

Any examination of what has caused the problems will focus on where utilities have been investing. During the 1990s, utilities invested heavily in new IT applications. For distribution operations, much of the investment was in GIS, work management, and financial and asset management applications. The primary purpose of these investments was to automate existing manual processes and reduce costs.

One element of the programs to develop such IT systems was usually a data val-

idation and cleanup activity. Maps and other records, even those in electronic form, were found to contain many errors and inaccuracies. But, as they were frequently maintained by staff who worked on the network and so had intimate knowledge of what was there, the errors and inaccuracies had only a minor impact on the business. Nevertheless, utilities grasped the opportunity to create accurate records. For some, this was a mammoth task. (One utility, while implementing a GIS, invested more than 200 man-years in cleaning up its geographic data.)

But once the applications were implemented, a fatal assumption was frequently made. The assumption was that the new automated processes would also maintain data accuracy. Data validation processes were put in place, and code was written in the update and maintenance routines to check the input data. But often all this code checked was whether the data in a given field had the right format and whether it was consistent with the fields to which it was connected. It didn't necessarily check whether the data was accurate or complete. Consequently, data quality and reliability degraded, but in such a slow and insidious way that the problems being caused were not recognized as problems.

The Problem Becomes Critical

As long as there remained staff with working knowledge of the network, the

> *Paper maps taken into the field may be marked up to show inaccuracies in the data, but getting those corrections back into the GIS is complicated.*

problem of eroding data quality was containable. An engineer receiving an inaccurate map or asset description would recognize it as such and act accordingly. Really conscientious engineers, of whom there are legions, would correct the hard copy documentation in the hope that it would be used to correct the computer records.

However, utilities were facing a new set of issues:

- Many engineers and technicians were leaving the utility, taking their knowledge with them.
- More work was being subcontracted to external maintenance organizations whose staff lacked an intimate knowledge of the network on which they were working.
- The pressure on staff to become more productive resulted in their having less time to report data errors—something they were not incentivized to do.
- Time and resources were being devoted to minimize customer minutes lost, with the unforeseen consequence of even less attention being paid to data quality.

For all these reasons, we believe data quality—which in many utilities is not only poor but declining—is an accident waiting to happen.

Potential Impact

Consider some likely impacts of inaccurate, incomplete data:

- A contractor receiving inaccurate data could have his field staff arrive at the wrong site, with the wrong tools and materials, and with skill sets that don't match the job requirements. Recent research in a leading U.K. electricity distributor has found that 10 percent of all jobs require revisits due to inaccurate data. Many other utilities believe this figure to be low.
- Delays in completing work will reduce customer service levels, potentially incurring performance penalties.
- Pressure to complete delayed work, when combined with inaccurate information, encourages shortcuts. This can have two impacts: failure to complete the work correctly, and increased health and safety incidents.
- Lack of accurate data contributes to poor network investment decisions. Failure to report the state of the network accurately will impair the asset manager's ability to plan maintenance and refurbishment programs, thereby reducing the network's operational performance.
- Litigation risk will increase. For many utility executives, the fear of being sued by employees, customers and third parties is ever present. If any of these litigants can prove that the information provided to them, or the information on which a key decision was made, was inaccurate, they will have a strong case. In extreme cases, jail terms are possible.

Addressing The Problem

No simple solutions exist. The starting point, however, must be a new attention to the quality and accuracy of data that makes up the physical network's asset register.

In the past, utility mapping departments have been required to manage data quality, which is problematic when those departments lack direct contact with the physical assets. Paper maps taken into the field may be marked up to show inaccuracies in the data, but getting those corrections back into the GIS is complicated. In addition, since the field crew's job is maintenance of the network and not maintenance of GIS data, these crucial data-update operations are performed when the crews have the time and inclination. That is not an effective way to maintain a quality asset register.

While current staff can be used to maintain data, locating and fixing data problems is not cost-free. Technology can provide cost-effective ways to make this happen. Mobile data-collection technology, in particular, is vital.

The challenge is to get data into the field so that it is accurate and usable. Field crews attending to data quality need the most current view of the data they can possibly get. If they are auditing data that is a week old or older, the chance of changes to the GIS representation of the data increases. The quickest way to get data to field crews is to give them electronic extracts. To manage this properly, the data must be extracted from the GIS and disconnected from the server so it can be loaded onto mobile devices.

Disconnected Versions

The principal challenge with mobile computing is managing disconnected versions. A "version" of a database is a subset of the data and a collection of changes to that subset. When a version is disconnected from the server database—which is the case with mobile applications—any changes made to the version are made without knowledge of what other users may be doing to the data as it exists in the server. Changes that different users make may conflict with each other. For mobile computing to work at all, this problem needs careful handling.

Working with spatial data adds a twist to this problem. Spatial data relies on a clear graphic display to make it truly valuable. When data is presented in a visually effective way, users are more productive. To take the data to the field in mobile units (such as handheld devices), the software must allow

both for effective management and effective display of the data. Unfortunately, handheld devices lack the computing power of desktop or even laptop computers.

Complicating this fact is the complexity of GIS information. This data tends to have rich data models with a great deal of detail about the network and connectivity stored within it. This makes the data valuable, but also presents challenges for disconnecting it from the server. How the handheld application manages GIS data, specifically in terms of limiting the information to what the field user truly needs, is critical to the system's success.

On the positive side, however, is the typical GIS's capability of handling changes that could be made in this environment. All of the GIS software on the market today has some notion of a long transaction, a paradigm where changes to the data can be stored and considered for long periods of time (weeks, months or even years) before being committed to the master database. With the support of long transactions comes support for managing conflicts, which some systems handle better than others.

The Supporting System

The disconnected version, however, is only part of the technology puzzle needed to make these types of solutions truly beneficial.

First, the mobile unit should have the ability to display data spatially and utilize GPS technology when appropriate.

Second, the structure of the data in the mobile unit should be somewhat independent of the structure of the data in the GIS. Trying to create a full replica of the GIS display in the mobile units is unnecessary and wasted effort. Handheld unit displays are often quite restricted in size and resolution, so the information being displayed should be relevant to the problem at hand.

Third, the software should provide built-in QA/QC processes that ensure accurate data during the collection stage. Ideally, automated data-validation and data-enhancement processes should be

Clearly, utilities must decide it is worth their while to spend additional money to create quality asset databases.

run on the data gathered in the field.

Fourth, the system should be able to communicate changes made to the data back to the GIS in a way that will not corrupt the GIS data itself, and do so in a way that supports user confidence in the data. Since field units will not take a complete GIS image with them, and

since field technicians are tasked with collecting field observations rather than maintaining GIS data, this implies there is a server-side function of the software that can cast the changes observed by the field crews into GIS-ready changes.

The Future

As this technology becomes available, it will become more cost-effective to consider field operations. Utilities, strapped as they are for human resources, should continue to look to outsourcing these functions for data collection and correction to bring their data up to a quality level that will support reliable operation. With the right mobile applications, knowledgeable people (whether contracted or employed by the utility) can gather valuable data as part of their work process.

Moving forward, data must be not only collected but maintained, an additional task that forms part of the data-maintenance process. Suitable software and hardware will continue to play an important role in allowing field personnel to keep the quality of network data at optimal levels. ∎

Dr. Robert Thurlby is Managing Consultant at Syntegra (U.K.). He has more than 20 years experience designing and developing information systems and business improvement programs for utilities worldwide. He can be reached at robert.thurlby@syntegra.com.

Mark Brulé is Chief Technology Officer of the Data Services business unit of Osmose Utilities Services, Inc., which provides innovative software, services, and data solutions to utilities. He can be reached at 315-433-1010 or visit www.osmose.com.