Real-time data collection technologies: Enhanced decision-making and cost savings January, 2005

TL Rust, H.Q Vu
Tetra Tech EM Inc
1881 Campus Commons Drive, Reston Va.

ABSTRACT

Handheld computers, Geographic Information Systems (GIS), and wireless communication devices are rapidly replacing traditional methods for field monitoring and data collection. Although pencil and paper remain important means of data transcription, field technicians can now use Personal Digital Assistants (PDA) to record their field notes and monitoring data. As data are uploaded wirelessly from the field, decision-makers can view real-time reports and maps that identify sample locations and monitoring results. The combination of PDAs, wireless communications, and Web-based GIS provides field personnel and decision-makers many benefits throughout the life cycle of a project, including improved data consistency, real-time transfer of data from field locations to centralized databases, input validation, elimination of transcription errors, and cost savings. Concerns have been expressed however, about investing in hardware, software, and training for a new technology. This paper, based on several years of experience using wireless technologies for dozens of projects, is focused specifically on two case studies. The first case study is a large lead removal site in the Midwest at which real-time data collection technologies were used throughout the project to collect thousands of data points. The second is the Hurricane Katrina/Rita emergency response requiring rapid data collection under extraordinary circumstances. At both sites, the use of real-time data collection technologies significantly improved the data management process which reduced overall costs and increased efficiency. These results could not have been achieved using traditional data collection procedures. The oral presentation will focus on the advantages and disadvantages of the real-time data collection technologies, lessons learned, and planning considerations. A live demonstration, following a typical data collection scenario in which data are collected and plotted on a GIS map in near real-time, also will be provided.

INTRODUCTION

During the last two decades, geographic information systems (GIS) and global positioning system (GPS) have revolutionized environmental data management. These technologies have resulted in vast improvements in data collection, data storage, visualization and analysis. Despite the widely agreed upon benefits of these technologies, many data managers struggle to remove bottlenecks in the flow of information from the point of field data acquisition to the point that data are transformed, cleaned, validated and loaded into a GIS. Many environmental data mangers continue to promote the field book as the primary means of data transcription in the field. Handheld computers such as Personal Digital Assistants (PDAs) have been available for years, but many professionals hesitate to use these tools in the field for a variety of reasons such as high costs, training requirements, durability, and the potential for data loss.

Although the risks and costs are real, data managers are beginning to realize the benefits of collecting data in electronic format. The terrorist attacks of September 11, 2001, the Space Shuttle Columbia disaster, and the anthrax scares proved that government and industry needed to invest more heavily in data collection technologies that would enable decision makers to have access to accurate environmental data more quickly. In recognition of this need, government and private industry began to reevaluate data collection practices and invested in technologies that improve the flow of information from the field to decision makers. The environmental industry has benefited from data management innovations that combine PDA devices, wireless communication, GPS, and GIS technologies. Correct application of these technologies can result in efficiency gains and costs savings for emergency response operations as well as traditional environmental site characterization and cleanup projects.

This paper discusses the application of a field data collection system that combines PDA technology, wireless internet communications, a web database and GIS technology. This system has been implemented it on hundreds of projects. Most recently, the system was used in New Orleans and the Gulf Coast to support the Environmental Protection Agency (EPA) in response to both Hurricane Katrina and Rita. This paper begins with a discussion of functional requirements and the major system components. It then offers lessons-learned and insight into data management techniques that can benefit any organization that wants to improve efficiency and reduce costs.

FUNCTIONAL REQUIREMENTS

During an emergency response, data collection begins the moment that field personnel arrive on site. Often, multiple government agencies and contractors work together under on incident commander. Data managers have daunting responsibilities from the beginning. They must coordinate data collection activities of multiple contractors and government agencies. Data managers often work around the clock, while providing reports, data, and maps to decision makers on demand. Therefore, “flexibility” is the primary requirement for any field data collection system. The system must be user-friendly and compatible with not only government mandated databases but also the information management systems used by other contractors. High-level goals and requirements include:

1. Save time and money on overall data management process;
2. Minimize logbook entries and handwritten transcription errors;
3. Maximize collection of data in electronic format;
4. Integrate PDA devices with GPS devices;
5. Use wireless technology to transmit data directly to the field to a centralized database;
6. Develop failsafe backup systems that allow data to be stored, recovered, and managed in the event that wireless connection is unavailable or an internet connection is unavailable;
7. Enable decision makers to have password protected access to a centralized web database;
8. Produce accurate GIS maps showing sample locations and associated data in near-real time;
9. Integrate with the EPA’s Scribe software;

These goals formed the basis for development of a system that consists of three major components: a Pocket PC PDA device, a desktop data management tool, and a centralized web management console. Data can be moved between each of these components using an internet connection (wireless or land line). In the event that an internet connection is not available, data can be transferred directly from the PDA device onto a laptop computer. The system provides data collection templates that comply with EPA requirements and standards. For example, it contains templates that are modeled after the EPA Scribe data set. Recognizing that data collection requirements are different on almost every project, the system does not limit data collection to a set of predefined templates. The user can create new data collection templates accounting for site specific conditions and requirements.

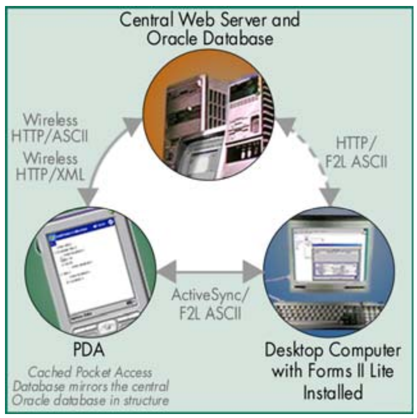

Fig. 1. Major system components

A beta version this system was field tested in 2003. During this period, field personnel were required to practice basic skills such as using a stylus pen to enter data into the PDA device, recharging batteries, transferring data from the PDA onto a laptop computer using ActiveSync software, and transferring data wirelessly by connecting a cell phone to the PDA. The "beta testing" period in 2003 resulted in number of improvements based upon user feedback. In particular, a new desktop software component was developed, enabling field personnel to manage all data on site without the need for internet connection for data transfer. In 2003, the beta version of the system was implemented on a major lead removal project in Viburnum, Kansas. This project involved multiple field teams collecting thousands of records of data over a 1-year period. Also in 2003, EPA successfully implemented the system at a large-scale emergency response exercise, which involved a fictitious satellite that impacted in the Kansas City area. The system proved to be highly effective in transferring field data to the central command post in real time.

Having extensively field-tested the beta version of the system, version 1.0 including documentation was released offices in Philadelphia, Atlanta, Houston, Chicago and Denver. These offices began to use the system on local projects after staff received around 4 hours of hands-on training.

WASHINGTON COUNTY LEAD DISTRICT

The system has been implemented at a major removal site located in the Richwoods Area Superfund Site in Washington County, Missouri. Since June of 2005, field team collected over 2,000 soil samples from 359 residential properties in and around Potosi, Cadet and Mineral Point as part of the investigation.

Staff working on the Washington County Lead project have used the system on a daily basis to collect information such as: property information (owner names, address info, etc.), occupant information (names, ages, etc.), sample location information (GPS coordinates, study area number, date, time, source area, etc.), screening data (XRF in situ results, date, time, location information, cell number, etc.), and water sample information (sample date, time, matrix, drinking water, treatment system, sample location, etc.).

Data collection requirements at the Washington County Lead project is similar to many other large-scale removal projects managed under EPA's Superfund Technical Assistance and Response Team (START) contract. These projects require contractor staff to work within tight budget constraints while striving to maintain data quality.

Cost Savings and Efficiency

The Washington County Lead project demonstrates that the combination of PDAs, wireless communications, and Web-based GIS provides field personnel and decision-makers many benefits throughout the life cycle of a project. Cost savings and efficiency gains are achieved on the "back side" of the data management process resulting in improved data quality, reduced rework, and more efficient map production and reporting. However, an investment in hardware and training places higher demands on field staff, who may initially resist the use of new technology complaining that entering data into PDA or other handheld computer device takes more time than recording data in a field book. Unless they are properly trained and informed about the benefits, field personnel might draw the conclusion that PDA devices are an impediment to their work rather than a benefit.

Although entering data into a PDA can be time consuming initially, field personnel supporting the Washington County Lead project learned to enter data as quickly or more quickly than they could previously using pen and paper. Field personnel become comfortable with the use of PDA technology after they have worked with it for a while. Data managers and GIS specialists, on other the hand, are more likely to appreciate immediately the benefits of collecting data in electronic format:

- Transcription errors are minimized. Before implementing the system, data managers manually transcribed log book records into a spreadsheet or Access database. This process was not only time consuming, but it also resulted in data loss and data corruption due to human error. At the Washington County Lead project, the system automatically compiled data from multiple PDAs into one data set. These data were then imported into an EPA Scribe database. After a few weeks of practice, data managers typically needed only minutes to complete the daily routine of transferring data from the PDAs to the Scribe data set.

- GPS coordinates are tied directly to field data. Prior to using the system, field personnel followed standard procedures for management of GPS data. Location data were recorded in electronic format, but associated attribute data were recorded separately in field books or spreadsheets. To ensure that the link between location data and attribute data would be maintained correctly, it was a common practice to write GPS coordinates on paper along with other data recorded at each sample location. At the Washington County Lead project, the new system improved the efficiency of data management by linking the Garmin GPS Map 76 directly to PDA devices. Data were transmitted from the GPS to the PDA over a connecting cable. Sample coordinates and attribute data were automatically stored in the same electronic file on the PDA. As a result, the integrity of location data was maintained with the highest possible integrity and no data loss occurred.
- Data are maintained in a centralized database accessible to off-site data manager, project managers and other stakeholders. Although Scribe served as the official end-point for all data collected at the Washington County project, Scribe was not accessible in a shared network environment. To overcome this limitation, data managers loaded data into centralized database accessible through a website. Both the EPA and the Missouri Department of Natural Resources (MDNR) were provided with secure logins accounts enabling them to monitor progress and download data for review as needed.

By utilizing the system to collect field data, data managers successfully decreased the time needed to process field data before importing it into Scribe. The PDA component of the system provides a consistent template for data entry. Input validation reduces the amount of transcription errors and greatly improves overall data integrity. Data are collected in electronic format in the field and are transferred directly from the PDA to the Scribe database at the end of each work day. Because Scribe is desktop software, it cannot be viewed over an internet connection. The data contained in it are not visible to decision makers and stakeholders, who are often located off-site. The system compensates for this limitation because data are transferred not only into Scribe, but also to the centralized web database.

Lessons Learned

- Despite the advantages of maintaining data in Scribe and also in a central web-accessible database, the maintenance of duplicate data is problematic. Edits to the Scribe data set are not automatically reflected in the central database and visa versa. Although records updated in Scribe can be exported to the central database, the process is both complex and time consuming. Experience at Washington County Lead and other sites demonstrated that complete synchronization of the duplicate data sets is not feasible.
- Although the PDA has proven to be a reliable way to collect and store data, there is an open debate about appropriateness of replacing field books with PDAs or other computing devices. Consequently, field personnel are compelled to continue writing data into their paper field books in addition to entering the same data into the PDA.

HURRICANE KATRINA/RITA NATIONAL DISASTER RESPONSE

The September 11, 2001 terror attacks and the subsequent anthrax scares motivated government and industry to invest in better training, equipment and software. In the years that followed, new data collection technologies were tested at small-scale emergency response events. The first true test of improvements in data management and communication came not from a terrorist attack, but from the forces of nature.

The most destructive hurricane in U.S. history, Hurricane Katrina, first made landfall on the south US coast in early morning hours of August 29. With peak winds at more than 175 mph, the storm caused massive damage and flooding to extensive areas of Alabama, Louisiana, and Mississippi. Katrina is now believed to have killed more than 1,300 people many of whom were swept away in the tidal surge or drowned in flood waters. Damage estimates range from $70 to $130 billion in damage. Subsequently, Hurricane Rita and Hurricane Wilma hit Texas, Louisiana and Florida compounding infrastructure damage and increasing fatalities. In the aftermath of the storm, environmental hazards still continue to pose a threat to human health. Entire neighborhoods in New Orleans and much of the Gulf coast have been reduced to piles of debris contaminated with a toxic sludge, industrial wastes, and sewage.

As soon as Katrina hit the Gulf coast, EPA mobilized its On-Scene-Coordinators (OSC) and emergency response teams. Data managers immediately began working with the EPA OSCs to develop a data management strategy. Meanwhile, other data management staff worked with the EPA Emergency Response Team (ERT) based in Edison New Jersey to configure an EPA Scribe database to store sampling and monitoring data for point locations at which field teams would discover thousands of displaced waste drums, leaking tanks, and other toxic wastes. Although cellular and landline communications were down across the Gulf coast, EPA was able to use satellite phones to continue coordination between data management experts located on site and in support office located across the country. Within less than 24 hours after the storm hit, field teams were on the ground collecting data using the system to gather data at sampling points and other locations where displaced drums, tanks, and containers were identified.

Cost Savings and Efficiency

Performance metrics for data collection in an emergency response are quite different from those of a routine site characterization or remediation project. It is difficult to weight the monetary costs of investment in technology against its ability to provide potentially life saving information to the people who need it. In an emergency response situation, reliability and speed tend to outweigh costs.

Hurricane Katrina presented EPA with data management challenges that it had never faced before. Unlike the September 11, 2001 attack or the anthrax scares, the extent of environmental disaster could not be easily delineated around specific containers, facilities, or property locations. The disaster caused massive environmental damage to an entire region. As a result, data managers scrambled to define data management needs and requirements. The system, which combines PDA devices, GPS, a central web-database and Scribe, proved to be invaluable particularly in the first two weeks of the response, successfully overcoming several of the unique data management challenges of the Katrina/Rita response:

Challenge #1: There is no time for extensive planning after the disaster strikes.
It takes time to develop a data management plan that meets the specific requirements of particular site. Yet, the process of data collection usually begins before data managers have had an opportunity to finalize their plans and configure databases to store all the relevant environmental data. Data managers were able to configure data collection templates and load those templates onto PDA devices in less than 24 hours. No programming was necessary to create data collection templates which were customized according to the data collection requirements dictated by EPA on-scene coordinators. The system was successful in large part due to the flexibility of its design, which allows an extremely high-degree of customization without sophisticated re-programming.

Challenge # 2: Multiple agencies and contractors use different platforms for data management.
In a multi-region emergency response, it is unrealistic to assume that one database management and GIS system will be shared between multiple agencies and contractors. While data centralization and coordination is always the goal, it simply cannot be achieved overnight. Therefore, data management tools that are used in emergency response must be compatible with an array of data management systems that are being used by other agencies and contractors. It must be possible to move data from one system to another with minimal hassle without data loss or corruption. The system overcame this challenge due to its compatibility with existing EPA software such as Scribe and Forms II Lite and its ability to export data in multiple formats including XML, MS Excel, MS Access, and comma-delimited text. As a result, field personnel were able to use the system to begin collecting data at the Katrina response without concern over how that data would be transferred into other systems.

Initially, EPA chose to use a Scribe database to manage monitoring and sampling information collected by field teams deployed in EPA region 6 (New Orleans and the Gulf Cost in Louisiana) and Mississippi (EPA region 4). However, in the weeks and months that followed, data management needs and requirements have evolved and new data management systems have been developed to manage household wastes removed from over 30,000 homes scheduled for demolition, to track pumping of sludge from thousands of manholes, and manage data related to contamination at industrial facilities. The system has potential benefit in all of these instances simply because of its ability to export data into whatever data management system is developed for each of these environmental cleanup activities.

Challenge # 3: Decision makers need rapid, near-time access to data for operational decisions, worker safety and public information. In addition to requiring rapid processing of data to support operational and public health decisions, EPA senior management needs data to be published on the Internet in as close to real-time as possible. As occurs in almost any emergency response situation, decision makers depend heavily on data managers to provide maps and reports that allow them to rapidly assess conditions on the ground as events unfold. Field teams work long shifts mostly during daylight hours, but the data managers typically work around the clock. Data managers are faced with the daunting task of compiling all data collected during the day, loading it into a central database, and providing ready-made maps and reports on demand. Fortunately, the system provides several features that reduce the burden on data managers. Data can be loaded into a central web database through multiple avenues, increasing the probability that data can be made available online even when many communication links are unavailable.

For several weeks in early September, nearly all communication infrastructures on the Gulf cost was completely destroyed or inoperable. Therefore, satellite phones were an essential communication tool and sometimes the only way for field personnel to communicate with support staff located outside the range of 2-way radios. PDA devices can be attached to satellite phones enabling data to be transferred directly from the PDA to the central web database. Therefore, satellite communication proved to be important not only for voice communication, but also transfer of digital data files.

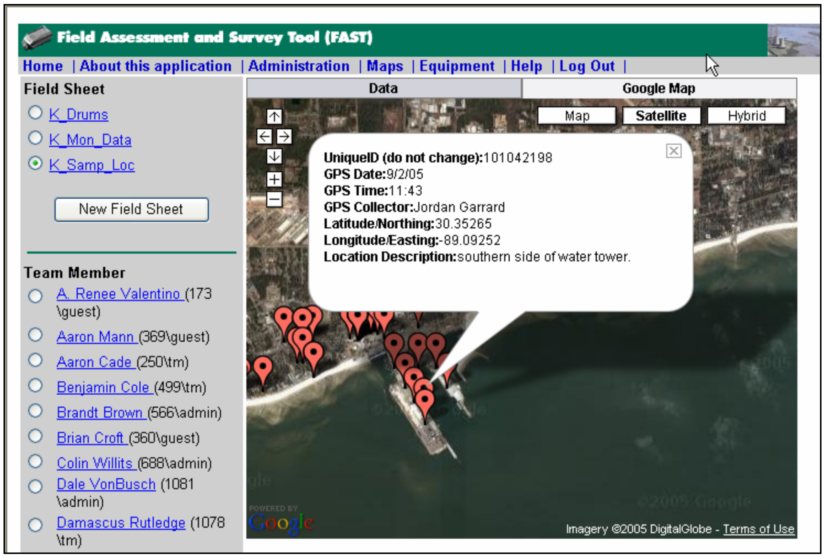

Fig. 2. Web Management Console

Challenge #4: Decision makers need to see GIS maps in near real time. Environmental disasters are inherently "geographic" in nature. The challenge of Katrina was the wide geographic scope of the disaster. GIS specialists needed to rapidly acquire base maps and satellite imagery. Making matters more complicated, most available satellite imagery was outdated and showed pre-flood conditions. Katrina dramatically altered the landscape and GIS managers needed access to imagery acquired after the storm and showing post-storm conditions. The system provided an immediate tool for presenting data on updated maps in near real-time. The web management console takes advantage of satellite imagery provide by the Google map server. Google offers a free software development kit that enables software programmers to utilize their web-based mapping tools. The advantage of this approach is that Google has already compiled detailed satellite imagery for most urban areas or other areas of high population density. Integrating with Google maps, the system displays sample locations on whatever satellite imagery Google provides. Fortunately, Google provided relatively detailed imagery for large segments of the Gulf Coast. Google imagery for New Orleans is highly detailed and had already been updated to show post-storm condition. Anyone with a password to the management console could login, view updated maps and satellite images of the Gulf coast, and review data associated with each sample point plotted on the map. In the early phases of the Katrina response, the management console enabled decision makers to begin viewing maps showing the density and distribution of environmental hazards discovered by field teams. These maps were available immediately and showed data updated on a daily basis.

As a follow up the first phase of the Hurricane Katrina response, data managers were asked to compile a list of lessons-learned. A common theme in the comments of data managers is that success depends on adequate investment in planning and training:

- Provide in-depth training to field personnel. Training should include *hands-on* use of new software, hardware, and cover basic data management practices. Failure to provide training to any member of a field team can result in a “bottleneck” that temporarily stops data flow.
- Conduct detailed requirements gathering before creating input forms to improve data quality and integrity. Although detailed requirements gathering is difficult or impossible in most emergency response situations, data collection requirements can be formulated in advance through contingency planning.
- Use “dropdowns” and default fields to improve data quality. It’s difficult to enter detailed notes into a text field on a PDA. It’s much easier and more efficient to select options from a dropdown list.
- Determine hardware needs before start of project or response. Contingency planning improves the chances that field teams will be deployed with the correct combination of equipment. For example, many responses will require PDA devices, cell phones, and GPS devices that are “intrinsically safe” or can be decontaminated properly without damaging the hardware.
- Interoperability is essential. Rapid data collection from diverse teams that are widely distributed requires robust synchronization capabilities. Federal agencies and contractors are investing in development of new systems that are intended to manage the entire data workflow. The danger is that too many competing systems are in-place, resulting in essential data being stored in different places and formats.

CONCLUSION

This paper has provided two specific examples of real-time data collection technologies used for enhanced decision making and costs savings. In the first example, the system was used on a large scale removal action in Washington County Missouri. This project demonstrates that an up-front investment in training, software, and hardware results in an overall gain in efficiency. However, it must be recognized that more time and energy is spent on the “front-end” of data management. Field personnel may find that they have additional responsibilities and require new skills for recording data in electronic format, understanding how to maintain and configure new hardware and software, and follow procedure that minimize the chance that data will become corrupted or lost. The additional time spent on electronic data collection is offset by time saved in data compilation, validation, and reporting.

The Hurricane Katrina emergency response posed data management challenges much different from those experienced at the Washington County Missouri project. Although it is more difficult to quantify costs savings in an emergency response, it is clear that wireless communications, PDA, and web technology were essential tools for provided crucial information for operational decisions, health and safety, and public communication.