Application of Six Sigma Methods for Improving the Analytical Data Management Process in the Environmental Industry

by Christopher M. French and Neno Duplancic

Abstract

Honeywell applied the rigorous and well-documented Six Sigma quality-improvement approach to the complex, highly heterogeneous, and mission-critical process of remedial site environmental data management to achieve a sea change in terms of data quality, environmental risk reduction, and overall process cost reduction. The primary focus was to apply both qualitative and quantitative Six Sigma methods to improve electronic management of analytical laboratory data generated for environmental remediation and long-term monitoring programs. The process includes electronic data delivery, data QA/QC checking, data verification, data validation, database administration, regulatory agency reporting and linkage to spatial information, and real-time geographical information systems. Results of the analysis identified that automated, centralized web-based software tools delivered through Software as a Service (SaaS) model are optimal to improve the process resulting in cost reductions, while simultaneously improving data quality and long-term data usability and perseverance. A pilot project was completed that quantified cycle time and cost improvements of 50% and 65%, respectively.

Contaminated site characterization, cleanup, and long-term monitoring and operation and maintenance processes generate large quantities of data that need to be collected, analyzed, managed, reported, and stored. In addition to traditional geotechnical testing, ground water, surface water, sediment, soil, air, and/or waste samples are chemically analyzed to assess the nature and extent of contamination.

The number of parameters for which samples are tested varies from site to site and can be as high as several hundreds. The federal Clean Water Act, an amendment to the federal Water Pollution Control Act of 1972, establishes the basic structure for regulating discharges of pollutants into the water of the United States. Section 307 defines a list of 126 priority pollutants for which the U.S. EPA must establish ambient water-quality criteria and effluent limitations. The EPA's list of Priority Chemicals (PC) under the Waste Minimization program, for example, contains 31 parameters. The list of PCs includes 28 organic chemicals and three metals and their compounds. Another common grouping, the Appendix IX Ground Water Monitoring List cited in 40 CFR 264, contains 228 metals and organic compounds. The U.S. EPA and other government agencies have had little time to ascertain which of the thousands of manufactured chemicals and their by-products pose the greatest risk to public health and the environment. As toxicological and epidemiological studies continue and detection instrumentation improves, it seems a foregone conclusion that the number of lists and/or the number of chemicals on each list will increase in the coming years.

These processes have led to an explosion in the amount of data that engineers and others must manage and evaluate. This is where the Six Sigma process, when applied correctly, can make a difference. Six Sigma is a data-driven management system with near-perfect performance objectives that has been used to acclaim at leading corporations like Honeywell, General Electric, ABB, Black & Decker, Dow Chemical, Dupont, Federal Express, Johnson & Johnson, Kodak, Motorola, SONY, Toshiba, and Amazon. This study describes what Six Sigma can do for the complex world of environmental data management at companies with a large portfolio of contaminated sites.

The name Six Sigma is derived from the eye-catching statistical target of operating with no more than 3.4 defects per 1 million chances. This principle can be applied in businesses of environmental data management to routinely improve data quality, decrease reliance on multiple consultants, achieve other positive results, and ultimately reduce the cost of environmental liability. The study demonstrates that, with certain modifications, Six Sigma could be a very
valuable tool for companies to improve their environmental data management.

The well-publicized objective of Six Sigma is to achieve practically perfect quality of performance, and this is indeed an ambitious objective in the complex world of environmental data management. Every major decision relative to cleanup and monitoring of contaminated sites hinges on quality of analytical and other data. The opportunity to degrade the quality of that data is significant in every step of the process—from the field sample collection to final regulatory agency quarterly report submittal.

Figure 1 presents a more detailed chart of the entire data flow that might occur in a remediation project. The goal of the chart is to describe the data management steps needed to properly administer an environmental remediation project from cradle to grave. The data management flow diagram depicts activities, as well as items being transferred between participants. The process is quite complex, showing a series of interdependent functions spanning event planning, laboratory analysis, analytical, geotechnical, and geospatial data management, verification and validation of results, database administration, data processing, culminating in the presentation of a formal document to the regulatory agency. The highlighted portion of Figure 1 provides an approximate delineation of the detailed Six Sigma analysis that is the subject of this paper.

The primary focus of the study was to examine the application of Six Sigma methods to improve electronic management of analytical laboratory data generated for environmental remediation and long-term monitoring programs. The examination is unique in that no prior body of work has applied Six Sigma methods to analytical data. Proper management of analytical data is mission critical over the protracted life cycle of site cleanup. The process, including electronic data delivery, data checking, verification, validation, database administration, regulatory agency reporting, and linkage to spatial information by means of Geographic Information Systems for purposes of interpretation and reporting, can be a very complex and costly undertaking, with many different players involved.

In the past days of environmental data management software — that is, before “Software as a Service” (SaaS) arrived on the scene to save the day — getting a big piece of new software running inside a company had numerous pitfalls and pain points. Consultants rang up big bills, and programs from different silo systems or vendors did not always work well together. In the end, the implementation often cost more and delivered less than anyone expected.

SaaS is a new way of delivering software as a service through the Internet. In traditional software, if a company wants to run, for instance, an accounting package, they have to first license the software, then buy the necessary hardware, then have their IT department get everything up and running, and hope it will work. The process can take months and cost thousands of dollars, and delays are common. With SaaS, though, companies use a Web browser and the Internet to connect to the software, which the

---

**Figure 1.** Environmental data management flow diagram. (Drawing courtesy of Parsons Corporation)
providers maintain for the company on their computers. The cost of SaaS can be 10% to 20% what the traditional approach costs, at least initially.

The poster child for the environmental industry SaaS is the program used in this study called the Environmental Information Management (EIM) system developed as a part of the environmental web portal LocusFocus™ by Locus Technologies from Silicon Valley, California. EIM™ has controlling market share of analytical data management in the United States. The company markets its approach with a totally automated approach for analytical data management (see Figure 2) and emphasizes ease of use: by opening a browser one can right away start tracking their analytical data, upload laboratory Electronic Data Deliverables, prepare output reports, and many more things.

Both qualitative and quantitative Six Sigma tools have been applied in the study. Qualitative tools assist in defining process variability, resultant process defects, and the design of controls to address the defective process. These include process mapping, benchmarking exercises, and the performance of stakeholder surveys to gather baseline qualitative and quantitative information about existing procedures. Quantitative Six Sigma tools measure how effectively defects and variations are eliminated from the process. The primary quantitative tools used are cause-and-effect matrices, failure mode and effect analyses, and a cost analysis.

Results reveal that automated, centralized, web-based tools are optimal to improve the process of analytical data management and store both primary chemistry results and analytical quality control (QC) metadata. These systems are also more likely to use procedures that check data integrity and QC relationships. This is significant because preservation of QC metadata is an essential element of long-term risk management.

Figure 2 presents a graph of normalized cost (cost per record) as a function of log (N1999-2001) for manual, hybrid, and automated analytical data management systems for the period of analysis from 1999 to 2001. Costs for automated data entry systems are juxtaposed against those for manual and hybrid data entry systems. Both cost functions appear to exhibit decay according to a power function. As the number of data records in the project database increases, the normalized cost per record decreases. The hybrid/manual cost curve presents a steep decay function. On the other hand, the slope of the cost decay function for automated systems is comparatively low, and the average value for data management is lower. Per-unit costs appear to converge at slightly more than 10,000 records. However, automated systems are more likely to store both the primary chemistry and QC data and provide greater functionality. The cost curves do not reflect the systemic costs associated with maintaining “silo” custom-designed or proprietary data management systems or the costs to access data through the consultant.

On a final note, with respect to cost, 80% of vendors indicate that laboratory invoicing is managed through the corporate laboratory contract management program. As currently implemented, the laboratory contract management program provides no mechanism for competitive spot bidding. Anecdotal evidence suggests that savings of 15 to 40% may be realized through implementation of a competitive bidding process for analytical services.

A cost function is derived to predict normalized costs per record for data management as a function of the number of records in a database. The function allows quantification of cost inefficiencies associated with nonautomated processes and provides a means to estimate potential savings associated with implementation of a centralized data management process.

Application of Six Sigma methods provides an effective means to drive process improvements. A control plan is presented to address the root causes of process defects and reduce variability in cost, quality and cycle time, and deviation from regulatory requirements. A pilot project quantifies cycle time and cost improvements of 50% and 65%, respectively.

Biographical Sketches

Christopher M. French, R.G., C.E.G., is a remediation manager for Honeywell International Inc. In addition to his remediation management responsibilities, he directs the data management and analytical services efforts for Honeywell’s Remediation and Evaluation Services. He is a licensed engineering geologist with 20 years industrial and consulting experience in environmental assessment, remediation, and management. He received a B.A. in geology and German literature from Amherst College and an M.S. in environmental science from Rutgers University. He may be reached at Honeywell International Inc., 101 Columbia Road, Morristown, NJ 07962; (973) 455-4131; fax (973) 455-3082; chris.french@honeywell.com.

Neno N. Duplancic, P.E., has more than 25 years experience as a senior executive in application of computer science to civil and environmental services companies. He is a founder, president, and CEO of Locus Technologies in Walnut Creek, CA. As a
Carnegie-Mellon alumnus, he pioneered the application of Internet-based technologies and Software as a Service (SaaS) model to the environmental industry. Under his leadership, Locus Technologies developed industry-first web-based applications, Locus-Focus™ and EIM™, for automated environmental data management and reporting. He may be reached at Locus Technologies, 1333 North California Boulevard, Suite 350, Walnut Creek, CA 94596; (925) 906-8100; fax (925) 906-8101; duplanci@locustec.com.

NGWA Interest Group Invitation to Participate

Participate in an NGWA Interest Group and keep abreast of the latest developments in science and technology. Interest groups are focused on a variety of topics and offer an excellent opportunity to network with other professionals in your field of expertise. Interest groups are open to members from any of NGWA’s four membership divisions.

Current interest groups include:

- Developing Countries
- Geothermal Energy
- Ground Water Availability
- Ground Water Law
- Ground Water Modeling
- Horizontal Wells

- Internet Ground Water Data
- Microbial Ground Water Quality
- Regulators
- Transboundary Ground Water
- Wellhead Protection

If you have an interest in participating in any of these groups or would like to discuss the potential for a new topic, contact Sandy Masters at smasters@ngwa.org or call 800 551.7379 ext. 502 for more information. Visit www.NGWA.org to learn more about this exciting NGWA member benefit.