BIOGRAPHICAL INFORMATION

Deb Bradbury
Principal Consultant
Enspiria Solutions, Inc., an Osmose Company

Specific Responsibilities

Founding member of Enspiria Solutions, Inc., established 2003. Responsible for providing GIS subject matter expertise and functional leadership for Enspiria Solutions and its clients. Utilizes expertise to consult, gather requirements, design, develop and implement optimal software solutions. Performs data model and application design, development, and configuration for customers in the gas, electric, and water markets. Optimizes data models to create a foundation to support applications such as graphical work design and analysis, survey, and inspection applications. Major project responsibilities have included supporting the Northeast Utilities Utility Asset Information System (UAIS) Phase 2 project by gathering requirements and writing field collection and data integration specifications, and assisting with the Pilot data acceptance phase of the project.

Past Experience

Extensive and diverse domestic and international experience in GIS data conversion and migration, consulting, and software development, and systems integration. Employment background includes positions at Public Service Company of Colorado, PlanGraphics, Convergent Group, and SchlumbergerSema. Experience includes:

- Contributed significantly to the successful completion of projects at: CPFL (Brazil), Gaz de France, Northeast Utilities, Allegheny Power (Pennsylvania), Jackson Utility District (Tennessee), Auckland City Council and Metro Water (New Zealand), Ontario Hydro (Canada), CILCO (Illinois), and Alliant Utilities (Iowa).
- Key member of the team that developed Convergent Group's Solutions Workbench data model called ENOM (Energy Network Object Model)
- Part of the Public Service Company of Colorado team that received Smallworld's Fast Tracker GIS implementation award

Educational Information

B.A., Communications, University of Denver, Denver, Colorado (In progress; completed 3 years)

Data Quality Triage

Deb Bradbury
Enspiria Solutions, Inc., an Osmose Company
10475 Park Meadows Dr, Suite 200
Littleton, CO 80124

ABSTRACT

Many utilities have failed to realize the full benefits expected of their GIS and GIS-enabled applications due to the lack of quality data. The pressure to acquire better data is further increased today due to the mandate of many public utility commissions for greater accuracy in the utilities' data, for example in the data to support joint use and streetlight billing. For most utilities, particularly those that have a legacy of multiple data sources and disparate systems, performing a full field data inventory to update the GIS database is too costly and time prohibitive. Fortunately, not all data errors are created equal. Rather than approaching a field inventory from a purely statistical measurement of existing data quality, utilities can prioritize the data errors by the impact on business, the cost and time to alleviate those errors, and data integration logistics. A significant savings in data conversion time and cost has been experienced with this process of data quality triage. This paper presents a successful approach to improving data quality and using recent projects with Northeast Utilities as case studies, discusses the lessons learned.

OBJECTIVES

1. Solve data quality issues cost effectively.
2. Reduce time to benefits realization.
3. Align and integrate GIS to realize greater enterprise benefits.

INTRODUCTION

Triage

Triage is a system used by medical or emergency personnel to ration limited medical resources when the number of injured needing care exceeds the resources available to perform care so as to treat the greatest number of patients possible. The word triage comes from the French word *trier*, which means "to sort." Credit for the system of modern day triage has been attributed to Baron Dominque Jean Larrey, a famous French surgeon in Napoleon's army who invented a method to quickly evaluate and categorize the wounded in battle and then evacuate those requiring the most urgent medical attention. He utilized these practices while the battle was in progress and triaged patients with no regard to their military rank.

In a broader sense the concept of Triage is the process of assigning priority order to projects on the basis of where funds and resources can be best used or are most needed. The following key concepts of Triage have been successfully integrated into a new methodology to cost-effectively improve utility company asset data quality:

- Assessment – to judge or evaluate in context, each facility object and attribute and its impact or usefulness to related GIS applications.
- Prioritization – to rate in order of importance, value or urgency any data update.

The first step to utilizing Data Quality Triage methodology is for the utility to develop an overall Data Strategy. An initial evaluation of the impact of data errors can reveal impacts to the following:

- Accuracy of regulatory reporting
- Revenue to the company
- Productivity of operations and dispatch
- Customer response time
- Crew safety

Quality requirements also vary depending on the target applications such as:

- SCADA/DMS
- System Planning
- Transformer Load Management
- Joint-use
- Maintenance
- Inspection

For example, the following general data types can be analyzed and rated based on their importance and impact to various utility functions or applications. See Table 1 for details

Data Type	$/Customer	Priority	Function
Device/Customer connectivity	$2 – 5	Critical	- Switching order - Outage management
Full 3-phase model		High	- Outage analysis and reporting - Phase balancing analysis
Geographic Landbase		High	- Crew dispatch - Truck maps
Schematic maps	< $2 add'l	H (urban) L (rural)	- Switching planning and switching orders
Switching cabinet model	Depends on source	High	- Switching - Outage management
Conductor and equip model	< $1 add'l	Medium	- Loading and voltage analysis
Intelligent landbase	Depends on source	Medium	- Non-customer calls or customer with missing data - Reliability analysis and reporting
Secondary networks	> $5 add'l	Low	- For North American systems, customer – transformer connectivity may suffice

Table 1 – Distribution and Outage Mgmt System Data Conversion Priority Example. (Source: Hatfield et al, "Data for DMS/OMS – How Much is Enough", GITA 2000.

When this overall system view of applications is taken, their critically can be rated. Customer to Transformer Linkage may be a top priority to support an Outage Management System (OMS) and to support Transformer Load Management. Often a legacy system will only provide on average a customer connectivity of 42% of the utilities' total customers. After a first pass of conversion, this percentage may rise to 76% connectivity to the right branch of the distribution network and 31% to the right transformer. This type of analysis can reveal that the percent of accuracy is not satisfactory and a decision may be made to conduct a new Field Survey to improve the accuracy of the customer to transformer linkages.

<u>Overview of CL&P Projects</u>

Connecticut Light & Power (CL&P), a division of Northeast Utilities, first implemented a legacy GIS in production in the early 1990's which was then converted and migrated to a Smallworld GIS platform in 2000. Realizing that certain key data needed by their GIS based applications (such as Outage Management) was missing or incomplete, NU undertook the challenge to take significant steps to improve data quality.

The need to improve data quality was additionally spurred by the Connecticut Department of Public Utility Control (DPUC) streetlight billing ruling in 2001. The ruling stated that CL&P must respond within a specific time frame of four months from the date the letter was received at CL&P to a town's request for a verification or audit of their municipal streetlights. This ruling permitted CL&P to collect back payments for street lighting in under-billed areas as well as requiring the utility to reimburse towns for over-billing.

It was determined that a streetlight only inventory would need to be immediately performed for 5 towns in Connecticut who had already requested an audit. CL&P did some of their own audits in 2001 and early 2002 and found that they did not have the manpower to verify the data for their entire system in the timeframe they wanted, so they bid this work out.

This streetlight data for the 5 towns was collected in tabular form and delivered to CL&P in an Access database format. The project was originally estimated to be complete in 20 months and was actually completed ahead of schedule in 15 months.

NU evolved this project to include contracting for a full field inventory and data collection to update their GIS. They also faced a significant challenge to integrate the new information seamlessly into their operational environment without disrupting on going work. To achieve this goal, the following factors had to be considered:

- Physical collection of field data in a fast, comprehensive, and precise manner

- Creating a systematic process to perform quality assurance and acceptance of the newly collected data
- Updating the utility's asset systems with the verified data without compromising existing operations

Starting in 2003, the CL&P division of Northeast Utilities (NU) began the system wide update of field information. NU decided to split the project into two phases; the primary goal of the first phase was to collect streetlight information, and the second phase was to perform a full system inventory of overhead and underground primary and secondary electric distribution. In addition physical connection of the customer and physical location of secondary conductors and the serving transformer bank, commonly referred to as customer-to-transformer physical "connectivity," was required in Phase 2.

The Phase 1 effort involved the field inventory and data collection of all of CL&P's streetlights, over 700,000 pole locations. Survey data, collected by Osmose, a subcontractor, was processed into a format that supported DPUC reporting and town audit requests. The data repository was a standalone Smallworld GIS that positioned the data to be migrated to the client's central GIS repository. This project required a pilot effort to prove the collection methodology and ability to achieve high accuracies in data collection and delivery. Ultimately, almost two hundred deliveries were processed and delivered over the fifteen month period.

To complete Phase 1 of the project the following three key tasks were performed:

1. Identifying key data affecting billing.
2. Field verify and data capture to improve accuracy.
3. Integrating collected information to take corrective billing measures.

The intent of the Phase 2 project was to conduct a full Field Inventory and Data Integration processing to inventory and verify overhead and underground distribution electric facility objects and attributes detailed in the specification document. Phase 2 of the project is in progress and will ultimately achieve the goal to enhance the accuracy and completeness of the data in NU's corporate GIS.

DATA QUALITY TRIAGE METHODOLOGY

Assessing and Prioritizing Data for Collection

Field Inventory is an excellent means to verify existing facility data, capture any changes or corrections and to collect missing data. At a high level, this process is comprised of the following steps:

- Extracting source data for verification
- Gathering necessary digital and paper source maps for reference
- Field collection and verification, comparing source to field conditions
- QA/QC of field work
- Data integration
- QA/QC of office work
- Delivery of data to staging environment where the utility performs Data Acceptance Testing
- Integration of missing and updated data to the production GIS

Rather than approaching this field inventory from a purely statistical measurement of existing data quality however, as discussed previously, utilities can prioritize the data errors by the impact on business, the cost and time to alleviate those errors, and data integration logistics. This approach is the basis of Data Quality Triage and is based on the concepts of Assessment and Prioritization.

These concepts were practically used during Phase 1 and Phase 2 of the NU projects. First, a series of initial requirement workshops were conducted that engaged the affected stakeholders in an interactive manner to detail the needed requirements and integration points.

Considerations included the fact that NU has an outage management application in production. This put greater emphasis on assuring the integrity of the primary three phase backbone paying particular attention to the electric protective control equipment. Other key influencing factors that helped workshop participants assess and prioritize what data was collected and how it was collected were to consider the cost, benefits and impact to:

- Electric outage management
- Electric circuit analysis
- Graphical work Design
- Safety issues
- Customer billing
- Business needs
- Data reconciliation
- Data integration

Additionally the quality of available sources of paper maps, digital references and the actual GIS data had to be considered.

The following provide two examples of electric facilities that were determined during the workshops to be of value to field verify and what information was collected: Streetlight and Capacitor Bank.

Streetlight

The following attributes of a streetlight were collected to update the database to support accurate streetlight billing.

- Type
- Lamp size
- Number of Lights
- Fixture type
- Owner
- Pole number and location
- Pole description

It's important to note that not all the possible attributes of a streetlight were targeted to be field collected or updated; only those that helped drive the billing rates. Attributes such as Mast Arm Type and Mast Arm Length were not collected.

Capacitor Bank

The following attributes of a Capacitor Bank are being collected to update the database to accurate electric circuit analysis:

- Company Number
- Phase
- No of Cans Per Phase
- Fuse Size
- Location (of the Capacitor Bank)

Additional attributes of the Capacitor Bank such as Connection Arrangement and Normal State were not deemed to be worthy of field collection using the Data Quality Triage methodology.

It is interesting to note that some data that is being field verified for Capacitor Banks will not actually end up being stored in the database. For example, Capacitor Banks that do not have Cutouts protecting them, which is an obsolete construction standard, will only be reported to NU. Additionally, Capacitor Banks with Cutouts that are open will also only be reported. Since open Cutouts in this condition do not cause an outage, NU can only find out about this situation from a field survey.

Specifications and Procedures to Assure Project Success

A set of data quality triage specifications and procedures were developed, based on the assessment and prioritization of the data. They described in thorough detail how the data would be pre-processed, collected, post-processed, integrated, and accepted. The Solutions team then went through a series of validations with stakeholders and refinements to the specifications before software and processes were created.

The following suite of four deliverable documents was developed, and accepted by NU, to assure project success. to define these specifications and procedures:

- Requirements Specification – documented the system requirements and target data model and environment.
- Data Scrub & Source Control Procedures – documented the scrub procedures and delivery process of source documents.
- Field Inventory and Data Integration Specifications – documented the scope and detailed object and attribution collection and integration rules to integrate the updated data to the production GIS.
- Data Delivery and Acceptance Procedures – documented the QC and QA methodologies used for both the field inventory and data integration processes.

When priorities are properly established and the logistics are carefully planned, an aggressive field collection and integration schedule is feasible.

RESULTS

A successful approach to improving utility asset data quality has been developed utilizing the inherent assessment and prioritization concepts of medical Triage. This methodology has been proven to achieve the following three objectives.

1. Solve data quality issues cost effectively. When the proper analysis is performed using Data Quality Triage, the optimum solution can be developed to solve the quality issue efficiently.
2. Reduce time to benefits realization. Data Quality Triage definitely reduces the time to benefits realization by creating an effective and efficient system to correct data quality issues.

3. Align and integrate GIS to realize greater enterprise benefits. This is an expected result of improving data quality. There are intangible benefits as well such as user confidence in the system which increases the GIS system use.

In summary, utilities have often failed to realize the full benefits expected of their GIS due to the lack of quality data. Since the cost of a full field inventory to collect and correct data is prohibitive for most utilities, a successful approach to field collection of data quality triage has been developed to address data affecting key business processes. A significant savings of conversion time and cost can be realized using this methodology.

REFERENCES

The following Enspiria Solutions, Inc., SchlumbergerSema, and Osmose Utility Services, Inc. proprietary and confidential documents prepared for Northeast Utilities were utilized in the preparation of this paper:

- NU UAIS Group 0 Schedule A Scope of Work
- NU UAIS Phase 2 Statement of Work
- NU UAIS Group 1 & Group 2 Requirements Specification
- NU UAIS Group 1 & Group 2 Data Scrub & Source Control Procedures
- NU UAIS Group 1 & Group 2 Field Inventory and Data Integration Specifications
- NU UAIS Group 1 & Group 2 Data Delivery and Acceptance Procedures
- “Data for DMS/OMS – How Much is Enough”, GITA 2000, Hatfield et al

Additional source information regarding medical triage can be found at:

http://en.wikipedia.org/wiki/Triage

http://www.emsmagazine.com/articles/emsarts/triage.html