QUALITY CONTROL FOR ELECTRICAL UTILITY DATA CONVERSION

Kelly Jean Fergusson, Ph.D., P.E.
Consulting Civil Engineer
224 Fernando Avenue
Palo Alto, CA 94306
(415) 855-9770
email: kj@cive.stanford.edu

ABSTRACT

In conjunction with its basemapping and utilities conversion project, the City of Palo Alto has developed quality control methods to check digital electrical distribution data as it is delivered from the conversion vendor. The topics of error and data quality as presented in the literature are introduced, as well as a typology of six error categories including sufficiency (completeness), classification, position, attribute value, data structure, and cartographic errors. A description of the basemap creation and electrical distribution conversion procedure is provided in order to provide a context for the quality control (QC) discussion. Finally, the QC methodology itself is described, including five individual quality control checks combining automated, computer-assisted, and manual procedures.

INTRODUCTION

The City of Palo Alto, California is converting its manually-maintained utilities engineering records to digital format for use with its high accuracy GIS. The municipally-owned utility provides water, gas, wastewater, and electrical services to 55,000 residents. The utilities data conversion is being performed by a North Carolina-based vendor, which was also responsible for the basemapping effort. Palo Alto sees quality control as an integral part of the conversion process to ensure that the data delivered by its vendor meets both contractual requirements and long-term user needs.

The quality control effort for the electrical distribution data is being designed and overseen by consultants to the city, including Chip Eitzel of Geodesy, and the author. Preceding the conversion of the electrical distribution data, planimetric basemapping was substantially complete, and conversion of the water, gas, wastewater, and stormwater systems was about 50% complete. The electrical distribution system, the focus of this paper, is one of six electrically-related themes being converted.

The quality control methods for the electrical distribution system include automated, computer-assisted, and manual techniques. Workpower for the quality control checking is provided by city staff, to be supplemented by college student engineering interns as needed.

BACKGROUND LITERATURE: DATA QUALITY AND ERROR TYPOLOGIES

In order to design an effective quality control program, an understanding is needed of the types of errors that can be found in both source documents and converted digital data. Since each quality control check is designed to catch one or more particular types of errors, the severity and consequences of the different kinds of errors should be considered. Thus, a quality control program can meet the budgetary goals of both the conversion effort and the GIS project as a whole, including the application development step subsequent to data conversion.

Various authors have discussed types of data quality requirements and errors. In the automated mapping / facilities management (AM/FM) community, conference paper authors have emphasized error types and quality concerns of most relevance to conversion of existing graphic engineering documents.

Thorpe (1984) discusses absolute and relative spatial accuracy, and reviews basemapping procedures geared to avoid position errors. Kauffman (1984) discusses high relative and absolute positional accuracy as fundamental to achieving digital maps which support engineering and operations functions of gas and underground electrical utilities. He refutes the notion that high positional accuracy utilities mapping is expensive, suggesting that map content is a more influential factor in determining the cost of conversion.

Sheehan (1984) lists seven tests which information products should meet:

"1. Accuracy - Free from mistakes and errors. Whatever is displayed, is correct.
2. Completeness - All necessary elements are displayed.
3. Timeliness - The data is current, up-to-date.
4. Credibility - Different sources agree.
5. Validity - The impact of changes is predictable.
6. Convenient - The required information is easy to use.
7. Readability - The required information is easy to read."

He mentions field misidentification, transcription, and data entry as error types, and suggests using computer programs to identify blatant attribute errors.

In the domain of groundwater quality, Scott (1994) addresses attribute error. She uses statistical quality control techniques such as univariate and bivariate distribution plots, regression analysis, and trend surface analysis to identify outliers that warrant additional field verification or research. These techniques are computer-assisted, in the sense that visual information is presented in charts in desktop windows, with simultaneous display of map graphics, enabling a groundwater scientist to spot data attribute anomalies by way of visual inspection.

Buttenfield (1993) defines the Spatial Data Transfer Standard's (SDTS's) five components of data quality: Lineage, positional accuracy, attribute accuracy, logical consistency, and completeness. Detreköi (1994) adds classification as a sixth characteristic. Buttenfield emphasizes the importance of tracking these characteristics at a metadata level. However, she describes evaluation and assessment of data quality as a major impediment to its representation. In other words, we lack methods to determine if a set of data meets quality standards, assuming such standards have been set.

In another broad discussion of data quality, Wu and Buttenfield (1994) point out that the purpose of a data quality evaluation determines the appropriate evaluation methods used, and that data quality is characterized by "fitness for use." In the utilities industry, the purpose of data quality evaluation is to ensure that source documents are converted according to contractual specifications, and that the needs of the users of the source documents will continue to be met by the digital data derived from manual sources. In particular, the level of digital data quality must continue to support engineering, construction, and maintenance activities in a manner equal to or better than the support provided by traditional documents.

Fergusson and Eitzel (1995) propose a typology of six error categories, each with several subcategories. The typology arose from our work in infrastructure data conversion, but is meant to be general enough to be applied to any GIS mapping project. That paper describes each category and subcategory of error, and provides comprehensive examples in the domain of utilities data conversion. The categories followed by their subcategories are listed briefly here:

° Sufficiency (completeness, superfluousness, redundancy);

° Classification (theme, feature, and attribute);
° Position (absolute, relative);
° Attribute value (wrong value, value not in list, missing mandatory value, non-unique key value);
° Data structure (spatial data type, GDS data type, association, unnecessary vertices, misplaced vertices, feature justification)
° Cartographic representation (symbol, style, annotation placement, congestion, scale, rotation).

We also propose twelve quality control checks, each designed to address one or more of the subcategories of errors listed above. It is this typology which forms the underpinnings of the methodology proposed below.

CONVERSION METHODS

Planimetric Basemap and Cadastre

The planimetric basemap data were captured using low altitude (2100 feet) aerial photography from which curb or street edges, poles, manholes, vaults, and valve covers were stereocompiled. The manholes, vaults, and valves were not classified and painted per utility system prior to the flight, nor were they distinguished from each other during stereocompilation. The planimetric data were specified to meet the American Society for Photogrammetry and Remote Sensing's (ASPRS's) standard for absolute positional accuracy of 0.4 foot root mean square (RMS) for well-defined points.

Parcel lines were drawn by Palo Alto staff based on the coordinate geometry (COGO) of records of survey, subdivision records, and other parcel documents. Each set of records yielded "blocks" of approximately 10 to 50 precisely formed parcels. Unfortunately, the historical parcel records did not contain surveyed points which tied into the City's current surveyed network, and funding was not available to ameliorate the situation. Therefore, the blocks of parcels were placed in a best-fit manner between curbs lines. This method has proved to be quite satisfactory while simultaneously meeting budget constraints of the project. In most cases the parcel line positions are expected to closely approach the 0.4 foot RMS accuracy specification.

Electrical Distribution Conversion

Both the basemapping effort and cadastre input were substantially completed before commencing the full scale production phase of the electrical conversion. The overall objective of this conversion is to duplicate the information on the source documents, in terms of spatial representation, content and cartographic form. Maintaining the look and content of the current source documents helps cultural acceptance of the digital data by the utilities staff, and does not presume to change the format of documents that have proven useful for decades. The source documents are distribution maps at scales of 1" = 40' and 1" = 100' depending on the density of equipment in the area.

For digitizing, the property lines on the source document are registered to the COGO property lines of the GIS. The source documents show rights-of-ways and property lines, but not curb edges. Their electrical features are known to have good relative positional accuracy with respect to the property lines because traditionally they have been placed in the field and drawn on the map by measuring from the parcel lines. Of course, in the field, a "property line" is often a hedge or a fence which may or may not align with the legal description. Likewise, the property line hand drawn on the source document is also subject to positional uncertainty. Despite the subtle positional accuracy issues with the property lines, both with the curb-placed COGO as well as the hand drawn cadastre on the source documents, this is the best available feature class for registration. Registration is performed on a block face by block face basis, splitting any discrepancy between the digital and manual property lines equally along the block face. This provides a consistent method for registering both overhead and underground areas. Digitization and text input then occurs to faithfully reproduce the source document in vector format.

Poles and other utility features are not being used for registration, for reasons of consistency and because of the risk of utility feature misclassification. Using poles for registration in the overhead areas and property lines in the underground areas would create unnecessary methodological inconsistencies, and assumes confidence in the relative positional accuracy of the poles on the source document that has not been systematically validated. The issue of misclassification is described below.

Despite the fact that the positions of utility features such as manholes, vaults, and poles were captured photogrammetrically, the conversion vendor was not responsible for moving the digitized feature locations to match them. There were several contributing factors to this decision. First, the density of street features belonging to a variety of utilities increased the chances of misclassification. Secondly, the technicians performing the conversion work lack the engineering judgment to make repositioning decisions. For example, suppose a sewer manhole was misclassified as an electrical manhole and a digitized electrical line was repositioned to connect to it. Such an error obviously could have disastrous consequences in the field. The vendor is providing a digital drawing containing all the photogrammetrically captured feature positions. Once all of the City's utility systems have been converted, this theme will be viewed simultaneously with all the utility systems, and appropriate repositioning will be performed by knowledgeable and experienced city utilities staff.

Source documents are submitted to the conversion vendor in geographically coherent phases. Once the documents have been returned to the City, any changes to them are made in red ink. Thus, upon final delivery and acceptance of the digital data, the redlines are input to bring the digital data completely up to date.

A data dictionary encompassing all the electrical distribution features classes was painstaking developed by examining the source document and interviewing electrical engineering and operations personnel. Over 100 features are included, each with mandatory attributes of source and date, plus between zero and six additional attributes. Electrical distribution is a substantially more complex theme than, for example, a gas distribution system.

QUALITY CONTROL METHODOLOGY

A combination of proactive (i.e. before data delivery) and post data delivery methods are being employed to ensure the quality of the electrical distribution data.

Proactive Methods

Interpretation guide. Using 11" x 17" photocopies of representative portions of the source documents, one or more examples of each feature listed in the data dictionary are identified graphically, and appropriate attribute values are noted. The photocopies are accompanied by a document listing all the features, further notes regarding the features, and references to the page numbers in the 11" x 17" document on which corresponding examples are shown. These two documents comprise the interpretation guide, which can be used to train both in-house and the conversion vendor's personnel regarding how to read and classify features and attributes on the source document. Using the guide, personnel become familiar with the correspondence between the source document symbology and the digitally represented feature names. From a database design perspective, the exercise of compiling the guide surfaces any missing or extraneous features and attributes in the data dictionary. Creating the guide is a very time consuming process and perhaps is suited only to complex themes with which the conversion vendor has limited experience.

Prototype area. A conversion prototype is developed to test all aspects of the conversion process. Several iterations of prototype deliveries and client evaluation and feedback are to be expected. In my experience, completing a satisfactory prototype takes longer than the entire production phase of the conversion. A good prototype ensures the following:

° Completeness of the data dictionary;
° appropriateness of symbol sizes and line and character style choices;
° that the conversion methodology is producing an acceptable digital product;
° that translation tables (if used) are functioning correctly;
° that quality control procedures are functioning as expected and in-house personal are trained;
° that expectations within the utilities organization are aligned with the deliverable.

These points all contribute to a smooth conversion effort and clean, application-ready data, which in turn will make application development cleaner, more generalizable, easier, and cheaper. The data itself will be easier to maintain in the long run. During the prototype phase, project participants are under pressure to approve it and proceed with production. Both the utility owner and the conversion vendor should resist the pressure to proceed until the prototype is truly satisfactory. This benefits all parties in the long run.

Data Delivery Review Methods

Automated check. This is an automated computer check to ensure that the digital data conforms to the data dictionary specifications. The delivered data is programmatically scanned, and each feature is compared to the data dictionary's corresponding entry. We check feature names, line and character styles, existence of mandatory attributes, and for the presence of redundant features. Appropriate ranges of attribute values, or conformance with a pre-specified list could also be checked at this point. Digital error symbols are placed automatically on a digital quality control drawing. This check is particularly useful during prototype development; by the time the production phase starts, the types of errors caught by an automated check have already been systematically addressed.

Visual inspection. This is a computer assisted check to identify classification errors. The technician views all the electrical distribution features in grey on the computer screen, then one by one, highlights a single feature class in yellow. The technician's familiarity with the utility components enables him or her to quickly recognize whether features have been misclassified. Text features are more commonly misclassified than symbol and line features. Theme misclassification is not a problem in this case since only the one utility appeared on the source document. Errors are annotated digitally on the same quality control drawing as was used for the automated check described above. The markups appear in red on the plots generated for feedback to the conversion vendor.

Light table check. The purpose of this manual check is to identify relative positional accuracy errors and completeness errors. A digital plot is created with content and extent corresponding to the source document, and is overlaid on top of the source document. Both documents are placed on the light table. The parcel lines are used to register the two documents together in the same manner as that used by the conversion vendor. Any discrepancies are marked in red.

As a supplemental check, the electrical distribution data can be viewed on the computer screen simultaneously with orthophotos and other utility data for a common sense check that relative spatial relationships are reasonable. Follow-up field checking for suspicious areas is performed by in-house staff since relative position between utilities not appearing on the same source document is outside of the conversion vendor's scope of work.

Cartographic check. This manual check for completeness and cartographic clarity uses the same documents as the light table check, and is somewhat redundant with it. Placing the source and digital documents side by side, the technician ensures that all features on the source document are represented on the digital output in a cartographically acceptable manner.

Attribute check. This check combines the use of automated programs with manual checking effort. A program scans the electrical distribution theme and generates a temporary text block containing the attribute values for each non text feature. The temporary text is placed at the mid-point of linear features and the center point, or "hook" point of symbol features. Using another program available by menu pick, technicians move the temporary text block for legibility. An arrow is automatically drawn from the new text location to its original position so visual linkage between the temporary text and what it is annotating is maintained. The technician plots out this view of the data, and checks it against the source document for feature and attribute completeness, as well as correct values of attributes. Manual markups are made on these "attribute plots" when errors are found.

This set of quality control checks is weighted toward catching the most important types of errors: Sufficiency (including completeness), classification, position, and attribute values errors. Data structure errors should be evaluated thoroughly at the prototype phase, but it may not be cost effective to check for them throughout the course of the conversion effort. However, data structure errors can cause difficulties and delays in subsequent application development. Cartographic errors, though minor from a database design and application development standpoint, interfere with the core communicative purpose of data in paper map form. If severe, such errors can inhibit or prevent cultural acceptance of the data within the utilities organization, which could be a crippling or fatal blow to a young GIS.

After these checks have been completed, the digital and manual markups are returned to the conversion vendor for the corrections to be made. The markups are returned to the City upon final delivery of the data, at which time City staff checks to ensure the corrections have been incorporated.

CONCLUSION

The reader has been presented with a brief introduction to Palo Alto's conversion project and to the topics of error and data quality as presented in the literature with which I am familiar. A description of the basemap creation and electrical distribution conversion procedure was provided in order to provide a context for the quality control methodology. Finally, the quality control methodology itself was described, including five individual checks.

A quality control process is an integral part of any conversion effort. No matter how great the tolerance specifications or how simple the theme, a second look at the data before it is released for day-to-day use will enhance the credibility of the GIS as a whole in the organization. Quality control is approximately 10% of the cost of conversion or data collection, and is time and money that provides an excellent return both through ensuring that contractual specifications are met, as well as instilling confidence in and acceptance of the new data by current and future users and funders of the organization's GIS.

REFERENCES

Buttenfield, Barbara P., *"Representing data quality,"* Cartographica, Vol. 30, Numbers 2 & 3, summer / autumn 1993, pp. 1-7.

Detreköi, Ákos, *"Data quality management in GIS systems,"* Computers, Environment, and Urban Systems, Vol. 18, No. 2, 1994, pp. 81-85..

Fergusson, Kelly Jean, and Chip Eitzel, *"Quality control methodology for infrastructure data conversion: a municipal approach,"* Proceedings of the 33rd annual Urban and Regional Information Systems Association conference, Vol. 1, July 1995, pp. 383-397.

Kauffman, Russell S., *"AM/FM accuracy requirements from an electric and gas utility viewpoint,"* Automated Mapping Facilities Management Conference Proceedings, AM/FM International, August, 1984.

Scott, Lauren M., *"Identification of GIS attribute error using exploratory data analysis,"* Professional Geographer, Cambridge, MA, 46 (3), 1994, pp. 378-386.

Sheehan, David E., *"Basic map accuracy —a systematic approach,"* Automated Mapping Facilities Management Conference Proceedings, AM/FM International, August, 1984.

Thorpe, John. A., *"Land base accuracy: a look toward the future,"* Automated Mapping Facilities Management Conference Proceedings, AM/FM International, August, 1984.

Wu, C. Victor, and Barbara P. Buttenfield, *"Spatial data quality and its evaluation,"* Computers, Environment, and Urban Systems, Vol. 18, No. 3, 1994, pp. 153-165.

ACKNOWLEDGEMENTS

The author gratefully acknowledges Dave Matson, Glenn Loo, and Patrick Valath at the City of Palo for their leadership in Palo Alto's conversion effort. Chip Eitzel of Geodesy developed many of the ideas and implementations described in this paper, and has provided unflagging encouragement. Robert Zeien and Barbara Fergusson gave me the support and personal space to finish the paper during what was planned to have been special dedicated family time. Thank you all.

BIOGRAPHY

Dr. Fergusson earned three degrees from Stanford University, including an M.S. and Ph.D. in civil engineering. As a fellow with the U.S. House of Representatives Public Works and Transportation Committee, she worked on issues related to GIS and the Clean Water Act. Work experience includes facilities and construction management, as well as computer system management, training, and application development. She currently provides GIS consulting services primarily to utilities-oriented clients. Dr. Fergusson is a registered civil engineer in the state of California.