UTILITY OWNERS' APPROACHES TO CONVERSION QUALITY CONTROL

Kelly Jean Fergusson, Ph.D., P.E.
Independent Consultant
224 Fernando Ave
Palo Alto, CA 94306

ABSTRACT

Make the most of every data conversion dollar by performing cost-effective quality control (QC) in-house. Build cultural acceptance of digital data by involving engineering staff in the QC checking at appropriate steps in the conversion process, with minimum time commitments. Target your quality control efforts on the types of errors that affect the results you'll be seeking from AM/FM/GIS analyses. Ensure the conversion vendor is performing to specifications.

Most utilities entering the AM/FM/GIS arena perform conversion of their traditional engineering records to digital format. Conversion of utilities and infrastructure drawings and data occurs at the beginning of system implementation, and typically is the most expensive and critical component of creating a AM/FM/GIS system. Ironically, at this time of project initiation, utility owners are generally new to the AM/FM/GIS field and are uncertain of cost and benefit trade-offs in the conversion process. This paper gives utility owners a better understanding of the categories of data errors, their potential consequences, and structured procedures and methods to perform quality control checking of the conversion data. These principles are applicable to the conversion of water, wastewater, stormwater, gas, electrical, street lights, communications, and planimetric data.

INTRODUCTION

The consequences of errors generated during a data conversion project can range from life-threatening accidents to minor annoyances. But with limited staff time and funds, how do utility owners prioritize conversion quality control (QC) efforts and target error checking toward the types of errors that will have the most serious consequences to the organization's anticipated use of GIS?

To answer this question, I will start by introducing a typical multiple-utility conversion project to give the reader a baseline context for the rest of my remarks. I then bring up some trade-offs regarding in-house staffing levels for quality control checking. Next, I will briefly outline six types of errors in GIS graphic and attribute data. The body of the paper will contain some observations and theories regarding the priorities of various types of QC checks given varying project objectives and circumstances. I will conclude with a discussion of the importance of well thought-out and realistic conversion specifications to ensure a smooth conversion and a minimum of QC effort.

This paper assumes a basic knowledge of conversion terminology and methods. For an introduction to these, please refer to (Fergusson and Eitzel, 1995).

CONVERSION PROJECT SCENARIO

A gas and electric utility is a typical customer of a conversion project. The utility's management has decided it wants to use advancing GIS technology to help achieve more competitive facility management and better customer service. Historically, perhaps the gas company acquired the electric company, and its mylar or linen engineering maps. Both sets of maps, gas and electric, have probably continued to be manually updated in some fashion, though updating may have suffered due to periodic staff and budget constraints. Perhaps the map sets are at different scales, the gas at 1" = 40' and the electric at 1" = 100'. The gas maps may include street curbs, and the electric maps may show property lines. They may cover different but overlapping territory.

How would you approach this conversion project? A tried and workable solution would be to hire one or more vendors to perform the steps of surveying, aerial flight and photogrammetry, stereodigitizing, and board digitizing. Graphic components would be created by a vendor photogrammetrically capturing the visible curb lines, gas valve covers, and electrical appurtenances with a stereodigitizing process. Property lines could be input using coordinate geometry (COGO) techniques from records of survey and subdivision plats. Each of the map sets would then be digitized. As each map set is digitized, the operator typically categorizes the various map features so that the map becomes "layered" or "intelligent", according to a pre-determined data base schema. The operator also inputs attribute data such as pipe diameter or conductor material from reading map annotation. This is a fairly typical and straightforward conversion project.

As the digitized graphics start to come into the utility organization's hands, it is the responsibility of the utility to quality control check the data to make sure it meets with the specifications of the contract. This is equivalent to a having building inspection when you are buying a home. The inspection protects the buyer from losses due to faulty workmanship and ensures the owner gets what s/he pays for. Like a first-time home buyer, the utility owner is typically fairly naive regarding what to check for. Unlike the home buyer, the utility may need to perform the checking or inspection process itself.

QC STAFFING CONSIDERATIONS

Quality control checking of converted utilities digital data can be performed by regular in-house staff, temporary staff, a quality control consultant firm, or any combination thereof. The total staffing hours and mix the utility decides on will depend on the scope of the conversion, the competence of the conversion vendor(s), and the day-to-day job demands on in-house staff (in addition to QC checking) during the acceptance time frame. 10% to 15% of the conversion budget is a reasonable figure to budget for quality control.

Giving in-house personnel the responsibility for checking the new digital maps will build their confidence in the correctness of the maps, and they will be more inclined to use the new digital product as it becomes available rather than sticking to the old ways. However, engineers and technicians that would hopefully be assigned these quality control tasks are often too busy to take on this additional though temporary workload. As an alternative, I have had success with using local college engineering students as interns to perform certain tasks, while only a few essential tasks are performed by the more experienced permanent staff. Interns or temporary employees can use the permanent staff as a resource when questions or ambiguities arise. As another alternative, the entire quality control process can be out-sourced to a company specializing in this area. In this case you gain the advantage of an expert audit, but lose the benefits of building grass-roots confidence in the data in-house. In all cases, a person very familiar with the conversion project specifications should define and coordinate the quality control process and should be knowledgeable and experienced enough to resolve any issues that arise.

TYPES OF ERRORS

There are six different categories of error on a conversion project. Each main category has several sub-categories. Although only the briefest description of the main categories and some examples are listed here, a complete and in-depth typology of these errors is given in (Fergusson and Eitzel, 1995). The intent of this current paper is to build on the previous paper by analyzing the categories with respect to the cost-and-consequence trade-offs inherent in identifying and correcting errors of different types.

Completeness (Sufficiency)

A completeness (sufficiency) error has occurred when something should (according to the specifications) exist in the converted digital data, but is missing. For example, a feature or a piece of text annotation which exists on the source maps does not appear in the digital data.

Classification

A classification error has occurred if a feature or attribute exists in the digital data but is classified incorrectly. For example, a gas feature appears on the electrical drawing, a valve is classified as a fitting, or an installation date is classified as an abandoned date attribute.

Position

Absolute position errors exist if digital map features are not located within expected tolerances with respect to their true positions. A relative position error exists if features are mislocated with respect to each other. For example, if the source document shows a conduit running four feet from the curb, and the digital shows the same conduit running 10 feet from a curb, a relative position error is present.

Attribute Value

This type of error exists when annotation text and/or attribute values misrepresent the true value. For example, if digital attributes should (according to the specification) match attributes values listed on the source document, an attribute value error has occurred if the digital attribute for a pipe diameter is 3", but it is listed as 2" on the source.

Data Structure

Data structure errors concern how data are stored and how data elements relate to each other. Pipe connectivity, pipe directionality, and logical associations between graphics and annotation are some of the areas where data structure errors occur.

Cartographic Representation

These errors are present when graphics standards have not been followed. Ideally, the digital graphics should match the look of the source map. In this case, errors exist when conventions used in the source document for annotation placement, line patterns and thicknesses, etc. are not followed in the creation of the digital representation.

ERROR CONSEQUENCES, QC METHODS AND QC COSTS

This section analyzes the data error types by discussing their potential consequences, and makes some observations regarding the methods, benefits, and staffing levels and costs associated with identifying them.

Completeness (Sufficiency)

The completeness errors have the potential to cause the most serious consequences from a safety and risk point of view. Cutting into a gas main or service in the field because it wasn't on the map can be a life threatening accident if a spark ignites at the break. Anything we can do to reduce the number of line breaks in the gas industry is worth it. USA and careful work by well trained equipment operators provide some safeguards, but complete maps and engineering plans can also help prevent these accidents.

As another example, a valve feature that is missing from a map can make a big difference in emergency response. Valves frequently get paved over in the field. When a line break does occur, the difference between a quick, fast response in which few customers are affected, and a slow, poor response in which many customers are affected can be the ability to quickly locate that paved-over valve.

Quality control checks for completeness depend on the method used to generate the feature in digital form. For photogrammetrically ("stereo-captured") features, the check consists of comparing the x, y, and z coordinate points from an ASCII file generated from the stereo-capture process with the features in the final delivered digital data. This check is recommended particularly if all or part of the data has gone through one or more digital translations where features are prone to being dropped. This check aids in positional accuracy checking as well. It is a fairly quick check, but may require an expert user to perform it. The possible extra step of comparing the coordinate points with a high resolution digital orthophoto is not worth the effort unless the points file itself is suspect.

To check digitized features for completeness, plots of the delivered digital data are compared to a copy of the source document to determine omissions. Ideally, the boundary extents of the plots, if possible, should match the extents of the source document in order to eliminate the confusion inherent in matching a north-south oriented rectilinear grid with the street pattern grid of the source document set. By highlighting features on the copy of the source document as the corresponding features are found on the digital map, errors of omission are easily identified as being the un-highlighted features left over when the check is complete.

Some industry experts might recommend that a utility owner only digitize selected features from the source documents rather than all the features, but I disagree with this approach for four reasons. First, the process of capturing only certain features is prone to generating completeness errors. If some features are supposed to be digitized while others are not, invariably digitizers or

QC checkers will omit features or add features that do not belong. Secondly, capturing everything on the source document makes digitizing a clear and well-defined task, and it makes quality control checking for completeness a clear and well-defined task. Digitizers don't have to spend valuable time deciding what to digitize and what not to digitize. QC checkers have a much easier time identifying omissions. Thirdly, information on the source documents is valuable, or it wouldn't have been put there in the first place. This is your one and only chance to capture information on the source documents. It doesn't cost much more (if any) to digitize all of the information rather than just some of it. The fourth and final reason is cost. As serious as the consequences for completeness errors can be, if the project specification calls for all features on a source document to be converted, completeness checking of digitized data can be performed by less experienced and lower-paid staff.

Utility owners that are converting a single utility subject should be concerned primarily with ensuring that their data is complete. For utility owners of multiple utilities, errors of classification, discussed next, and positional accuracy also may have significant consequences.

Classification

Classification errors can have consequences as serious as those of completeness errors, as well as less serious implications. When two or more utility subjects are being converted from the same source document, errors occur when digitizers or document preparation personnel classify features in the wrong subject. Such an error is similar to a completeness error when the subjects are plotted separately, but could be less serious if the subjects are always plotted together on the same output document. For example, consider a source document containing both water and sewer utility graphics. If a 4" water fire service is incorrectly classified as a 4" sewer lateral, fire department personnel later using the maps could lose precious seconds or minutes locating the water source in the field. Gross errors, such as the sewer layer containing all the fire hydrants, can occur through errors in a translation process from one digital system to another.

Subject classification errors can be detected by plotting each utility subject in a different color. Experienced utility personnel will be able to quickly detect if any features are misclassified, because the color establishes the subject categorization for the brain, and looking at each subject, the QC checker can quickly distinguish if something does not belong.

The above example in which the fire service was confused with a sewer lateral can occur either through misinterpretation or because the source documents occasionally are ambiguous. If this is the case, determining the true field configuration may require a field check. However, field checks are extremely expensive compared to other quality control tasks. Save field check work for only the most risk-inherent situations. Otherwise, I suggest keeping an "Items for Field Checking" log book consisting of photocopies of the ambiguous source document graphics . So long as you keep track of these ambiguous situations, it is advisable to defer them because you need to focus all your staff time on the most important and cost-effective checks. Field check items can be checked as time and budget permit well after successful completion of the conversion project.

Classification errors also occur within a single subject. In practice, this seems to happen most frequently with text features. For example, service description text might erroneously be categorized as main pipe text. Similarly, attribute classification errors may also occur, whereby an abandoned date is confused with a installation date, or a diameter with a distance measurement. These types of classification errors can impact the smooth development of applications that are intended to help you manage and keep track of your utilities. Application developers are limited in what they can do when there are classification errors. They have to build in more routines to handle different exceptions, and more error trapping. Such code is more expensive to write, less elegant in design, and more expensive to maintain.

Certain classification errors, such as the use of an illegal name for a feature class, can be detected quickly, automatically, and inexpensively by programmatically comparing the feature names in the digital data with the list of legal feature names in the project's data dictionary. Although a specialist or an expert user will need to write the program, it can be executed by a beginning or intermediate user.

Classification is one of the most underrated and difficult tasks in conversion. The classification step generates many errors because (through map interpretation) this step must add the "intelligence" we expect from an AM/FM/GIS. In addition to the subjects-by-color and automated checks listed above, there are two other techniques that I have found helpful in reducing classification errors.

The first is an additional QC check. In this computer-assisted "visualization" check, all features in one subject are displayed on the screen. Using a menu pick, the QC checker selects one feature class at a time to display in a bright yellow color. The QC checker determines whether all the highlighted features truly fall into the selected category. For example, if fire hydrant valves are the selected feature, only fire hydrant valves should appear in yellow on the screen. If a regular main valve shows as highlighted, the checker flags it as an error. The personnel performing this check must be thoroughly versed in the utility system they are checking because the brain relies on contextual cues in this check. Therefore, utility technicians or engineers are recommended for this check rather than less experienced interns.

The final technique is to develop a source document interpretation guide at the project's specifications development stage. In conjunction with designing the project's data dictionary, examples of every type of feature and every type of attribute information are explicitly identified on copies of the source documents. This guide becomes an invaluable reference to the vendor's digitizing team to aid in classification of features and attributes. As a supplement to this approach, consider defining and including an "unknown" feature class for each of lines, map symbols, and annotation text. These "unknown" feature classes will ensure that features that cannot otherwise be classified are at least captured, thus meeting our goal of completeness.

Position

There are several consequences of position errors. In general, they reduce the usefulness of the digital basemap for engineering purposes such as design. This is a consideration for utilities that do a significant amount of in-house design. *Absolute* position error (or a too-loose absolute accuracy specification) makes it difficult to integrate GPS points with the utility graphics. Logical consistency suffers most with *relative* position errors. When the basemaps don't make sense, people loose confidence in the GIS in general.

On the positive side, high position quality of both types reduces surveying needs. And when survey and GPS measurements *are* taken, they integrate well into the digital graphics. Engineering designs are more accurate, reducing expensive field changes and construction change orders. Quality positional data enables logical consistency to be better maintained, particularly when different utilities subjects from different sources are overlaid together in digital format. This is important for studies such as risk assessment analyses of underground gas and electric utility adjacencies in seismic zones.

It also bears emphasizing that "low accuracy" has a different meaning than "positional errors". A low accuracy map may contain zero out-of-specification errors. In absolute positional accuracy, higher accuracy costs more. It costs significantly more to do planimetric mapping at high accuracy scales (approx. 0.5 to 2.0 ft. rms) than lower accuracy scales; however, advancing photogrammetric technology is dropping the price of high accuracy planimetrics.

Relative accuracy, however, can be achieved simply with careful and intelligent digitizing techniques. You are paying for the digitizing anyway; you might as well have it done in a manner that preserves relative accuracy.

A phenomenon that I have observed is that utility features generally have good relative positional accuracy with respect to curbs or property lines on source documents. This is probably because when the drafter placed the line originally on the manual drawing, s/he probably did so given fairly precise measurements between the curb and the pipe. Thus the positional accuracy of the utility relative to the curb or property line may be better than the overall inherent accuracy of the map according to a USGS- or ASPRS-style specification.

So when utilities are being digitized from more than one source, it is important to register maps using the same methodology for each set. Recall the gas and electric utility scenario introduced at the beginning of this paper. To maintain relative accuracy during digitizing, the gas maps are registered to the curbs on a curb-face by curb-face basis, and the electric maps are registered to the property lines on a block-face by block-face basis. In this manner, the data sets, when viewed together, have a better chance of mimicking the true topology in the field, and maintaining logical consistency. Registering the map once per page is not enough because of discrepancies between the curb and property lines on the map and those same lines as they appear in the planimetric or COGO data.

To check for absolute positional accuracy, use the QC check for photogrammetrically captured features described in the completeness section. Make sure the final feature locations match the ASCII coordinate positions. To check for relative position errors, lay a plot of the digital data over the source document on a light table. Using the same registration procedure as the digitizer did, use a red pencil to mark any features that do not overlay correctly for later adjustment. If some features have been moved to photogrammetrically-captured locations, these features should be shown differently than digitized features on this particular plot. For example, the captured features could be shown as filled symbols and the digitized features as hollow, by using the source attributes (digitized or photogrammetric) of the features to create a "thematic" map. This way, the QC checker can determine if discrepancies exist for a legitimate reason. This light table check can be competently performed by an intern or junior technician.

Clearly, high positional accuracy is most beneficial to engineering-oriented multiple utility owners, and when underground work is common. However, if you have a single above ground utility such as an overhead electrical distribution utility or an above-ground gas transmission line, spatial accuracy isn't nearly as important, because people can clearly see components and they tend not to get paved over!

Attribute Value

Unless attribute value errors are quite frequent and widespread, they generally have the potential to cause relatively minor information and decision problems. You could end up with slightly misleading results to database queries. Of course, these results could impact operations and maintenance decisions to varying degrees. More widespread and frequent errors have the ability to impact network modeling results in which, for example, attribute values of pipe diameter and material may be important.

Attribute errors are commonly used to measure conversion accuracy. However, this may be because they are easy to count rather than a really good measure of overall conversion process performance!

To check for attribute errors, programmatically read the attributes for each feature and create a text block containing the values. Place each text block on or near (with a leader line) the feature it describes. Compare the plot (done on a subject by subject basis) with the source document to identify discrepancies. This check can be performed by technicians and bright, careful interns.

This check can be complicated by using more than one source document for attribute data. Deriving information from more than one graphic source document is prone to human error. The digital data that results is also more difficult to check. For example, a water utility might have its detailed 1" = 40' block maps, and a 1" = 1000' distribution map, both containing diameter and material attributes. I might simply convert both sources and then do an automated comparison of the attribute discrepancies.

Data Structure

Data structure errors are often software platform-dependent. The ability to perform application development is affected strongly by data structure errors. Some types of data structure errors can be detected or even corrected automatically with software; others require more expensive manual intervention. Network connectivity is not software platform-dependent and is an especially important aspect of data structure integrity.

Cartographic Representation

Cartographic representation errors impact the legibility of output products and their initial acceptability to people accustomed to working with the source documents but not with GIS data and maps. (This can be a very large percentage of a utility organization.) From an application developer's point of view, such errors are of negligible importance, but they will certainly influence the perceived success of the conversion project on the part of people who are not intimate with it. Undeniably, these opinions can influence the future course of the AM/FM/GIS project as a whole.

Conformance to the look of the source documents is the ideal standard of performance in this area. The graphics of the digital map should match the look of the source map. Assuming conformance to the source document was specified, errors exist when source document conventions for annotation placement, line patterns and thicknesses, etc. are not followed in the creation of the digital representation. Naturally, some accommodation should be made for the reduced ability to take "artistic license" in digital graphics vs. manual graphics. In addition, it would probably not be cost effective to require the conversion vendor to avoid graphic conflicts (especially regarding annotation) between utilities graphics derived from different source documents.

PROJECT MANAGEMENT CONSIDERATIONS AND CONCLUSIONS

In the above discussion, I have attempted to demonstrate how approaches to quality control activities are dependent upon the anticipated uses of the AM/FM/GIS, and how sensible project specifications are central to achieving a level of data quality that is suitable to those uses. The better your conversion project specifications match the anticipated uses of your AM/FM/GIS, the better your chances of a successful implementation.

I have suggested that different types of errors can have different consequences for your utility organization. When developing a conversion specification with your conversion vendor, consider specifying different acceptable error rates for each type of error, depending on the importance of quality data in each of these areas to your conversion goals and to your organization as a whole.

Several examples showed that it is easy to make a conversion project more difficult without getting a reasonable pay-back for that extra work. Keep conversion methods and quality control procedures simple, understandable and consistent. Perform a pilot conversion on a small area to test that the results from your specifications and the conversion process will meet your organization's expectations before contracting out the bulk of the work.

Finally, budget sufficient permanent and temporary staff resources to do a thorough job of quality control checking. Not only will this ensure that the utility owner gets what it pays for, but it will also build acceptance and confidence in the digital data in-house which in turn will build enthusiasm and smooth the transition to use of the new data on a day-to-day basis.

ACKNOWLEDGMENTS

The author gratefully acknowledges Chip Eitzel of Geodesy in San Francisco, CA for his intellectual and practical contributions to some of the ideas and methods expressed in this paper.

BIBLIOGRAPHY

Fergusson, K. J., and C. Eitzel, *"Quality Control Methodology for Infrastructure Data Conversion: A Municipal Approach,"* proceedings of the Urban and Regional Information Systems Association (URISA) conference, San Antonio, TX, July 16-20, 1995.

McLure, G. W., *"Quality Control Data Checking,"* proceedings of the AM/FM International conference X, Snowmass, CO, July 20-23, 1987.

Montgomery, G. E., and H. C. Schuch, GIS Conversion Handbook, GIS World Inc., Fort Collins, CO, 1993.

Stefanini, J. J., *"QA/QC: Keeping Your Data Accurate,"* proceedings of the AM/FM International conference XIII, Baltimore, MD, April 23-26, 1990.

Thorpe, A., *"Keys to Successful Water Conversion,"* proceedings of the Urban and Regional Information Systems Association (URISA) conference, San Antonio, TX, July 16-20, 1995.