

PALO ALTO'S DATA CONVERSION EXPERIENCE

Kelly Jean Fergusson
Independent Consultant
168 Oak Court
Menlo Park, CA 94025
Phone (650) 327-4533
Fax: (650) 327-4523
<http://www.fergusson.com>
kj@fergusson.com

Dave Matson
GIS Coordinator
City of Palo Alto
P.O. Box 10240
Palo Alto, CA 94303
Phone: (650) 329-2296
Fax: (650) 329-2299
dave_matson@city.palo-alto.ca.us

ABSTRACT

Palo Alto's conversion objectives include replacement of traditional engineering maps, and application development and use. The city's most important application is that used to maintain the data. Data maintenance enables a lifecycle approach to information and facilities management. The conversion process utilized by Palo Alto is described in broad terms, with a more detailed discussion of data dictionary design, prototype development and quality control of the data delivered by the conversion vendor.

INTRODUCTION

Once an organization decides to use AM/FM/GIS technology, data conversion becomes an ongoing effort. Palo Alto began implementation of GIS in 1989 and has realized substantial benefits since its system's inception. Data conversion continues to be a demanding, yet rewarding component of overall AM/FM/GIS implementation.

Like an independent utility company, the City of Palo Alto delivers electrical power, gas, and water to over 28,500 commercial and residential customer accounts. In addition, it provides traditional municipal services such as stormwater and wastewater collection, pavement maintenance, and street light and traffic light operation to the city's businesses, residents, and visitors.

AM/FM/GIS technology is now an indispensable part of Palo Alto's infrastructure management efforts. Its graphic and tabular digital data are used daily for such tasks as capital improvement prioritization and planning, engineering design, drafting of construction drawings, and service to customers planning property improvements or needing general information about their property.

Like any business in an increasingly competitive environment, Palo Alto seeks to provide a high quality level of service to its residential and business customers. Being adjacent to Stanford University and located on the north end of Silicon Valley, its 57,000 residents are upper income and highly educated. Its businesses are in the booming yet volatile world of high-technology research, product development, manufacturing, and services. Both residents and businesses demand the best in utility service.

CONVERSION OBJECTIVES

The purpose of the conversion effort has been two-fold. First, the new digital maps were needed as a replacement for the manually-maintained engineering records that traditionally have been used to record the components of and changes to the city's utility systems. Secondly, Palo Alto wanted to utilize the new digital data in a wide range of computer applications to better manage its infrastructure and provide improved service.

Replacement of traditional engineering maps

The new digital maps maintain the look and feel of the source documents by using similar line styles, character styles, and annotation placement. Every item of graphics and text appearing on the source document is duplicated in the digital version. Maintaining this familiarity has helped tremendously with cultural acceptance of the new product. Paper plots can even be generated which have the same extents and orientation as the original block maps. Having complete sets of paper maps will be essential for emergency response crews to restore operations in the event of a moderate to large earthquake.

Another similarity between the source documents and their digital replacement is scale. The high relative spatial accuracy of source documents was carried over into the digital world by selecting a conversion process that included low altitude aerial photography and board digitizing. In principle, digital utilities derived from large scale source documents (i.e. small, detailed areas on each sheet), when overlaid, result in sensible spatial relationships between the mapped components. For example, a sewer line runs past, not through, an electrical vault. On the other hand, small-scale source documents (i.e. a large area on one sheet) inherently contain more spatial error, and are usually more schematic in nature. Digital utilities derived from small-scale source documents cannot be expected to overlay and portray sensible spatial relationships.

Palo Alto's combination of the high *relative* accuracy of the different utilities' source documents with the high *absolute* accuracy of the stereodigitized aerial photography *does* enable the overlay of multiple digital utilities derived from different source documents. This level of spatial accuracy allows engineers and technicians to use the digital maps as a starting point (graphics of curb and gutter, existing utilities, etc.) for the design of new project. Another factor contributing to the use of the digital maps as the basis for capital project design and construction drawings is the continuity of the digital maps.

One of main differences between the old manual records and the new digital records are that the digital maps are continuous. In the manual world, the maps were drawn on a sheet-by-sheet basis. Designing a construction project at an intersection might have required edgematching and piecing together information from as many as four different sheets. Now, with a continuous digital map, a technician or engineer pulls up a seamless intersection on-line. There is only this one coherent and definitive place to go for the spatial information.

A second major difference is that the new digital maps contain intelligent features. This means that every single feature on the source documents, including annotation, has been *categorized*. Deciding on these categories (i.e. feature classes), and storing and tracking the categories in a *data dictionary*, comprises a substantial portion of the data modeling effort. Examples of feature classes include sewer manholes, electrical vaults, transformer text, etc. Text features often need to be further classified into types of text, e.g. transformer voltage text, load text, tap text, etc. Though categorization is one of the most challenging aspects of data conversion, a well-defined data model is the foundation for all the applications of the AM/FM/GI system.

These two conversion characteristics of 1) intelligent map features, and 2) a map which is continuous, allow for an accurate inventory of the infrastructure. The seamlessness of the map means that features cannot be repeated on two adjacent map sheets as they could be on the old manually maintained map sheets. Thus, there is one digital feature on the new map for every one physical feature in the utility system in the field. Every valve, every pipe segment is accounted for once and only once. An accurate inventory of the utility system, in turn, allows its valuation based on its physical components. Such a valuation then can be used to budget or allocate the optimum level of funds for utility maintenance and operation.

Once there is one and only one graphic feature in the digital data for every one physical feature in the field, the graphics can be successfully linked to other existing relational databases of system components or customer information.

Application Development & Use

Having digital utility records is just the beginning. Employing software applications that utilize this data is another major benefit of the conversion. Palo Alto is already utilizing the converted data with its customer service application and stormwater management system. Many more applications are under development or in the planning stages for the coming months and years.

Applications require very “clean” data. During the conversion, Palo Alto used many hours of staff and consultant effort to ensure that the data it was receiving from the conversion vendor conformed to the data dictionary requirements and other specifications.

From its inception, a fundamental purpose of Palo Alto’s system was to not just *display* the city’s engineering maps, but to *maintain* them as well. Even the best quality digital map will degrade over time and eventually become worthless if a rigorous procedure is not put in place to record updates and changes. As construction projects are completed and emergency repairs are made, the new information must be incorporated into the digital maps and related tabular information in a systematic way.

Palo Alto’s approach to this challenge of data maintenance was to contract for the development of a data maintenance application called the Feature Editor. Probably the most important application at Palo Alto, Feature Editor enables authorized users to update the graphic and tabular data. Users add, modify, and delete map features in a captive, controlled, easy-to-use interface which ensures that the relational database and digital graphics remain synchronized. For example, if the user moves a manhole, Feature Editor keeps the connecting pipes connected. If the user changes the diameter of a pipe, the diameter is changed in the relational database, and the text feature annotating that pipe is also changed to reflect the new value. Data is the most expensive component of implementing a GIS so maintaining it makes clear economic sense. Using Feature Editor, the city’s large investment in data is safe from inadvertent corruption.

This data maintenance application allows Palo Alto to practice true life cycle information management. For example, suppose an analysis program highlights the need to replace a certain valve and pipe. An engineer copies the existing digital view of the project location with all other utilities shown and designs the new components. If the project involves multiple utilities, the engineers assigned to each utility share the same graphics, and can coordinate to solve any potential spatial conflicts. Once the project is approved then built, the construction drawings are marked up on the job site to reflect the as-built condition. The as-built drawings are routed back to the engineering office, and a technician uses Feature Editor to update the master graphic and tabular data. Thus, the same data is used in planning, engineering design, drafting of construction documents, and source map updating via as-builts.

Proven benefits to Palo Alto from this lifecycle approach to information management include cost savings from several different sources. At the beginning of a project, there is no longer any need to send out paper source documents to be digitized. Not only does this save the cost of digitizing, but projects can start faster and be completed faster. Secondly, the ability of multiple engineering disciplines to share data during the design phase results in better design and fewer interferences in the field. These benefits result in fewer construction change orders and thus lower project costs. At the foundation of all these benefits are conversion data that have been painstakingly created and thoroughly quality checked for conformance to a detailed data model.

CONVERSION PROCESS

Conversion is a complex, time consuming, and mind-stretching undertaking for any organization. “Fast track” approaches have a reduced chance of success. Setting the expectations of upper management ahead of time with respect to realistic schedules, costs, and staff commitments will enable smoother management of the project when difficulties inevitably arise.

An organization that is starting a conversion effort must first ask itself who will perform the conversion. Will it be done by in-house staff or by a conversion vendor? The advantages of performing the work in-

house are that in-house staff knows the data best, and they will take the most pride and care in ensuring its correctness. As a result, the final product is more likely to be error free and culturally acceptable. There are fewer “transaction costs” working in-house than working and coordinating with an external vendor. The disadvantages of in-house conversion are the challenges of resources: having sufficient space, staff time, staff expertise, and specialized equipment to undertake the effort. Conversion vendors are generally better-equipped with these resources. However conversion vendors may also have large backlogs which must be taken into account in a conversion project schedule.

Palo Alto started its data conversion effort with a small pilot project, mapping several different utility and planimetric subjects in an area of just a few city blocks. Having the vendor capture the breadth of subjects and features within each subject uncovered many challenging issues. The pilot project was a “throw-away” because, from a mapping perspective, the final result did not meet the expectations of the engineering staff. However, the pilot was successful in that it served as a basis for specifications for the ensuing conversion Request For Proposals.

Selecting a conversion vendor

Selecting a conversion vendor takes several steps. The first is to develop a Request for Proposals (RFP). So as not to reinvent the process, Palo Alto staff started by obtaining and reading RFPs for similar projects from other municipalities. Because of the complexity of its conversion needs and the different process steps anticipated (ground control survey, aerial photography, etc.), Palo Alto hired an outside consultant to write its RFP. The RFP responses were evaluated, and the several best firms were selected for interviews. Contract negotiations were undertaken with the successful bidder. Throughout this process, Palo Alto did its best to communicate its priorities to the vendors. Ideally, a vendor is selected that has done other projects as similar as possible to the proposed project.

An essential aspect of developing the RFP is defining the scope of the conversion effort. Palo Alto required application-ready data for twenty subjects based on seven sets of source maps covering the entire geographic extent of the city (27 square miles). To do this, Palo Alto needed to have a land base stereodigitized from aerial photography. To create the photography, a survey-quality ground control network needed to be established. Digital orthophotography was a by-product of the aerial photography and digital elevation model. Each of these major steps to be performed was outlined in the RFP.

Specifications and populating the data dictionary

Conversion specifications are developed as part of the contract negotiations and as part of preparation for the prototype area. They are more detailed than the preliminary specifications in the RFP, and are often written on a subject by subject basis. They describe both methodological and data model issues. The step-by-step conversion methodology addresses how planimetric features will be captured, how digitizing registration will be performed, how map details will be handled, and how source document conflicts will be resolved. The data model specifications catalogue every type of feature on the source document.

Deciding upon, storing, and tracking the categories (i.e. feature classes) in a *data dictionary* comprises a substantial portion of the data modeling effort. A general rule-of-thumb is that there is one feature class for every graphically different symbol or line style on the source document. For example, the dashed lines on an electrical distribution map may represent the feature class of ductbeam, while a solid line may represent an overhead conductor.

Categorization while digitizing is one of the most challenging and error-prone aspects of data conversion. In order to do a good job of categorization, the workers performing the digitizing must be able to read and understand the source document maps, particularly when categorizing annotation. Features need to be categorized by subject (for example, water vs. wastewater), and then by feature type (for example, main valve vs. hydrant valve). Annotation also needs to be further categorized by type. For example, the text block ‘CU’ would have a type of ‘material’, and the text block ‘4”’ would have a type of ‘diameter’. Some conversion processes rely on automated methods for categorization. The degree of success achieved by automated methods is correlated with the simplicity and clarity of the source documents.

For a project like Palo Alto's, the aerial photography and stereodigitizing will have been performed while the specifications are being developed and the data dictionary is being populated. Once aerial photography, stereodigitizing, and specifications were complete, work on the prototypes could commence.

Selecting and Checking the Prototype

The purpose of the prototype is to make necessary changes to the conversion methodology and data dictionary's contents before full-scale conversion production is undertaken. A prototype area must be completed for each set of source documents included in the scope of the conversion effort. Select a small area that has a wide variety of features. Expect at least two iterations of changes before the prototype is accepted and the utility owner gives the go-ahead for full production. Palo Alto's prototype of its digital water, gas, wastewater and storm subject, based on a single set of source documents, took two years and eight iterations before it was accepted. Prototypes cycles of electrical distribution, streetlights, and coax network were progressively fewer and faster.

When the first iteration of the prototype is received from the conversion vendor, it must be thoroughly reviewed. First, compare a paper plot of it to the source document. Is the content complete? Are the symbology, line styles, and character styles and placement as we expected? Make a detailed list of anything that needs to be changed. Next, look at the prototype on-line. Are the features categorized correctly? Do the features match their specifications in the data dictionary? Note any deviations. Check polygon closure and network connectivity. Are expectations being met? Overlay different utilities. Are relative accuracy requirements being met. Compare the online coordinates of ten to twenty planimetrically captured features with the surveyed coordinates of the same features in the field. Are absolute spatial accuracy specifications being met? The prototype also serves as a basis for negotiation and agreement with in-house stakeholders in the system.

Conversion Production and Quality Control

As each prototype is approved, the vendor can commence full-scale production for the prototype's subjects. Typically, the converted data will be delivered in large geographic segmented areas of the city on a staggered schedule. Palo Alto was divided into seven delivery areas for this purpose. The staggered schedule allowed Palo Alto staff to check the data in the first delivery area while the vendor worked on the second, check the second area while the vendor incorporated changes to the first and worked on the third, etc.

At the point when source documents are sent to the conversion vendor they must be "frozen" for two reasons. First, quality control checkers need to compare the delivered digital data to the source document *as it was when it was sent to the vendor*. Checking against an updated source document leads to confusion as to what is and what is not the vendor's responsibility. Secondly, after the maps are sent to the conversion vendor, engineering updates that used to be incorporated on the manual maps should be saved for updating the delivered digital data. Not only would it be a wasted effort to update the manual maps as they are being retired, but these legitimate changes to the infrastructure could be lost. Because of the extended duration of the conversion of its block map subjects (water, gas, waste water and storm water), Palo Alto did continue to update the manual maps. However, red ink was used for the changes drawn after the source documents had been sent to and returned from the vendor. These red-lines constitute the first changes to be made to the digital data once it has been accepted.

Quality control checking includes using manual, automated, and computer assisted methods to review the data. A detailed typology of error types and their consequences is given in [Fergusson 95] and [Fergusson 97], but is beyond the scope of the current discussion.

Manual methods include completeness, attribute, and light table checks. For the completeness check, the source document and a plot of the corresponding area of the digital data are placed side by side on a table, or on a light table. The checker makes sure that all the features on the source document appear on the digital plot, and marks up any exceptions. The attribute check is performed on a subject by subject basis,

regardless of the number of utilities on the source document. It requires data base attributes for each feature in a subject (e.g. diameter, material, gauge, etc.) to be displayed as text blocks in the digital view. These are then plotted out, again using an extent that matches the source document, and compared to it to determine any inconsistencies. Light table checks for correct spatial placement of digitized features are awkward and time consuming with little payback. Palo Alto only used light table checks during the prototype development.

Completely automated checks are used to check each feature in the digital data against the data dictionary requirements for that feature class. Mandatory attributes such as sources and date stamp must be populated for each feature. The line style and or character style specifications for each feature must be met. "Hook" points, otherwise known as rotation points, must be located within the graphical bounds of a feature. If a feature deviates from the data dictionary specifications for its feature class, an error message containing the feature class, the type of deviation and the coordinates of the offending feature is appended to a report.

Computer assisted checks include the visual inspection check and the network topology check. The visual inspection check reviews the "intelligence" that has been assigned to the map features. The user highlights one feature class at a time in yellow by picking the feature name from a menu. S/he makes sure that each feature highlighted on the screen actually belongs to the feature class that was picked. In other words, if water main valves only are highlighted, an error is noted if a hydrant valve also appears highlighted – it belongs to a separate feature class and has been misclassified. The visual inspection check is especially useful for reviewing the classification of text features. The network topology check ensures that pipe segments only break at nodes such as valves, tees, couplings, and services. (Palo Alto has generally used link-node models when designing its topology for the various utilities.)

Sometimes ambiguities arise that cannot be resolved by examining the source documents or by rules of data precedence. Because it is so expensive, Palo Alto decided not to do any field checking as part of the data conversion project. Instead, ambiguous areas are conspicuously marked with a "field check required" symbol. These problem areas, numbering about 150 to 200 for the entire city, will be field checked and/or cross checked against construction and repair orders as need arises and as funds become available.

The members of the Palo Alto conversion team are not great believers in the philosophy of an acceptable percentage of error. This is because the types of errors that generally are measured by percentage, such as attribute error, are much less important than, for example, completeness errors. Omitting a gas lateral from the digital map is unacceptable; the cost of an explosion caused by an equipment operator digging up an unmapped gas line could be thousands of times the expense of thoroughly checking the completeness of the digital data as compared to the source documents.

Palo Alto used a combination of technician staff and interns to staff its quality control effort. Technical drafting staff performed the visual inspection checks, since these require a greater knowledge of the utility systems than the other checks. Junior and senior undergraduate college students enrolled in civil or electrical engineering majors were hired as intern staff, and were able to perform all of the remaining checks. In addition to consulting personnel, Palo Alto utilized three of the more experienced interns as supervisors of the technician and other intern staff.

Data Acceptance

Data must meet the specifications. That said, the reality is that sometimes it doesn't. In order to keep the momentum of the project rolling, certain data deliveries were accepted and paid for, with the recognition that some items in the delivery were systematically incorrect. The problems were identified by the quality control personnel, but the conversion vendor did not correct them. Sometimes the Palo Alto consultant had not made the specifications crystal-clear, particularly for the earlier subjects that were converted.

Palo Alto's strategy has been that, after several iterations of pointing out a systematic problem (e.g. all water meters are rotated incorrectly by 180 degrees) and not having it corrected, we simply put the issue on a list to be dealt with on an "in-house changes" basis. Additional issues were added to the in-house changes list because the conversion personnel recognized further detailed data requirements after the

specifications were set, or simply because the “changes” required engineering judgement that did not exist in the conversion vendor’s organization. An example of such a change is when pipes on adjoining source document map sheets did not meet. Palo Alto did not want the vendor to make an arbitrary decision regarding which pipe was in the correct position and which one should be moved. Therefore, this change will need to be researched and corrected in-house.

Data was accepted when all edits had been incorporated and all other data errors and required “in-house” changes had been identified.

NEXT STEPS

Palo Alto’s data maintenance software application, Feature Editor, has rigorous requirements with respect to data structure. For example, all the pipe discontinuity corrections, as described above, need to be made before Feature Editor can manage the edits. Palo Alto has undertaken an intense effort to make such changes over the coming few months so that the Feature Editor can be fully implemented.

Feature Editor will then be used to update the digital data from the maintenance records, construction drawings, and source document red-lines that have accumulated during the conversion.

Meanwhile, the conversion effort continues. Additional subjects slated for conversion include electrical distribution circuit schematics, traffic signals, and the hardwire system.

Palo Alto is currently developing or has recently completed numerous applications that will utilize the data. Some examples include:

- FLO for managing the stormwater collection system
- the customer service application to help residents and businesses find information regarding utility service and connection, and basic parcel information
- project setup for creating a base map for an engineering design and construction project, allowing for shared information between utility engineering departments
- pavement management for inventory and prioritization of city-wide pavement conditions and planned improvements
- TREE KEEPER for inventory and maintenance of the 40,000+ city-owned trees in Palo Alto.

At the core of all of these applications is complete and clean data that is an outcome of the data conversion efforts.

CONCLUSION

Data conversion has been an ongoing effort for the past several years at Palo Alto. But the tenacity and perseverance of city personnel and the conversion vendor is now yielding substantial rewards to the city’s citizens, businesses, and staff. If only we had had all this experience and knowledge about conversion when we’d first started! It is indeed a pleasure to have a forum such as this AM/FM conference to share our case study and some of the lessons we’ve learned.

REFERENCES

Fergusson, K. J., “*Utility Owners’ Approaches to Conversion Quality Control*,” proceedings of Automated Mapping/Facilities Management (AM/FM) International Conference, March 1997.

Fergusson, K.J. and Eitzel, C., “*Quality Control 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.