

BIOGRAPHICAL INFORMATION

Gail M. Trask
Director of Software Development
Byers Engineering Company

Specific Responsibilities

Joined Byers Engineering Company in 2000. Responsible for the development, certification, and delivery of all of Byers Engineering's legacy systems, such as EWO (Engineering Work Order) and Map Viewer. Also responsible for developing customized applications that augment the legacy products or that provide additional functionality in a standalone manner.

Past Experience

Joined Cap Gemini Sogeti in 1994. Assigned to the Outside Plant Engineering Design System (OPEDS) account with BellSouth. Functioned as a systems architecture consultant during development of the system and managed the conversion of the data into the new system.

Data processing experience spanning thirty years including economics research, supervising the collection of street name and address data for the cities of New Haven and Meriden, Connecticut and fifteen surrounding townships for the Census of 1980, the automotive industry, and the computer-aided systems engineering industry.

Educational Information

B.A. – Foreign Service and International Politics, Pennsylvania State University

Professional Memberships

GITA

SILO-BUSTERS: A GEOSPATIAL SOLUTION TO SYSTEMS INTEGRATION

Gail M. Trask
Byers Engineering Company
6285 Barfield Road
Atlanta, GA 30328-4303

ABSTRACT

The legacy systems that companies deal with every day were developed over many years on platforms spanning several technology generations. As a result, it is difficult to mine data from these multiple systems into one common view. In addition, the current business environment deters major system replacement. However, the need for integration persists, and the benefits from such integration are attainable by breaking down the walls between legacy systems at the data layer and by capitalizing on the inherent relationships that geospatial data provides. Through the use of geospatial data, companies can attain increased workflow integration and collaboration among work groups without the disruption of major systems replacement. A geospatial data repository can eliminate the need for manual research, and it can provide an easy way to store inventory data gathered in the field during routine maintenance activities with GIS-based hand-held units. By breaking down the legacy systems at the data layer, the silo-busting approach offers fast payback on enhanced applications while moving towards the next generation of enterprise architecture – one step at a time.

PAPER

Everyone is familiar with these problems. Our workday lives are filled with trying to get data out of proprietary databases, trying to match data from one system to another, finding that the Network organization has its data organized by network feature, but the Billing department organizes its data by customer number, encountering reticence or downright refusal to modify the legacy systems, and no funds for new systems development.

How do we get past these problems? How do we make our organizations more efficient, our decisions based more on fact than on instinct? The silo-busting approach provides an answer.

The first step in the silo-busting approach is to identify all of the legacy systems that, ideally, need to be integrated. Ideally, there may be fifty systems that need to be integrated, but tackling all of them at the same time would be impractical. Resources to do the integration analysis aren't available, and funds to integrate all of them aren't available. So why bother to identify all of the legacy systems? The effort will result in a catalog of all of the systems that have the potential for integration, and the next step in the silo-busting approach is to pick and choose the systems from that catalog where integration will result in the highest return on investment.

The next step is to choose a geospatial database. A geospatial database is one that has special capabilities for handling spatial data types, such as Oracle® Spatial or IBM Informix® Spatial DataBlade Module. These data types support the definition of data as being points, lines, polygons, etc., via SQL that utilizes semantics familiar to DBAs.

These databases provide another advantage in addition to DBA-familiarity: they are non-proprietary. Therefore, specialized software is not required to interpret the contents of the database. Proprietary data is one of the problems mentioned at the beginning of this paper. It creates a dependency between the company purchasing the software and the software vendor, and that dependency creates a disincentive to change, thereby locking the purchaser into less flexible systems and data storage methods.

Step three involves selecting a good, stable landbase. There are a number of these available in the marketplace, such as Tele-Atlas and Geographic Data Technology (GDT), both of which provide smart centerlines, Topologically Integrated Geographic Encoding and Referencing system (TIGER®), which provides semi-smart centerlines, and Digital Orthophoto Quarter Quads (DOQQ) to name a few.

A good landbase provides a standard reference for the X and Y geographic coordinates that tie all of the rest of data being integrated together. However, commercially available landbase databases usually do not contain all of the geographic information that a company would like to see, especially geographic information that is specific to a company. For example, wire center boundaries are available commercially, but subdivisions within wire centers are not. In addition, commercially available landbase data may not be entirely current, particularly in areas of high growth, and they may not be entirely accurate. Streets may need to be added to the landbase for areas of high growth, commercially available boundaries may need to be adjusted, and other boundaries may need to be added. If multiple landbase sources are purchased, then a single projection system must be selected and the multiple sources projected to it. Similarly, conflation* of the multiple sources must be performed in order to arrive at one consistent landbase.

Armed with a catalog of systems to integrate and a geospatial database into which the systems' data can be integrated, and a good landbase, the silo-buster moves to step four, the most difficult

* Conflation is the process of combining the information from two (or more) geodata sets to make a master data set that is superior to either source data set in either spatial or attribute aspect. The objectives of conflation include increasing spatial accuracy and consistency, and updating or adding new spatial features into data sets, updating or adding more attributes that associate with the spatial features of the data sets, etc. The geodata sets in conflation are defined as maps or digital spatial files that cover the same area, describe the same information in different forms and vary in density and accuracy (Yuan and Tao, 1999).

and costly of the steps. Step four involves the identification of the legacy data to move into the geospatial database and how to geo-code it. Again, there are a number of sources ranging from facility maps (poles, cables, cross connects, conductors, fuses, transformers, pipes, valves, etc., etc.) to customer data, to trouble history data. The data may be in multiple formats, such as raster and vector files, flat files, and relational databases. The key to integration is assigning each piece of data the X and Y coordinates that position them in space.

Multiple sources demand multiple tools in order to get the data into a geospatial database. If street addresses are available, perhaps through an assignment system or through a trouble-ticket system, then the commercially available landbases include the capability to geo-code the addresses, i.e., translate them into a specific position on a specific side of a specific street. Of course, the ability for the software to do this is only as good as the data to be matched. If the commercially available landbase knows a street as “Wynwood,” but the data to be matched has the street as “Wynnwood,” it is not going to match. The geo-coding tools are of immense help in this process, but they are not guaranteed to eliminate manual matching, so when planning the geospatial database population effort, resources must be allocated to clean up what the geo-coding tools miss.

Geo-referencing is another method of capturing the needed X and Y coordinates. Indeed, geo-referencing is the process of locating an entity with real-world coordinates. If there is no other means of geo-referencing facility data, hand-held units equipped with GPS capability can capture X and Y coordinates for facilities during routine maintenance activities. If facilities are stored within a system that has inherent geographic data, such as vector files, then that data can be brought into the geospatial database although it may require re-projection to another coordinate system first.

Step five involves making data consistency and accuracy checks, such as pair connectivity for telephony applications, conductor connectivity for electric applications, and pipe segment connectivity for oil, gas, and water applications. Performing consistency and accuracy checks ensures that the new geospatial database can be used for the analysis that everyone craves: traces, forecasting, troubleshooting.

Finally, the geospatial database can be enhanced by adding data from other utilities, from the city, or from competitors. Such data can be obtained from permits or from information that cities provide. In addition, obtaining and using such data has been simplified in the wake of September 11 due to Homeland Security data sharing rules.

What do most of us have today? We have situations like the following:

Customer Name	Customer Address	DSL Subscriber
Mr. John Jones	1000 Willow Way	No
Mr. Dave Smith	1002 Willow Way	No
Ms. Sarah Smart	1004 Willow Way	Yes
Mr. & Mrs. Evans	1001 Willow Way	No
Mr. & Mrs. Owens	1003 Willow Way	No
Mr. Ben Cartwright	1005 Willow Way	Yes
Mrs. Victoria Barkley	1007 Willow Way	Yes

Table 1 DSL Subscriber List

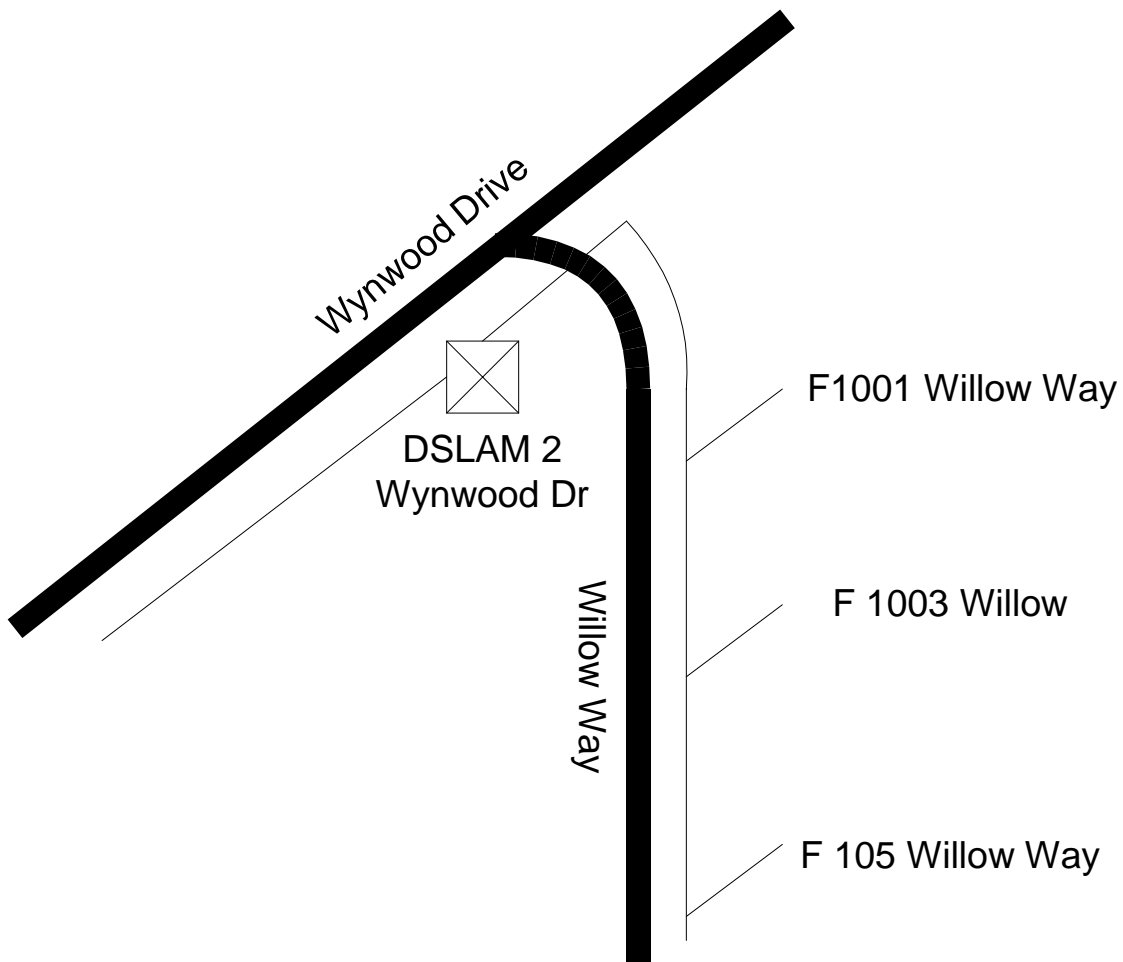


Figure 1 “Typical” network facilities map

What if we were to combine the table and the map, suppressing the distribution network?

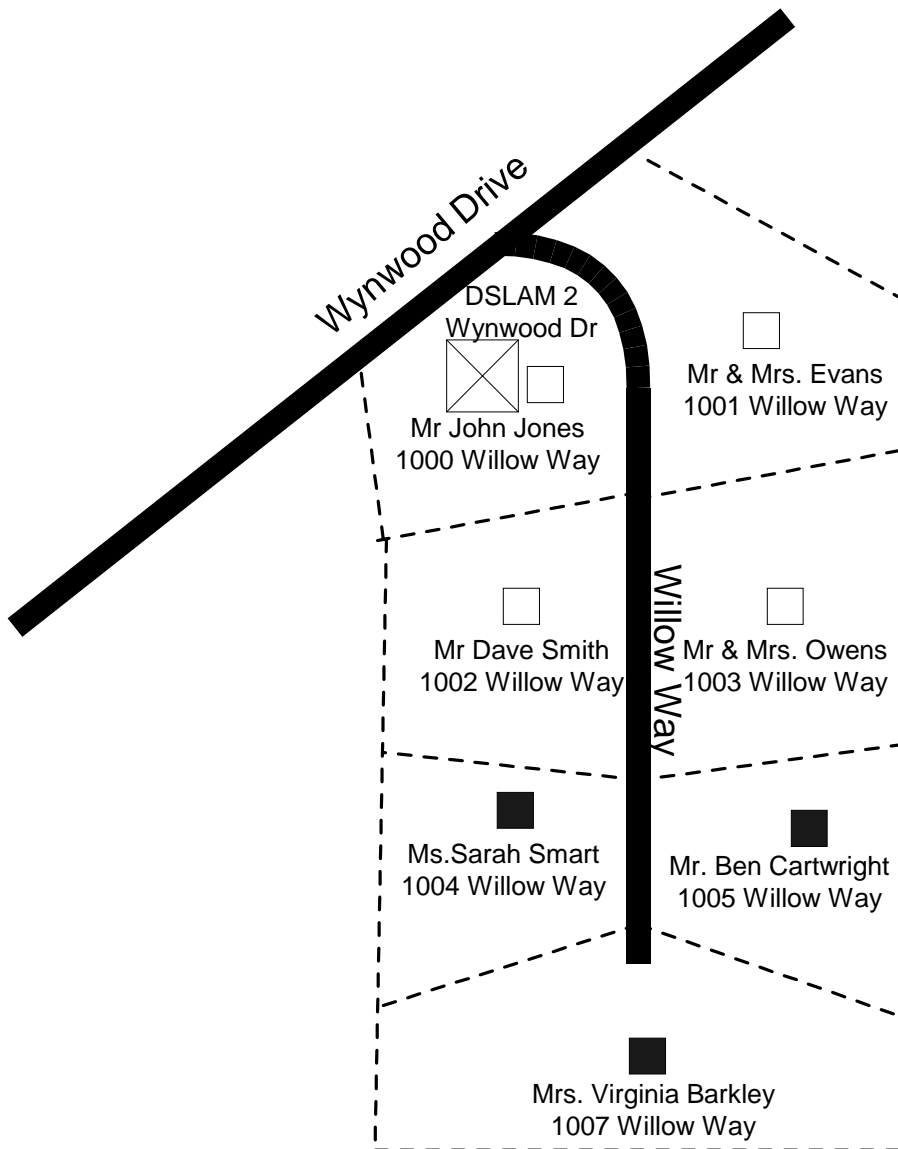


Figure 2 Mapped DSL Subscribers

Now we can see who our customers are in relation to the plant serving them, who subscribes to DSL (or any other service), and the addresses are correct. This map instantly conveys far more information than the table and the network map in isolation.

This is precisely the benefit of a geospatial database: it provides the power of place. The OpenGIS Consortium estimates that 80% of all data has a geographic reference:

- Geographic features (buildings, shopping malls, central offices, substations, pumping stations)
- Transportation (streets, railroads, bridges, highways)
- Hydrography (rivers, lakes, streams)

- Political boundaries (state, county, city, township, school district, fire district, tax district)
- Network facilities
- Customer addresses
- Trouble ticket data
- Joint use facilities (poles, conduit)
- Demographics (age, income)
- Competitor information
- Access points
- Dispatch locations
- Points of interconnection with other networks
- And more...

Integrating systems through a geospatial database had several advantages:

- It allows systems to be integrated incrementally – one at a time or several at a time.
- It allows far more analysis to be performed far more easily than on isolated, independent data stores.
- It provides the capability to view the entire facility life cycle, from forecasting, to planning, to design, to construction, to assignment, to trouble-shooting.
- It decreases the cost of building new applications because the data, once converted, is already available for use by another system.
- It allows companies to be prepared to protect their critical infrastructure, whether threatened by Mother Nature or something man-made.

Spatial data is the “great enabler.” It is visual, and its visual nature facilitates analysis, communication, and update. Its visual nature makes it easy-to-use, and that translates directly into productivity for those using the data. Greater productivity leads to a positive return on the initial investment of building the geospatial database.

BIBLIOGRAPHY

Yuan, S. and Tao, C., 1999, Development Of Conflation Components: Proceedings of GEOINFORMATICS '99, The International Conference on Geoinformatics and Socioinformatics.