DATA SUPPLY CHAIN FOR ARNOLD AND BEYOND[1]

James Meyer, Arizona Department Of Transportation

Introduction:
There has been much deliberation over the past few years about establishing and developing a supply chain for roadway data. Specifically the data need for building a comprehensive Linear Referencing System (LRS, a.k.a. All Roads Network of Linear Referenced Data: ARNOLD) and an address range roadway segments to meet the multitudes of business requirements, known and unforeseen. This paper will talk about the background and history of Arizona's LRS, important key concepts that need to be recognized, where commitments to work together need to be agreed to and supported in order to effectively support data supply chains and multi/intra organizational needs.

Background and History:
At Arizona Department of Transportation (ADOT) there has been a long history of maintaining and supporting road centerline and linear referencing system network for more than 25 plus years. The network has been supported by many different software systems and people, but it always had the goal for inclusion of all public roads. As Arizona began to rapidly grow in the 1980's, 1990's, 2000's, and continuing today the growth rate exceed ADOT's capacity to keep the network up-to-date. As a result, the network soon lost the capacity to keep all roads current in the system and ADOT was forced to focus time and resources on the roads which were owned and maintained by the State. The other roads in the State were left alone until a periodic update came from a county (Arizona has fifteen counties). When the new data came in from a local agency it was manually reviewed and washed to match its data schema, geometric qualities, and revise and format the road naming convention to match the states. After the county road data had been sufficiently cleaned up to meet the ADOT standards the painstaking effort to replace and stitch in the new road network would begin. For small relatively low road density and much slower growth counties this workflow met the needs for ADOT. However, for moderate to high growth counties this approach was much too time and resource consuming to be effective. For the moderate to large county datasets ADOT would choose to only do one or two counties every year at most and sometimes not even process a county for a couple of years. The result was the road network was becoming less and less current for the lower order roads.

In 2012 the United States passed MAP-21 Transportation bill into law. This law created the requirement for all states to report to the Federal Highway Administration (FHWA) a geospatial dataset representing all public roads for each and every state and territory. Subsequently FHWA released a memorandum to the State DOT's indicating the need to report this data. FHWA also gave this data report a name and called it the All Road Network of Linear Referenced Data (ARNOLD). With this memorandum FHWA also provided special funding provisions in the form of eliminating the State match on using federal funding. The special funding was especially key to the success of achieving a complete all public road networks because the timeline to completion was about 20 months from the time the memo was provided. This meant that ADOT could find and use resources quickly to achieve this tight deadline. The additional funding also freed up resources to accomplish this goal and codified the effort with a federal mandate to keep our all public road network up-to-date.

[1]Presented at the 54th Annual Meetings of the Canadian Transportation Research Forum, May 26 - 29, 2019 at Vancouver, British Columbia

With this new mandate and funding to accomplish the requirement ADOT started to work on a strategy to accomplish this effort. We reviewed how the process was done previously and decided that the level of effort required was to great to be sustainable so a new process was needed. We looked for the appropriate data suppliers for this all road network and determined that the Emergency Response (E911) road data was the best source to use. We came to this conclusion based on that E911's primary concern was responding to emergencies at a local level meaning they need to have up-to-date local roads. This meshed very well with the ADOT perspective of primarily focusing on the State Highway System and the off-state-system federal aid network (federal aid network refers to the Federal Functional Class System), effectively combining the top-down and bottom-up approach to roadway mapping together. Another factor in deciding on using the E911 road network was a recent development in emergency response called NextGen911. NextGen911 required emergency response call centers to use a Geospatial Data set (GIS based system) for locating where to send first responders and help route them to the proper location. This resulted in the E911 community scrambling to figure out how to accomplish the GIS based system goal for the new NextGen911 needs. It so happened that when ADOT began to outreach to the E911 community there were many positive responses to constructing a all road network for the state. There were also many not so positive responses to the required needs, but by and large the E911 community though working with ADOT would help them accomplish their missions too.

Our plans included gathering the E911 data throughout the state and develop a process to automate the geometric stitching process for network topology. The stitching process essentially connected road centerlines at points of ownership changes between networks. To automate ADOT developed the concept of demarcation points or points of agreement where two or more road centerlines would meet and be stitched together to ensure network topology resulted. The stitching process was automated by applying business rules to the demarcation points as such the rules allowed ADOT to repeat the stitching process every time the local agency would supply a new dataset to ADOT. This also meant ADOT's big effort in the beginning was the development of the demarcation points throughout the state. This effort took the better part of a year and led to having a robust process by which ADOT could stitch and re-stitch road networks from the Arizona Data Supplier.

As this new process matured ADOT started to think about ways to sustain this process. We quickly realized that we needed a way to communicate effectively about road centerlines a preserve the history of that communication for reference. To accomplish this we initiated the Phase II ARNOLD project to create a communication platform to review and comment on road centerline issues and preserve the communication. What made our approach very successful is the map-based format we use for communicating issues. The platform we developed resides in our Statewide Geospatial Clearinghouse (AZGEO) and connected to ESRI ArcGIS Online system so the areas of interest would be omnipresent in AZGEO and data supplier's ArcMap environments. The geospatial nature of the system allowed for the community to post an issue/question and based on geographic location the appropriate data owners would be notified of a problem via email and were able to then quickly review the place of the question and address it. The issues posted also are kept with all the communication threads to ensure the situation outcome is documented to prevent others from issuing a new issue ticket and capturing the outcome of the discussion.

The communications tool was well received by the local communities and has grown into the primary way by which ADOT and the local agencies communicate road centerline issues and provide suggested fixes. With the communication tool in place another phase of Data Supply Chain required development as a way to sustain the integration of local data into the statewide all public road network. To manage this workflow going forward an efficient methodology was needed to sustain the timeliness of the all road network. To initiate the next phase of this process we developed a mission statement with goals to support the data supply chain effort as seen in the next section.

Mission Statement:
To facilitate a statewide program that promotes collaboration on Roadway data, improves data quality and quantity, directly supports E911, and supports the needs of the State of Arizona.

Goals

1. Provide an easy to use transactional data update tool through AZGEO
2. Provide feedback and facilitated communication that addresses data quality
3. Build on existing workflows as to not create a new workflow and reduce the number of steps (actions) that are required now
4. Produce a statewide address data layer that is build from the combined E911 data providers for a seamless road network with addressing.
5. Collaborate with all statewide partners to align goals and objectives whenever and wherever possible.
6. Expand the State business capacity such as a Statewide Address Geocoder.

Perception:
Perceptions of a data supply chain can vary from person to person and organization to organization but it is largely based on traditional supply chain mechanics known primarily to manufacturers and distributors. However, the perception of a data supply chain can be misunderstood as being too complex to administer/manage due to its abstract nature. Other negative perceptions include; no benefits to my needs and too costly or draining on resources to be practical. To correctly understand and perceive the data supply chain principles it is better to consider this process workflow as an "Authoritative Crowd Sourced Data" whereby all contributors are data owners and are the final authority on their sourced data. This approach acknowledges the local agency authority, knowledge, and expertise with respect to the roads in each community.

Value:
A significant question asked when establishing a change or making a major shift in a business process is identifying the return on investment (ROI) or the value of doing something. Identifying what the ROI is for the data supply chain was very beneficial for gaining management support and communicating that to your business partners. In the Arizona example; the value of partnerships allows small organizations to readily share resources, data improvement by sharing and reviewing data, cross organizational standardization (establishes expectations), and reduction of costs by focusing resources on critical path activities.

Another element to be aware of, as the value increases for the data so does the complexity of it. This is a paramount reason for planning, developing systems, and committing resources to the data supply chain infrastructure. If done correctly the draw on resources (human hours and money) will decrease. There will be also be a highly visible and valuable resource created that has the potential to allow the organization(s) from having information to having insight and become more proactive in addressing issues or identifying opportunities. Essentially this concept formalizes and establishes an authoritative crowd sourced data built by the experts who own the source data.

Other less obvious value that comes from this approach is the added benefits of developing partnerships with data owners/suppliers. These partnerships if fostered can lead to greater collaboration/cooperation in this activity and other future endeavors. Working with your partners can also lead to new ideas and discover new ideas and approaches the elements of the data supply chain.

Other added values we have come to realize have been the usefulness and accuracy of the roadway data when combining regional, state, national layers. Through the combination of these road data sources the completeness and timeliness of the data has increased dramatically. Working with our local partners in the supply chain has spontaneously created a self-governing group that provides quality assurance and quality control on the roadway networks by increasing the number of new eyes on the data and combining uses of the data (i.e. Transportation Networks and Addresses). The Edge-matching process with Demarcations Points has required that network connections between road centerline data from different sources has forced agencies to better depict the extents of right-of-ways and "clean-up" where two networks touch. By combining business needs we have been able to produce a conflated and condensed output data source whereby the networks from the transportation business provide the foundation of the geometric network for address networks. This means there is a singular geometry representing the roads and has saved several local agencies from needing to maintain multiple roadway networks for the multiple business needs which saves money and time to focus on more important business needs. In working with our local partners one of our goals has been to Target for Zero Redundancy and eliminates duplicate data. To support the multiple businesses needs our effort supports modelling addresses and attributes are being maintained as GIS event layers to ensure the use of a common geometric road network. Other benefits include metrics on GIS responsiveness and provide data suppliers with offsite copy.

Organization:

To organize a structure in your data supply chain several actions or structures need to be in place. First of all in the State of Arizona we identified the primary data owners from around the state. The primary data owners (suppliers) are local E911 call centers and/or public works organization. The State DOT is also a data supplier as well as a consumer of their own data. The reasoning for utilizing the emergency responders for the data supply has to do with timeliness of data updates and a mandate to keep the data current. As an organization the E911 community relies heavily on data sharing and coordination for ensuring responders are able to find where emergencies occur, prevent false locating, and route emergency responders to save lives. The E911 organizations also have a primary focus area of the most local of roads (greatest level of access) in any given area of the state (i.e. where people live, work, and play). Conversely the state DOT has a prime focus on the higher order (greatest level of mobility) roads like Interstates, expressways, US routes, and State Routes. By combining these two focus areas the overall road/route network has a dual focus of top-down and bottom-up wrapped into one statewide system.

The second approach that was taken in Arizona involved direct communication with all potential data suppliers. A partnership was established between the State DOT, State E911 office, and the State Cartographer's office which combined goals and resources toward one purpose. With this partnership a foundation to coordinate communication about bringing the statewide centerline data together was established. In cooperation with the State E911 office and the State DOT office; representatives presented the business needs of developing a data supply chain that focused on road centerlines for ARNOLD to the stakeholders throughout the state. These presentations laid the groundwork for establishing data supply chains by communicating the need, the timing, and the vision for the cooperative efforts. It also created an informal partnering between state and local officials whereby understanding and trust are fostered.

As a result of the presentations to the local agencies and communicating the need for sharing data; these agencies started to supply their data to the State DOT. This data was used to build a statewide centerline through a process of automatic edge-matching utilizing a series of demarcation points that were assign smart behaviors(e.g. Snap to, trim, extend, etc.). Essentially the differences which are always present in schematic representations of real work situations were handled with only minor issues needing a manual solution. What we learned through this process is that a data standard was needed to minimize disparity in the data. For instance the issues that were presented with the various interpretations of dual carriageway were prevalent throughout the state. As a result, an effort was made to start trying to define dual carriageways from the aspect of public works/DOTs and emergency responders.

Other challenges to creating the supply chain were and are associated with communication between the data suppliers. We also faced organizational issues within the State DOT relating to system changes that management mandated a change of fundamental Linear Referencing System (ESRI Roads & Highways). These two challenges changed how we decided to support and organize the supply chain. After the initial effort that used the system of demarcation points to stitch together multitudes of road centerline and communicate to the local agencies in an ad hoc manner a formalized web-based data clearing house (AZGEO) was constructed to be used as the basis for future data communication, data submittal, and data reviewing tool box.

The latest incarnation of the data supply chain residing in AZGEO is in its first iteration going to be a transactional based data submittal tool which imports only the changes from the previous/existing data already available. The system also reviews the data for quality and consistency as defined in the data standards (Standards are based on NENA, Arizona DOT, and Arizona E911)[2] as it is loaded and provide feedback to the data submitter. The system allows the various data suppliers, processors, and consumers to indicate and notify data owners about potential errors or issues with any aspect of the centerline through a ticketing process.

Process/Workflow:

In the data supply chain process local E911/public works organizations maintain their centerline/route datasets natively within their county/regions. After polling the local organizations it was determined that these roadway datasets are managed not on a scheduled frequency but instead are typically updated as needed. While the local agencies manage their roadway datasets as needed the data supply chain process requires that they submit their roadway datasets to the unification process at least yearly with a preference for up to quarterly submittals. This allows for the statewide unified dataset to be as timely as possible.

Local agencies have been given access to "submit your data" tools that resides within AZGEO whereby the local agency can easily provide their data to the data supply chain process. The AZGEO submittal tools run a series of important validation and change detection processes on the submittal before ADOT is notified regarding new data being submitted. This data preprocessing strategy supports the local agencies in their submittal of correct roadway data with proper quality data and reduces the amount of employee time needed for accurate quality control for both the data supplier/local agency as well as the data receiver/ADOT.

The data submittal process on AZGEO is comprised of two major independent toolsets. These two toolsets can be run independently of each other but are meant to be run in sequence to facilitate a full data submittal. The entire set of AZGEO data submission tools is called the “AZGEO Centerline Data Validation System (CDVS).

The first of the two major toolsets is the ADOT Centerline Validation tools. This system ingests the local suppliers’ roadway data and processes it through a series of data quality assurance tests. The process includes both scalar and spatial tests and is intended to catch most of the data issues that have been identified during previous submittal iterations. Data quality assurance tests that are run include; proper field naming, proper field types, proper field values (null values, incorrect spacing, etc.), and spatial accuracy checks (overlaps, unlikely geometry, route bifurcations, etc.) This toolset also normalizes the input data sets to a standard that can be used throughout the remainder of the unification data submittal process. This normalization process is geographically dependent.

The ADOT Centerline Validation tools returns a Esri map package which included results of any data quality assurance issues found. This map package allows the data supplier to quickly ascertain and include these data quality issues into their native geospatial working environment for quick adjustment before resubmitting. When the roadway data supplied is found to have data integrity issues it is preferred that the local data supplier utilizes the data quality map package results to; adjust their master data, and re-submit to the data unification process. The toolset does internally adjust many of the data quality issues found so as to allow for the data submittal process to move forward without the need for the local data supplier to make all necessary adjustments.

The second major toolset is the ADOT Data Comparison tools. This system runs a change detection routine that compares newer (currently being submitted) centerline vintage against the older centerline vintage stored in the system. The review of the data detects any attribute (Scalar) and any geometric changes from old vintage to new vintage. The review process returns a comparison result to the data supplier that summarizes all the road name changes, geometric shape changes, and address range changes. This tool works with all of Arizona’s seventeen data providers.

Upon the completion of the data review the data supplier can review the change summary and decide if the detected changes meet expectations. If the results meet expectations then the data supplier can proceed with generating a transaction file (proposal file) representing only the changes from the previous vintage of road centerline. The proposal file is then transmitted to ADOT for review and integration into the all public road network LRS. Once the new centerline data is pulled together ADOT updates the LRS routes and produces a new address event layer. The address event layer is then published to AZGEO for the general consumption of a unified statewide address layer.

Alternatively the local data supplier can choose not to submit their data and use the provided tools to validate and check their own data as iterations, internal quality control and/or other local needs. It could also require additional edits and fixes to meet internal standards before the data can be shared.

Inside of AZGEO the change detection tools look for several very specific aspects of change between data vintages. The current list of change detection items are: *add, delete, merge, split, trim, extend, partial reshape*, *reshape*, *maintain geometry*, *road name change*, *and address change*. After the tools look at the data a "proposal" file is generated with 24 fields of data and rows representing each and every change aspect indicated.

Underway Developments:
Building on the foundation for the Data Supply Chain additional functionality is being readied for deployment. A significant new process to replace the old process of requesting changes to the Federally Functionally Classified (FFC) roads. The historical process largely involves using paper forms which the requesting agency would need to provide reasoning for the change e.g. traffic volumes and new construction. The form required a sketched map depicting the area of change and a comprehensive written justification. This process would take significant resources by the requesting agency with no guarantee of successfully making the change. Many local agencies would simply put this off until they had a significant number of requests before submitting them. These requests would then be received by the Arizona Department of Transportation (ADOT) for consideration and review with a subsequent back-and-forth the overcome discrepancies and inconsistencies. Once the ADOT review is completed for the whole state (i.e. all requests in a calendar year submitted together) would then be sent to Federal Highway Administration (FHWA) for final approval? This process required much effort on all parties and can take over a year to complete.

To overcome the challenges with this antiquated process an online change request application was built for use in AZGEO. The design of the tool leveraged some of the communication tool which was developed for communicating issues with road centerline modifications for ARNOLD. Largely the tool is built to communicate and in this case relays the changes desired by the requesting party. To make things simple the tool allows the requestor to find supporting information like the current AADTs, current Functional Classification, and adjacent road classifications. Once the selected road has been identified the requestor defines the extent of the road, requests a new classification, and then adds all the supporting documentation to justify the change with their own comments. With everything in place the requestor then submits their request electronically to ADOT where we review the request for supporting information and consistency with the FFC rules. If the request doesn't meet the necessary standards then the request is rejected and the original request or go revise or drop the request, if successful the request is passed on to FHWA for final approval where the same options for rejection or acceptance is done. Once the request has been finalized then the new FFC codes will be reflected in the master data set.

Future Developments:
During the calendar year 2019 ADOT is planning to initiate a fourth phase to this project to add more functionality and provide LRS support to other business areas. The first focus area for developing a new supply chain is with the travel demand model(s) (TDM) that the state and large MPO's need to support. This latest efforts goal is to eliminate the issues associated with having data on various roadway networks. Phase IV will do several things to remedy this issue.

First, the Travel Demand Modeler's typically use something other than an LRS enabled GIS. They also need several data items from the LRS data like traffic volumes and data related to capacity (Number of

Lanes). The current workflow requires that data moved between systems is required to undergo a geocoding or conflation process in order for the data to be used in each respective system.

Essentially this work will involve conflating the TDM to the authoritative all public road network using advanced conflation tools. By unifying the networks the movement of data from one system to the other is seamless in terms of geography. A second process to move the data will take the original business data from the LRS enabled GIS environment and run it through a processing tools/routine that sections the data according to TDM rules and produce a specific data file the pulls together the needed data items and puts it into a TDM shape file for consumption. Conversely the data in the TDM can be exported to a geometry that matches the authoritative all public road network geometry.

Other future develops will likely take the form of providing the LRS network as a service to be used by external customers especially at the county and city levels of GIS data management for Public Works organizations and E911 emergency responders. Additional work should be done to assist local agencies to reduce the number of road networks and help them move to one authoritative representation of the road network.