2010 NSDI CAP Final Report

Category 2: Framework Data Exchange through Automated Geo-Synchronization

Coordinating Local, State and National Data Stores

Framework Data Exchange through Automated Geo-Synchronization

Submitted By:

Presented To:

January 20, 2012

Cooperative Agreement Number G10AC00237

Organizations

Applicant Organization Information (Submitted on Behalf of Government Partner):
The Carbon Project / Arkansas Geographic Information Office (AGIO)
15 New England Executive Park, Burlington, MA 01803
Internet Address: http://www.thecarbonproject.com

Government Partner:
Arkansas Geographic Information Office (AGIO)
1 Capitol Mall, 2B 900 | Little Rock, AR 72201
Internet Address: http://www.gis.state.ar.us/AGIO_index.htm

Principal Investigator:
Learon Dalby, GIS Program Manager
Phone: 501-682-2929, Email: learon.dalby@gmail.com

Technical Lead:
Mark Mattson, Senior Software Engineer, The Carbon Project
Phone: 781-270-0674, Email: mmattson@thecarbonproject.com

Community Coordination Lead:
Jeff Harrison, CEO, The Carbon Project
Phone: 703-628-8655, Email: jharrison@thecarbonproject.com

Applicable Framework Themes: Transportation (Road Centerline Data) and other themes

Geographic Scope or Area: State of Arkansas and Nationwide

Executive Summary

The Carbon Project and the Arkansas Geographic Information Office (AGIO) are pleased to submit this 2010 NSDI Cooperative Agreement Program (CAP) Category 2 Interim Report for the Category 2, "Framework Data Exchange through Automated Geo-Synchronization" effort. Since kickoff in May 2010 the project developed and deployed a federated approach for exchanging local, state and national framework data using geo-synchronization services, easy-to-use shapefiles and common data models. In particular, the project has developed, tested and deployed a unique model for the NSDI that provides methods to ingest county transportation shapefile data into a geo-synchronization federation and synchronize county-level databases to state-level databases. To achieve this in a manner not disruptive to current processes, the system leverages common folder structures consisting of FTP shapefile updates from multiple Arkansas counties. The project team then developed a transformation capability to compare shapefiles to Web Feature Services (WFS) on ESRI ArcGIS Server, extract data changes and then transmit them to a GeoSynchronization Service (GSS) – and also output them as Shapefiles. The capability is based on a new change detection application called the The Carbon Project Change Validator (Figure 1). With this new capability our approach provides a bridge between current shapefile processes and a new generation of NSDI GeoSynchronization. It will also allow the state to advance data stewardship, and federal systems to automatically keep data sets up-to-date as needed. The solution is re-deployable to other sites through open APIs and free tools.

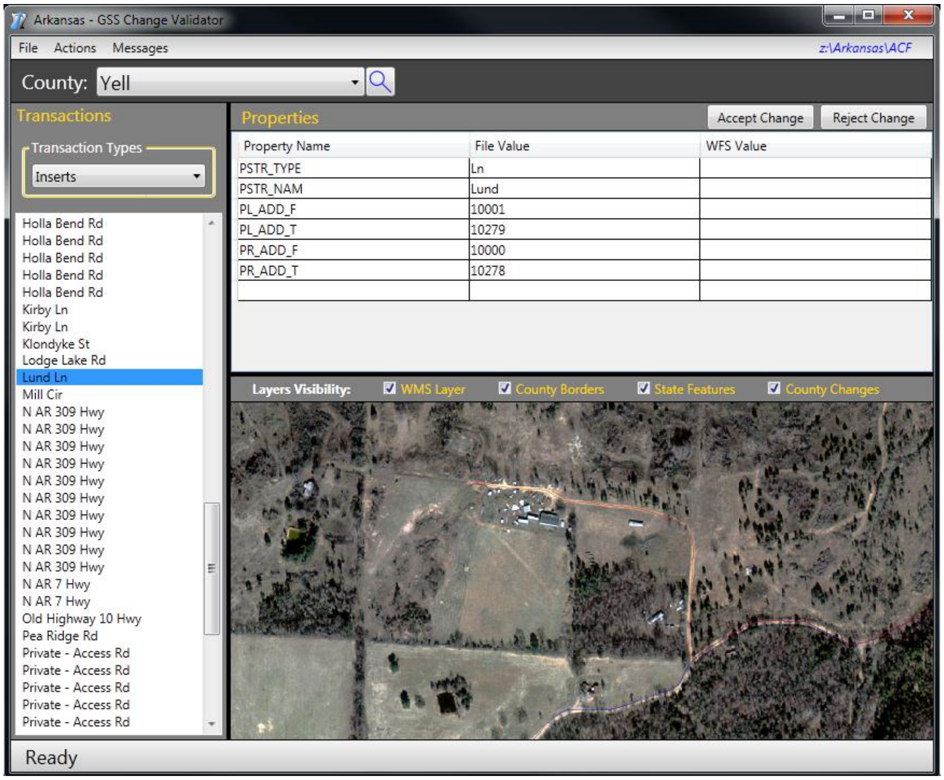


Figure 1 – This CAP effort developed a new capability, called The Carbon Project Change Validator. The Change Validator transforms county-based road updates from Shapefiles to geosynchronizable transactions - and also outputs 'change-based' shapefiles. The tool leverages OGC WFS for statewide baselining and is suitable for a federated National Spatial Data Infrastructure (NSDI) that acknowledges current production operations - and future GeoSynchronization.

Project Narrative

The goal of this project is to bridge the gap between the current needs of local and state data production organizations and the geo-synchronization vision outlined in the 2010 CAP grant. The result is a practical capability for exchanging transportation framework data between local, state and eventually national data stores through a geo-synchronization service as well as Shapefiles using common data models. To achieve this objective the Arkansas AGIO assumed the leadership role for this project and worked with The Carbon Project to deploy a geo-synchronization capability focused on easy to use ESRI Shapefiles. Deployment of the Arkansas county-to-state geo-synchronization system was in the form of a simple one-directional tree where the state's data is synchronized against all county layers, thus any operation committed against the county layers can be propagated to the state level (Figure 2) and also output as Shapefiles (Figure 3).

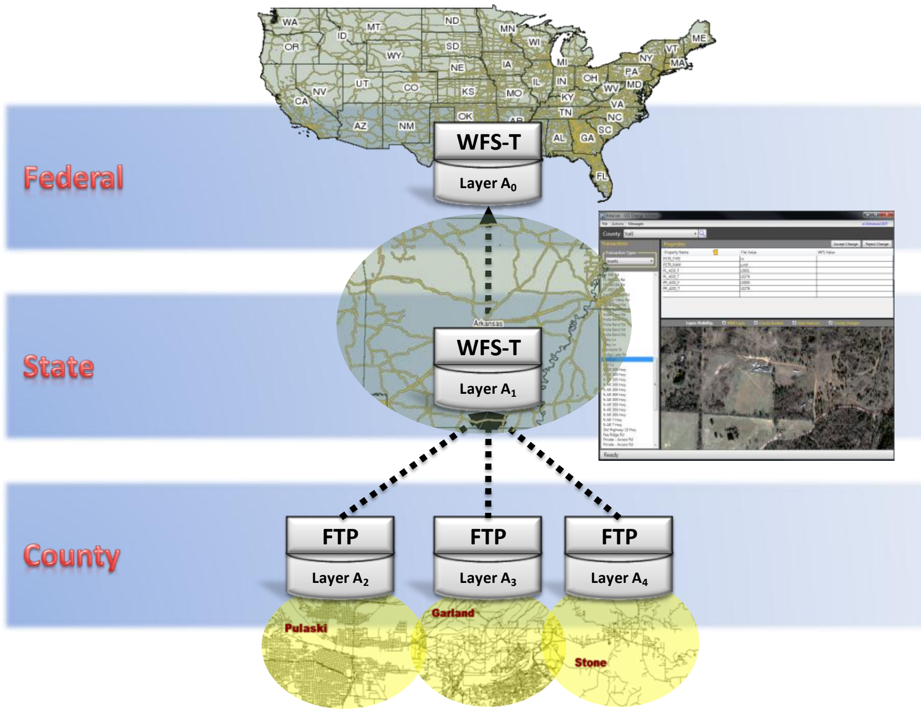


Figure 2 – This project developed a geo-synchronization capability for Arkansas, enabling exchange of framework data between local, state and national data stores through geo-synchronization.

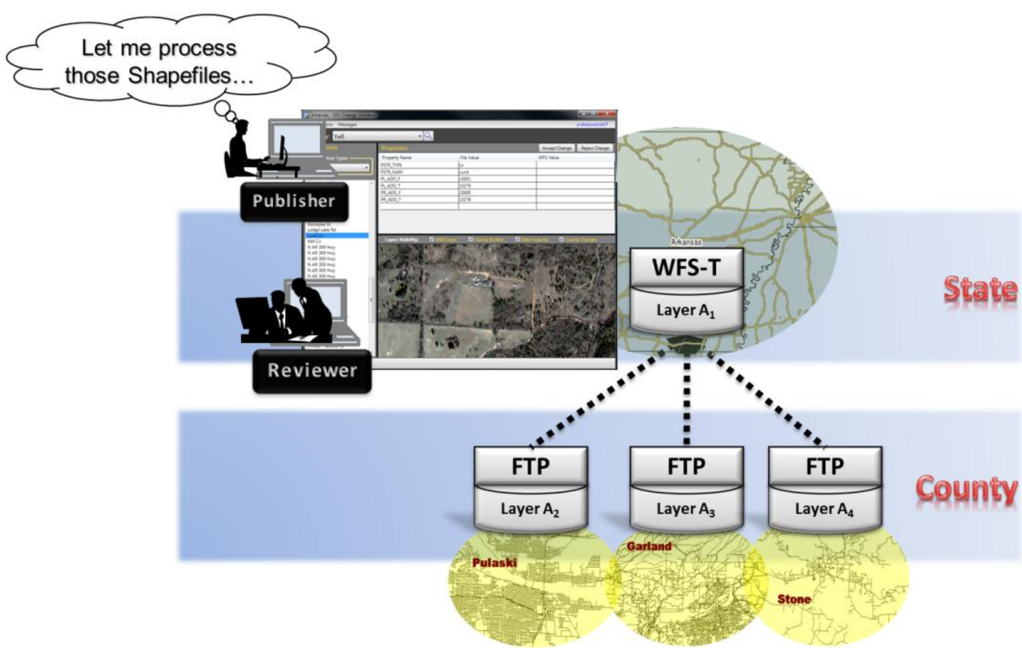


Figure 3 – The project also developed a geo-synchronization capability tailored for Arkansas needs - enabling exchange of framework data between local and state data stores via shapefile-based geo-synchronization.

The capability uses the geo-synchronization system from The Carbon Project (CarbonCloud Sync), ESRI's ArcGIS Server and two new 'Change Validator' tools from The Carbon Project - and provides a unique capability for NSDI geo-synchronization by:

1) Addressing methods to ingest County data into the geo-synchronization federation and;
2) Geo-synchronizing a local-level database to a state-level database.
3) Output change-based shapefiles for ease of use.

To achieve this first step we established a common folder structure consisting of contributed GIS updates via FTP from over 70 Arkansas counties. Our project then leveraged the Web Feature Server (WFS) capacity of ArcGIS Server 9.3 to establish a baseline to compare changes to, compares updates to the baseline using the Change Validator, and then outputs the results as either 'geosync' transactions or shapefile updates. With this capability we can bridge current Arkansas business processes with the "to be" processes outlined by FGDC. Without this recognition of current processes we believe geo-synchronization may not successfully address real-world operations. The solution is applicable (re-deployable) to other sites and problem domains through an open API, deployment on cloud-based services, and two free Change Validator tools so that it may be used by other Framework data exchange providers.

This approach was required because currently the Arkansas state-level GIS receives a drop of Shapefiles features data from counties on approximately a monthly basis. These files are ingested into the state level view, including all provided data changes and updates. Little or no resources are available to deploy WFS-T in each of the 70 counties. The enhanced system can accommodate current process and offers the option for future introduction of WFS-T to counties and states with minimal disruptions. Furthermore, this system will enable the Arkansas state-level to be updated with new transactional sources including citizen (or professional) provided updates (Gov 2.0).

Our technical approach is based on the Open Geospatial Consortium (OGC) draft for GeoSynchronization Services (GSS) and the OGC Web Feature Service (WFS). The GSS in this project uses The Carbon Project's CarbonCloud Sync[1] software. This solution provides a powerful capability agnostic to the underlying information infrastructure; it simply interfaces with WFS-T services or FTP common file folder, and includes the ability to analyze transportation Shapefiles data and transform changes into WFS transactions using Geography Markup Language (GML) or output Shapefiles. This approach enables our solution to support current and future NSDI production operations simultaneously.

There are two key components to the solution. The first component is the CarbonCloud Sync service that disseminates changes throughout a federated deployment of geospatial services, keeping any registered remote services up to date with authorized updates (Figure 4). The second component is a custom standalone application called the GSS Change Validator. The GSS Change Validator application ingests and analyzes the registered county layers (Shapefiles) and manufactures change-based transactions or shapefiles on county-level transportation framework data layers. This two-part approach meets the needs of state data production organizations and the geo-synchronization vision of the CAP grant.

[1] http://www.thecarbonproject.com/Products/cloudsync.php

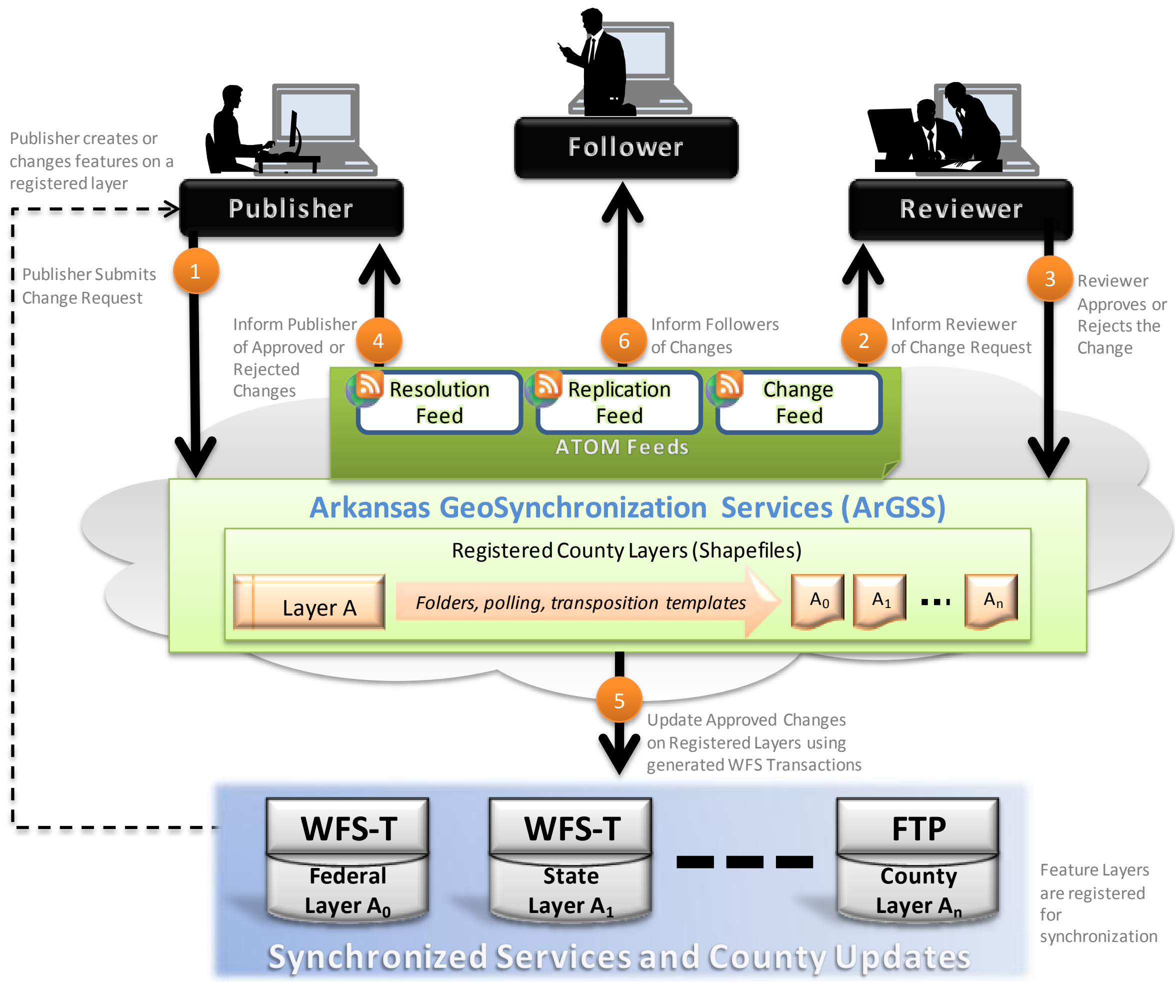


Figure 4 – The Arkansas Geo-Synchronization capability facilitates maintenance and update of NHD into authoritative State data layers, which can be provided for integration into the NSDI.

This solution required an initial setup. During this one-time process we created a layer template for the state data (a schema-based description of the data). We also generated county-level layer templates for each participating county. Once the layers are registered in ArGSS we will register the synchronization links by mapping counties to the state layers, and the federal WFS-T to the state's WFS-T. This process is performed using the CarbonCloud Sync's Web-based management client. The process is similar to any CarbonCloud Sync deployment with the exception of using static Shapefile fields to produce the layer's data templates instead of WFS-provided schema sources.

The project was conducted as a series of collaborative tasks to develop, deploy and sustain the Geo-Synchronization capability in Arkansas ***and*** engage government stakeholders. Specifically, multiple coordination meetings were held where the tools were demonstrated during development. Final demonstration and tool transfer were also provided. This approach allows Arkansas to implement shapefile-based geo-synchronization and WFS-based geo-synchronization with minimal change to existing technologies and planned development.

GSS Change Validator

The GSS Change Validator is a new standalone application for the NSDI community (Figure 5). The application is presented to the user as a simple, form-driven tool that digests and analyzes county level data and automatically determines changes. Its primary purpose is to take data compiled by each of Arkansas' counties, compare that data against existing statewide data, determine what has changed and make the changes by submitting them to a GeoSynchronization Service (GSS) called CarbonCloud Sync. The application is also able to access any to use any OGC compliant data source such as Web Map Services (WMS) and Web Feature Services (WFS) endorsed by the US Federal Geographic Data Committee (FGDC) Steering Committee (along with other standards developed externally to FGDC).[2]

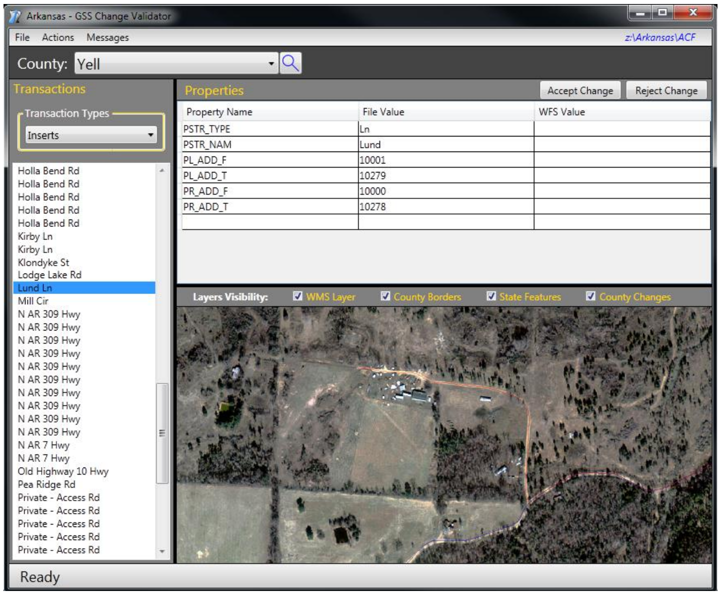
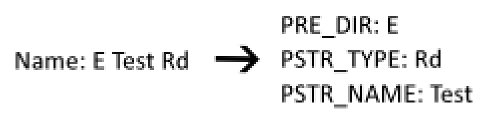


Figure 5 - The GSS Change Validator is a new standalone application for the NSDI community. The application takes data compiled by each of Arkansas' counties, compares that data against existing statewide data, determines what has changed and makes the changes by submitting them to a GeoSynchronization Service (GSS) called CarbonCloud Sync or outputs them as shapefiles.

Data Importing/Schema Transposition Process

County data comes in a variety of different information models. Our challenge was to convert the county schemas into the standard ACF schema that is used at the state level. The majority of the fields mapped one to one from county to state. These relationships are expressed in an XML file that can be updated at any time to reflect any changes in schemas. However, certain fields contained a complex arrangement of data. In those cases, it was required that we move the processing into the code level and mark that field with a generic FUNCTION attribute in the XML. Here is an example of data that required this type of processing:

[2] http://carboncloud.blogspot.com/2010/10/ogc-standards-officially-endorsed-by-us.html



Transposition from one field into 3 fields

Our current processing algorithms contain logic for prefix/suffix directions, all possible PSTR_TYPE values (Rd, St, Dr, etc.), some specialize highway rules and several other elements. After the schema is transposed into the proper format, the data is validated against the ACF rules and is modified (if necessary) to comply. Things like case and max length are checked here.

WFS Data Retrieval

When a county is selected for processing, we need to compare it against the existing WFS features. To do that, we must first query the server for all features from a particular county. We do this by sending a WFS Filter Encoding (FE) query to the server. However, because the county geometries are often quite complex, it's often not possible to send it as a Filter Encoding. So to augment the process we send the Bounding Box of the extent of the entire county to the server. The resulting data set if then checked to see which features actually belong with the county and those are do not are removed. The result is all the features that are contained or intersect with the current county geometry.

Comparison

Once we have the county data in the proper format and the WFS data from the server, we can initiate our comparison algorithms. We compare both the geometry and the properties from both the county and the state files. The property comparison is done based on the fields in the county file. The geometry comparison is done with a small variance built in (usually under a meter and you can adjust this variance in the settings if you need to.) The geometry comparison also takes into account the line direction, so it will match the geometry even if the lines are drawn in the opposite direction. Any features that are matched are immediately removed from both data sets. If either the geometry or the properties completely match, then we mark this as an update and remove it from both data sets. Features that remain on the county side are considered inserts. Features left on the WFS side are considered deletes. If an update has occurred which changes features and geometry, the system will be unable to match it with the state data and will result in both and insert and a delete. The resulting data (Inserts/Updates/Deletes) are presented to the user for inspection. It's up to the user to determine if the changes are valid and if they should be acted on.

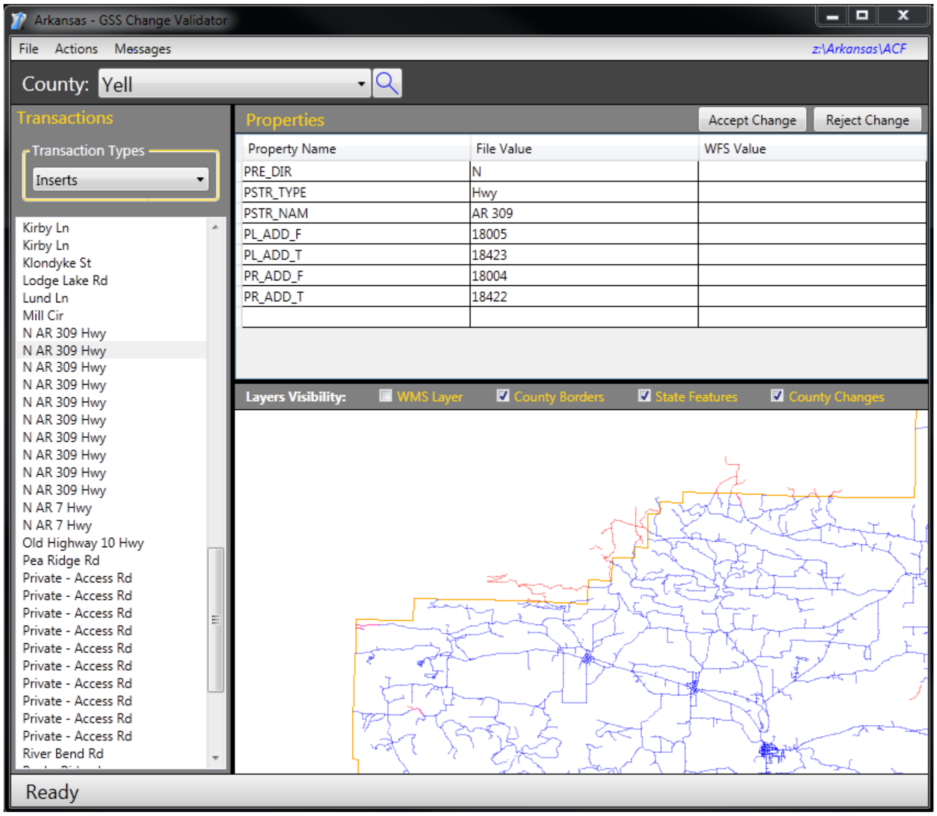


Figure 9 – The GSS Change Validator with inserts are shown on the map in red

In the preceding image, the inserts are being shown on the map in red. The current WFS features are displayed in blue. The user must select a feature in the left list, determine if it is valid and then click the accept change button to indicate that it's a valid change. We build in the ability to export changed as Shapefiles in the event it might be easier to use that change data.

Submitting Changes

After a feature change has been accepted, it can be sent to the GSS for processing. To do that, users simply select the *Actions->Submit Accepted Changes* menu item and the changes will be sent to the GSS service. The transactions will show up in a GSS aware client like Gaia for approval.

Use of WMS and WFS for Reference

Because we are using CarbonTools PRO, a geospatial interoperability toolkit, behind the scenes, we are able to use any OGC compliant data source.[3] In this case, we have Arkansas GeoSTOR 1 meter imagery WMS that we are overlaying the change on top of and Arkansas GeoSTOR WFS with Road centerline data for reference.

---

[3] www.CarbonTools.com

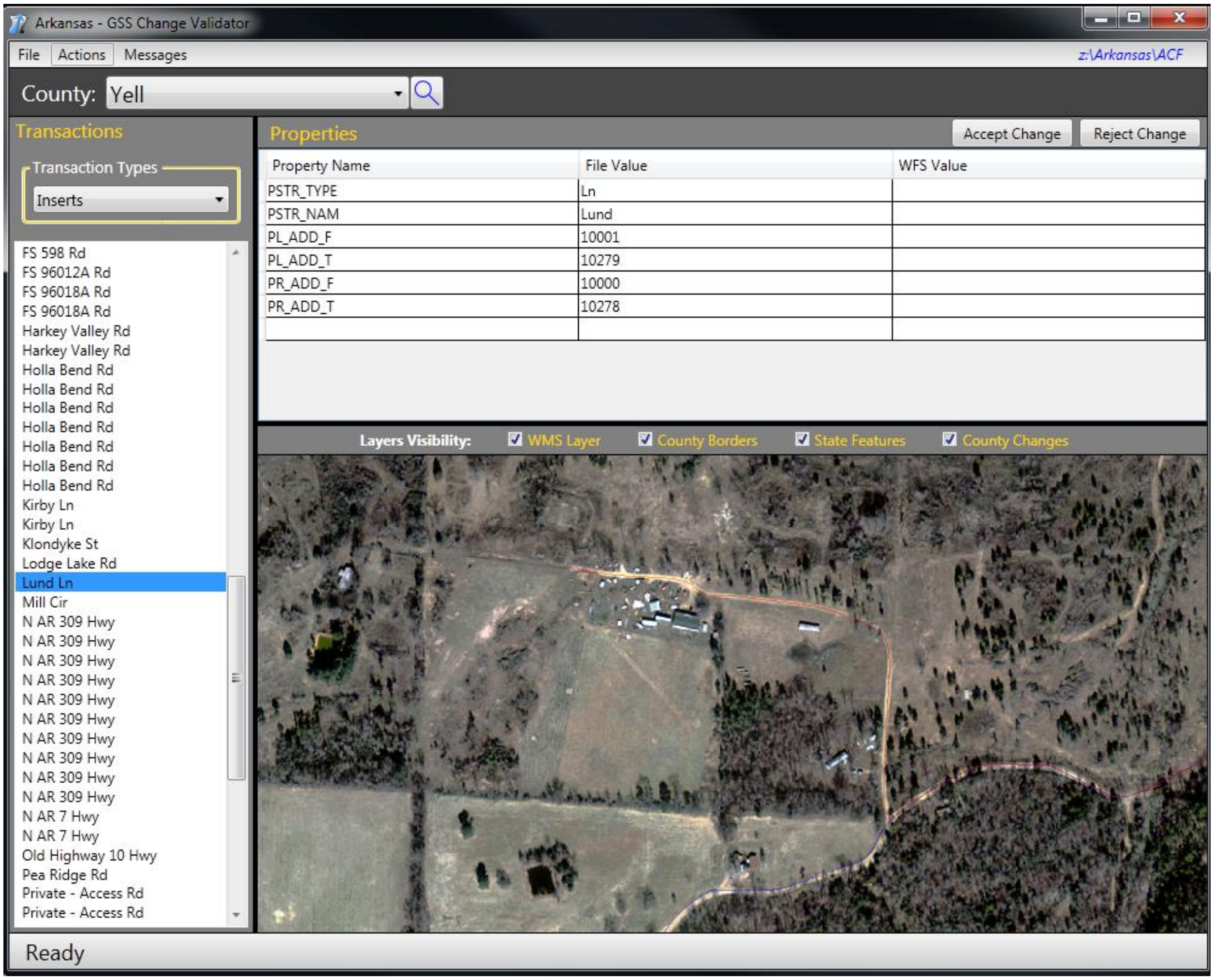


Figure 10 – The GSS Change Validator is able to access Arkansas GeoSTOR WMS and WFS for reference

This process bridges the gap of the current system's handling of transactions originated as static files and WFS-T operations in automated geo-synchronization. The system uses advanced processing and comparison techniques to generate the WFS-T like operations.[4] From there, the normal CarbonCloud Sync processing will take care of propagating the changes down the deployment chain. This process allows the state's current Shapefile and FTP driven exchange process to evolve to a geo-synchronized federation.

This project has already delivered distributed services and applications that can be widely used for GeoSynchronization. The project team has successfully avoided practices that would inhibit the use of the distributed data and deliver short term gains. This project has focused on interoperable SDI solutions and created a GeoSynchronization environment that will entice additional participation from NSDI community of practices.

[4] Made possible by deploying CarbonTools PRO (www.CarbonTools.com) on the server

Feedback on Cooperative Agreements Program

The NSDI Cooperative Agreements Program provided an opportunity for the project team to work with NSDI practitioners from Arkansas and other organizations on an initiative that has real world implications in much of the U.S NSDI. The program provides for the injection of new technologies and approaches into the geospatial community. The grant provided both research challenges and important collaboration experiences.

Strengths: The program reviews and funding decisions were made very quickly. We were also pleased to have the opportunity to prototype services and simulate deployment scenarios with other team members, an effort that came out of discussions and regular teleconferences with the CAP government team. The program's mixture of federal government, enterprise, and local government teams was also very beneficial. Overall, the program is making very good progress towards promoting key aspects of realizing NSDI services online. Continued emphasis needs to be placed on promoting an online infrastructure of standards-based location content across the nation that can flexibly support operational requirements, and governance of resulting standards for information sharing. With the progress on geo-synchronizaton exemplified by efforts such as those outlined in this report we can identify no technical impediments to advancing such an infrastructure. However, we suspect funding issues are holding back development of this online infrastructure – which is difficult to understand since geo-synchronization offers one of the most compelling methods to save government money by turning geospatial stovepipes into interactive community data networks.

Weaknesses: Although not a weakness, additional external Federal engagement (i.e. outside FGDC) in project continuation and partnering efforts should continue to be encouraged. This is occurring but agencies such as USDOT and others can benefit from CAP solutions such as those developed in this project - and should continue to engage more in the process. Specifically, the CAP needs to have continued strong liaison in operational aspects of these agencies and with other state, federal, and commercial interests.

The team had no program management concerns. The team received prompt responses to questions. Additionally the program management team’s format for meetings and communications greatly facilitated collaborations.