



## Quality Control at the US Marine Corps using GIS Data ReViewer

### Author:

Joe Simons, Data Services Manager at GEOFIWEST  
Technology Associates Intl. Corp.  
[joseph.simons@taic.net](mailto:joseph.simons@taic.net)

### Abstract:

The Western Regional Geospatial Data Center (GEOFIWEST) located at Camp Pendleton, California recently overhauled their quality control program which has resulted in significant improvements in data quality, business systems integration and reporting capabilities. By replacing a legacy, custom-built data quality control application with ESRI's PLTS Data ReViewer, and using ArcGIS Model builder to support business systems integration, annual data quality tests and required business systems integration requirements were met with greater accuracy using less manual effort than in previous years. Improvements were so great that the Data ReViewer and integration processes are becoming the standard across the enterprise at the Marine Corps. This presentation will provide a case study of quality control processes of an enterprise geospatial database, GIS integration with an enterprise asset database, and the combined result as measured through enterprise-level system and program effectiveness ratings.

### Overview:

The goal of the GEO*Fidelis* program is to provide accurate and up to date geospatial information in support of Installation management, to improve our stewardship of natural resources, protect the environment, and support the training of Operating Forces. The quality control measures put in place by the GIS staff at GEOFIWEST are a vital component towards achieving this goal.

The GEO*Fidelis* program is a regionalized program overseen by Headquarters Marine Corps, with regional data centers in the east at Camp Lejeune and west at Camp Pendleton. This document and proceeding presentation focuses on the quality control measures at the Western Regional Geospatial Data Center.

### Quality Control Measures:

Prior to implementing QC measures using GIS Data ReViewer started in 2009, quality control of GIS data at GEOFIWEST was loosely implemented using a series of solutions. These ranged from visual QC to a GOTS solution called the Data Model Checker that validated quality using similar QC measures as the Data ReViewer, but was more complicated to administer, lacked the ability to automate, had poor reporting options, and did not have built in corrective measures.

The "GEO*Fidelis* Readiness Ratings" (G-ratings) is a process by which Headquarters Marine Corps gauges and grades the Installations readiness to support mission critical GIS responsibilities. During the months prior to the GEO*Fidelis* Readiness Ratings for FY09, quality



control was performed on each GEOFIWEST Installation's feature data and the results were shared with the Installation GIS Manager. The mechanism that was used to perform the quality control was the PLTS GIS Data ReViewer toolset.

Although feature-level quality control can be enforced via several methods, they can be grouped into two categories: (1) real-time solutions to be run during an edit session, and (2) post-process solutions, to be run in a process after the data is already in the database. Both solutions should be implemented for a good quality control system however the focus of this document is the latter.

At GEOFIWEST, the post-process solution is accomplished using the GIS Data ReViewer's batch job checks, which are used to validate the contents of existing data. These are scheduled to run at times when the database is not in use, are used for delivering reports, and are ideal for an existing database that already has invalid values.

The GIS Data ReViewer allows you to configure several different types of data quality checks and store them for later use in a file known as a batch job file (.rbj). These checks can be run repeatedly, modified, and resourced to new datasets. At the 9.3.1 release of the Production Line Tool Set (PLTS), there are currently 39 types of checks that can be configured, ranging from SQL statement checks against the attribute data to geometry and database validation checks. As data quality inefficiencies are found by the data checks, records are written to a GIS Data ReViewer table (Figure 1) identifying the feature layer, feature record, type of check that failed, and additional information.

Figure 1 - GIS Data ReViewer table identifying the feature layer, record, type of check failed, and additional information

	B	F	H	J
1	OBJECTID	REVIEWERCODE	ORIGTABLE	NOTES
2	56	228	BA_BA_aboveground_storage_tank_point	ENVAST_ID is a mandatory attribute and cannot be an empty string or NULL
3	1746	228	BA_BA_aboveground_storage_tank_point	PROOCT_D is a mandatory attribute and cannot be an empty string or NULL
4	1747	228	BA_BA_aboveground_storage_tank_point	PROOCT_D is a mandatory attribute and cannot be an empty string or NULL
5	1748	228	BA_BA_aboveground_storage_tank_point	PROOCT_D is a mandatory attribute and cannot be an empty string or NULL
6	1749	228	BA_BA_aboveground_storage_tank_point	PROOCT_D is a mandatory attribute and cannot be an empty string or NULL
7	1750	228	BA_BA_aboveground_storage_tank_point	PROOCT_D is a mandatory attribute and cannot be an empty string or NULL
8	1751	228	BA_BA_aboveground_storage_tank_point	PROOCT_D is a mandatory attribute and cannot be an empty string or NULL
9	1752	228	BA_BA_aboveground_storage_tank_point	PROOCT_D is a mandatory attribute and cannot be an empty string or NULL
10	1768	228	BA_BA_aboveground_storage_tank_point	PROOCT_D is a mandatory attribute and cannot be an empty string or NULL
11	2558	228	BA_BA_aboveground_storage_tank_point	PROOCT_D is a mandatory attribute and cannot be an empty string or NULL
12	403	228	BA_BA_aboveground_storage_tank_point	VOLUME is a mandatory attribute and cannot be an empty string or NULL

## Implementation for FY09 Readiness Ratings

The first phase of the implementation of Data ReViewer was to create data checks that focused on addressing inadequacies that may be identified the GEOFIWEST Readiness Ratings for FY09. The intent is that the QC checks will be expanded over time to include complex geodatabase integrity checks, attribute checks, geometry checks, and combinations of each. However, outlined below, the initial implementation focused solely on "Common Installation Picture" (CIP) layers and Real Property layers (INFADS).

As of 9.3 SP1, GIS Data ReViewer has 39 out of the box validation batch checks. The Execute SQL checks are the ones primarily responsible for accomplishing the tasks of getting the attributes in order for the readiness ratings.



## Details of CIP and iNFADS Checks

Mandatory fields in each CIP layer (identified as such by the *GEOFidelis* Data Model 1.0 document) were checked using SQL statements based on their specific requirements. In most cases this meant that the field was checked for NULL values and empty strings. In the case of FACIL\_ID, the SQL statement checked for NULL values, empty strings, and fifteen total characters beginning with "NFA" followed by exactly 12 integers. Fields with FACIL\_ID values = "NFA" with trailing zeros were also flagged as errors.

## Configuration and Level of Effort

Configuration of the data checks took less than 40 hours for all Installations, and that time could have been minimized if not for the need to gain familiarity with the software. The level of effort to run the data checks and export them to a spreadsheet took a couple of hours initially for all GEOFIWEST Installations combined, however this was cut down to approximately an hour towards the end. This reduction in required time was because the batch job file incorporated a geoprocessing model which required only double-click to execute the check, and it automatically exported to a comma-separated values (CSV) file.

## Schedule of Data Reports

Starting in July 2009, the data checks were created and configured to match each Installation's available data. After each Installation's data checks were configured, the data was validated and the spreadsheets were distributed. This process was performed twice a month through the end of August. During this time, and through the end of the ratings period, several Installations requested data checks more often, such as immediately after performing corrective measures. The Installations were provided with new data reports as a result.

Starting in September, all GEOFIWEST Installations were provided with, at a minimum, once a week reporting. However, this number increased to several per week in some cases towards the end of the month as the deadline for Readiness Ratings submission approached.

Figure 2 shows an example of one Installation's improvement in data quality during this timeframe. Note that the number of violations does not equate to the number of features in violation; if a certain layer has ten mandatory attributes and only three are populated then there will be seven data violation entries. Also, one key area to highlight here is that the GIS staff introduced news errors after week 37/09, but were able to address them before the final ratings were issued.

**Figure 2: Error table**

<i>Week</i>	<i>Data Violations</i>
31/09	5269
34/09	3162
36/09	338
37/09	339
38/09	959
39/09	990
40/09	305

## Quality of Final Deliverable

In all cases, the Installation's data violations decreased significantly as they began to take corrective action. Furthermore, in all but one case, the Installation's data violations decreased from anywhere between 70 and 100 percent from week 31/09 to the final archive delivered at the end of week 40/09. Furthermore, each Installation received high grades in the FY09 Readiness ratings reports, which partially in large part was the result of the corrective actions that they took prior to the end of FY09.



## Future Plans for FY10

### Quality Control Measures Not Covered in FY09

Although the implementation of PLTS GIS Data ReViewer helped identify inefficiencies in attribute data and enabled GIS Managers to take corrective action much easier, more can be done to expand the quality control measures for the entire Installation dataset. Currently GEOFIWEST is expanding the types of checks that are being done and the extent of feature data to include all SDSFIE feature classes. The first phase of the expansion is focusing on invalid geometry, invalid domain values, invalid subtype values, and non-unique primary key values.

Most GIS quality assurance programs involve more than one individual to carry out specific tasks. These tasks typically fall into one of three areas of quality control processes: review, corrections, and verification of corrections. An entire suite of tools are available in PLTS GIS Data ReViewer geared towards this type of workflow, incorporating manual and automated quality control measures. These workflows are currently being underutilized because of a shortage of licenses. The goal is to increase the number of licenses that can be shared over Citrix so that Installations can take advantage of the built-in functionality in GIS Data ReViewer.

### Expansion of GIS Data Checks for FY10 (Currently Implementing)

During the FY09 QC process, we recognized a need to expand the checks in two ways: (1) expansion to the entire data model, and (2) different types of checks. The goal was to validate the entire data model with the best known available information. Because of this, the GEOFidelis Data Model 1.0 document was used as the primary source document. All feature classes in an Installation's database that been redefined to have Execute SQL statements run checking to verify that (1) mandatory attributes are populated and that (2) facil\_id fields are populated with correctly formatted values (i.e. NFA+12 integers). This document was also use to validate that the Primary Key designated fields have unique values, using the Unique ID check. This is a very labor intensive check, so it was limited to just the readiness ratings layers.

The Invalid Geometry, Invalid Domain, and Invalid Subtype checks were expanded check against all feature classes that exist in the Installation database. Below is a list of all used Reviewer Data Checks and a brief description of the function of each check:

**Invalid Geometry:** Finds features whose geometry is empty, nothing, or not simple as well as those with empty envelopes.

**Invalid Domains:** Validates coded value and range domains to ensure that all values meet domain constraints.

**Invalid Subtypes:** Searches for feature classes with improper or null (optional) subtypes.

**Unique ID:** Searches for duplicate values in a table or feature attribute field across the entire database.

**Execute SQL:** Finds features based on a SQL query WHERE clause run using feature attributes.



## **Conclusion:**

In conclusion, the implementation of GIS Data ReViewer for the FY09 Readiness Ratings was a success and yielded great results. All Installations decreased the number of their data errors significantly, and in part contributed to a higher rating for the entire region as a whole. The reporting schedule of once a week seemed to be a manageable timeframe for Installations to resolve their data issues. The tools in GIS Data ReViewer provide the user with several options to check the quality of the data, and GEOFIWEST in FY09 was just scratching the surface. In the future, much more quality control measures can and are being implemented by GEOFIWEST using GIS Data ReViewer.

## **References:**

### **GeoFidelis Program**

<https://www.hqmc-facilities.org/gis/index.cfm>

[http://proceedings.esri.com/library/userconf/ieug09/papers/geofidelis\\_frances\\_railey\\_program\\_brief.pdf](http://proceedings.esri.com/library/userconf/ieug09/papers/geofidelis_frances_railey_program_brief.pdf)

[http://proceedings.esri.com/library/userconf/proc07/papers/papers/pap\\_1983.pdf](http://proceedings.esri.com/library/userconf/proc07/papers/papers/pap_1983.pdf)

### **Technology Associates**

[www.taic.net](http://www.taic.net)

### **Reviewer:**

[http://webhelp.esri.com/arcgisdesktop/9.3/index.cfm?TopicName=An\\_overview\\_of\\_GIS\\_Data\\_ReViewer](http://webhelp.esri.com/arcgisdesktop/9.3/index.cfm?TopicName=An_overview_of_GIS_Data_ReViewer)

### **Free Online Workshop - Introduction to GIS Data ReViewer**

[http://training.esri.com/acb2000/showdetl.cfm?DID=6&Product\\_ID=900](http://training.esri.com/acb2000/showdetl.cfm?DID=6&Product_ID=900)

## **Author:**

Joe Simons is the GIS Data Services manager for GEOFIWEST and is an employee of Technology Associates. He has a bachelor's degree in Geography from the California State Polytechnic University at Pomona. He has been directly supporting Marine Corps enterprise GIS program for the past year. Prior to GEOFIWEST, he was an employee at ESRI for eight years as a product specialist and ArcSDE Administrator.