SAN JOSE INTERNATIONAL AIRPORT

GIS DATA STANDARD COMPLIANCE QC/QA PROCEDURES

TABLE OF CONTENTS

1.0 INTRODUCTION

1.1 Scope

This document details the quality assurance (QA) procedures that are to be carried out on GIS data submitted to the Mineta San José International Airport (SJC). GIS data encompasses the virtual representation of real-world assets, infrastructure elements, cadastral boundaries, temporal events, airspace surfaces, topography, etc. using points, lines and polygons on a computer screen or in printed output. GIS data also includes attributes (alphanumeric information) that describe these geographic features and metadata that provides information about the geographic and attribute data itself.

The procedures described in this document should be applied to all GIS data submitted to SJC so that the quality of this data can be assured. Some GIS data that is submitted will have been produced by surveying means. For these submittals, the Survey Data Quality Assurance Procedures detailed in SJC-ACM-AIMS-3240 should also be applied.

1.2 Purpose

GIS data can be used for a variety of purposes, but only if its quality meets the expectations of those who will be using the data. The purpose of this document is to detail procedures that can be applied to assure the quality of all GIS data submitted to SJC. For data that is submitted by SJC staff, contractors and consultants who are required to adhere to GIS Standards outlined in CADD and Mapping Standards and Procedures (SJC-ACM-AIMS-2200) these procedures will also serve as an acceptance test for the data. If the test fails, the data should be returned to the provider for correction and clean-up. For agencies and others who are not required to adhere to SJC standards, these procedures should be applied as a means of establishing whether or not the incoming data adheres to SJC standards.

This document is written as an instruction manual for airport staff or 3rd parties tasked with reviewing GIS data submittals. The procedures are not intended for data creators who should be following quality control procedures designed to make sure that data collected meets the specifications required by the airport.

1.3 Contact Information

This document was prepared by the Carter & Burgess (C&B) Team. If you have any questions, please call Lysée Moyaert at (408) 501-7712, or David Tamir or Behzad Mohammadi of the C&B Consulting Team at (818) 784-7585.

2.0 OVERVIEW OF THE QUALITY ASSURANCE PROCESS

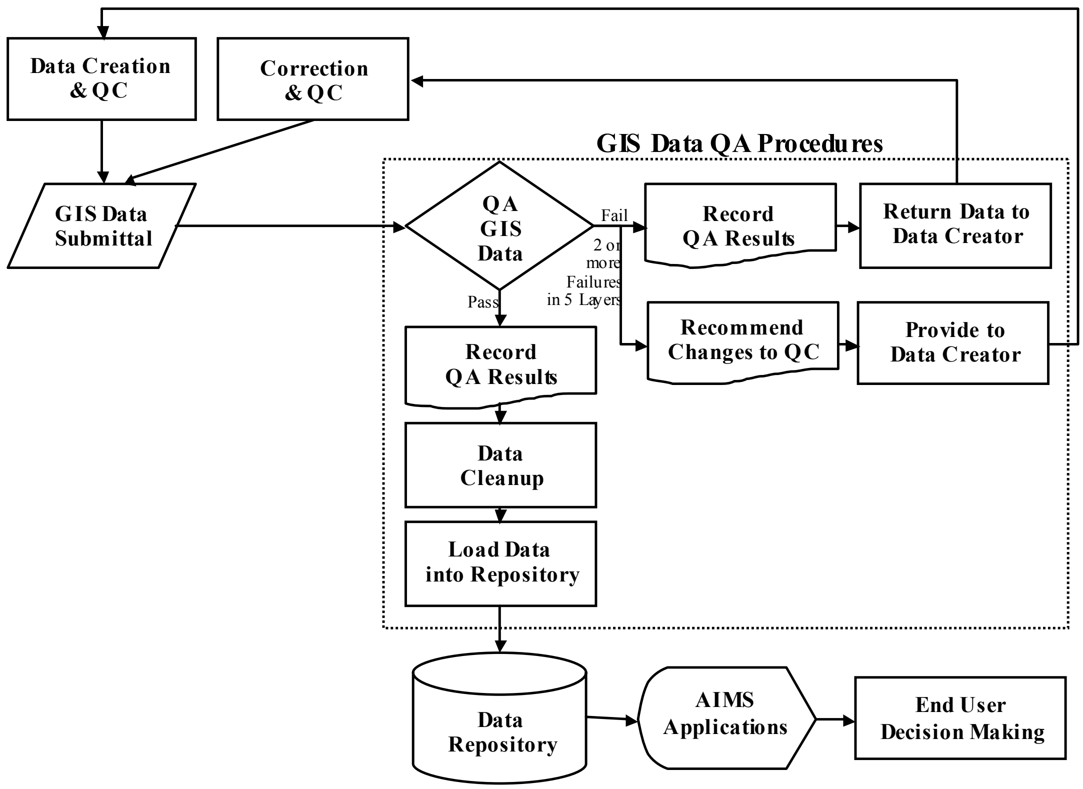
Quality Assurance (QA) serves as a final check of the data, to catch any problems that may have been missed by Quality Control (QC) procedures carried out as the data is created. QA also serves as a regular test of whether or not the production and QC processes are producing data of the required quality. QC, on the other hand, occurs during the development of GIS data that will be submitted to the airport. These definitions are based on definitions used in ISO 8402:1994 Quality Management and Quality Assurance.

QA procedures will be performed on all GIS data submitted to SJC. The QA procedures for GIS data submittals have been developed independently from any QC processes that have or will be developed by data creators and are based solely on the GIS standards adopted by SJC. QC procedures need to be developed by data creators as an integral part of each data production process. They should be administered using checklists on a daily basis by the staff and contractors hired to collect data, as well as other agencies and private firms who collect GIS data.

QA procedures are to be carried out by airport staff or an independent 3rd party reviewer. For GIS data, most of these procedures involve manually checking a random sampling of data elements against a predefined set of acceptance criteria. GIS submittals in which errors are found beyond an acceptable limit will be rejected and returned to the submitter for correction and re-submittal. All data that passes the acceptance criteria can be loaded into the data repository as specified in document SJC-ACM-AIMS-4140.

Figure 1 presents a generalized process for performing quality control (QC) and quality assurance (QA) on drawings.

Figure– 1 – Overview of QC/QA Process



3.0 GIS DATA ACCEPTANCE CRITERIA

The acceptance criteria for GIS data submitted to SJC are defined by the CADD Standards and (SJC-ACM-AIMS-2200). This document details the layering standards GIS data must meet. In addition, documents must be submitted along with metadata, a requirement specified in the metadata standard (SJC-ACM-AIMS-2500).

4.0 DETAILED QUALITY ASSURACE PROCEDURES

4.1 Informal Review

An informal, cursory review of submitted GIS data is recommended prior to applying the rigor of the formal QA procedures. This cursory review should focus on obvious discrepancies between what was submitted and what was supposed to have been submitted. Data layers that are of the wrong type (i.e. point, line or polygon), obvious misalignment with the Airport Grid coordinate system, lack of metadata, etc. are among the issues that can be identified at this stage.

Since this check is informal and prompted by case specific requirements, it is not effective to outline a definitive procedure for this step. Instead, it should be defined on a case-by-case basis by a trained GIS technician who is familiar with the type and content of data being submitted. Such personnel are more likely to spot the obvious errors that this step is intended to identify. If such errors are found, the more time consuming and costly formal QA steps can be avoided. If no such errors are found, then QA can proceed as follows.

4.2 Random Sampling

Since QA procedures are labor intensive, 100% validation, is not cost justified nor it is necessary. Experience has shown that testing a statistically valid random sample of data elements is sufficient to assure quality (i.e. conformance with specifications/standards) at a 98% confidence level. This means that 98% of the data elements contained in a data set can be statistically assured to adhere to SJC GIS standards. For the application of this procedure to GIS data submitted to SJC, a data set is equivalent to a related collection of GIS layers stored in a GeoMedia Access Warehouse or ESRI shape files.

4.2.1 Determining Sample Sizes

The American National Standards Institute (ANSI) MIL-STD-105E sampling standards to assure data quality is being used to define the method and sample size of elements to be tested. The size of the sample to be inspected during the quality assurance testing is determined by the lot size, which is the total number of features on a given layer. Table 1, Sample Sizes and Allowable Error Rates, is derived from the ANSI standards document. This table illustrates the random sample sizes and allowable number of errors based different lot sizes. For example, under Normal Inspection criteria, a layer with 20,000 features to be tested (lot size) would produce a Sample Size of 125, with a maximum of three allowable errors. If a data set produces three or less errors during quality assurance testing, then it will pass the testing with a rated accuracy of 98%. Even though the data set had passed, any errors that were identified are to be corrected before the data is converted and loaded into the spatial repository.

If two out of five consecutive layers in a data set fail, then tightened inspection criteria are used until five consecutive layers in a data set pass inspection. Under tightened inspection criteria, the number of allowable errors is reduced. The far right column in Table 1 indicates the number of allowable errors under tightened inspection criteria. After five consecutive datasets pass tightened inspection, then normal sampling procedures will be re-instituted.

Table 1 – Sample Sizes and Allowable Error Rates
(Derived from ANSI/ASQC Z1.4-1993)

Lot Size	Sample Size	# of Allowable Errors – Normal Inspection	# of Allowable Errors – Tightened Inspection
2 – 15	2	0	0
16 – 25	3	0	0
26 – 90	5	0	0
91 – 150	8	0	0
151 – 280	13	0	0
281 – 500	20	0	0
501 – 1200	32	1	0
1201 – 3200	50	1	0
3201 – 10,000	80	2	1
10,001 – 35,000	125	3	2
35,001 – 150,000	200	5	3
150,001 – 500,000	315	7	5
500,001 and over	500	10	8

4.2.2 Determining Lot Size

The application of this procedure to determine the sampling size requires knowledge of the number of features depicted on a layer (i.e. lot size). To determine lot size, the number of features can be queried using GIS software.

4.2.3 Generating a Random Sample

Based on the guidelines described above, the next step in the QA process is to randomly select the appropriate number of features based on the sample size indicated in Table 1. To do this in an unbiased manner, a computerized random generator is recommended. Perhaps the easiest random generator to use is the RAND() function in MS Excel. Using the steps below, this function will yield a list of randomly generated numbers. Since SJC GIS Data standards require a numeric unique id for each feature, these randomly generated numbers can be used as pointers to a randomly generated list of features in the data layer. The following steps can be used to generate this list:

1. Determine the lot size and sample size as indicated in Section 4.2.3 and 4.2.1.
2. Using GIS software, determine the highest numeric value of the numeric unique id used as the key field in the GIS layer.
3. Open a new worksheet in MS Excel.
4. In any cell, type the following formula =INT(RAND()*[Highest ID Value]). This will generate a randomly generated integer between the value of 0 and the Highest ID Value determined in step 2.
5. Copy the formula down a column in the MS Excel spreadsheet so there are at least twice as many entries as required for the sample size.
6. Sort this list using the ‘Sort’ function from the MS Excel ‘Data’ menu item.
7. Use this list to select GIS features based on their Unique ID, skipping 0s or repeated values.
8. Continue selecting Unique ID values until the required sample size has been reached.

The steps listed above assume that the submitted GIS data conforms to SJC GIS Standards. If the data has been provided by a source not required to meet these standards, it is possible that numeric unique ids may not exist in the GIS layer. This would render the steps above useless. In their place, the sample features can be selected manually. When manual procedures are used, care should be taken to obtain as random of a sample as possible while insuring that the features selected are well disbursed throughout the extent of the data.

4.3 Checking Proper Placement of Features on GIS Layers

Insuring that the proper features are on the proper GIS layers is critical to the ability of the data conversion processes to correctly load the GIS data into the data repository and ultimately to the functionality of the AIMS applications that will be used to access data in this repository. The following steps are therefore required to assure geographic features are placed on the correct GIS layers, based on the SJC GIS layering standards detailed in SJC-ACM-AIMS-2200.

1. Overlay the layer being tested on top of GIS basemap layers and an orthophoto for outside domain data, or on top of an accurate floor plan for inside domain data.
2. Use the GIS software to identify each feature in the sample set. Review the adjacent and underlying features shown in the basemap, orthophoto or floorplan to insure that this feature represents an instance of the type of features for which that layer is intended. For example, is the layer being review is for roads (i.e. TRVEHRDS) and the roadway feature being checked corresponds to a river, wall, runway or other linear feature than this feature fails the test. If the feature appears to be a road, but may or may not be in the correct location, then it passes this test.

3. Repeat step 2 for all of the randomly selected features. If the number of allowable errors is exceeded, stop and record that the drawing has failed due to placement of features on wrong layers.
4. If all randomly selected features are checked and the number of allowable errors has not been surpassed record that the drawing has passed feature/layer placement quality assurance.

4.4 Validation of Attributes

An important component of GIS data is the attributes related to specific features. To assure that attributes are properly recorded for each feature follow the steps below for each layer as well as for each of the randomly selected features that have attributes.

For the entire layer, check for outlying, missing or obviously incorrect attributes using the following steps.

1. View the attribute data in tabular form using the GIS software.
2. Scan each attribute column, sorting the table if necessary, to identify values that are outside of their domain (i.e. list or range of acceptable values).
3. Scan for values that are not formatted properly (e.g. invalid date formats).
4. Scan for text values that are misspelled.

For each of the randomly selected features in the sample set, perform the following steps.

1. Review each attribute value and confirm that it is correct. This will likely require looking up correct values in other documents, addressing systems, etc. Familiarity with the data and the airport property itself will greatly reduce the labor involved in this task.
2. Repeat step 1 for all of the randomly selected features. If the number of allowable errors is exceeded, stop and record that the drawing has failed due to inaccurate attributes.
3. If all randomly selected features are checked and the number of allowable errors has not been surpassed record that the drawing has passed attribute quality assurance.

4.5 Checking Absolute Positional Accuracy

Critical to the functionality of any spatial data repository is that all features are accurately represented in 3 dimensional geographic space so that they can be layered on top of one another and queried in relation to one another. Absolute positional accuracy refers to the difference between a feature's location in the GIS versus its true position on the earth's surface. In this case, the Mineta San José International Airport Grid is the geographic space used to represent the earth's surface, so the purpose of this step is really to check the feature's location to its true Airport Grid location. More information on Airport Grid can be found in document SJC-ACM-AIMS-2400.

For data sets that are provided from sources not required to adhere to SJC standards, it is likely that the data will not be in Airport Grid. In these cases, the data must first be transformed into Airport Grid using the transformation procedure described in SJC-ACM-AIMS-2400.

To validate that the absolute positional accuracy of submitted GIS data meets SJC accuracy requirements, the following procedure should be used:

1. The GIS layer being tested should be compared with a source of higher accuracy. For drawings of elements that are outside (i.e. the outside domain), the SJC orthophotos should be used. For drawings of elements that are inside buildings (i.e. the inside domain), a floorplan whose spatial accuracy exceeds SJC accuracy requirements should be selected. Once the appropriate source of higher accuracy has been identified, the data being tested should be over overlaid on top of it.
2. For each of the randomly selected features, measure the distance between the position shown on the submitted data versus the position shown on the source of higher accuracy. Refer to CADD & GIS Accuracy Specification document (SJC-ACM-AIMS-2611) for accuracy requirements.
3. Repeat step 3 for all randomly selected features. If the number of allowable errors is exceeded, stop and record that the drawing has failed due to positional accuracy errors.
4. If all randomly selected features are checked and the number of allowable errors has not been surpassed, record that the drawing has passed positional accuracy quality assurance.

5.0 REPORTING OF QUALITY ASSURANCE RESULTS

The results of each QA procedure should be recorded as a means to communicate identified errors to the individuals who will be loading data into the data repository, in the case of data that passes QA tests, and to staff or contractors who will be required to correct drawings, in the case of data that fails. The results should identify the file name, format, submitter, date submitted, GIS layer, and feature ID, nature of the error and any relevant notes. The form shown in Appendix A should be used for this purpose.

APPENDIX A – Sample Form for Reporting of QA Results

NORMAN Y. MINETA SAN JOSE INTERNATIONAL AIRPORT

GIS Data Submittals
Results of Quality Assurance

File Name ______________________

Format
- ☐ GeoMedia Warehouse
- ☐ ESRI Shape

Submitted By ______________________

Date Submitted ______________________

QA Performed By ______________________

Date QA Performed ______________________

Results

☐ PASS

☐ FAIL

Layer Name	Feature ID	Nature of Error	Notes