Overview: **Spatial Quality Anomalies Diagnosis (SQUAD) Tool for ArcGIS**

1. Introduction and Background

Knowledge about health facility locations is important in addressing HIV, maternal and child mortality, and other issues. As a result, there has been a rapid growth in large geospatial data sets, such as master facility lists (MFLs). An MFL and other similar lists typically contain locations of health facilities as well as attributes of the facilities—including name, address, or the administrative unit in which the facility is located.

Assessing the quality of these data sets can be challenging, because there are two types of possible errors: **spatial errors** and **attribute errors**. Assessing spatial errors involves looking at such things as the presence of a coordinate, whether it is properly recorded, and the accuracy of its location. Assessing attribute errors involves determining whether attributes such as site name or site ID are correct.

When you work with large spatial data sets, manually reviewing each record to validate both location and attribute information can be prohibitively time-consuming. A more effective approach would be to look for anomalies in the data that may indicate data quality issues. MEASURE Evaluation's **Spatial Quality**

Anomalies Diagnosis (SQUAD) Tool—https://www.measureevaluation.org/resources/tools/geographic-information-systems/squad-tool —identifies six types of anomalies in spatial data:

Anomaly 1 Missing coordinates	Anomaly 4 Duplicate facility name
Anomaly 2 Coordinates truncated; possible typos	Anomaly 5 Site falls outside of expected district, but near a district boundary (> 2 kilometers)
Anomaly 3 Duplicate coordinates	Anomaly 6 Site falls far outside of expected district (> 2 kilometers)

2. Prerequisites to Use the Tool

Two inputs are required to use the SQUAD Tool. Both should be stored as feature classes within a geodatabase. (A properly formatted sample geodatabase is provided with the tool.)

The tool works with ArcGIS v10.5 with an advanced license.

Here are the details about each of the two required feature classes:

Sites. Should be a point feature class of sites that contains separate fields for:

- Unique identifier (can be the auto-generated "ObjectID" field, if necessary)
- Name of site
- Name of administrative unit in which the site is supposedly located

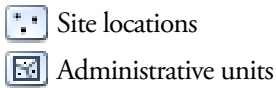- Separate fields for X and Y coordinates (longitude and latitude) of site

Administrative units. Should be a polygon feature class file for administrative units in the country or region of interest. The file should contain a field for the name of each administrative unit (e.g., "district"). These names should come from the same standard list of administrative unit names that the sites use.

If the administrative unit feature class contains especially detailed polygons, it may be necessary to use the "Simplify Polygon" tool to eliminate some of the finer detail, so that the tool will not run excessively slowly. (The recommended geodatabase size is 2 megabytes or less.)
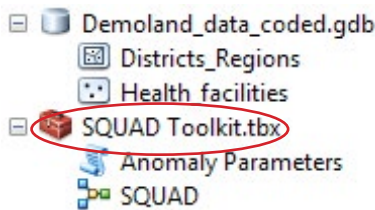
3. How to Use the Tool

1. Create a new .Mxd file in arcgis.
(An example .mxd file, SQUAD_Demo.mxd, has also been provided to permit users to test the tool.)

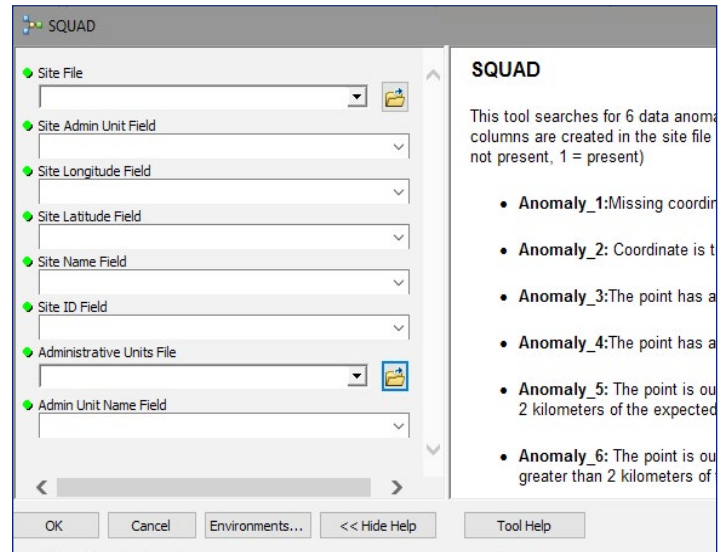2. Load the two required feature classes.

    ⬚ Site locations

    ⬚ Administrative units

3. Add the "SQUAD Toolkit" toolbox to the toolbox pane.

- ⊟ 🗄 Demoland_data_coded.gdb
  - ⊞ Districts_Regions
  - ⋰ Health_facilities
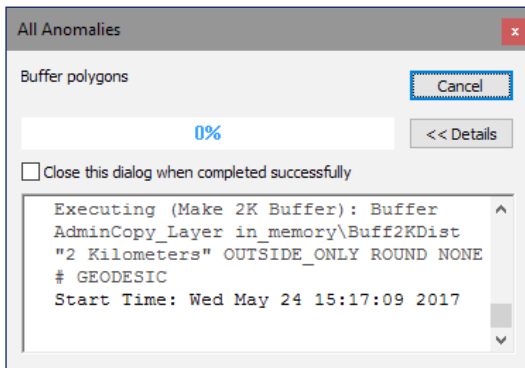- ⊟ 📦 SQUAD Toolkit.tbx
  - ⧉ Anomaly Parameters
  - ⧉ SQUAD

4. Double-click on the "SQUAD" model to run it.

A dialogue box will open. You must identify the inputs to the tool, as illustrated here:

SQUAD

This tool searches for 6 data anoma columns are created in the site file not present, 1 = present)

- **Anomaly_1:** Missing coordin
- **Anomaly_2:** Coordinate is t
- **Anomaly_3:** The point has a
- **Anomaly_4:** The point has a
- **Anomaly_5:** The point is ou 2 kilometers of the expected
- **Anomaly_6:** The point is ou greater than 2 kilometers of

4. Results from the Tool

While the tool is running, a progress box will appear. When the script has run, the box will say "Completed," and you may close it.

Fields for each anomaly code will be added to the attribute table of the sites feature class. A value of "1" means the anomaly exists for the record, while a value of "0" means there is no anomaly.

Records that have an anomaly can then be investigated to determine the quality of the data. For instance, it could be all right for sites to be in the same location or to have the same name. If an Anomaly_5 is found, the error could be inaccuracy of the administrative unit boundary.

OID *	Shape *	CID	UID *	Anomaly_1	Anomaly_2	Anomaly_3	Anomaly_4	Anomaly_5	Anomaly_6
1	Point	1	17ed905f-a5fb-4e10-8a31-b9562106b4ad	0	1	0	0	1	0
2	Point	4	bd47fc21-0f85-4da2-8c52-0329e9826fa9	0	0	0	0	1	0
3	Point	9	ad7c3763-c736-47e9-822a-9674d3ce9070	0	0	0	0	0	0
4	Point	12	7b75a398-af2a-43a1-ba0a-92b48a37e103	0	0	0	0	0	0
5	Point	12	fbc0584c-52ea-408b-bdba-e4bd95f2890e	0	0	0	0	1	0
6	Point	16	d1d3baef-5d1a-45a4-97fa-60e400d1dca3	0	0	0	0	0	0
7	Point	16	d62ab22e-2fd7-4400-a6a2-75d6f4822cbf	0	0	0	0	0	0
8	Point	18	911394ed-2708-47a6-8e3e-4250f7a5999e	0	0	0	0	0	1
9	Point	20	092d0702-1450-414a-a025-0a6b14d06561	0	1	0	0	0	0