Driving data quality improvement by a rules based method

Anna Williams
Driving data quality improvement by a rules based method
Examples of data quality rules for Critical Monitoring Points:

- There should be no more than one live CMP per PA
- Reference of the CMP should match reference of the DPA e.g. CMP 0010621CMP_01 belongs to DPA 001-06-PA-02
- The data logger reference identifier must be populated within 30 days of CMP creation in GIS
- The data logger reference must have a prescribed format e.g. H13228_01

Driving data quality improvement by a rules based method
Driving data quality improvement by a rules based method

THAT DATA IS ILLOGICAL
Driving data quality improvement by a rules based method

Data issue identified

Data team works with SMEs to identify data sources for analysis

Joint team develop data quality rules and metrics

Rules implementation via workflow

Metrics results are stored and Data Quality dashboard created

Data stewards and users collaborate to resolve issue

Data quality and rule management process
Driving data quality improvement by a rules based method

**GIS Clean Water Asset - Data Quality**
Weekly View for Data Stewards

<table>
<thead>
<tr>
<th>Criticality</th>
<th>Category</th>
<th>DQ Metric Name</th>
<th>DQ Dimension</th>
<th>% Passed</th>
</tr>
</thead>
<tbody>
<tr>
<td>High</td>
<td>Critical Monitoring Points</td>
<td>dkm_gis_cmp_one_per_dpa_accuracy</td>
<td>Accuracy</td>
<td>100.00%</td>
</tr>
</tbody>
</table>

**Accuracy**

- **% Passed:** 100.00%
- **# of Tests:** 3,576
- **Total Failures:** 0

**Issues**

- **362 Issues**
  - 2018 Q3: 13
  - 2018 Q4: 28
  - 2019 Q1: 35
  - 2019 Q2: 14
  - 2019 Q3: 24

- **0 Issues**
  - 2018 Q3: 0
  - 2018 Q4: 0
  - 2019 Q1: 0
  - 2019 Q2: 0
  - 2019 Q3: 0

To drill into the details of the failing records click on the required 'DQ Metric Name' from the table below:

<table>
<thead>
<tr>
<th>DQ Metric Name</th>
<th>Dqm Name</th>
<th>Ind Critical</th>
<th>DQ Dimension</th>
</tr>
</thead>
<tbody>
<tr>
<td>dkm_gis_cmp_one_per_dpa_accuracy</td>
<td>Live Critical Monitoring Points (CMPs) should never have more than one Discrete Pressure Areas (DPA)</td>
<td>High</td>
<td>ACCURACY</td>
</tr>
</tbody>
</table>
Driving data quality improvement by a rules based method
Driving data quality improvement by a rules based method

### Detail Dashboard

#### Rule Name:
dkm_gis_cmp_one_per_dpa_accuracy

#### Failing Records:
253

### DMA Reference

<table>
<thead>
<tr>
<th>1st Report Key</th>
<th>1st Key Source</th>
<th>2nd Report Key</th>
<th>2nd Key Source</th>
<th>Rule Name</th>
<th>Table Name</th>
<th>DMZ_NAME</th>
</tr>
</thead>
<tbody>
<tr>
<td>44166338</td>
<td>FEATURENUMBER</td>
<td>201-11</td>
<td>DMZID</td>
<td>dkm_gis_cmp_one_per_dpa_accuracy</td>
<td>GIS_CMOSW1T11</td>
<td>SOUTHPORT</td>
</tr>
<tr>
<td>44147239</td>
<td>FEATURENUMBER</td>
<td>199-09</td>
<td>DMZID</td>
<td>dkm_gis_cmp_one_per_dpa_accuracy</td>
<td>GIS_CMOSW1T11</td>
<td>SOUTHPORT</td>
</tr>
<tr>
<td>44224807</td>
<td>FEATURENUMBER</td>
<td>220-06</td>
<td>DMZID</td>
<td>dkm_gis_cmp_one_per_dpa_accuracy</td>
<td>GIS_CMOSW1T11</td>
<td>BURY</td>
</tr>
<tr>
<td>44226460</td>
<td>FEATURENUMBER</td>
<td>220-06</td>
<td>DMZID</td>
<td>dkm_gis_cmp_one_per_dpa_accuracy</td>
<td>GIS_CMOSW1T11</td>
<td>BURY</td>
</tr>
<tr>
<td>44220765</td>
<td>FEATURENUMBER</td>
<td>253-04</td>
<td>DMZID</td>
<td>dkm_gis_cmp_one_per_dpa_accuracy</td>
<td>GIS_CMOSW1T11</td>
<td>ROCHDALE</td>
</tr>
<tr>
<td>44207655</td>
<td>FEATURENUMBER</td>
<td>253-11</td>
<td>DMZID</td>
<td>dkm_gis_cmp_one_per_dpa_accuracy</td>
<td>GIS_CMOSW1T11</td>
<td>ROCHDALE</td>
</tr>
<tr>
<td>44228657</td>
<td>FEATURENUMBER</td>
<td>253-04</td>
<td>DMZID</td>
<td>dkm_gis_cmp_one_per_dpa_accuracy</td>
<td>GIS_CMOSW1T11</td>
<td>ROCHDALE</td>
</tr>
<tr>
<td>44273873</td>
<td>FEATURENUMBER</td>
<td>215-13</td>
<td>DMZID</td>
<td>dkm_gis_cmp_one_per_dpa_accuracy</td>
<td>GIS_CMOSW1T11</td>
<td>ROCHDALE</td>
</tr>
<tr>
<td>44478868</td>
<td>FEATURENUMBER</td>
<td>220-04</td>
<td>DMZID</td>
<td>dkm_gis_cmp_one_per_dpa_accuracy</td>
<td>GIS_CMOSW1T11</td>
<td>ROCHDALE</td>
</tr>
<tr>
<td>44278682</td>
<td>FEATURENUMBER</td>
<td>220-04</td>
<td>DMZID</td>
<td>dkm_gis_cmp_one_per_dpa_accuracy</td>
<td>GIS_CMOSW1T11</td>
<td>ROCHDALE</td>
</tr>
<tr>
<td>44276864</td>
<td>FEATURENUMBER</td>
<td>220-04</td>
<td>DMZID</td>
<td>dkm_gis_cmp_one_per_dpa_accuracy</td>
<td>GIS_CMOSW1T11</td>
<td>ROCHDALE</td>
</tr>
<tr>
<td>44131039</td>
<td>FEATURENUMBER</td>
<td>217-07</td>
<td>DMZID</td>
<td>dkm_gis_cmp_one_per_dpa_accuracy</td>
<td>GIS_CMOSW1T11</td>
<td>BOLTON</td>
</tr>
<tr>
<td>44144503</td>
<td>FEATURENUMBER</td>
<td>214-03</td>
<td>DMZID</td>
<td>dkm_gis_cmp_one_per_dpa_accuracy</td>
<td>GIS_CMOSW1T11</td>
<td>BOLTON</td>
</tr>
<tr>
<td>44415209</td>
<td>FEATURENUMBER</td>
<td>057-09</td>
<td>DMZID</td>
<td>dkm_gis_cmp_one_per_dpa_accuracy</td>
<td>GIS_CMOSW1T11</td>
<td>LANCASTER</td>
</tr>
<tr>
<td>44437396</td>
<td>FEATURENUMBER</td>
<td>025-12</td>
<td>DMZID</td>
<td>dkm_gis_cmp_one_per_dpa_accuracy</td>
<td>GIS_CMOSW1T11</td>
<td>KENDAL</td>
</tr>
</tbody>
</table>

Download result to CSV, open in Excel, filter by DMZ, order by DMA etc.

Feature Number that can link items back to GIS

United Utilities
Helping life flow smoothly
Driving data quality improvement by a rules based method

Benefits:

• Easy to use Dashboard 😊
• Data ownership is clearly defined
• GIS Data Maintainers and Network Performance Engineers more aware of their roles when it comes to GIS data quality and impact on others
• Less data issues generated and quicker corrections at source
• The data quality rules are written in SQL and can be fine-tuned, also additional rules can be added over time
• Better quality GIS data for:
  - other teams e.g. modellers, incident managers
  - systems that are used in decision making and reporting
• Easier to interlink systems which is a positive step towards our Systems Thinking vision