

**POC Offering: DATA QUALITY OVERVIEW**

Purpose & Objectives:	<p>This POC will use a limited set of data to demonstrate our ability to:</p> <ul style="list-style-type: none"> <li>• Evaluate the status of a dataset in relation to data quality.</li> <li>• Provide insight to the type and magnitude of specific data quality issues.</li> <li>• Highlight areas requiring further investigation.</li> <li>• Outline the approach for detailed analysis, cleansing &amp; remediation, as necessary.</li> </ul>
Compelling Issues:	<p>Data quality analysis is most commonly undertaken as a precursor to data cleansing and is very effective in guiding the cleansing exercise. Data cleansing is necessary when poor data quality begins to negatively impact the business of an organisation.</p> <p>Data cleansing is also commonly performed prior to migrating legacy data to a new system, or transitioning into a data centre / cloud environment.</p> <p>Organisations are usually aware of data quality issues through a combination of anecdotal evidence (complaints by business users and customers that there are data problems) and specific issues such as problems with reconciliation.</p>
Steps:	<ol style="list-style-type: none"> <li>a) Workshop with key stakeholders and SMEs to define objectives.</li> <li>b) Install Datalynx profiling tools on client's Server.</li> <li>c) Work with client SMEs to define business rules to be evaluated</li> <li>d) Configure Datalynx software to perform statistical analysis and process rules against sample data</li> <li>e) Produce reports.</li> </ol>
Outputs:	<ol style="list-style-type: none"> <li>a) Data Analysis system report showing results of profiling</li> <li>b) Report on findings highlighting issues, a recommended approach for remediation and suggested next steps.</li> </ol>
Next Steps:	Detailed analysis of the complete dataset, applying relevant statistical and business rules to identify all issues.
Constraints:	<ol style="list-style-type: none"> <li>a) The POC will be undertaken over an elapsed period of three weeks.</li> <li>b) A sample dataset of up to 1 million records will be used.</li> <li>c) A number of criteria (rules) will be defined to identify examples and counts of the types of issues that exist within the dataset. Up to five days are included for rule specification.</li> </ol>
Client Commitment:	<p>An appropriate Microsoft Windows server must be available, with access to the relevant dataset, which will house the analysis tools. (Server minimum specifications will be provided).</p> <p>A Client SME (with an understanding of the dataset and the business use of the data) must be available for up to five days to work with the Datalynx analyst.</p>