21, rue d'Artois, F-75008 PARIS
http : //www.cigre.org

CIGRE US National Committee
2013 Grid of the Future Symposium

Automated Aggregation of Data for Asset Health Analysis

K. PHILLIPS, C. SCHNEIDER, P. CAMBRAIA, M. MUNSON
American Electric Power
United States of America

SUMMARY

Automated analysis of asset health for the purpose of efficient asset management has become a major focus for electric utilities in recent years. One of the key reasons that automated analysis has long been unachievable is the lack of a dataset that truly allows for automated aggregation and analysis. Data has typically been stored in separate, disconnected systems in a manner that does not allow for linking the many parameters associated with a substation asset. Because the sophistication of data networks and computing resources have historically trailed behind the sheer number of station assets and activities performed, organizing and collecting the data in a way that allows for automated analysis has never been a primary concern. Analysis of the data requires a dataset that can be automatically aggregated with minimal human input and maintenance. Organizing the data into a common information model (CIM) tailored to asset health is an obvious necessity. Collecting and storing the data so that it can be pulled from multiple sources into the CIM format is also essential. The paper will briefly review the current state of industry standards for asset health data models. Suggestions will be made for the future standardization and collection of data that will allow for automated aggregation of data. A real time asset health data historian currently in development at American Electric Power (AEP) will be discussed. The preliminary results and lessons learned from the data collection methods employed for AEP's asset health program will also be reviewed.

KEYWORDS

Asset health, common information model (CIM), data aggregation, data historian

kdphillips@aep.com

Introduction

Asset Health Analysis

Electrical power transmission and distribution utilities have long been tasked with upgrading, replacing, and maintaining an interconnected fleet of substation equipment. Whether batteries, switches, circuit breakers, regulators, transformers, or any other numerous assets, utilities have had to balance the competing criteria of customer reliability, safety, Operations and Maintenance (O&M) expenses and capital investments when managing their fleet effectively. In the past, during times of rapid electrical grid expansion, much of this fleet was in a younger, healthier state. Currently, many utilities are facing both an aging fleet and downward pressure on O&M expenses. For instance, at American Electric Power, 33% of transformers are 50 years or older, and nearly 18% are 60 years or older [1]. The obvious implication is that utilities now must operate a program of asset management in the most effective way possible.

Historically, asset managers were able to operate at a local level. With years of experience and a deep knowledge of the equipment at a relatively small collection of substations, an asset manager was able to effectively administer the maintenance and replacement of his or her assets. As utilities have merged with one another, total assets have increased, operations have become centralized, and the control of the funding for asset management activities has moved further from those who interact with the assets on a daily basis. These factors create the need for a centralized asset management program that includes asset health analysis, condition based maintenance, and dynamic rationalization of capital and O&M spend.

The concept of asset health analysis is not a new one. Utilities have been recording measurements of asset health indicators for years, particularly on transformers, for the purpose of ranking replacements and maintenance. These practices have relied on manual inspection and analysis of the data by a trained engineer. The concept of using a calculated health index is discussed in [2]. New programs such as the Asset Health Center (AHC) at AEP aim to reduce human involvement with the initial assessment of incoming data, predict equipment failures and required maintenance, and provide evidence and visibility for the case to renew the grid. The AHC solution (co-developed with ABB) provides analysis on top of several data sources including: nameplate and population databases, historical test records, SCADA trends, as well as data from real time health monitoring systems. The AHC attempts to solve the problem of ineffective asset management practices with intelligent data feeds and algorithms that combine expert level knowledge and predictive capability.

Transitioning Data Collection Practices

As previously mentioned, utilities traditionally have not had a primary concern for collecting and storing substation equipment data in a way that would allow for much automatic analysis. Early methods included handwritten notes stored with the technicians who took measurements or in a file cabinet at a central office. Substation and equipment routine inspections were recorded in substation control building log books. Even with the advancement of computing technology and the availability of spreadsheets and databases, data was still hand recorded and manually digitized, allowing greater opportunity for error. As more and more data was digitized, separate, disconnected databases created silos of data that could not be corroborated against one another. It was not uncommon for original SCADA (Supervisory Control and Data Aquisition) databases and inspection record databases to use different names for the same substation, given the unique bandwidth and size limitations of SCADA data. Even the advancement of connectivity and bandwidth has not guaranteed easy access and interpretation of data; cloud storage now allows for utility data to be directly uploaded from test equipment to vendor websites and databases, which can greatly complicate the issue of automatic aggregation back to the utility. To confound the issue further, utilities are now beginning to augment the historical records with the implementation of online sensors [3].

Automated Aggregation of Data

Automatic asset health analysis implies just that: analysis that can be performed with minimal human interaction in the process. Most current attempts at asset health analysis are flawed in one of two ways. Either the process is structured on top of an imperfect data set and is therefore limited in its effectiveness, or it assumes a well-managed data set which actually requires extra human interaction to maintain said data set. For the automated asset health analysis to really accomplish its goal, the entire process must assume a holistic approach when considering the level of automation. Given the history of data collection techniques and the disconnected databases that are required to populate a model intended for asset health analysis, one cannot assume that an asset health data model (industry standard or otherwise) is readily available. In fact, it is proposed here that an *ideal* asset health data model is likely not easily aggregated from existing data sources, but instead must be automatically aggregated from new, intelligent data sources. **In other words, one must consider the need for automatic data aggregation when designing the collection, storage, organization, attributes, and history of each source database and datapoints that may be used for asset health analysis.** Of course, asset health analysis may not be the only use (automated or manual) of these data sets. While additional criteria must be considered during the design of data sets, this paper will focus on automated asset health analysis as the driving force. Figure 1 shows an example hierarchy of source data models, aggregated data models and analysis layers.

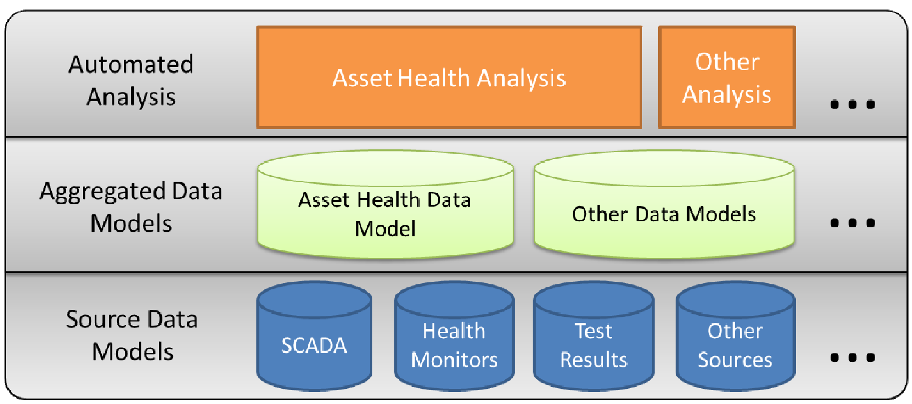

Figure 1 Hierarchy of sourced data, aggregated data, and analysis layers

<u>Data Sources</u>

Before discussing the best practices for data aggregation, it is necessary to review the available data sources for asset health analysis. Table 1~~Table 1~~ describes several possible sources.

Table 1 Possible data sources for asset health

Source / Type	**Description**	**Updates**
Nameplate	Typical nameplate data includes manufacturer, year of manufacture, voltage levels, power ratings, model number, serial number, etc.	Static
Manufacturer Specifications	Manufacturer specifications include suggested maintenance steps and intervals. Also included here is specific performance information such as contact travel speeds for breakers or factory bushing power factor for transformers.	Static
Operator Specifications	Operator specifications include utility maintenance steps and intervals (compliance based or otherwise). This data set also includes things like acceptable limits for measurements such as DGA levels.	Static
Accounting Data	Accounting data includes initial cost, cost of ongoing maintenance, depreciation timeline, availability of spare parts, cost of failure, etc.	Periodic

Source / Type	Description	Updates
Status and Network Topology	This data describes if the unit is in service or spare, where it is connected in the network, and tracks the history of changes.	Periodic
Test Results	Results from equipment tests such as insulation tests, oil quality, breaker timing, etc.	Periodic
Inspection Results	Results from visual inspections that cannot otherwise be automated.	Periodic
SCADA	Real time data used by operations such as loading, voltage levels, breaker status, etc.	Continuous
Health Monitors	Health monitors measure data in real time specific to asset health. Most monitors replicate or imitate traditional equipment testing or metering and also provide on-board intelligence.	Continuous
Relay Event Files	Event files can be triggered to capture information about faults or other abnormal events.	Event Based
Trouble and Failure Information	Data recorded about corrective maintenance or equipment failure. Failures can be subsystem failures or complete asset failure.	Event Based

As can be seen in the previous table, there is a vast array of data available for asset health purposes. The data originates from several different sources, and it is updated with varying frequency. The nature of these data sets provides several challenges when aggregated for asset health.

Challenges

Inadequate Source Data Models

Source data models themselves may be inadequate for the purpose of automated analysis. It may be possible that a data set is simply incomplete. For instance, if a measured test result is temperature dependent, the temperature at the time of the test must be recorded with the test result. If lab equipment can introduce error into the lab results, it is important to track the particular lab that performed the test. On the other hand, it may be the case that the needed data is recorded but must be pre-interpreted before being aggregated. For instance a data model for equipment failure might include a notes section that explains whether the failure was permanent or temporary, and whether a cause was identified. In this case, the data was tracked, but not in a useful way. One would have to read through a comments section attached to each failure report to further organize the individual failures. A fundamental assumption for any data aggregation tool intended for real world application must be built around the fact that the data to be gathered will be most certainly incomplete and to a certain extent inaccurate (or in some cases contradictory)

Discrepancies between Sources for Asset Identification

Different source data sets may refer to the same asset with unique, non-linkable, designations. For instance, a maintenance record database, a SCADA database, and a protective relay configuration may refer to the same substation with three different names. In this case, a translation table must be used to link the data from each of the three data sources to a single substation. These translation tables require manual intervention to match the names between the three databases. Additionally, the different data sources may use a different hierarchy for defining what an individual substation is. Where one system refers to a single substation, another system may refer to multiple substations divided up by voltage level or ownership (transmission vs. generation). This implies that the translation table will not exist as a one-to-one relationship between station names.

Identification by Position vs. Asset

When a data model intended for asset health refers to an asset, it is referring to the physical piece of equipment. The implication is that the piece of equipment could experience a failure, be moved, be refurbished, exist at different locations, become energized or de-energized, or any number of other status changes. The asset health data model tracks data against the asset itself, taking into account status changes. Other data models, such as SCADA models, track data against a position rather than an asset. This means that the history and trend of data reflect a location in a network topology, regardless of which asset filled that position. Another illustration of this challenge is the fact that a protective relay configuration may refer to a line position, but contain fault data that can be applied to a breaker or transformer. In both the SCADA and relay cases, the data can be attributed to a certain asset if the topology and status history are known. Without an adequate data model for topology and status, the data translation becomes not only a manual maintenance process, but also one that is perpetual and triggered by changes in the source data.

Undefined Data Model

Another challenge to automated aggregation of data is an undefined or changing data model. Data sets that were typically not considered for any automatic analysis were not constrained to a standard template. An example of this is a relay event report file. The triggers, data points, lengths, headers or precision may change from one relay configuration to another. Without a standard, automatic parsing becomes impossible. This issue is not limited to relay event files, but could exist with other data sets such as a battery conductance test report.

Suggested Best Practices for Asset Health Data Aggregation

<u>Industry Practices</u>

Common Information Model

An industry standard known as the Common Information Model (CIM) attempts to address the needs of data models for power systems. The CIM is a collection of three standards [4]:

- IEC 61970-301: abstract model of major elements in an electric utility system that are part of the operations of the utility
- IEC 61968-11: extends the model to additional aspects of the power system, including the distribution system, asset tracking, work scheduling, and customer billing
- IEC 62325-301: models data exchanged between participants in electricity markets

At the most basic level, these standards define a hierarchy of equipment models and attributes and allow the equipment to be linked together to form a topology. The models are based on classes and inheritance: equipment classes inherent attributes from parent classes. The standards employ the Unified Modeling Language (UML) to model this hierarchy. They also use XML as the language to describe the models. When designing a new data set, whether it is a source system such as a set of maintenance records, or an aggregated data set such as an asset health database, it is important to consider the models defined by the CIM standards.

CIM for Asset Health

The current CIM standards are not geared specifically towards asset health purposes. The Asset Health Focus Community is a distinct arm of the CIM users group tasked with developing the requirements of a CIM framework for asset health analysis and condition based maintenance [5]. The scope of this effort is to set requirements for both the input and output side of the asset health analysis. The input data is generally what has been discussed throughout this paper. The output data can include things such as messages, alerts, risk assessment, maintenance triggers, equipment orders, etc. Setting requirements for the output data will require further refinement of the use cases by the industry, and is outside the scope of this discussion.

The focus community will concentrate its initial efforts on defining the asset health related attributes of circuit breakers and transformers. The team will eventually deliver a UML that proposes enhancements to the current CIM models. One of the benefits of this type of standardization is that asset health software tools can be upgraded or replaced without affecting the underlying data structure. This adds a degree of flexibility for the utility.

Data Collection

Besides having the correct model for a source database, the methods in which the data is collected must be adequate enough for eventual asset health analysis. As stated earlier it must be assumed that the collected data will not be perfect. The key is how to extract that maximum amount of information from incomplete, inaccurate, and contradictory data. That is what the asset engineer does manually today. The data aggregation tool has to assess all the data and establish a confidence level with the data as a feeder to an expert system. The expert system can translate the mass of many thousands of data values and textual descriptions into a set of likely conditions on a limited amount of equipment that most needs the attention of the asset engineer.

Equipment Inspections and Testing

When manual inspections are performed, the inspector needs direct access to the source database and a data entry process that satisfies the following requirements:

1. Inspection data cannot be handwritten. The data must be entered directly into a tool that has real time access to the source database. Operator must be able to see previous data collected.
2. The equipment should be barcoded and the inspection tool should have the ability to link the barcode to the correct equipment in the source database.
3. The inspector should not collect duplicate data. The inspection database should have the ability to link real time data that is available for the equipment so that the user does not duplicate information such as loading or operations.
4. Measurement data must have data validation at the source. If the operator must input measurement data that is not otherwise available, the inspection tool should be equipped with high and low limits or other data limits that eliminate obvious user errors.
5. Subjective inspection data must be collected in a way that can be automatically assessed. The tool must present options that are clearly defined as responses to questions about subjective assessments such as the amount of rust or amount of oil leaking on an asset.
6. Equipment testing software and hardware should ideally deliver results directly to utility's maintenance database. If this is not an option, the test software needs to produce a file that is easily interpreted and able to be parsed. The file needs to contain a link to the equipment, through barcoding or some other method.

Event Based Data – Relay Event Reports

The creation of event based data is triggered by certain system operations or phenomenon rather than periodically. Data of this nature can be collected via a SCADA system or through relay event report files. Best practices for the latter case are given here:

1. The file must use a pre-defined standard format for triggers, wordbits, analog channels, length, and resolution.
2. The filename must link the file to an asset in the CIM.
3. The association with the asset must be understood from the CIM. For instance a relay may be associated with a set of breaker currents, but the currents could be applied to a breaker, transformer, line, or some combination based on the network connection.

Health Monitor Data

Health monitor data is collected by sensors connected to the substation equipment. These types of sensors are not new to the industry, but their use in a system that employs automated data aggregation is innovative. The following are some suggested best practices for dealing with the monitors and their data:

1. The amount of on-site data analysis should be limited. This level of monitor intelligence will depend on the particular application, but in general, there are only two reasons analysis should be done at the substation: (a) to calculate *immediately actionable* (operational) alarms, and (b) to process extremely large amounts of data into smaller data sets that will meet the bandwidth of the collection system. As much as possible, the raw data should be fed from the sensor to the collection database.
2. IEC61850 is the preferred communication method and protocol for communication within the substation. The standard is used to model substation protection and automation, and the technology should be applied to health monitoring.
3. The use of the data should determine the collection frequency of the aggregating system. It is assumed here that health monitors will be deployed at stations with a relatively large (>56kbps) bandwidth. Health monitors will invariably be competing with other technologies such as SCADA and syncrophasors for that bandwidth. The health monitoring system designer should let the use cases determine the frequency of data pull from the individual monitors. The designer must answer the question of how fast a failure can be predicted by asset health analysis. Figure 2 shows a composite gas measurement on a transformer that failed internally over a period of roughly two hours. In this case, the real time monitor database would have to poll several times an hour to predict the failure and take corrective action.

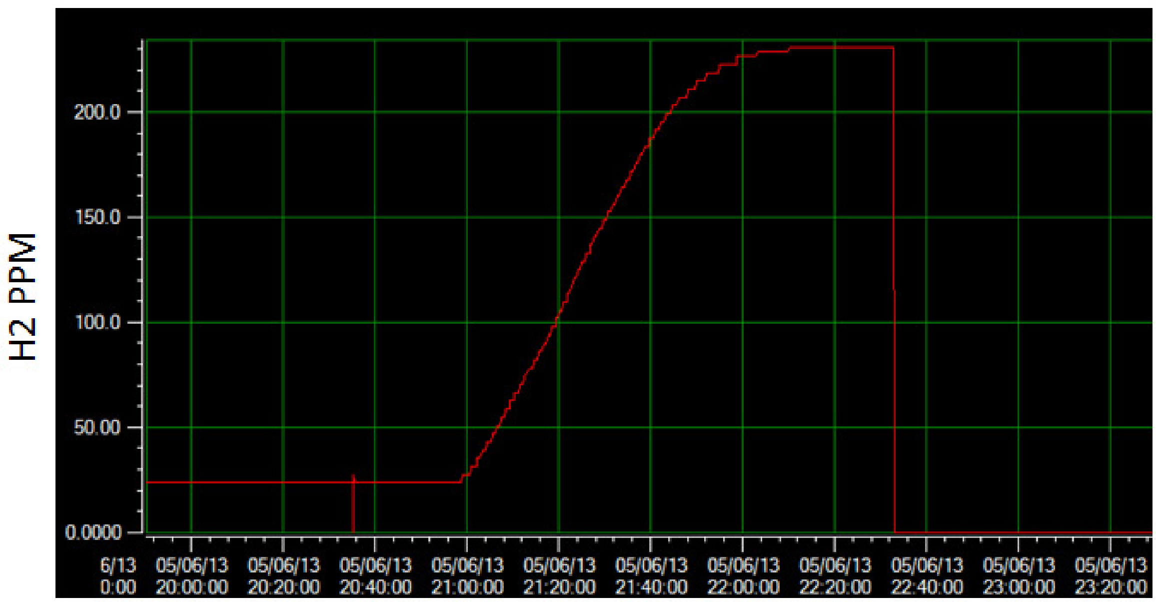

Figure 2 Real time composite gas measurement during internal failure of transformer illustrates an opportunity to predict transformer failures and gives guidance on how frequently to poll data and run algorithms.

Asset Health Data Implementation at AEP

Although a complete set of best practices could be identified and an ideal set of systems and processes could be designed, it is unrealistic for a utility to implement all of the necessary systems at once. It is therefore suggested to use a phased implementation. AEP has taken a phased approach for both the design of its Asset Health Center and the data systems that will support it. The goal of such a phased approach is to allow as much usability as soon as possible while keeping in mind future expandability.

<u>Transition to New Maintenance Tracking Software</u>

AEP has used a home-grown database and interface for tracking asset population, status, maintenance, and inspections for several years. Both the database and user interface have grown over the years with new data and functionality as requested by the business unit users. Recently, most of the budget for change requests has been limited to those that add compliance tracking capability. AEP is now beginning to transition to a new application that will manage this data set.

Because the development for the new maintenance tracking database overlapped with the development for the asset health analysis platform, AEP was forced to make a decision on how and when to bring in this existing data set. It was decided that an intermediate data warehouse would be built. The data warehouse pulled the needed fields from the existing maintenance database. The model for this data set was a copy of the existing data model in the home grown database. The justification for building the intermediate data warehouse is that upon completion of the new maintenance database, it could be connected to the intermediate data warehouse with minimal modification and impact to the asset health analysis database. See Figure 3.

As development for the new asset database continues, it has become apparent that there will be more work involved in connecting it to either the intermediate data warehouse or the asset health database than originally proposed. This is due to the fact the structure of the database will change significantly to meet many of the aforementioned data collection requirements. The lesson learned from this development is that the creation of the intermediate data warehouse was a benefit to the project, but not in the originally intended way. Although it will not minimize the amount of work to connect the new database, it did allow for up front work to be done in selecting the tables and fields in the original database that were important to asset health analysis. This work and documentation has proved useful in the planning of the new asset database. In hindsight, the intermediate data warehouse could have been created as a true bridge between the existing data model and a more improved data model, rather than a copy of the existing data model.

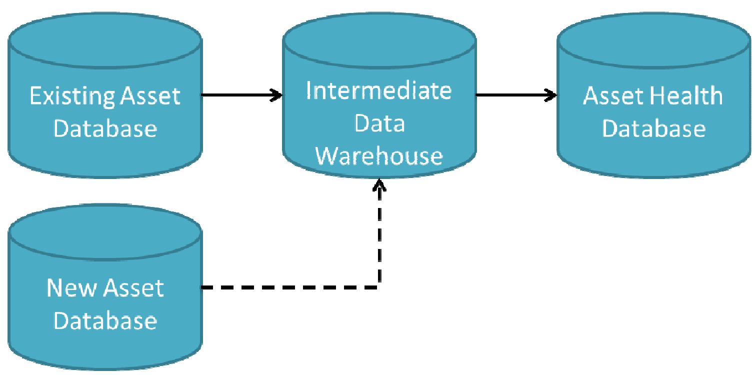

Figure 3 The intermediate data warehouse is used to transition from an existing asset maintenance database to a new database with more complex asset model.

<u>Connection of SCADA and Monitoring Data</u>

Similar to the implementation of the asset population and maintenance records, the implementation of the real-time data has taken a phased approach with a necessity for future expansion. Analog SCADA data at AEP is tracked in a data historian product which can store large amounts of analog trend data in a very efficient manner. It was decided that this same type of database would be used to aggregate both SCADA data and health monitoring data for asset health purposes.

The system currently being developed is the System Parametric Information (SPI) database. The system will be able to compile data from multiple SCADA systems, equipment health monitors, and station computers. Figure 4 shows the architecture for the SPI. The overall business definition of the SPI is to aggregate and store data that is non-operational (data that is not immediately actionable). In other words, allow the SCADA system and system operators to continue to receive data that is immediately actionable such as critical alarms, breaker operations, and loading information. The SPI will alleviate the load of the SCADA system by taking analog readings such as transformer Disolved Gas Analysis (DGA) measurements that require extra analysis before action.

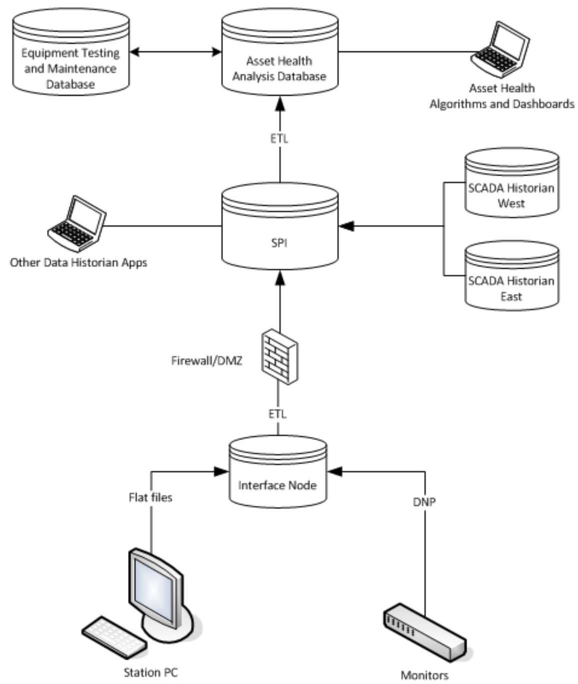

Figure 4 System Parametric Information (SPI) architecture. The SPI aggregates data from multiple SCADA historians, equipment health monitors, and station computers.

The initial phase of the SPI will concentrate on aggregating existing data from SCADA and new data from transformer monitors. The data from monitors will be extracted via DNP connection from the SPI directly to the monitor through the substation LAN. Connection to the SCADA historical data is accomplished through the proprietary connection method of the data historian software. This is easily achieved because the SCADA historian and SPI are using the same software package. The real challenge is organizing existing SCADA data which uses tag names that refer to asset locations rather than individual assets as tracked in the asset maintenance database.

To link the data back to the individual pieces of equipment, a mapping database is created that resolves the two datasets. The mapping database also assigns a "measurement code" to the datapoints of the SPI. The measurement code references a measurement code index that explains what the measurement is and what type of asset it can apply to. For instance, one of the individual measurement types for a single phase transformer is phase current, whereas there are three individual measurement types for a three phase transformer that track the phase currents. Figure 5 shows the translation table and how it fits into the SPI.

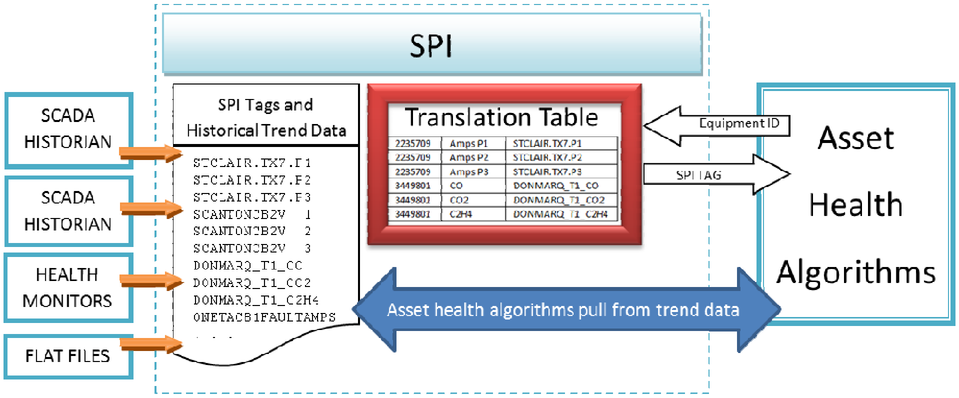

Figure 5 Translation table placement within the SPI: The translation table is used to bridge the gap between existing SCADA point tags and equipment serial numbers in the maintenance database.

Conclusion

The automation of asset health analysis for electrical substation equipment requires a well designed and aggregated data model that represents both current and historical attributes of the equipment. This data can come in the form of maintenance records, test results, SCADA data, online monitoring, event based data, and several other sources. Historical data collection and organization practices do not lend themselves well to automated analysis because they require manual intervention to interpret and compile into a useable asset health data model. The premise of this paper is that the aggregation of data itself must be an automated function that builds a data model fit for automated analysis. One must consider the need for automatic data aggregation when designing the collection, storage, organization, attributes, and history of each source database. Of course, there are other uses for the data in these source systems, but asset health analysis was highlighted as the primary use here.

Several challenges to data aggregation were presented here. These challenges include: inadequate source data models, discrepancies between sources, identification by position vs. asset, and undefined data model. Additionally some best practices were suggested. These include the use of CIM standards for data models, as well as several best practices for the collection of data.

While ideal data models and collection techniques can be theorized, immediate implementation is unrealistic. A phased approach that allows for future expansion and connection is the philosophy of AEP’s asset health team. In this paper, the ongoing implementation of a data system for asset health at AEP was reviewed. The creation of an intermediate database for equipment maintenance results opened an opportunity to study the existing data model as well as provide guidance for a new data model for a new maintenance database which would be implemented in the future. Additionally, the implementation of a SPI system for real time data aggregation at AEP was reviewed. The creation of this system allowed a way to link asset nameplates of the maintenance database to asset position in SCADA. The work done on this system is a base that allows for future expansion.

BIBLIOGRAPHY

[1] Fleeman, J., “AEP Launches Asset Health Center,” Transmision and Distribution World, Feb 1, 2013, http://tdworld.com/asset-management-service/aep-launches-asset-health-center.

[2] Scheibe, M.; Richert, J., “Data in Hand. Now What?,” Transmision and Distribution World, June 21, 2013, http://tdworld.com/grid-opt-smart-grid/data-hand-now-what?page=1.

[3] Jahromi, A.; Piercy, R.; Cress, S.; Service, J.; Fan, W., "An approach to power transformer asset management using health index," Electrical Insulation Magazine, IEEE , vol.25, no.2, pp.20,34, March-April 2009

[4] “Common Information Model Primer” (EPRI Report 1024449 2011 Program 161 IntelliGrid http://www.epri.com/abstracts/Pages/ProductAbstract.aspx?ProductId=000000000001024449)

[5] “Asset Health Focus Community Effort - Overview” (CIMug Asset Health Users Community May 2013 http://cimug.ucaiug.org/Focus_Comms/AssetHealth/Public%20Documents/Forms/AllItems.aspx)