REGULATORY COMPLIANCE DATA MANAGEMENT

Jonathan Cuff (Neo Engineering Consultancy Ltd), Richie Murray (Neo Engineering Consultancy Ltd) and Dominic Hollewand (Neo Engineering Consultancy Ltd)

ABSTRACT

Over the last decade, there has been a significant rise in data analysis and reporting along with the emergence of big data. However, it often appears that little attention is given to ensuring data is validated for its quality and purpose. In New Zealand, many regions have geographically remote water assets. Therefore, regulatory compliance data often must pass through many remote control and telemetry components such as media converters, RTUs, PLCs and gateways before finally being stored on a database within a corporate network. It is critical that this data's integrity can be verified from source to destination.

This paper analyses data management through the entirety of a data path. Management of all elements of a data path creates a data custody chain that allows the path to be traced end-to-end. This is critical to validating regulatory compliance data and reports. A data path begins with data generation and includes all data transfers and data storage. Data generation is instrument focussed and involves the correct calibration and setup of instrument parameters. Data transfer encompasses data scaling, the volume of data transfers and data transfer protocols. Data storage includes time stamping, time synchronisation, data compression methods, data types and sometimes multiple databases.

However, the initial setup or correction of a data custody chain is only one part of data management. This paper also explores the management of instrument and software modifications along the chain. Change management ensures transparency, traceability, and verifiability of modifications to data paths.

Ultimately, this paper tests the hypothesis that a report demonstrating consistently compliant results can be hiding wildly fluctuating non-compliant events. It will lead you to ask in future – what is the data path compliance validation behind this report? Are these reported results real?

KEYWORDS

Data management, data custody chain, regulatory compliance

PRESENTER PROFILE

Jonathan Cuff works as an electrical engineer at Neo Engineering Consultancy Ltd. His background is in the design of low voltage distribution & control systems and in PLC programming.

1 INTRODUCTION

In the water industry, managers rely on compliance reports to prove that their sites are meeting their regulatory obligations. These reports are usually made up from multiple data sources including SCADA historian data and laboratory test results.

The historian data that is used to create compliance reports compares field data from instruments with the requirements of the regulations to demonstrate compliance. However, over time it is possible that this regulatory compliance data loses its validation for quality and accuracy. When the path that data takes from sensing element to information store is not validated, it is not possible to be sure that the resultant compliance reports depict an accurate and true representation of what is occurring in the field.

The drinking water standards and grading process require that monitoring equipment is regularly calibrated, that alarming is reliable and that appropriate logging is undertaken. Therefore, it is crucial that regulatory compliance data's integrity can be verified and periodically checked from source to destination. This can be of particular concern when operating close to a regulation limit as non-compliances in the field may be masked by the system.

This paper provides a high-level overview of how a compliance data point is obtained, how this data can be stored or transferred incorrectly along its path to the historian and how these issues can be avoided. This paper suggests that for there to be true confidence in the output of compliance data monitoring the three steps of design, secure and control need to be in evidence.

2 THE PROCESS OF ATTAINING A COMPLIANCE DATA POINT

Compliance data can typically be passed through multiple devices after being generated by the compliance instrument before being stored in a historian. To help explain this compliance data path, and provide real-world context, this paper is based on an example path of turbidity compliance data from a small, rural water treatment plant. This site will be operated by a PLC with an RTU that sends the data back to a central SCADA via a UHF radio system. This example has been selected as it is considered somewhat typical and provides a good opportunity to demonstrate the challenges small rural sites face with onsite and intersite communications. Figure 1 shows graphically the compliance data being generated then transferred and stored through multiple devices before being stored in a historian.

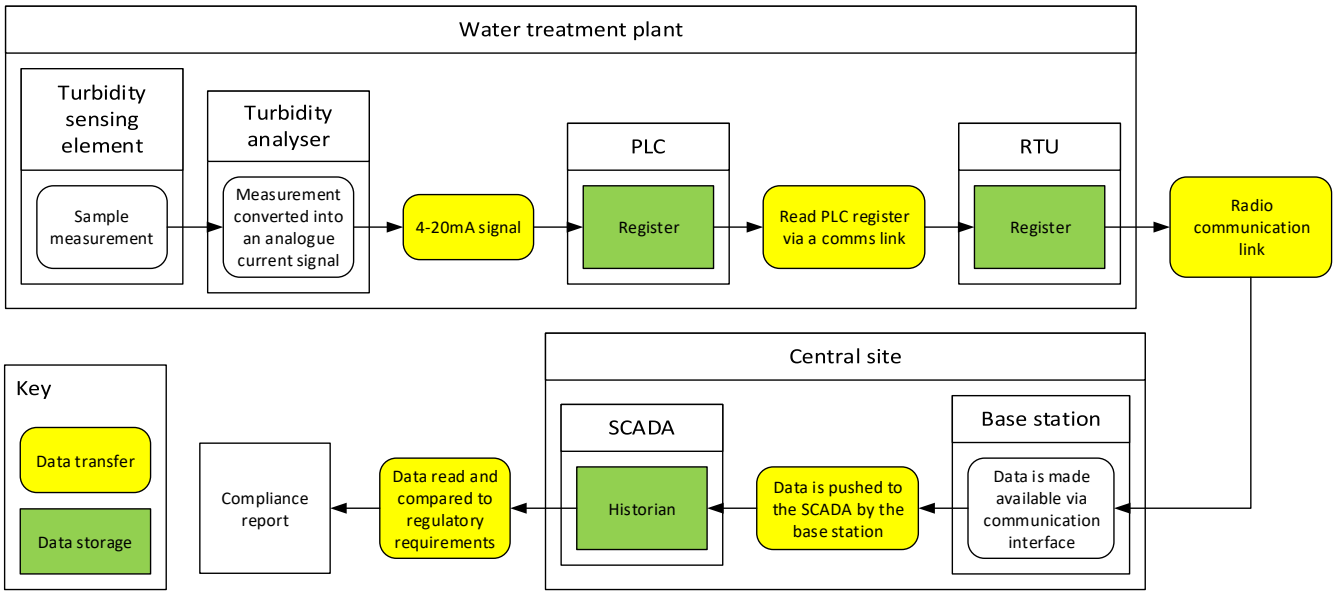

Figure 1: The compliance data path for a turbidity data point for an example rural water treatment plant

It is important to note that different systems to that shown in Figure 1 will still be exposed to similar risks that will be highlighted in this example.

2.1 THE TURBIDITY DATA PATH

The data is generated at the turbidity sensing element and analyser at the water treatment plant. The sensing element measures the turbidity content of the water and passes this on to an analyser. The measurement is converted by the analyser into a 4-20mA signal so that it can be read by the site PLC. The signal is based on a 0-5NTU scale as defined by a user in the analyser configuration.

The PLC is able to determine the value measured by the compliance instrument based on the 4-20mA value it receives. This data is stored as a floating point number within the PLC to ensure that there is no loss of data precision. However, storage in the water treatment plant PLC is fleeting; each data point is only held within the PLC until it next updates within a few milliseconds.

An RTU is used to read the turbidity value in the PLC via communication link e.g. Modbus RTU. As with the PLC, data is stored as a floating point number within the RTU. However, in contrast to the PLC, the RTU timestamps and stores the data that it reads from the PLC. This ensures that: 1) regulatory compliance data is able to be referenced to the time it was measured; and 2) the data is not lost in the event of a communication failure. The RTU is capable of storing data from the last few days.

Turbidity data between the treatment plant and the central site is transferred via a UHF radio link between the site RTU and the central site base station. Communication between the sites is generally initiated by the base station, however, the treatment plant can raise an immediate "unsolicited response" if it needs to send an alarm to the base station. When the base station initiates communication with the treatment plant, all of the treatment plant's RTU's data is sent to the base station. The RTU then clears its memory of all sent data.

The turbidity data at the base station is made available to the SCADA directly through a communication interface. This interface pushes data to the SCADA upon event i.e. if the data received from a remote site is different from that previously read. The SCADA, in turn, stores the data to a historian database that is used to create compliance reports typically via a third party software application.

3 WHAT COULD GO WRONG ON A COMPLIANCE DATA PATH?

In the example presented in section 2, there are at least eight identifiable interfaces where errors can be introduced. Some of the most common error causes are explained below. A single example is provided where required to provide some context. However, it is important to note that, with the exception of measurement drift, there is a risk of introducing these errors at most or all data storage locations and data transfers along a data path.

3.1 SENSING ELEMENT MEASUREMENT DRIFT

If periodic maintenance, cleaning and calibration is not performed on an analyser and sensing element its measured value may drift from the true value. This will fundamentally affect the quality of compliance data being generated and will not give a true value. The drift can either lead to higher or lower levels being recorded. The worst case is that the data shows that compliance is being achieved but the measured variable is actually out of the compliance range. Measurement drift is not possible to monitor through traditional means and is thus difficult to detect and therefore regular maintenance is essential.

3.2 SCALING MISMATCH

Scaling refers to the range and resolution of data that is set up in the devices along the data path. The greatest probability of a scaling mismatch occurring is when the instrument's settings are updated or the instrument is replaced. In the worst case, this error may be difficult to detect as the data values recorded can be reasonable.

Figure 2 shows an example where a new turbidity instrument has been scaled between 0 and 5NTU. However, the PLC has been scaled between 0 and 3NTU. This will result in higher than actual turbidity data being read by the PLC e.g. a reading of 1.5NTU at the instrument will be read as 0.9NTU at the PLC.

Instrument	
Reading	1.5
0% Scale (4mA)	0
100% Scale (20mA)	5
Measurement unit	NTU

→

PLC	
Reading	0.9
0% Scale (4mA)	0
100% Scale (20mA)	3
Measurement unit	NTU

Figure 2: Example of a mismatch between a turbidity instrument's scaling and its PLC's scaling

3.3 UNIT MISMATCH

A unit mismatch occurs where one device is set up to accept a data value with a certain unit but receives the data scaled for a different unit. As with scaling mismatches, the greatest probability of a unit mismatch occurring is when the instrument's settings are updated or the unit is replaced. This form of error will produce a step change in the data value. If trends are being regularly observed this step change may be noticed and the error resolved.

As a unit mismatch is particularly unlikely for turbidity, Figure 3 shows an example where a new flow transmitter has been scaled between 0 and 5L/s. However, the PLC has been scaled between 0 and 5m3/hr. This will result in much lower than actual flow data being read by the PLC e.g. a reading of 2.5L/s at the instrument will be read as 2.5m3/hr at the PLC.

Instrument	
Reading	2.5
0% Scale (4mA)	0
100% Scale (20mA)	5
Measurement unit	L/s

→

PLC	
Reading	2.5
0% Scale (4mA)	0
100% Scale (20mA)	5
Measurement unit	m3/hr

Figure 3: Example of a mismatch between a flow instrument's unit and its PLC's unit

3.4 DATA LOCATION MISREFERENCE

When two devices communicate via a communications protocol, data could be read from an incorrect register. This form of error is most likely to be introduced during PLC or RTU programming updates. This could potentially have a significant effect on the data value recorded in the historian as the data recorded will have no correlation to the actual instrument reading. As the data will most likely be vastly different from what is expected, it is usually easy to spot these errors. However, there is a small possibility that the data in the incorrect register provides a reasonable reading.

An example of this form of error is the site RTU software is written and the turbidity register referenced in the site PLC is accidentally typed as 40010 instead of 40001. This results in a PLC register with different data i.e. not turbidity data, being read and recorded by the RTU.

3.5 DATATYPE MISMATCH

It is possible to lose data precision if an inappropriate datatype is used as the data passes through the path. This form of error is most likely to be introduced during programming updates to devices in the data path. While the effect of this could be severe, the unusual resultant trends on SCADA should make this error simple to recognise.

An example of this is a floating point number being stored as an integer within the site RTU. This will result in the loss of any information to the right of the decimal place in the data e.g. 1.5NTU becomes 1NTU.

3.6 TIMESTAMPING AT BASE STATION

It is possible in some old RTUs that data is stored but not timestamped locally. Data is just stored sequentially with the RTU and the timestamping is performed when the data reaches the base station. This becomes problematic when radio communications are lost. If the example water treatment plant used one of these RTUs then all turbidity data stored in the RTU during a communications outage would be time-stamped incorrectly.

This would occur due to the delay between actual data measurement time and the time when the data is received at the base station.

It is important to note that it is possible to time shift the data at the SCADA based upon the RTU's data recording frequency. However, if there is an onsite issue e.g. some readings are not taken, the time-shifted data will be offset incorrectly if the time(s) of the issue cannot be identified.

3.7 DEVICE "LOCKUP"

Sometimes devices on the data path might "lockup" which results in the data not updating on the device. This is identifiable as a flat line on a SCADA trend. These lockups occur randomly and are generally not predictable although it is possible over time to identify the risk of different devices locking up based on the frequency of previous lockups. As with datatype mismatches, the effect of this could be severe. However, the resultant flatline SCADA trends make this error easy to identify although this requires that trends are observed. Alternatively, watchdog alarms may be set up for lockup prone devices to reduce the reliance on trend monitoring.

3.8 COMPLIANCE REPORT GENERATION

Compliance data generation may be performed in-house or by a third party supplier. There is an opportunity for errors in compliance reports if data must be manually manipulated for some reason. This could result in a relatively inconsequential or major issue depending on the nature of the mistake made.

4 HOW CAN COMPLIANCE DATA PATH RISKS BE AVOIDED?

Compliance data path risks can be avoided by the adoption of a formal process that requires the path be consciously engineered, documented and safeguarded. The following three stages are recommended:

1. Design the data path by implementing, commissioning and documenting a data custody chain.

2. Secure the devices and software involved in the data path.

3. Control changes to the data path.

4.1 DESIGN THE DATA CUSTODY CHAIN

Compliance data path issues can be avoided by designing and testing data custody chains. These custody chains involve documenting how data will be stored and transmitted between each device for each data path. Through the calibration verification process, this can be tested to provide end to end confidence.

The data custody chain in Figure 4 shows all of the critical information through the compliance data path of the turbidity data point for the example water treatment plant. It should be defined prior to the implementation of the data path and verified at the time of commissioning. This ensures that each device will be set up appropriately and that the data in the historian is correct by ensuring that any maintenance personnel (instrumentation, PLC, RTU, SCADA) have a known, documented design to work with.

Instrument		PLC				RTU		RTU base station		SCADA		Historian	
Internal		Input		Output		Internal		Internal		Internal		Internal	
Last calibration	1/01/2019												
Next calibration	1/04/2020												
0% Scale	0	0% Scale	0							0% Scale	0		
100% Scale	5	100% Scale	5							100% Scale	5		
Unit	NTU	Unit	NTU	Unit	NTU	Unit	NTU	Unit	NTU	Unit	NTU	Unit	NTU
Analogue scale	4-20mA	Analogue scale	4-20mA										
		Data tag	NTU_in	Data tag	NTU_out	Data tag	NTU_in	Data tag	SiteX.NTU	Data tag	SiteX.NTU	Data tag	SiteX.NTU
		Data type	REAL	Data type	REAL	Data type	REAL	Data type	FLOAT(32)	Data type	FLOAT(32)	Data type	FLOAT(32)
		Words used	2	Words used	2	Words used	2						
		First register	30001	First register	40001	First register	40001						

Figure 4: Example custody chain for the example water treatment plant turbidity data path

Once this framework has been implemented, a data custody chain should be created for each compliance instrument.

4.2 SECURE THE DATA PATH

Following the design and implementation of the data custody chain, the devices and software on the data path need to be secured to prevent unauthorised modifications. This can be achieved by considering both physical & software security and human factors.

Key physical and software factors to consider when securing the data path are:

1. The transmitter should be set up so that its parameters cannot be inadvertently modified.
2. The PLC, RTU, radio, etc. should be secured within an enclosure.
3. The base station, SCADA PC, historian server, etc. should be located in a secure location.
4. Backups should be performed and tested so downtime is minimised should there be an issue.
5. Appropriate IT security should be in place e.g. network separation, firewalls, etc. to prevent unauthorised external access.

Human factors that need to be considered are:

1. Access should be controlled on the basis of competent personnel only.
2. Rules should be implemented about when entry/access can and can't be effected.
3. Understand who holds keys to access secured devices.
4. Personnel who undertake works on the data path should be trained and briefed around expectations around access.
5. Automate data paths where possible to minimise human interaction.

4.3 CONTROL CHANGES TO THE DATA PATH

Change management relates to people, process and technology. It is necessary to consider all of these factors when undertaking change management. In particular, a thorough process is critical to ensure that changes are carried out correctly and verified. An example is as follows:

1. People involved in data paths should understand what is required of them, have the necessary skills to make changes and know the change management process.
2. A control process should be defined that requires at a minimum:
    - Application to make a change.
    - Review and risk assessment of the change.
    - Authorisation to carry out the change.
    - Communication and recording of when the change occurred or is planned.
    - Confirmation that the change has been successful and proven end to end.
    - Confirmation that the data custody chain and all other documentation has been updated.
    - Closeout and acceptance at the end of the change.
    - Periodic audit of the process.
3. The available technology needs to be considered to determine whether the right tools are being used to:

- Prevent unauthorised changes.
- Confirm changes.
- Provide backups.
- Automate data paths.

5 CONCLUSION

All compliance data paths need to be designed, secured and controlled. This can be achieved by implementing data custody chains for all compliance data paths to provide confidence in compliance reports. Additional security and wrap around change management are required to ensure ongoing verification of the data paths.

To be confident that a best practice approach has been adopted, a water manager or compliance manager should be able to see the following elements:

1. The data path has been defined for each measured parameter along its custody chain.
2. Security of all critical elements is in place and prove this is being maintained.
3. Change management procedures are in place and can be demonstrated.

If, as a water or compliance manager, you can use the “show me” test to compare documentation with device settings and have confidence that all of the above elements are in place, you can have confidence that what your reports and trends are showing has roots in reality.

GLOSSARY

NTU	Nephelometric turbidity units
PLC	Programmable logic controller
RTU	Remote terminal unit
SCADA	Supervisory control And data acquisition
UHF	Ultra high frequency
VPN	Virtual private network