Data integrity principles

By **Dr Peter Johnson**, CEO, WaterOutlook.

As CS Lewis once said: "Integrity is doing the right thing even when no-one is looking." When it comes to data integrity, the right thing is ensuring there is overall completeness, accuracy, consistency bringing transparency and auditability.

Gone are the days where simply writing recorded values for the operation of plant into a diary or spreadsheet for reporting is accepted as best practice; where SCADA historian and reporting systems are the final repository for data; where labs are the final data holder; and where operations and maintenance contractors get to hold and ration data to their client organisations.

The world has moved on to one of electronic storage of all operational data. Every aspect of our lives is moving to electronic systems that, in part, record data and make it transparently available to a plethora of other systems/software/people.

Data comes into organisations and is used for many purposes. It may be generated within that organisation or outside it. It's used for reporting, compliance, added to other data to form complete information sets, modelling, budgeting, and strategy, and many other purposes.

So, with billions of data points flying around and through the internet of things at any given moment, how can an organisation ensure that its data has integrity?

Two core principles:

1. Cut down the chain of custody of data. By reducing the number of people handling a piece of data as it enters any central data system, the chances of error or manipulation of that data are reduced. The objective should be whoever or whatever generates the data inputs it and whoever or whatever needs the data takes it out.

2. Use the concept of the 'Single Source of the Truth' in connected systems. The core principle here is to ensure that the source data remains the same, and any use of that data can be traced through a transparent audit system to the raw data feeds.

Today there is a clear trend for more data to lie outside, and beyond the control of, an organisation. In a local government organisation, this ‘outside’ data comes from electricity companies, laboratories, contractors, weather organisations, etc. It is often supplied by specialist consultants or suppliers. Therefore, the amount of data a Council has that is native to their systems is more limited.

Not all data is created equal. The concept of operational data versus compliance data is firmly entrenched in contemporary best practice. Operational grade data is used for day-to-day operations. Compliance is a measurement against some imposed benchmark.

Typically, to become compliance data, operational data needs to undergo a series of events from collection, cleaning/calculation to reporting.

This means that most, if not all, compliance data is derived or calculated values as opposed to simply measured data. To maintain data integrity in compliance data, there needs to be an audit trail that can track the raw data set, and the way the calculations were performed to get to the derived data output. This must be done in a way that auditors such as drinking water assessors and Regional Council compliance officers can easily understand and measure.

There is a growing international trend in the emergence of independence in the data systems. From the auditor’s perspective, there must be confidence that the data holder has no vested interest in the data. If the data holder is also an advisor, the audit process needs to consider conflict of interest.

High integrity data systems do not have an advisory component. For example, Xero is not your accountant and Orion Health is not your doctor. As soon as an advisory component is introduced into data management, alerting, and reporting it again ventures into the realm of conflict of interest.

Central to integrity is ownership and awareness of data and accountability for it. This is particularly important in local government where public health is concerned. Everyone needs to be confident that accurate data will be there when needed.

Data integrity is vitally important to any organisation. Trusting that data is independently retrieved and transported, that it is accurate and auditable means local government has a plethora of up-to-date information at its fingertips. Data can be continuously sent and received between a Council and its regulators, consultants, contractors, power companies and others all in near real time.

Having a single, independent data repository for trusted data that can come from many sources, can then be easily used as a portal for change, whether it’s a changeover of contractors, for shared services or preparing for amalgamation. WNZ