

Data Cleansing and Annotation Tool



National Technology Reference
Group (NTRG) Workshop

Presented by Ian Scott, Water Corporation
March 2010

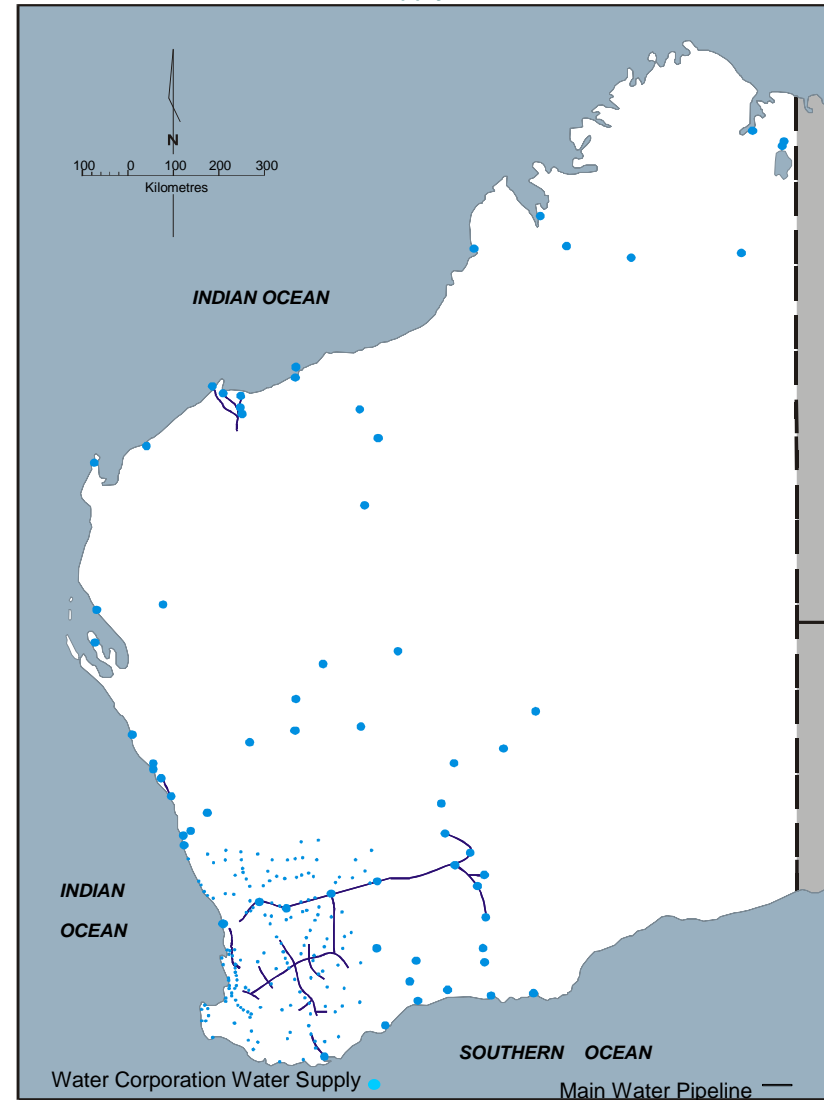
Agenda

- Water Corporation
 - Overview of the Company
- Need for Data Quality Improvement
 - Strategic Challenges
 - Changes to Data Gathering
- Solution
 - Needed solution
 - Architecture of the Solution
 - Business Process Change
- Example Cases of Improved Data
 - Cockburn Sound
 - Canning Dam
- Summary

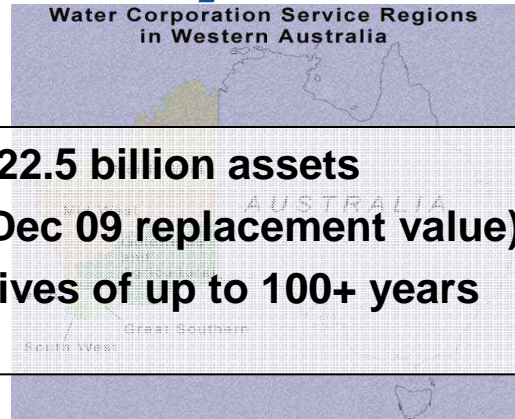
Water Corporation at a glance

- One of the Australia's largest water service providers
- State-wide responsibility
- Area of over 2.5 million km²
- World's largest water supply area
- +\$22 Billion Asset Base
- Serves population of 1.8 million
- More than 2,000 employees
- Annual revenue \$1.5 Billion
- Capital program - \$700M+ / yr
- Water to +300 cities and towns
- Wastewater to 100 cities and towns

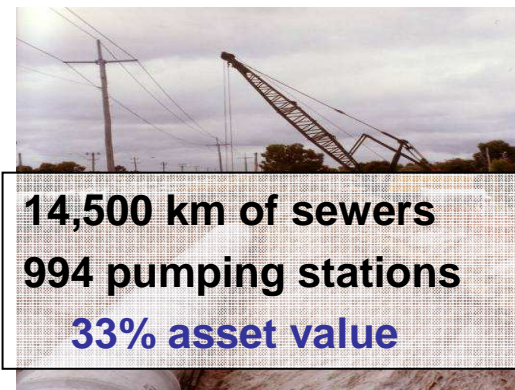
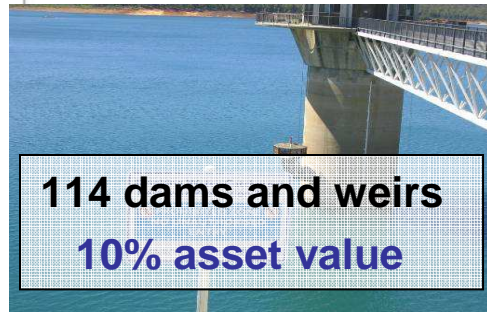
Western Australia - Water Supply



Our assets and systems



**Approximately 75% of assets
(by value) are buried**



Strategic challenges

- Our organisation is in the “Forever Business”.
- Drives for Sustainability
 - Drying climate, increasing population, minimise environmental footprint, etc
- Changing economy
 - Significant capital program
 - Difficulty in accessing skilled staff
 - Increasing costs
- Increasing customer and stakeholder expectations
- Increasing Regulator requirements
- Pressures for changes in industry structure



Changes to Data Gathering

- To meet some of the challenges SCADA was seen as cost effective way to control and monitor the distributed asset base.
- OSIsoft PI software is the system of storage for SCADA sourced data for the Corporation.
- Primary data collection method uses flow totalisers to gather volumes over time.
- Manual data collection will eventually be replaced with SCADA.
- Growth in SCADA exploded from 10% asset coverage in 2005/06 to a projection of 80% by 2012/13.
- Errors have been identified in the data.
- SCADA installs have been so quick that the business has struggled to implement data quality processes and proper commissioning procedures.
- Challenge for data delivery to BoM.



Data Delivery Breakdown

- High frequency data set only

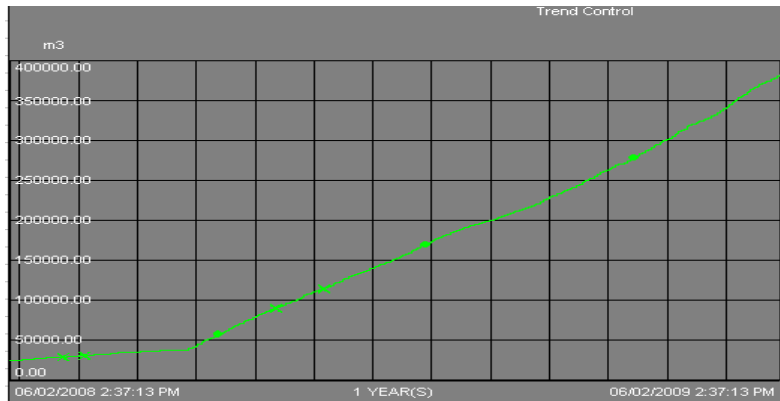
Regulation	Site Count	Category
Reg 3.1	335	Water Course
Reg 3.2	2,290	Groundwater
Reg 3.3	290	Surface Storage
Reg 3.4	110	Weather Data
Reg 3.9	1,950	Basic Water Quality
Total	4,975	

- Over 90% is not “quality” gathered.

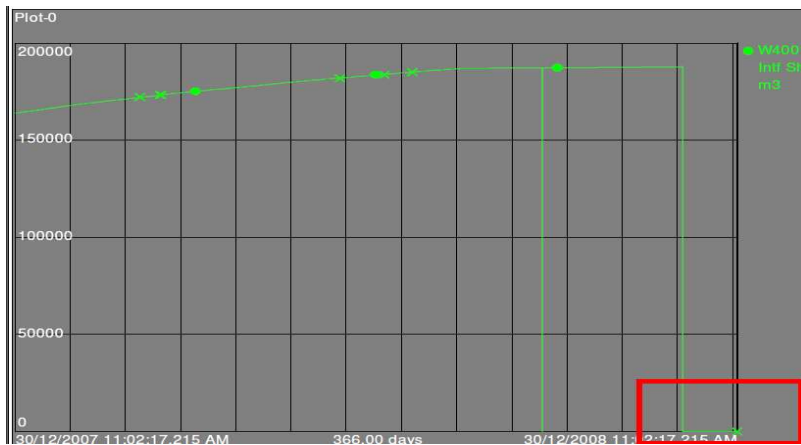
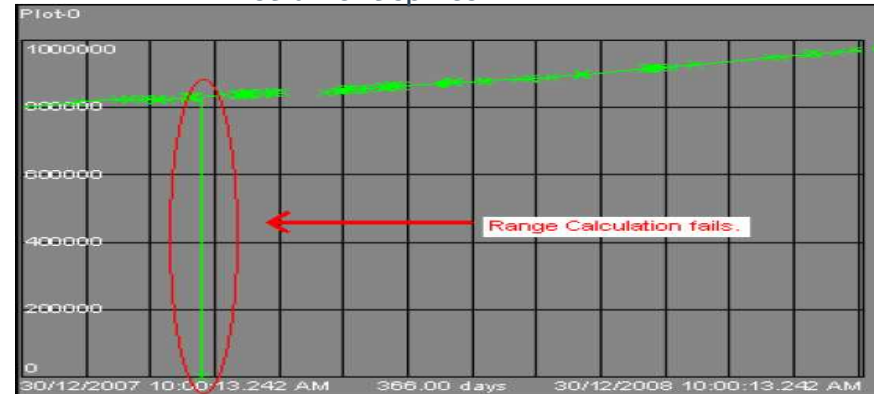


Data Error Examples

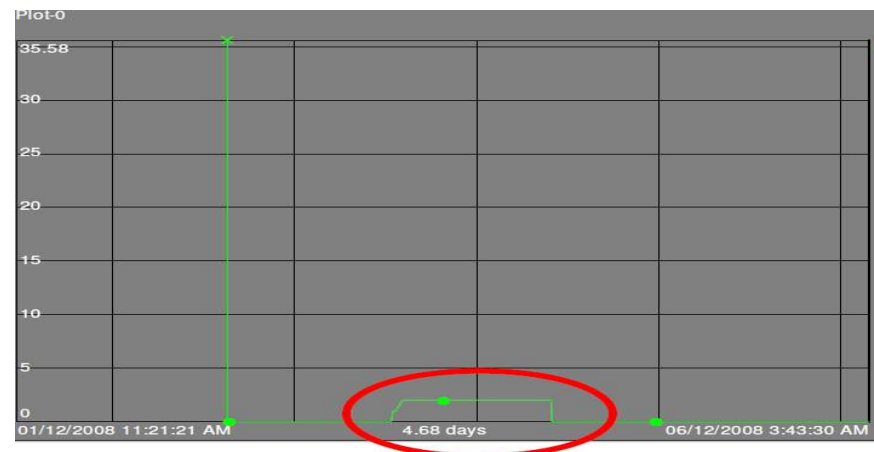
Local RTU scale changes



Instrument spikes



Scales set to zero

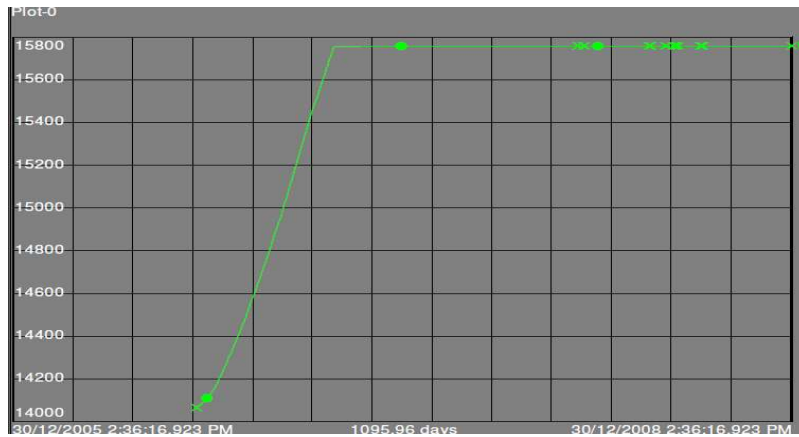


Totaliser not functioning correctly

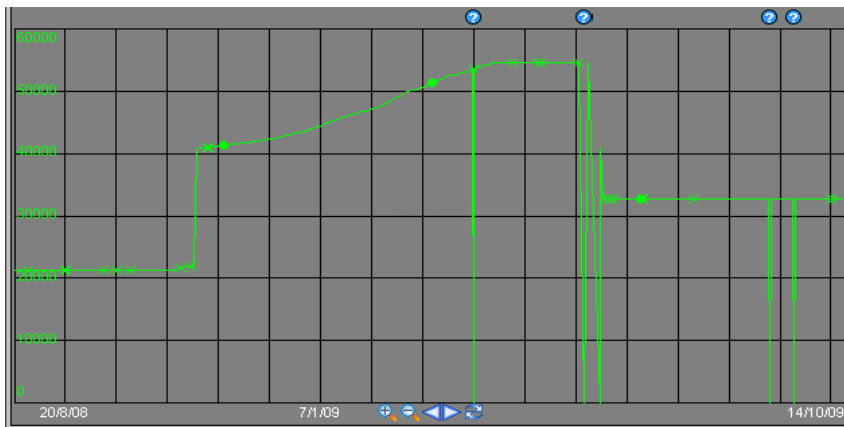
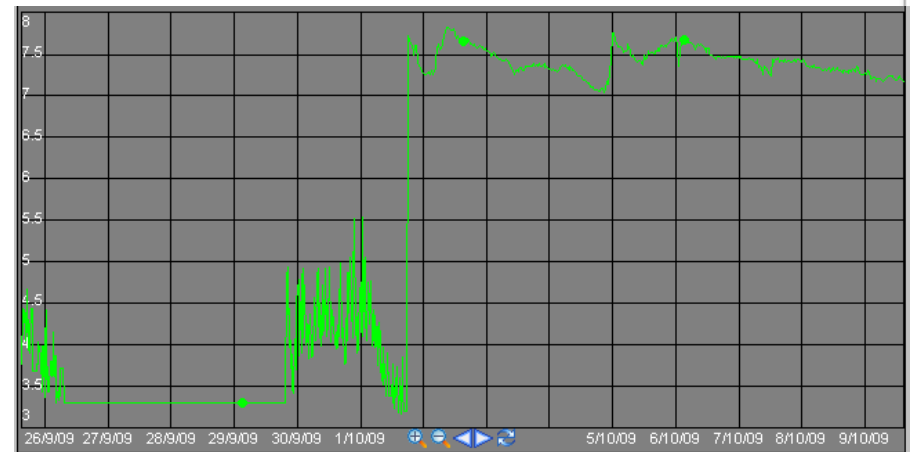


Data Error Examples

Totaliser stalls

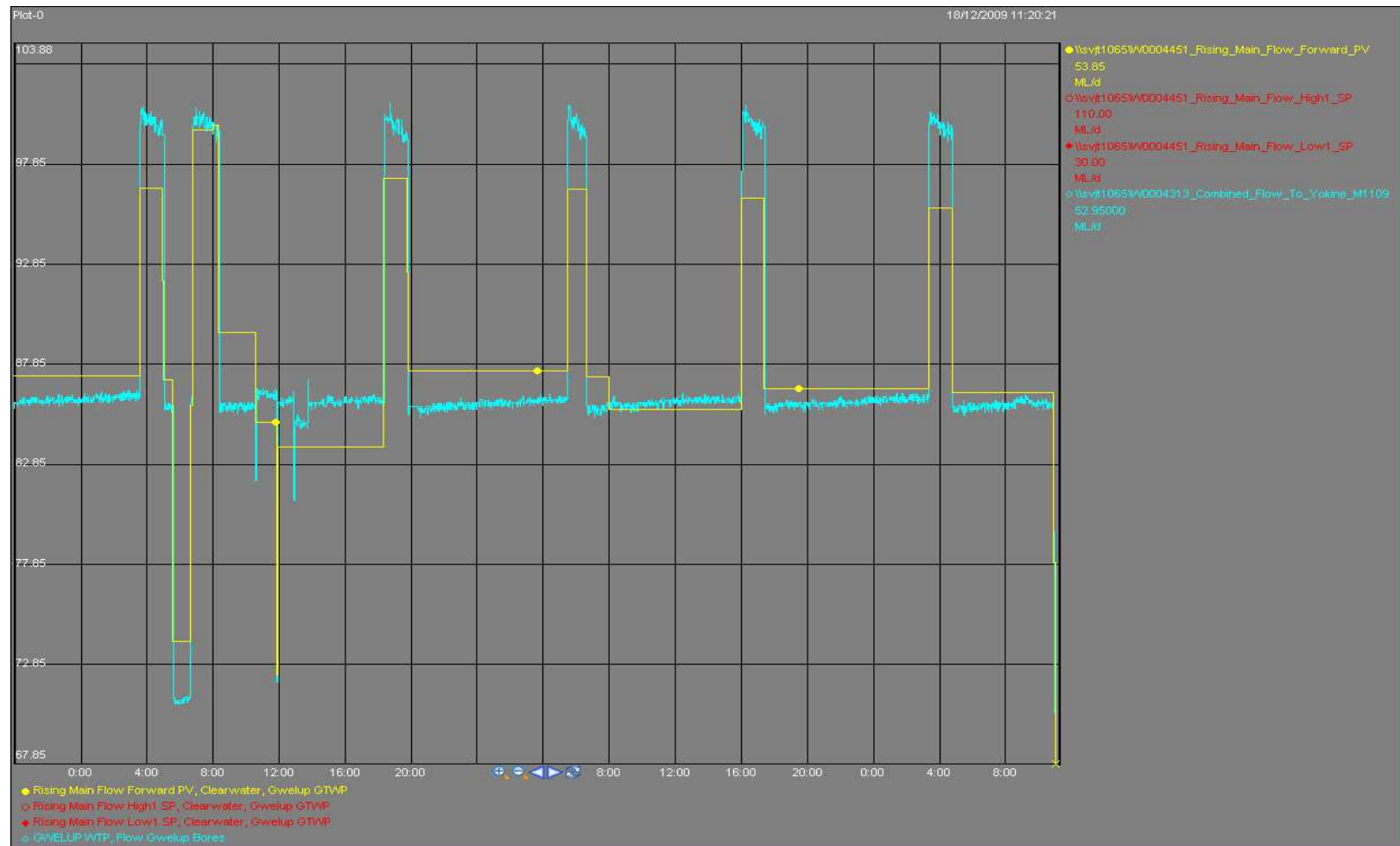


Instrument calibration shift



RTU change out without end to end checks

Poor Data Resolution



Needed solution

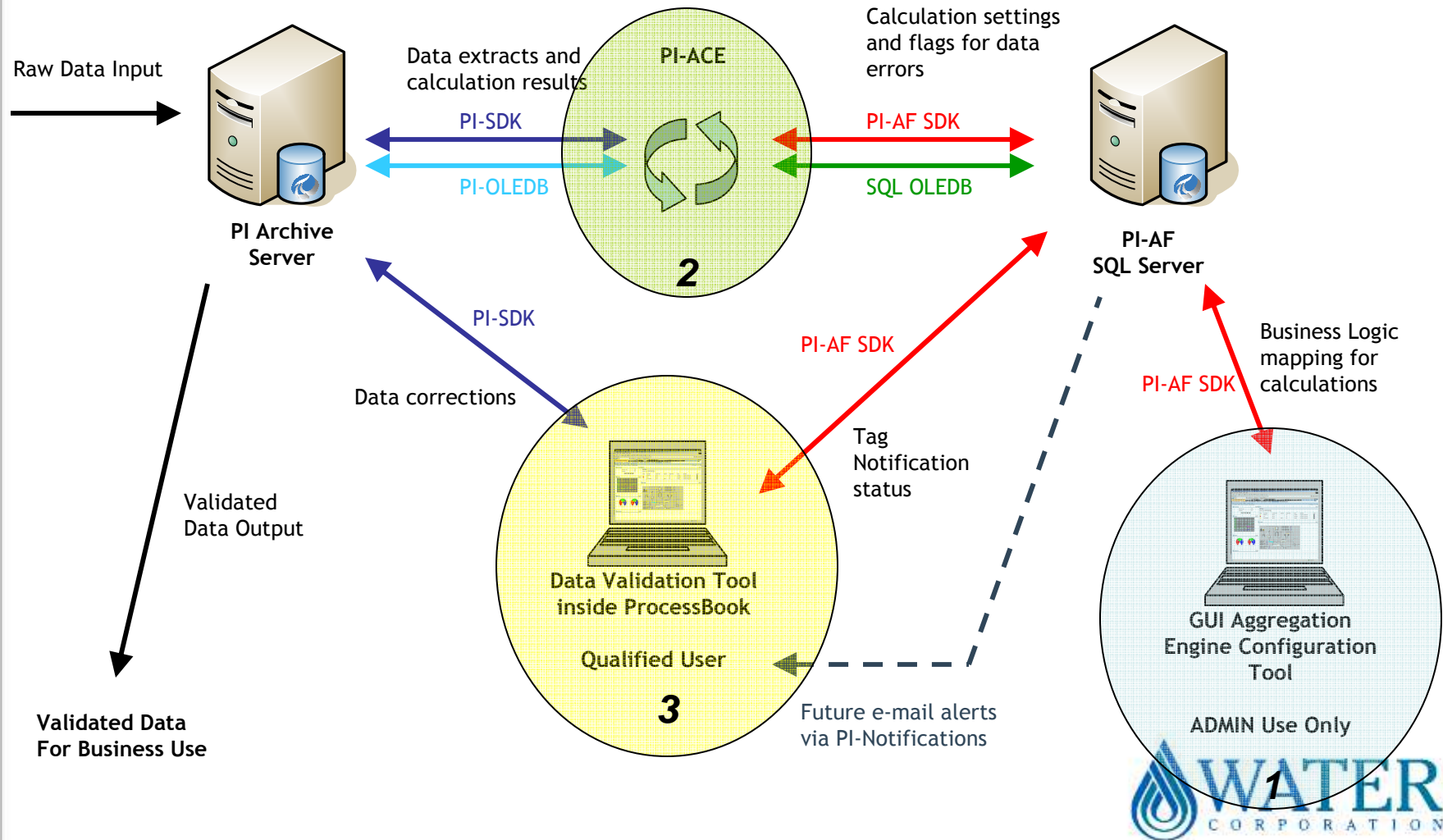
- Raw data must be kept safe and secure
- Automated way to identify data errors, including notifications
- Editing capability, both automatic and manual
- Quality tagging not considered as data is either bad or good.
- Comments used to indicate levels of confidence.
- Modular and expandable. Customisable to any environment.
- Must use core PI software to minimise future support and maximise longevity of solution
- Workable for any PI-based system
- Find funding (Thank you BoM!). Part of the funding was to share the solution.



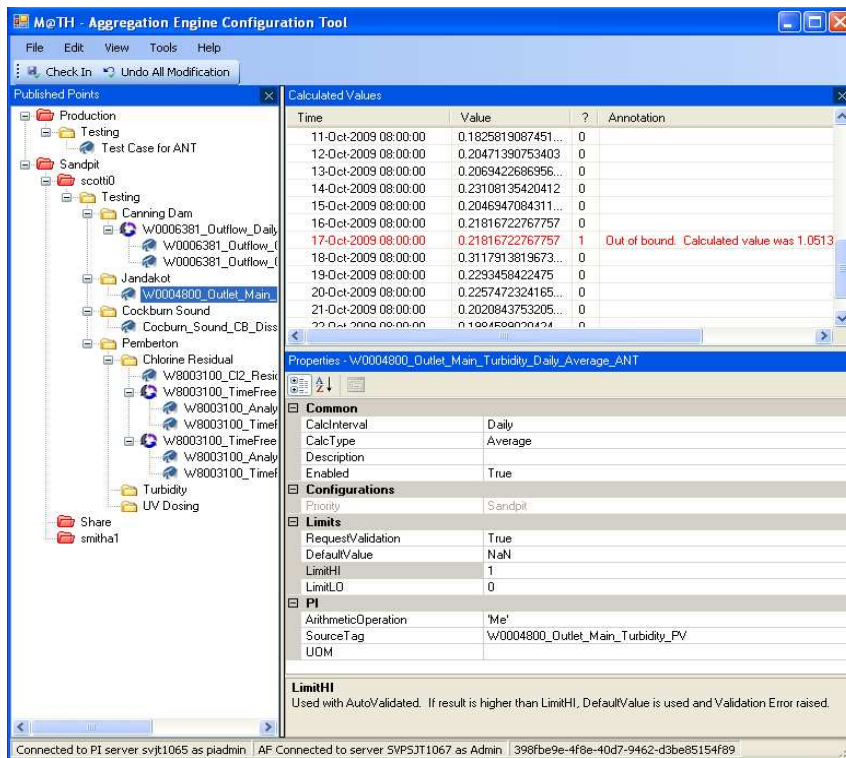
Solutions (nearing implementation)

- PI-AANT (Aggregation, Annotation and Notification Tool)
 - Consistent analysis approach to data
 - Audit both business logic modification along with the data
 - Aggregate the data into Corporately acceptable intervals.
 - Auto analyse the raw data for quality issues (out of limits, missing data, etc) and flag the aggregated data as questionable.
 - Allow users to modify the questionable data to establish the Corporate truth.
- Process modifications to fix root cause
 - With SCADA comes a culture of data complacency
 - Develop new process threads embedded in the Planning phase to deal with data quality along with asset creation
 - Engage with the business areas to define new process threads and determine what it means to them.
 - Proposal to add personnel to monitor and co-ordinate data solutions, which include SCADA instrument fixes.

Solution Architecture



Solution Deliverables 1 – Aggregation Engine Configuration Tool



- GUI tool to interact with PI-AF 2.1 and the PI Server
 - Production/Sandpit areas
 - Create/rename tags
 - Base Points
 - Reporting Points
 - Configure the calculations, set alert levels and establish complex formulas.
 - Define hierarchies to arrange information
 - Ability to copy existing rules to new tags
 - Uses weak references in PI-AF
 - Manual recalculation of any rule in the hierarchy
- Audibility functions
 - Configuration changes stored in tables in PI-AF, using native version control.
 - Three PI tags used per calculation for data auditing.

Validated tag with one value per timestamp
 Audit tag with multiple data per timestamp
 Annotations stored in string tags to allow multiple entries per timestamp and scalable for HA

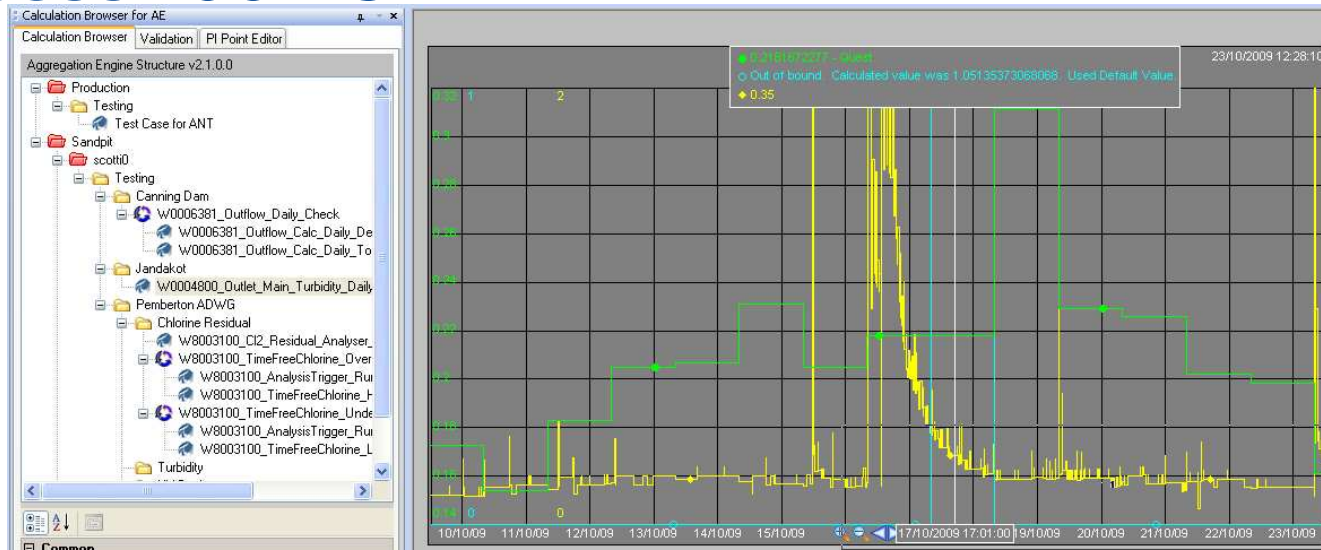
Solution Deliverables 2 - PI-ACE

Aggregation Engine

- Aggregation Engine Features
 - Modular construction of the ACE code to allow for future calculation modules.
 - Reads the configuration of the PI-AF system for tag configurations.
 - Makes for easy expansion of new calculations
 - Regularly scans the PI-AF structure for new sites and setting changes without stopping the system.
 - Main summary statistical functions in place
 - Average, Min, Max, Standard Deviation, Delta
 - BatchON and BatchOFF allowed to calculate event frame times
 - Calculation intervals of 15mins, 30 mins, hourly and daily
 - Code allows for mathematical functions on base data.
 - **Hierarchical calculations ensure all Reporting Points are calculated when Base Point data is modified.**

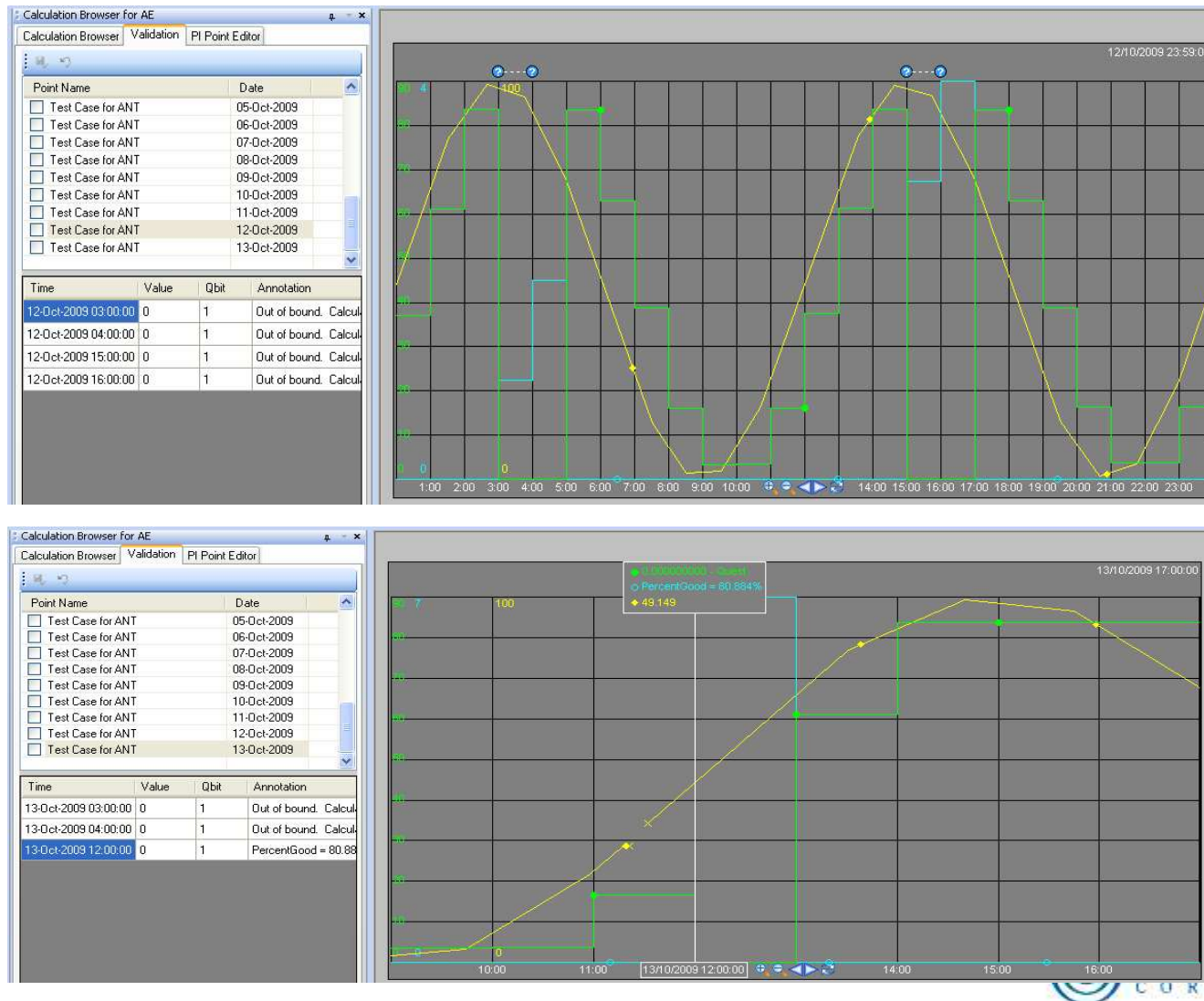


Solution Deliverables 3 –Validation addin to ProcessBook 3.2

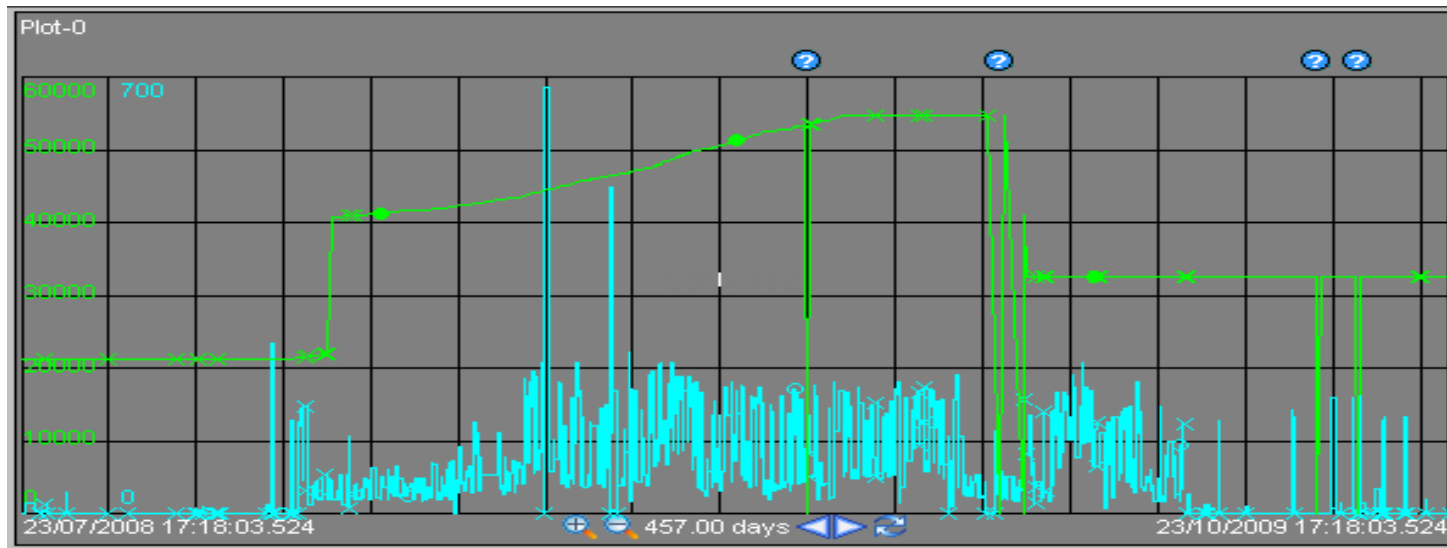


- ProcessBook addin features
 - Ability to scan the hierarchy for the tags of interest
 - Configuration panel shows the business logic for the tag
 - Trend displays source, validated and comment tags for Base points. Reporting points include source tags for calculation.
 - Validation Panel to focus on the data with quality issues.
 - PI Point panel for cleansing any tag.

Data Validation Steps



Example Case – Canning Dam Flows



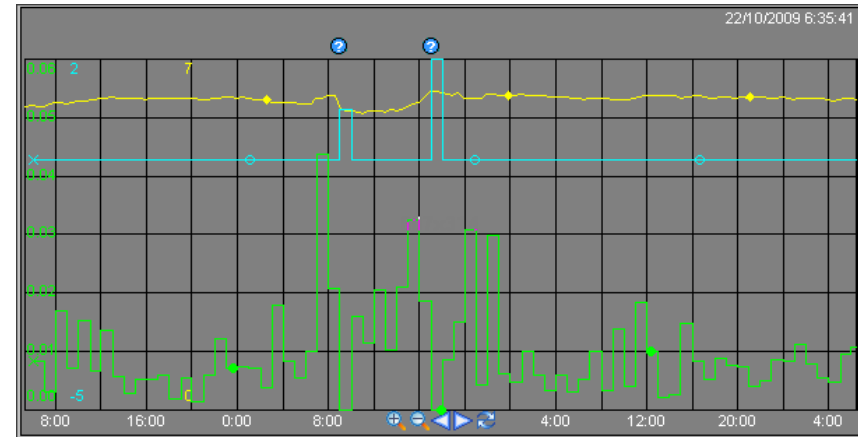
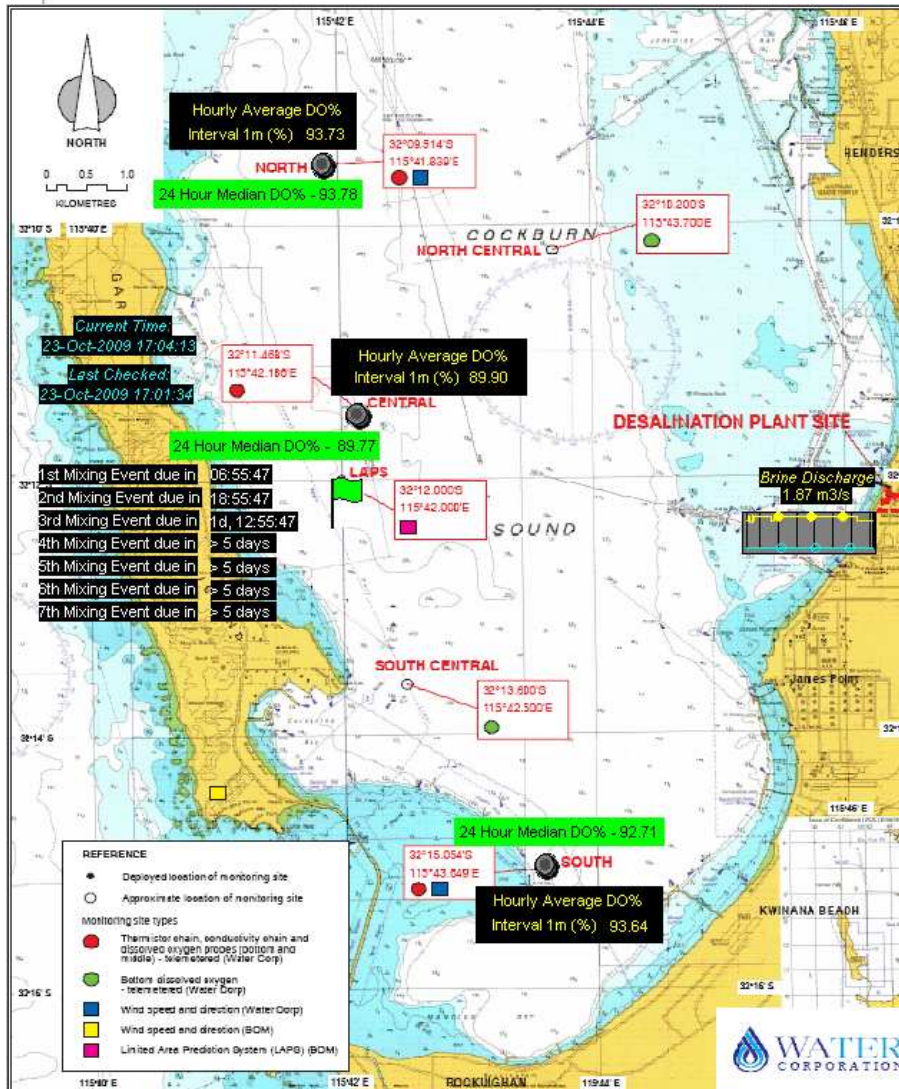
Problem

- Monitoring of flow volumes critical to report to the Department of Water for licence purposes.
- Poor change management of SCADA equipment resulted in wrong data being stored in our reporting systems.

Solution

- Have daily figures calculated from both the flow rate and the flow totalisers.
- Compare the two daily figures for variations greater than 5%.

Example Case – Cockburn Sound Monitoring



Requirements

- Critical parameter monitoring of dissolved oxygen in Cockburn Sound required for the Perth Seawater Desalination Plant.
- Operating License has comprehensive parameters that can cause the Desalination Plant to be shut down if exceeded.
- Vital to know when the instruments are failing.

Solution

- Use Std Dev to identify when DO levels drop too quickly.



Summary

- The Corporation has a changing environment for data gathering.
- Instrumentation is changing manual data collection.
- Too much data to manually check.
- Without the human element we need automation to monitor the automation.
- The solution aggregates, validates and notifies Corporately acceptable data against business logic.
- Business culture has to move with technology.
- Business needs to focus on outcomes at implementation.
- NWI has been a catalyst to review poor data practices.
- Entering discussions how the solution can be distributed.



Thank You



Questions?