The Use of Spreadsheets in the Era of Big Water Data

You think you're compliant. Can you prove it?

AQUATIC
Informatics

AQUATIC
Informatics
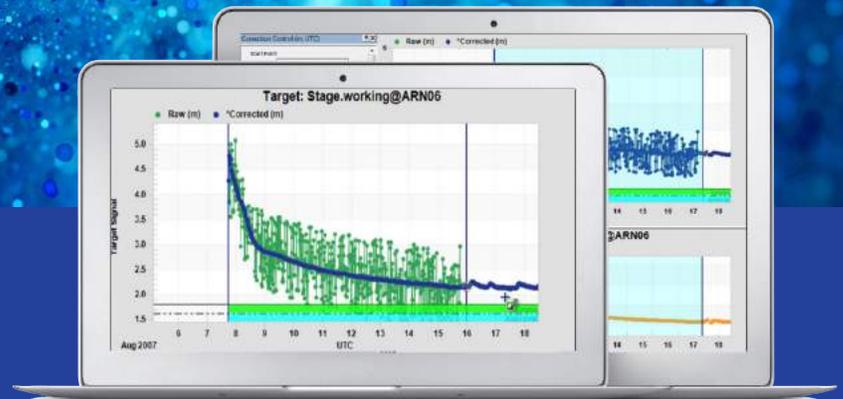
Introduction

How do you manage the water data that you base your decisions on?

As technology advances, water data becomes easier to generate. That means water resource managers, whether managing environmental, municipal, or industrial water, have increasingly large and complex data sets to manage and interpret. Increasingly, regulators, stakeholders, and the community expect water managers to make informed decisions based on high quality data.

How do water data sets get so big? If you consider that one continuous sensor can collect four data points an hour – a relatively slow rate for sensors – 24 hours a day, 365 days a year – that equals 35,040 data points per year! And that's just for one site. Given that modern multi-probes can facilitate 4, 5, or 6 sensors, data sets can become unwieldly in a very short amount of time.

Then there is the variety of data. There are data collected from sensors, physical water quality data generated from the collection of grab samples, and data from the analysis of grab samples provided by the laboratory. And that is just the water quality data! There is also information coming in from flow meters and weather stations, and river stream gauging stations report water levels every few minutes. Biological sampling generates different data types entirely. The question then becomes how to manage all the data, and how to turn raw data into information on which to base decisions.

What do you need in a water data management system? First and foremost, you need one that can securely manage your data throughout the data cycle. A modern system should also manage:

1. Acquiring and importing data
2. Validating and analyzing data
3. Visualizing and modelling data
4. Reporting and sharing data

What data management systems do most organizations use to manage their water data? It turns out that over 60% of them use spreadsheets and/or instrument vendor software, and data often end up scattered across multiple spreadsheets or software packages. Documenting your decisions could become a nightmare. If your regulator or auditor requested records of your compliance monitoring, would your data management system provide the surety they require?

Water managers should consider the cost of non-compliance in terms of risk – business risk, reputation risk, and legal risk. You must balance these significant risks against the cost of investing in timely, accurate, and defensible water data. By examining the strengths and weaknesses of spreadsheets, the value of an industry-specific software solution becomes clear.

1.

Acquiring and Importing Big Water Data

The first step in the data cycle is collecting and importing data into a system that will store the data securely. Data can come from many sources:

• SCADA systems

• Remote monitoring stations via telemetry

• Data downloaded onto mobile devices

• Laboratory data management systems (e.g. LIMS)

Data entry into a spreadsheet

With spreadsheets, all data must be entered manually. This has obvious issues with potential copy/paste and/or transcription errors. With different people entering data, version control becomes challenging. Data may end up in older versions, jeopardizing chain-of-custody documentation, wasting time, and introducing the potential for double entry. Decisions may potentially be made based on outdated data in older versions of the spreadsheet.

Spreadsheets also struggle to handle different data types from multiple sources. New sheets must be generated to deal with each parameter or each sampling event.

As the data set grows and updates are applied, there is also the potential to lose historical data.

Spreadsheets do not maintain a permanent record of original data points or metadata, nor were they designed to store data points collected over time.

Data entry into an industry-specific software solution

With an industry-specific software solution, users can rapidly enter field or lab results, automate telemetry imports, and integrate with external sources. New data can be automatically processed in near real-time, saving hours of work daily.

The software eliminates time spent exporting, importing, and amalgamating data from various systems. Modern web service Application Program Interfaces (APIs) make it easy to integrate with:

• Existing systems

• Data acquisition tools

• Data repositories

• Business intelligence (BI) tools

A permanent record of original data points is maintained, even while the data is systematically quality controlled for accuracy. Errors are detected and corrected, with all changes documented in an audit trail.

2.
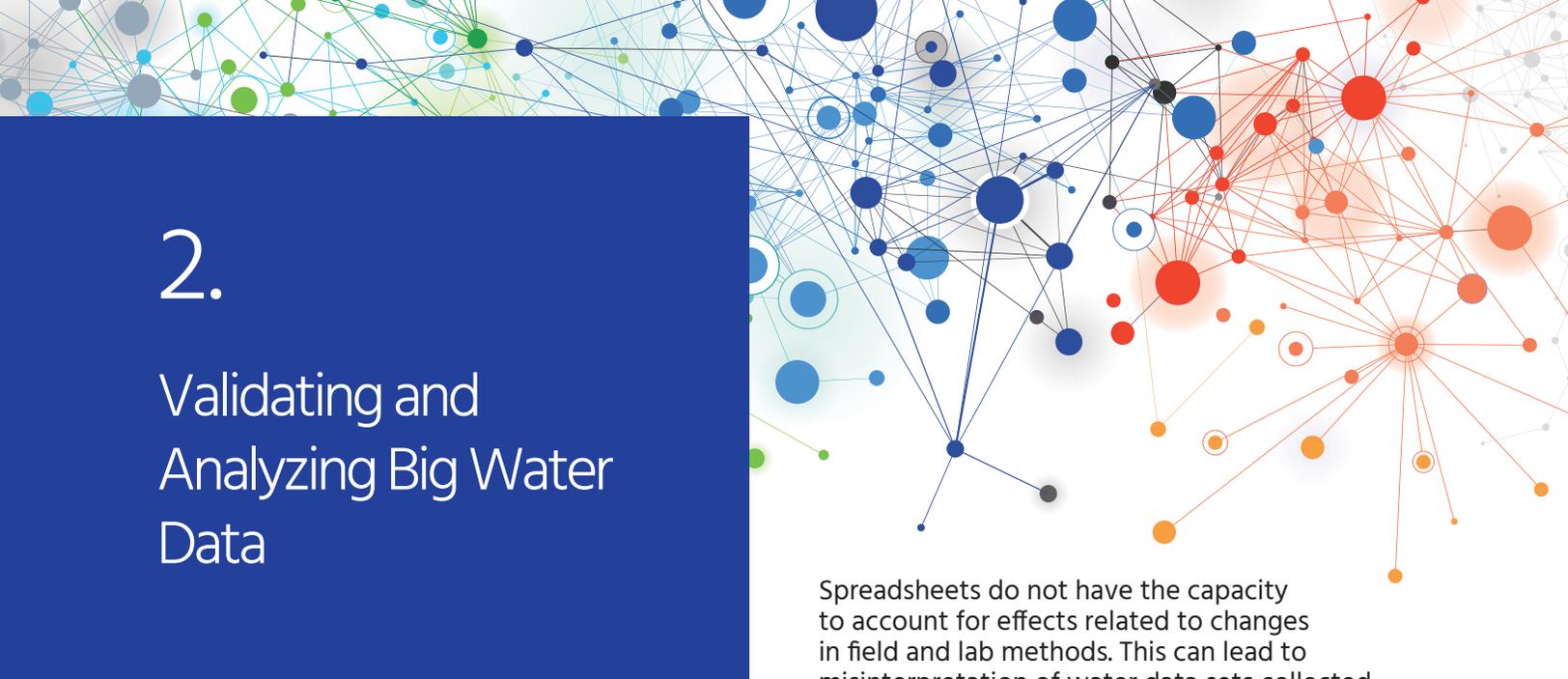
Validating and Analyzing Big Water Data

Once in a data management system, the data must be validated and analyzed to provide usable information for you and your team. Information coming in from sensors and remote monitoring stations often contain anomalous or erroneous data that can be confusing or misleading for end-users of the data.

Non-detects are a separate but significant issue with data from laboratory analyses. When a sample contains so little of a contaminant that the analytical machine is not able to detect it, it is called a "non-detect." Non-detects are given a value as less than the level of detection (e.g. <0.05 µg/L). Because it is a non-numerical value, a non-detect can't be included in data to be analyzed, graphed, or reported.

Thus, all data require systematic review for error detection and correction to assure a clean data set for visualization and analysis.

Validation and analysis in a spreadsheet

With a spreadsheet, data must be manually processed to deal with non-detects, which can lead to a loss of metadata. Any changes made are not traceable and data may not be of use for future analysis. Additionally, it is difficult to analyze the impacts of sensor drift from data probes.

Spreadsheets do not have the capacity to account for effects related to changes in field and lab methods. This can lead to misinterpretation of water data sets collected over time.

Significantly, because of these types of validation issues, data held in spreadsheets may not be accepted by regulators or be defensible in court.

Validation and analysis in an industry-specific software solution

An industry-specific software solution automates quality assurance (QA) processes. An audit trail catalogues all changes to original data points and tracks changes by date, time, and user.

The software automates complex computations, such as discharge derivation, evapotranspiration, and sediment loads. Calculations can be chained together to solve difficult water problems specific to your organization's mission.

A robust data management system automatically flags data outside of acceptable ranges and sends notifications to users by text message or email. In addition, sampling events can be triggered and/or tracked based on your quality control flags and you can add custom alerts that warn operators before regulatory standards are exceeded.

Acceptable ranges can be based on your regulatory standards or international standards. Industry-specific software solutions come with international standards already imbedded. Automatic updates of standards then keep your organization up-to-date with all changes.

3.

Visualizing and Modelling Big Water Data

With data entered, validated, and analyzed, the next stage in the cycle is to use that data to assess current conditions and to predict future conditions based on past performance. This requires visualization and modelling to unlock the story behind the numbers.

Visualize and model in a spreadsheet

When data is held in spreadsheets, standard formats and graphical elements can be preformatted. However, each addition of new data requires manual processing, updating, and formatting. This can make manipulating large data sets extremely difficult.

It is relatively easy to generate a wide variety of charts and graphs in a standard spreadsheet. Choosing the most appropriate type, however, is not always self-evident. Your team should be focused on the information provided by the visualization and should not be wrestling with the graphing package itself.

Spreadsheets do not readily allow analysis of different data types. Qualitative vs. quantitative data or time series vs. discrete sampling often cause problems, and data may need to be exported to other software packages. Unresolved issues with data quality can influence the visuals created.

The limited modelling capabilities of spreadsheets still require an expert to manually process, update, and format each new data set. Because models are dependent on the quality of information provided, poor quality data can lead to erroneous models. In fact, bad data often only become apparent when it's time to prepare reports for regulators. You want your team getting the best possible information about critical elements in your operations before that happens.

### Visualize and model in an industry-specific software solution

Industry-specific software systems deliver advanced, responsive analytics for evaluating and visualizing big data sets. Maps, graphs, and charts provide the complete picture, empowering users to instantly spot outliers and identify trends.

An industry-specific software solution also generates graphs and charts automatically and in near real time. Customizable online dashboards enhance visualization. Your team will have the information they need at their fingertips.

Models are only as good as the data used as input. With clean, validated data, model outputs are more robust and allow for integration across platforms.

4.

Reporting and Sharing Big Water Data
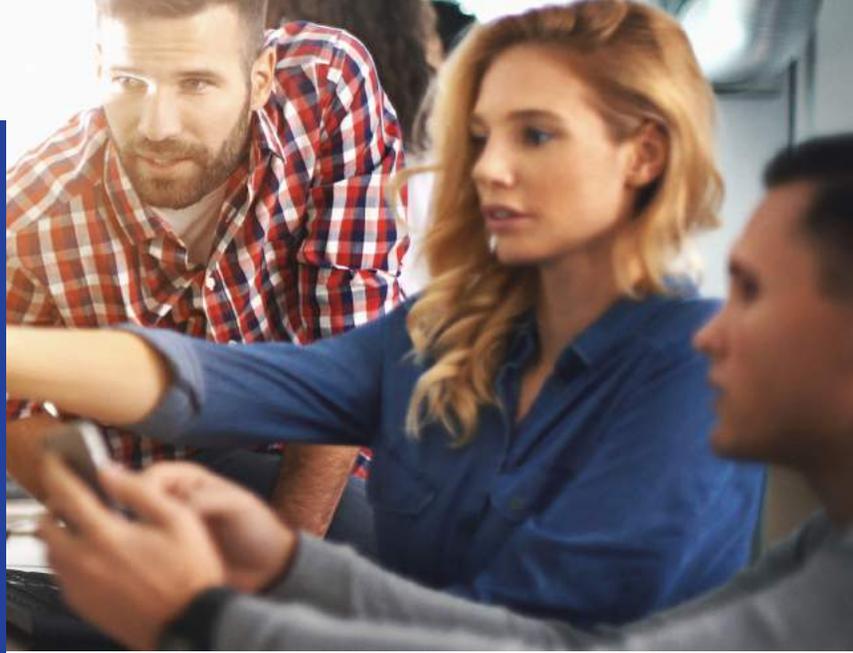
The final stage of the data cycle is to report and share the insight generated by the data. You have regular operational and compliance reports to generate. You need to share water data, and the information generated, to internal and external stakeholders. The data management system you use will determine the ease with which you can report and share your data.

Report and share from a spreadsheet

Using spreadsheets for water data management complicates reporting and sharing. The data required to prepare operational and compliance reports can end up dispersed across multiple spreadsheets. The data could even be on separate machines or servers. Finding and collating the data can take days or even weeks – time that your team could better spend solving problems and addressing issues.

Sharing the information generated from your monitoring program within your organization is critical. However, when that information is held in spreadsheets, sharing isn't easy. Updates sent to team members via email may end up lost in their inbox. Your team needs to stay on top of relevant information and they don't need to waste valuable time and resources looking for data.

Perhaps the most alarming aspect of using spreadsheets for water data management is access. It is difficult to control access to critical data and analysis tools. With access questionable, your data will again be difficult to defend to regulators or in court.

Report and share from an industry-specific software solution

An industry-specific software solution streamlines reporting. With automated report generation, you and your team will be able to generate reports instantly. Report publishing then becomes routine, increasing productivity.

Software also facilitates sharing across your organization and with stakeholders. Concurrent users can access the data from multiple locations and computers. Web publishing and integration with BI systems are part of the package.

Importantly, in an industry-specific software solution, user access permissions and rights are centrally controlled by the application administrator. Access control, combined with the audit trail, means your data is defensible.

Conclusion

Effective information management can help mitigate millions of dollars in risk exposure. Knowledge derived from water data empowers organizations to effectively manage the risks of:

- Regulatory fines
- Legal action
- Engineering infrastructure mishaps
- Unplanned extreme environmental conditions
- Health hazards

Quality water data and the information it provides can ensure the sustainable use of water for maximum social, economic, and environmental benefits. All that information and knowledge must be built upon high quality data.

Examining the data cycle reveals the shortcomings of spreadsheets as a data management tool:

- Acquiring and importing data into spreadsheets is labour-intensive and data is not secure, and there is the potential to lose valuable historical data during updates.

- Validating and analyzing data is time consuming, with complex calculations that are hard to automate. Inappropriate visualizations can lead to misinterpretation of trends.

- Visualizing and modelling data in a spreadsheet can require expert attention and requires manually processing, updating, and formatting with each new data input.

- Reporting and sharing from a spreadsheet is also time consuming, with data scattered across multiple sheets and access difficult to control. Sharing is not straightforward.

A modern, industry-specific system transforms monitoring data into a highly valuable asset – unlocking the enormous potential for increased transparency. Environmental water, drinking water, and wastewater are linked through the water cycle, and data points across the cycle have the power to change the way water is managed.

Reduce risks to your organization – business risk, reputation risk, and legal risk – with timely, accurate, and defensible water knowledge stored in an industry-specific software solution.
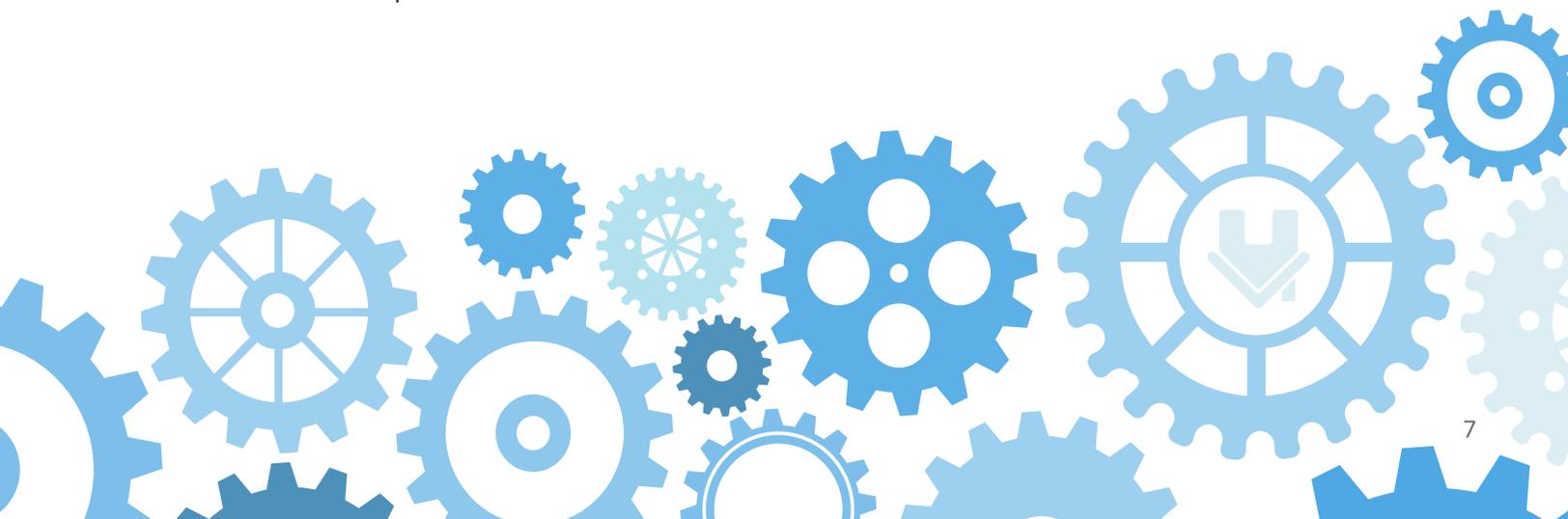
"In reality, what often passes as a 'system' would be better described as a 'collection' of disparate applications held together by the heroic efforts of passionate and dedicated experts. The retirements of these experts make it difficult to maintain such systems. Deferment of critical decisions has led to many such systems being pushed to the point of imminent failure. It is time to look forward and consider a modern water data management system."

-Stu Hamilton, Senior Hydrologist

Are you ready to modernize your water data management system? Not sure how to transition from spreadsheets to an industry-specific software solution?

Contact Aquatic Informatics today to discuss your water data management needs and sign up for a free demonstration!

BOOK A CONSULTATION

Contact us today to learn more about what our water data management platforms can do for you.

www.aquaticinformatics.com