THE BEST WATER DATA POSSIBLE!
5 KEY REQUIREMENTS FOR MODERN SYSTEMS

Whitepaper

Water Data in a Modern World

Hydrometric data providers have been rapidly exploiting modern technology to keep pace with the changing needs of modern society for timely, accurate, and defensible water data. The past two decades have seen an almost complete turnover in field technology. Most monitoring networks today use continuous monitoring and real-time communications technologies, requiring a modern approach for data handling.

A Once in a Generation Change

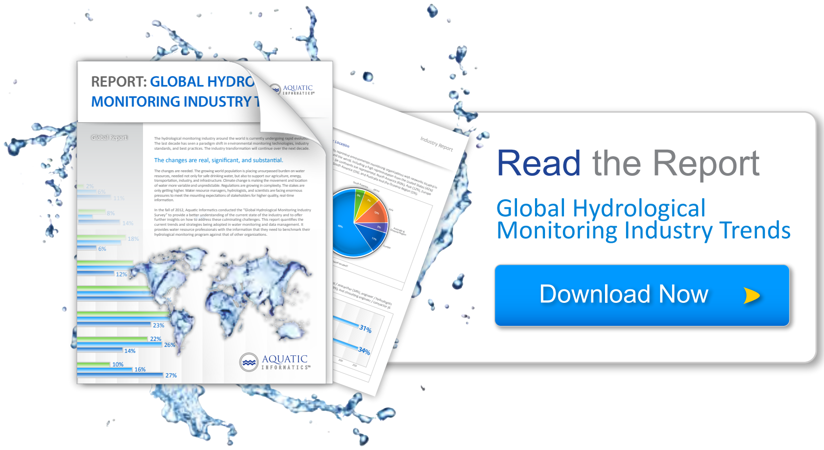
Hydrometric data processing is inherently repetitive and tedious to do by hand. Most of the work is ideally suited for automation, so when computers became affordable in the 1980s almost every hydrometric agency dedicated engineers to the task of making the work more efficient. The scope of these software applications continued to grow as new efficiencies were found in every aspect of the data production process. Some of these solutions were privatized as governments around the world struggled to control deficit financing in the 1990's.

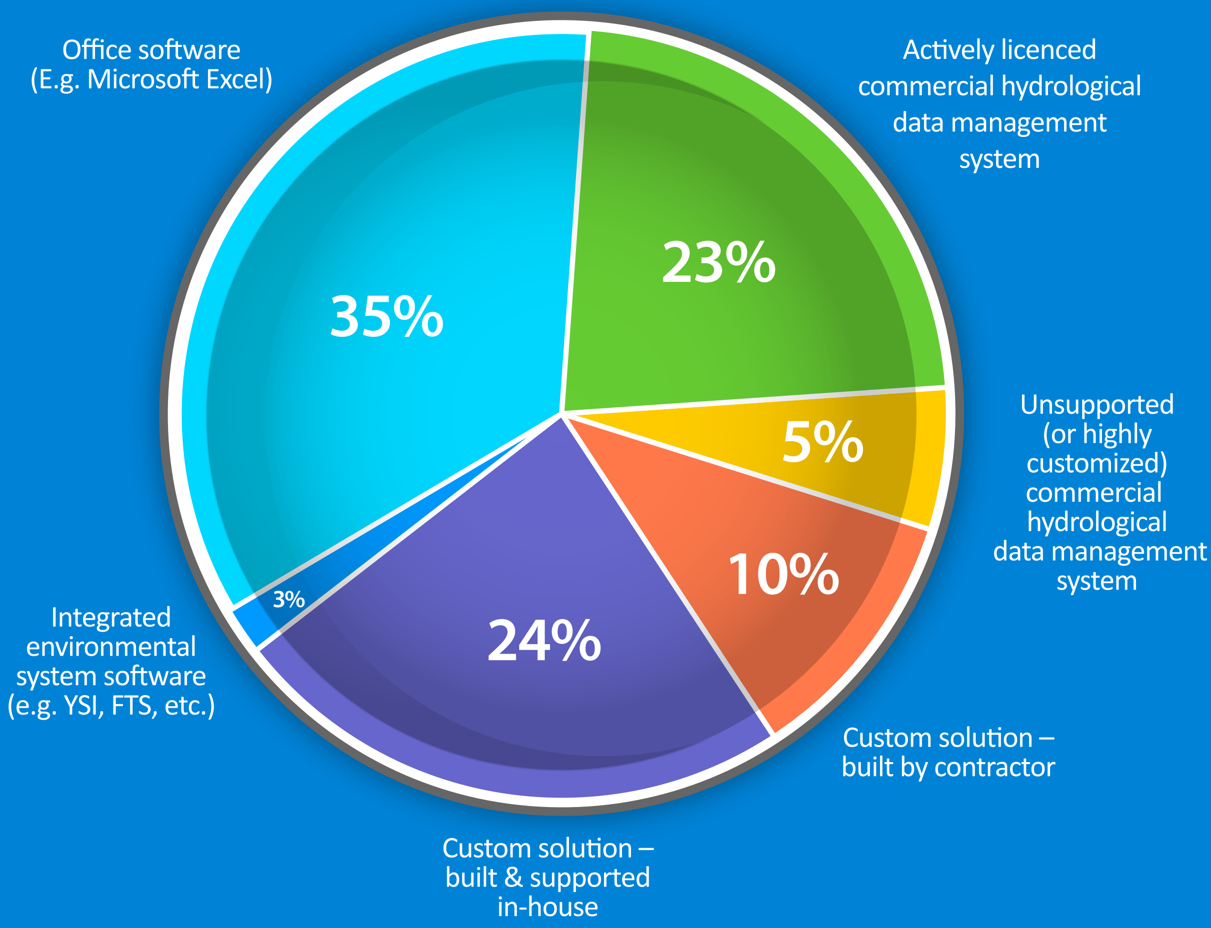
Small-scale monitoring programs developed prolifically through the 1990s, primarily to address local scale environmental concerns. Lacking the resources of the national hydrometric programs this sector became dependent on adapting generic software to their hydrometric needs. It is typically only when reaching the limits of spreadsheet applications that custom software was developed. These solutions tend to be highly specific and marketed to a regional niche market. Reconciling the increased demand for timely data in the context of new and increasingly complex data sources and these aging data systems is forcing every hydrometric data provider to re-examine their data management framework.

AQUATIC Informatics

Disparate Data, Disparate Systems, Desperate Hydrologists

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. Vital components of these ‘collections’ are often either command line driven applications or extremely complex spreadsheets that can only be maintained by the original author. 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.





Which data management system does your organization primarily use? | Read the report at www.aquaticinformatics.com

5 Key Requirements for a Modern Water Data Management System

In the search for sustainable water data management solutions all scales are relevant. The smallest operator and the largest national hydrometric program face uniquely different challenges, but the solutions to their many and varied problems share many common features. Regardless of the starting point, convergence on a shared solution is now possible.

Modern hydrologic data management systems are commercially available that support current industry standards for water data management and that meet the evolving needs of water resource managers, hydrologists, stream hydrographers, and engineers. The world's preeminent environmental monitoring agencies, including the United Geological Survey (USGS) and Water Survey of Canada (WSC), rely on these systems every day to build better rating curves and to publish more credible and timely environmental data. These very agencies are helping shape the design and functionality of modern data management systems to ensure the solutions continue to meet their maturing needs and emerging industry best practices.

The remainder of this whitepaper outlines 5 key requirements for selecting a modern data management system to achieve the best water data possible.

Modern hydrologic data management systems are commercially available that support current industry standards for water data management.

1. Secure, Centralized Data Management

Timely, accurate data is desperately needed to yield answers to difficult water resource management issues. Modern water data management systems allow data to be securely managed in a coherent framework of quality control procedures and accompanying metadata that are needed to establish confidence in the data.

Centralized Environmental Data

Properly integrated data management systems give water resource managers centralized access to all of their hydrological data by making it easy to import, enter, and integrate hydrometric data from various sources. The system must be extensible and seamlessly access and process externally managed data and metadata (e.g. data from other corporate databases). Data is automatically ingested to a centrally managed database, complete with source tracking, ensuring that both timeliness and data provenance are managed with little or no human intervention for repetitive tasks.

Secure Data

Business continuity and disaster recovery planning are crucial considerations for valuable data assets. Centralized data storage simplifies data backup, enabling off-site replication and hot failover, which means more secure data.

Defensible Audit Trail

Centralized data management provides for an unbroken chain of custody from the raw data through to the final product, including a complete history of who did what, when, how, and why. Hydrometric data are complex requiring corrections for gauge instability, sensor drift, and spurious anomalies. These corrections require human guidance and mistakes are possible. Data derivation involves subjective interpretation of field visit data to develop rating curves, manage the effects of dynamic channel conditions, build empirical models, and in-fill missing periods. A single, secure repository for data provenance is needed to meet best practice standards for handling of evidence.

Centralized data management is as important for the smallest agencies as it is for the largest. The largest agencies can suffer from inconsistent data processes across geographical locations and from inefficiencies and service reliability issues resulting from numerous servers handling data. The smallest agencies – those that have data stored in collections of files and spreadsheets – can suffer from fragmentation of datasets, disconnection of data from metadata, formidable barriers to automation of repetitive tasks, and inefficient data retrieval.

Consolidating data into a properly designed relational database solution solves these problems. These systems are scalable for future requirements and provide integrated data security, improving the credibility and defensibility of all data products and reports.

2. Evidence-Based Analysis

The heart of any modern hydrometric data production system is interpreting, understanding, and predicting relationships amongst variables. In a large majority of locations, stage is used to predict discharge.

The shape of the stage-discharge relationship is unique to the specific conditions that control the flow of water past the gauge. This is challenging for several reasons. The controlling features of the stream can be highly sensitive to dynamic geophysical and biological activity. The field measurements of stage and discharge, used to calibrate and validate the relationship, have error and it cannot be assumed that the error is normally distributed and free of bias. The range of measurements does not span the range of data, so extrapolation of the relationship is a requirement. The measurements that have the highest leverage on the extrapolated segments of the relationship are typically the ones that have the highest error (due to adverse measurement conditions) and the least replication (due to the infrequent occurrence of extreme conditions).

> A single, secure repository for data provenance is needed to meet best practice standards for handling of evidence.

Better Rating Curves

Fortunately, the shape of a stage discharge relation is a function of well-known hydraulic principles. The interpretation of the relation can therefore be guided by an appropriate combination of science and inspection of physical evidence. Field observations including summary statistics, measurement details, notes, and photographs are of crucial importance in this analysis. The time-series history informs interpretation of channel modifying events and model parameters can be used to constrain critical analysis, leading to defensible conclusions. Well-integrated rating curve development tools allow hydrographers to focus on the evidence needed for interpretation of hydrologic and hydraulic conditions, ensuring that the curve represents the reality at the gauge and is not just a statistical artifact of heteroscedastic data.

Advanced Data Modeling

In some cases, such as the development of a stage-area relation as an input to an index-velocity model, the solution is physically constrained by channel geometry. In this case the relationship can be mathematically derived from imported channel cross-section surveys with little human interaction.

Complex Relations Require Powerful Diagnostics

In other cases, such as index-velocity to mean-velocity, the relationship is empirical. The discovery of an empirical relationship with predictive power requires iterative, experimental reduction of structure from the model residuals to achieve the simplest model with the least systematic error. A modern data management system delivers a simple user experience for this process with a rich set of residual plots against any choice of predictor variable, accompanying statistics, the ability to add or remove variables from a multiple linear analysis, and the ability to develop compound relations.

The process of evidence based analysis includes the interpretation and exploitation of all relevant sources of information. The outcome of this process is well documented, credible, defensible data.

3. Automated, Real-Time Data Processing

There has been a rapid evolution from hard-copy data publication (batch processed on a quarterly and annual production cycle) to the continuous web-based publication of real-time hydrometric data. Timely, efficient, and quality-focused data production requires a fundamentally different set of tools than were originally developed for batch processing.

With continuous communication of data, from gauge to web page, there is high risk of disinformation: spurious data that looks real, valid data that looks spurious, and valid stage with corresponding invalid discharge. Validating data in real-time requires many site specific considerations such as the local weather, controls on runoff processes, hydraulic characteristics of the channel, currently deployed technology, and the fault history of the gauge.

It is far too onerous to use traditional techniques for data scrutiny to meet the required interpretation and analysis of every gauge, all of the time. Modern data management systems allow the hydrographer to provide training for the computer that accommodates site-specific process understanding. At the time of acquisition, the data are evaluated point-by-point using the interpretive wisdom of the hydrographer captured in parameter settings.

Automated Quality Assurance & Quality Controls

Automated quality controls provide a preliminary sanity check of the data by filtering out technological faults (e.g. spikes and flat lines).The system differentiates between valid rates of change and spurious excursions to allow the corrective model to adapt to rapid fluctuations of the target signal. The order of operations has to be configurable (e.g. trim outliers before evaluating the rate of change), and corrections to derivation models need to be configurable (e.g. shifts to rating curves).

Supervision of the automated corrections can be triggered by notifications or simply be systematic review of the data. The hydrographer becomes part of a continuous

Modern data management systems fully support traceability and reversibility.

“Quality has to be caused, not controlled” – Phil Crosby

learning/optimization process whereby correction parameters are systematically updated and improved to progressively eliminate the occurrence of Type I (false positive) and Type II (false negative) errors.

Modern data management systems fully support traceability and reversibility. Automated and manual markup of the data, auditable event logs, source tracking, and the ability to undo/modify/redo any corrections all become metadata that are securely managed providing complete data history. Provisions are needed for automated and manual markup of the data, auditable event logs, source tracking, and the ability to undo/modify/redo any corrections.

Real-Time Alerts & Notifications

Notifications are an important part of any real-time system. Is something happening that the hydrographer needs to know about? Did something not happen that the hydrographer should know about? What are the most urgent priorities for work today? Are the station health indicators at a gauge trending toward imminent failure? Are there early indicators of a hydrological event that the hydrographer will need to respond to?

“Quality has to be caused, not controlled” – Phil Crosby

A modern hydrological data management system can be configured to automatically provide timely information needed to cause quality. Quality caused is work avoided.

Automating quality controls requires a predictive approach rather than relying on a reconstructive set of tools. The focus is on verifying site-specific assumptions implicit in the predictive parameters. Planning and decision making is evidence-based and datadriven, resulting in responsive and accountable actions. Modern water data management systems provide warning of potential station health issues and indicators for preventive maintenance and early intervention, thus improving visit scheduling and technological choices.

Timely, Dependable Data

By automating control processes, modern systems improve quality assurance and the rigor of final approved data. The hydrographer spends less time on manual, tedious, data manipulation, which allows for more time in the field and for performing higher level analyses. In other words, the computer does what the computer does best and the hydrographer does what the hydrographer does best. Disinformation is diverted from the gauge to web page communication. Unit value data are made available as a final product in the quality management framework. At every step in the process, the best available data are immediately available to end users. Final approved data are available much sooner.

4. Data Standards & Procedures

The quality management mantra of “say what you do, do what you say” is as true for data computation as it is for every other step in the data production process. There is no data transformation that should be done by ‘magic.’ The algorithms used in data computations must be easily discoverable. Is daily discharge calculated from the arithmetic mean of stage readings or is it the integral of unit value discharge divided by duration of time? The use of normative reference documents reduces the likelihood of systematic errors, improves the inter-comparability of data from different sources, and ensures that data will stand up to the closest scrutiny.

A community approach to data sharing is required in a modern world.

Standard Operating Procedures (SOPs)

As a best practice, computation methods should be published in an internationally accepted standards document. In a recent global industry survey, 65% of water resource professionals indicated that they use or plan to start using the USGS techniques and methods papers. Those who prefer to use the WMO Manual on Stream Gauging will find that the description of methods for computations is taken almost directly from the USGS Sauer (2002). By working with the same modern data management software as the agencies setting the standards, environmental monitoring agencies of all size can truly demonstrate compliance with the published standards and ensure consistency in how data are handled.

Data Visualization

Having software that does the right thing is one thing, having hydrographers that consistently use it the right way is another. The design of the user experience in any data production system is a key factor in the production of consistent, reliable data. One of the key differences that any user will notice with modern data management systems is the sophistication of the graphical user interface, which drastically improves productivity. Data visualization dominates the screen with intuitive controls that are made obviously available for any data manipulation.

Training

Hydrometric data are complex, but the tools for managing the data do not need to be. Most new users of modern systems are surprised at how intuitive and easy the graphical interfaces are to use. Off-loading the burden of software training to a vendor with a rich and sophisticated suite of resources means that more time and effort is available for training in best practices approaches to improve overall productivity, quality, and reliability. Adoption of software that is widely used improves recruitment because the best and most talented hydrographers want to work for agencies that use software that they already know and trust.

5. Water Data Dissemination

Hydrological data is most valuable when available to the right people at the right time for the right place. To optimize the usability of water data, it must be searchable, discoverable, and sharable. But water data are inherently complex and this complexity has been a barrier to usability.

Data Sharing

Most data users have developed a one-to-one relationship with their data providers. This relationship may have taken the form of receiving hard-copy data summaries, telephone queries, flat file data exchange, or intimate familiarity with the use of a web site. When data are needed in a machine readable format, specialized parsers have been developed and one-off data push/scrape communication pathways built. These solutions are brittle and inflexible, requiring the assistance of IT support in both agencies for any change in data sharing requirements.

A community approach to data sharing is required in a modern world. The mandate for monitoring is evolving with increasing ownership at local levels of governance. There is rarely a single authoritative data source. As data sources become more fragmented, integrating technologies are increasingly necessary to make data searchable, fitness for purpose discoverable, and for the data to be easily machine readable. Fortunately, a very versatile and flexible alternative for data search, discovery, and access is now available and supported by modern data management systems.

Data Inter-operability

The Open Geospatial Consortium (OGC) is an international consortium of companies, government agencies, and universities, which have developed publicly available interface standards for information and services. A Hydrologic Domain Working Group within the OGC has successfully implemented a new standard for water data called WaterML 2.0. In combination with other OGC standards, such as Web Feature Services, finding water data is becoming as easy as locating the nearest coffee shop on your smart phone. Open any internet browser, do a map based search for water data, discover the metadata for all datasets identified, and find a link to access the data. Furthermore once a connection is made to a dataset, the data are self-describing so that there is no longer a need to develop a different parser for data from every data provider.

A modern hydrometric data management system supports OGC standards for interoperability. The main benefit for data providers is that end-users are able to self-serve, greatly reducing the need for stand-alone data dissemination systems and the accompanying burden of providing repetitive, time-consuming support to end-users of the data.

Reporting & Publishing

WaterML 2.0 supports general data access, but there is still a need for highly customized dissemination of information. Modern hydrometric data management systems come with standard reports, templates, and tools to configure specialized reports as needed. The production and distribution of these reports can done ad hoc or be automated so that end-users receive timely information either on an event-triggered notification basis or on a regular schedule. Organizations can also publish data in real-time to the web in a comprehensive, virtual earth-map user environment.

Conclusion: Making the Move to a Modern Data Management System

A lot has changed since the last major episode of investment in hydrometric data management systems. In the past many operators could build their systems in-house or cobble together piece-meal solutions using a mix of scripting and commercial software. Most hydrometric data providers have reduced their professional staff to the bare requirements for data production and have little flexibility for redeveloping those systems. Even if they had hydrologists they could commit to the task, the design of modern graphical software is now so complicated that whole teams of specialists are required.

There are three typical excuses not to modernize that are widely shared across the hydrometric community.

"Change is Hard" The adoption of any modern hydrometric data management includes: migrating data, adapting systems and methods, and training staff. Selecting the right system and services makes change easier. Considerable flexibility is required in mapping custom business processes and workflows to a new system. The best systems are adaptive, not prescriptive. None of this is easy. But change only becomes harder as systems age, key expertise is lost, and 'quick-and dirty-fixes' proliferate.

"Money is Tight" Inefficient, unproductive data management is not a good solution for tight budgets and increasing data delivery expectations. Modern data management systems streamline data production ensuring timely, accurate, reliable data and information services with a vastly reduced number of human touches on the data. A USGS study on efficiency has shown that the adoption of a commercial rating curve tool alone allowed the agency to save 59,000 hours per year (USGS Fact Sheet 2007).

"The Future is Uncertain" The future will not look like the past. The time to defer decisions about modernization has passed. Spreadsheets, custom solutions, and legacy systems simply cannot keep up with the growing demands of today's water data consumers and stakeholders. Progressive, forward-looking vendors have a vested interest in anticipating industry trends – they are dedicated to ensuring that solutions for new and emerging technologies are continuously available.

The 5 key considerations for timely, accurate hydrological data management collectively provide a compelling argument for modernization. The three barriers to modernization are red herrings. Industry leaders, large and small, have overcome these fears and are embracing the efficiency, effectiveness, and reliability afforded by modern hydrometric data management systems. From the very smallest to the very largest hydrometric operator, a shared data management solution is now possible.

Published by Aquatic Informatics

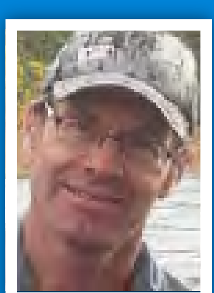
To watch a VIDEO demonstration of AQUARIUS or to learn more, please visit www.aquaticinformatics.com.

Aquatic Informatics™ Inc. provides leading software solutions that address critical water data management and analysis challenges for the rapidly growing environmental monitoring industry. It understands the challenges of environmental data management. Its flagship product AQUARIUS is carefully engineered to ensure a smooth transition to modern best practices for hydrometric data management.

AQUARIUS is the time series data management software used by the largest and most advanced hydrological and environmental monitoring agencies, including the USGS and Water Survey Canada, and it is scalable to fit the needs of any size of monitoring network. AQUARIUS has a simple design that combines an intuitive and efficient user interface with the latest hydrological science and techniques. AQUARIUS allows hydrologists and technicians to manage the data they collect more quickly and to a higher level of quality, so they can deliver more effectively on the evolving demands of stakeholders.

AQUARIUS is fully configurable to adapt to and support any Quality Management System. AQUARIUS excels in retrospective data analysis as well as real-time continuous work flow management. It has a unique portfolio of features for real-time sanity checking, error detection, data cleaning, data flagging, automatic bias corrections, and rating shift management to streamline quality controls. AQUARIUS automatically builds an auditable data processing, correction, and editing log, ensuring data defensibility. The AQUARIUS rating development tool is engineered to support the latest global standards set by the USGS, ISO, WMO, and OGC, to ensure the highest confidence in calculations of flow. Its flexible reporting and publishing tools include an extensive list of industry standard report templates and an integrated report template builder for customized reporting.

The AQUARIUS architecture is designed to manage and synchronize data from multiple networks for fast reliable solutions to difficult water management issues. Advanced support for data migration ensures continuity with legacy systems and its state-of-the-art architecture ensures your system is secure, scalable, and integrated. The intuitive AQUARIUS toolboxes combined with comprehensive training and support resources ensure rapid deployment and implementation in any operational environment.



Meet the Author

Stu Hamilton is dedicated to improving the science of water monitoring. He has 17 years of field experience in northern Canada, and nearly as many years in research and development and operational management with the Water Survey of Canada. Stu is an Associate Expert in Hydrology with the World Meteorological Organization (WMO); the Canadian Liaison with the Hydrometry Committee (TC 113) of the International Standards Organization (ISO); the current President of North American Stream Hydrographers (NASH); and a Member of the Open Geospatial Consortium (OGC) Hydrological Domain Working Group. Since 2009, Stu has been a Senior Hydrologist at Aquatic Informatics. For more insights from this author, you can read Stu's Hydrology Corner blog here.