METHODS FOR AUTOMATIC STORAGE, VISUALIZATION AND REPORTING IN DATALOGGING APPLICATIONS

Andres Thorarinsson
CEO, Vista Data Vision, Hofdabakki 9c, IS-110 Reykjavik, Iceland, Email: andres@vista.is

ABSTRACT. The handling of data in geotechnical and environmental datalogging applications is quite different from typical industrial data handling. This is partly because industrial users are most interested in the present situation while researchers often value long and uninterrupted series of measurements. There is, however, increased interest to incorporate the best parts of industrial systems into geotechnical datalogging applications, namely database storage, visualization of data, alarm services, web service and the latest communication and interconnection technologies between software applications using open standards. The purpose being to cut the cost of data handling and reporting, perhaps as much as 80% compared with the older tabulated data handling, while giving new meaning to the word "logged data".

INTRODUCTION. After 15 years of work with a wide variety of Datalogging systems, it has become clear to the author as to which are the cornerstones of a successful data handling system. Based on this experience, this paper describes key functions which the author considers a powerful data handling and visualization system should include. An Important part of such a system is an easy-to-use Web Service that is popular by users and which greatly simplify data analysis. A great deal of the ideas presented here originates from industrial systems for public utility operations, sewer and water services, and energy management.

1. DATA ACQUISITION SYSTEMS

1.1 Data Acquisition Systems

There are many systems available, some more sophisticated than others. A good system is able to connect to sensors via analog inputs, digital inputs, counters and serial communications. If flexible inputs to the datalogger are important, even more so are all its accessories like a wizard for automatic creation of datalogger programming code, general programming language, battery backup, communications initiated by alarms, data retrieval software with scheduling and time synchronization with logger internal clock, and all possible communications options available including direct line (serial, bluetooth), RS485 multidrop network, dedicated line (modem on private line or leased line), radio link (public, shared or private channel), dial-up line (land line, cellular modem), IP Network connection (network connection, dial-up, cellular IP modem) and satellite.

1.2 Communications

In the past, the primary option for communications between a datalogger and Data Center were landline dial-up connections or a radio links. In either case, site preparations were needed and installation could take some time. Newer dial-up cellular modems changed all this. When the cellular modem was configured, the datalogger was able to communicate as soon as power was on. These dial-up cellular modems have totally changed the way data is collected from dataloggers and made the datalogging industry much more flexible.

A special type of communications is IP communication using the new and popular cellular IP modems. These modems are fast and reliable and easy to implement and use either a fixed or dynamic IP number. However, even if IP cellular modems are reliable, users should be prepared for possible connection problems during peak periods.

For the Data Center to initiate communications with a cellular IP modem, the IP address of the modem needs to be known. As the number of unique IP addresses is a limited resource, the phone company most commonly utilizes a dynamic IP numbering to cellular IP devices, issuing a new IP number automatically each time a device is turned on, and dropping an IP number if device is inactive for some time. In the case of dynamic IP numbering, the Data Center has no way of knowing the IP number of the device and therefore no data collection can take place. The solution has been to pay the phone company an extra fee in order to have a fixed IP number for the cellular IP modem.

In order to make IP communication to the datalogger easier and of lower cost, the solution is to program the datalogger to send the cellular modem's IP address on regular intervals to the Data Center. The Data Center receives the IP information and inserts it into the data collection engine that collects data from the data logger. The benefit of this procedure is that communications and data collection can be set up without making any special IP address arrangements, at extra cost, with the local phone company.

Another solution is to have the datalogger use its cellular IP modem to send data messages to the Data Center at regular intervals, either to write a file into the Data Center's FTP site or to send the data as email. In this case the Data Center will continually monitor the FTP site for new data or, to check the Mail Server for new emails. Once the Data Center has read the new data, it is handled the same way as data collected by data retrieval initiated by the Data Center. The drawback with this solution is that it is not possible to connect to the datalogger to adjust, trouble shoot or reprogram. If this is a problem, a solution is to include the IP number of the cellular modem in each message and thereby the Data Center will always know the IP address of each datalogger.

2. DATA HANDLING AND VISUALIZATION

2.1 Database storage

Since the early days of data logging, data has been stored in binary files or text files. Binary files have the advantage of being small and fast but they have the disadvantage of being hard to read as the data is stored as binary values ~~but~~ not as readable ASCII-characters. Furthermore, if its translation key is lost then the data will no longer be decoded and is therefore lost forever. Text files are more solid but larger in size and have various formats where csv (comma separated values) data is most common in English speaking countries. Every one who has been working with dataloggers for some time have a great number of data files and increasing effort is needed to manage it. Often, there are multiple copies of data files from the same site, sometimes with interleaving data, therefore, in order to work with older data, investigation work is needed before useful results can be achieved.

It is here that a database solution comes to the rescue by storing all data in a single system. The database is a perfect solution to store a variety of data in an organized manner. A database also has much more capability than just storing the data. A well-designed database system for logged data is the foundation which all other data handling and services are based, including data visualization, analysis, reporting and distributed service as well as web services.

When planning for the database storage of logged data, the most important feature is the organization of its data tables. A strong database structure will limit each database table to a limited set of related variables. The price for the database developer is a complex database structure but the reward for the user is a flexible system with a fast response time that will stay fast remain that way for years to come.

2.2 Data Visualization

The most common form for the visualization of data is to display it as trend lines on graphs. A graph can include several variables from the same datalogger as well as variables from any other datalogger in the system for cross-reference. A page may include several graphs with some tens of variables on display at the same time. It should be easy to group related pages together for easy overview. The user should be able to easily select any time period and to browse data back and forth in time. A graph should also be able to display simultaneously variables on different time periods like 10min, 15min and 1hour. Data should also be available as tables as well as graphs and the user should be able to export any data from a graph as text file to be used in other data handling and management systems.

Scaling of the Y-axis of graphs is of greatest importance. Often a fixed Y-scale is well suited, as the user can easily sense the value by its position on the graph. If the Y-

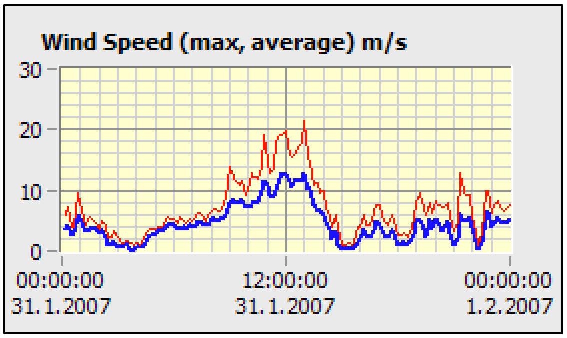

Fig. 1: This graph for wind speed has fixed Y-scale, bold blue trend line for average wind speed and narrower red trend line for maximum wind gust.

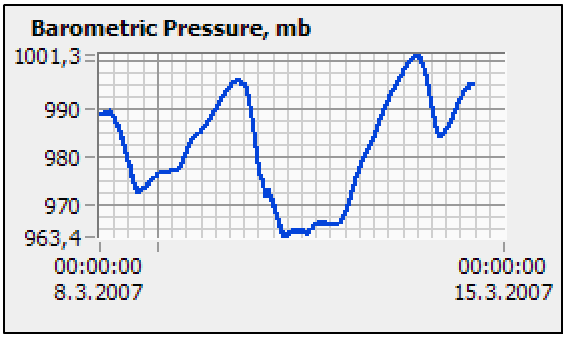

Fig. 2: This graph for Barometric Pressure has auto Y-scale and bold blue trend line.

scale is fixed, it needs to be set for the maximum deflection expected, like 0-360 for wind direction or 0-100 for RH%. Sometimes, a graph gives a better overview by using a fixed value for one end of the Y-axis and using auto-scale for the other, for example, wind speed, where lower end of the Y-axis is set to 0 and the upper end to auto-scale. There should also an option to set both ends of the Y-scale as auto, so the values on display will always fill the graphs. When using auto-scale, a sensor error (resulting in logged value of +6999) will make the correct values become unreadable, and therefore minimum automatic data validation is sometimes found useful to correct such errors.

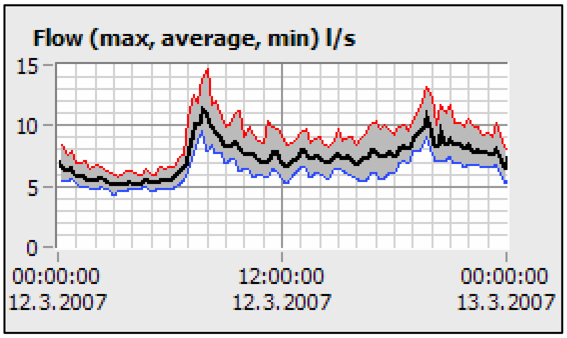

Fig. 3: In this graph for flow, it is easy to recognize the average flow (bold black line in center) as well as well as max (narrow red line) and min (narrow blue line) flow.

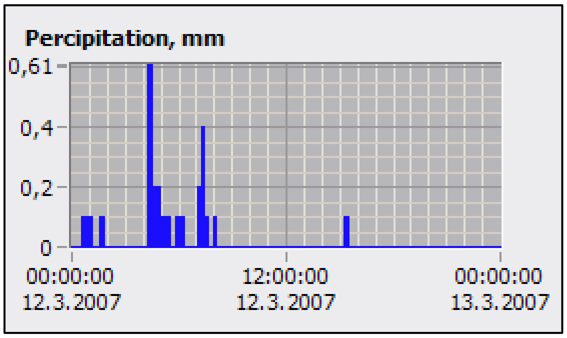

Fig. 4: This graph use bars instead of trend lines to display percipitation.

Other nice graphical features include inverse Y-axis, adjustable color and thickness of graph lines, showing data as bars or areas, changing background colors and the ability to set the Y-axis to a logarithmic scale

2.3 Web Service

Once data is in-house and the data visualization system has been configured, there may be an interest among co-workers and clients to have access to the data. Until recently, the only option was to install copies of the data visualization software on each and every PC to access and analyze data. Even worse, users outside the closed network would have limited or no access to the data. All this is now changing with the advent of advanced internet/web techniques.

Nowadays is easier to implement web service than ever before, and a complete data visualization system can operate as a stand alone web service, giving outside users full access to all data, graphs and reports. Running web service means that access control becomes important, where each account may be adjusted as needed for individual needs. As an example, a guest access to data may have limited possibilities for features and sites while a supervisor may have access to all features and all sites. Using access control, each user can have access only to those graphs and those features that fit his account. Web-access is perfect to give in-house co-workers access to data, trend lines and reports, as well as giving data service to customers and the public.

The bottom line is that once the general handling of logged data is set up, the web access to data for clients, outside co-workers and public becomes the single most important part of the data service and the most useful one.

2.4 Web Map

While graphs are probably the most common way of presenting data, Web Maps take different approach. On a Web Map, generally only the latest value of each variable is present on the display. However, an important feature is that a web map can include a background map, a flow diagram or a photo, showing measured values at the correct location. And perhaps the most important feature is, that background color for each variable may change depending on the alarm state for that variable, for example become Yellow for H (high) and L (low) alarms and Red for HH and LL alarms.

Now, the operator only has to take a quick look at the web map for that site. If all background colors are green, all is well. However, yellow and red background colors may indicate problems which the operator will have to deal with, much like as is incorporated in standard industrial SCADA systems.

Web maps represent an important addition to a powerful data visualization system.

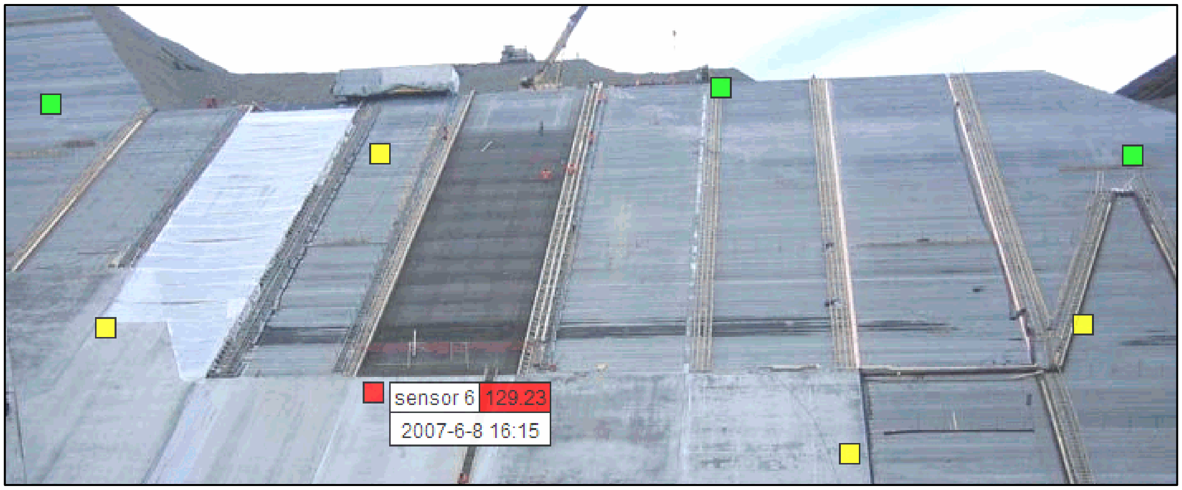

Fig. 5: A web map gives great overview of complex instrumentation. In this figure, a simulated set of sensors show normal status (green), H alarm status (yellow) and HH alarm status (red). The actual readings are visible by clicking on one or more of the sensors (see sensor 6 as example.

2.5 Real-time Systems

In real-time data management systems, measured values are read from a datalogger in real-time. Most often, "real-time" means reading data every second or every few seconds. Until recently, real-time systems were hard to build, often because of communication difficulties. A real-time connection required a dedicated communications line or radio link from Data Center to the datalogger site. Nowadays, system builders have two new options for communications, like a direct line ADSL IP communications and the new cellular phones featuring GPRS/Edge. Live IP communications are available everywhere where there is a telephone exchange nearby or a cellular network.

When using ADSL communications, baud rates up to 100kB are possible. With such a fast connection, not only may data be retrieved in real time, but additional data like data streams from dynamic sensors, still pictures from cameras and live video may also be brought to the Data Center in real time.

All this data brings constraints to the Data Center as a special capability is needed to handle the real-time data.

Real-time data service includes alarms, live data on web pages and even control where web based users can flick a switch on a web page to start a process in the

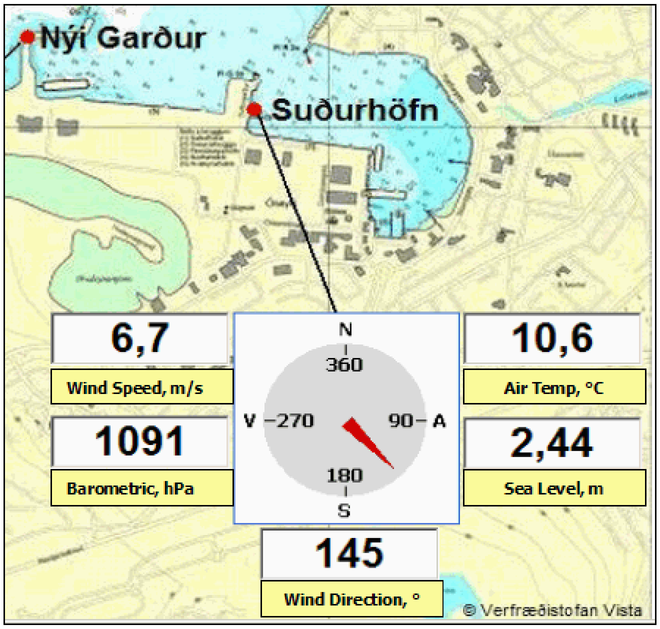

Fig. 6: Live weather is brought to a web site, with weather data being updated every 5 seconds. A flicker-free image is a great help to (the Harbor) staff which now has easy access to the most vital operational information; wind speed, wind direction and tidal level.

datalogger, for example like starting a pump or initiate a command to start a calibration process. Latest developments in data web pages do allow values on the web pages to update every few seconds without refreshing the whole web page, therefore behaving much like an ordinary display on a SCADA system.

3. AUXILIARY DATA SERVICES

3.1 Virtual Variables

One of the advantages of having logged data is that new calculated variables can be created as a function of measured variables. As an example, dew point is a function of air temperature and relative humidity, chill factor is function of air temperature and wind speed and water flow may be a function of water level. The possibilities are endless. Often, the idea of creating a calculated variable is not born until after a datalogger has been deployed and further programming of the datalogger is not feasible.

In order to overcome this difficulty, a data visualization system may incorporate a feature that allows the user to create a virtual variable, i.e. a value that is calculated as a function of any variable already in the database. Now, instead of having the datalogger calculate new variables, these calculation formulas can be created afterwards or as needed and stored in the database. Furthermore, a formula can be tested and verified until proven to be working correctly, without any interference with the program in the datalogger. The virtual variables can be displayed on graphs like any other variable, and can be included in all reports.

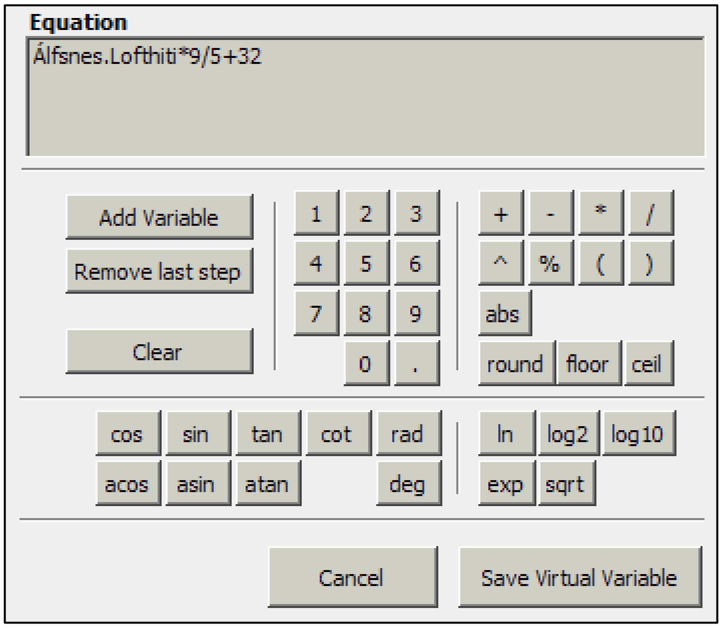

Fig. 7: Virtual Variable is built in an equation editor. Here, an air temperature calibrated in °F is converted into °C. The equation editor allows for complex equations with any combination of measured values from any field station using a variety of operands.

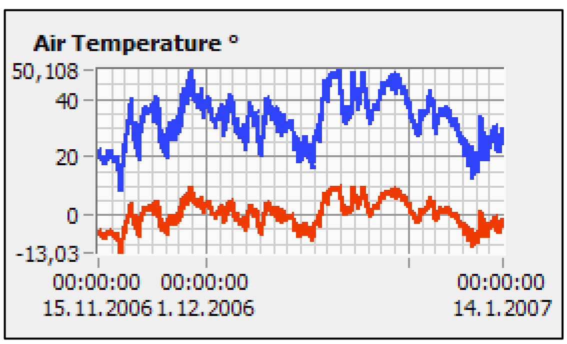

Fig. 8: This graph shows the result of the conversion from °F to °C, see also fig. 8. The lower trend line (red) is the original variable while the upper trend line (blue) is the new virtual variable calculated according to the equation in fig 8. While both trend lines are shown her for comparison reason, most commonly only the virtual variable would be on display in actual use.

Virtual Variables represent a great step forward in making data visualization systems more powerful than ever.

3.2 Alarm Services

Data logging has different meanings to different users. While some users are mainly interested in seeing what has already happened, others would also like to be notified if measured values suddenly fall outside alarm thresholds. Such alarms could be caused as a result of high water levels, ambient measurements to be out of limit, low pressure etc. Users in need of alarms need a specific alarm service.

There are several different aspects of alarm service:

A) Most common for alarms are the threshold levels that generate an alarm if exceeded. Such levels may be LL, L, H and HH, where L=Low and H=High. These threshold levels are checked by the Data Center once data has been collected, and an alarm sent as email and/or as a SMS text message via Cellular phone if threshold is crossed, either to announce a new alarm situation or to clear an alarm situation. There are two groups of staff members that should receive alarms, e.g. the Maintenance Group and the Operators.

Alarms for low battery voltage or power outage may be sent to Maintenance while alarms for high level or excess movement may be sent to Operators. The message set-up may incorporate work shifts and allow some staff-members to work dayshift and others to be on duty outside normal working hours. An important feature is an automatic reminder that sends reminder messages once every day to those on duty.

Site: Alfsnes WeatherSt

	Monitor	Low Low	Low	High	High High	Send Low & High Warning	Send LowLow & HighHigh Warning	Deadband
AirTemp	☐	0,000	0,000	0,000	0,000	Both	Both	2 %
Battery	☑	11,500	12,050	14,000	20,000	Both	Both	2 %
Battery2	☐	0,000	0,000	0,000	0,000	Both	Both	2 %
DiffVolt	☐	0,000	0,000	0,000	0,000	Both	Both	2 %

Fig. 9: Each and every variable in any datalogger is a candidate for an alarm service. In this example, the battery voltage is monitored for LL, L and H values. The HH value is set out of range. Notice the Deadband which is adjustable.

B) Then, there is an alarm situation if data is not being collected because of some failure, whether it is caused by loss of power, drop of communications or system failure. As an example, if data is expected to be collected every 4 hours but new data has not shown up for 5 hours, then an announcement should be sent alerting that the data collection is out of schedule.

C) Another cause for alarm is response time. Is it sufficient for an alarm situation to be first detected when data is being collected by the Data Center or should it actually be detected onboard the datalogger? As an example, if data is collected every 4 hours, and an alarm situation occurs 1 minute after data has been collected, then there will be another 4 hours before data is collected again and the alarm situation detected.

If the datalogger is programmed to detect an alarm situation, there are two options regarding distribution to users: 1) to notify the Data Center that an alarm situation is detected and the Data Center distributes the alarm messages according to its recipients list, or 2) to send SMS alarms direct to users. Either way, on-board detection of alarms opens doors for Early Warning Systems.

D) Finally, the operator needs a classical overview of alarm status for all stations as a simple text display with each alarm as separate text line and with the latest alarm on top of list. New alarms should blink with a yellow or red color, and the color should remain steady after an operator's acknowledgement and the text line should disappear when alarm situation is cleared. A log list should show the history of alarms. Preferably, the alarm handling should be handled from a web display as with all other data handling and browsing, allowing the operator to look into his alarms from any location with a web browser, whether it is at work, while traveling or at home.

Alarm Overview | Alarm Setup | Data Update Alarm

Active Alarms

	Site	Variable	Alarm Time	Alarm	Limit	Current Value	Current Time	Confirm
	Sewer Pump Station	Pump 4 Flow	2007-06-08 15:15:00	low	0	0	2007-06-08 15:15	Confirm
	Sewer Pump Station	Pump 1 Flow	2007-06-08 09:45:00	low	0	0	2007-06-08 15:15	Confirm
	Sewer Pump Station	Pump 3 Flow	2007-06-08 07:45:00	low	0	0	2007-06-08 15:15	Confirm
	Sewer Pump Station	Pump 2 Flow	2007-06-08 03:00:00	low	0	0	2007-06-08 15:15	Confirm

Fig. 10: This alarm panel is accessible on an access controlled web site. The operator can access the alarm overview from any computer on the internet, acknowledge any or all alarms, re-adjust any alarm level and create a new one. Such a front end to a logged data management system helps distributed operations.

3.3 Reporting

Reports incorporated into a data visualization system need to be powerful and speedy. A general report generator should allow users to manipulate graphs and get reports momentarily. This is extremely useful in order to better understand what is happening at the site and how values are related to each other. A report on the fly can include sums, averages, max and min, and be capable of loading for any variable on display, as well as more complex reports like sort-by-size, histogram, XY-graph and polargraph. The XY-graph can be very useful, allowing the user to plot any measured value against any another, for example temperature vs. depth, flow vs. level, wind direction vs. wind speed etc. A specialized version of the polargraph is a 2-variable version called a (wind) rose, which may display any variable as a function of a (wind)direction, like wind speed, air pollution, rain or temperature. The (wind) rose report generator can be a very powerful tool to analyze behavior of measured data.

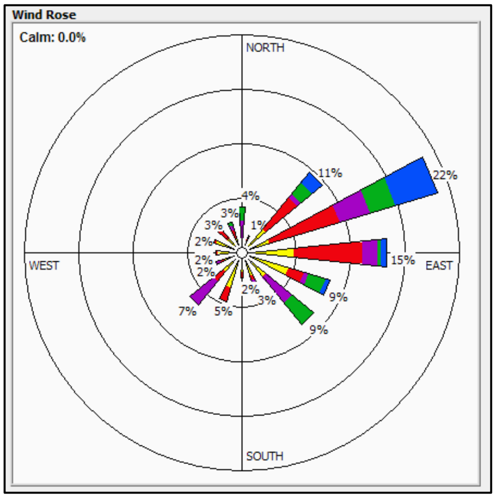

Fig. 11: An user configurable web-based Wind Rose is an outstanding report tool. New is any measured variable can be displayed as function of wind direction, like rain, temperature and %RH.

Then, there are more formal documents, such as daily, weekly or monthly reports for any measured data, which should include a header text, general info text, tabulated data and a graph of the reported variables. An effective method is to have a report generated automatically as soon as the day, week or month has ended and sent by email to the supervisor or owner for approval before being issued in final format report.

Reporting can be a complex issue and it is difficult to imagine that any reporting system will cover all needs of all users. As the data is stored in a database, the user also has the option of using SQL-queries to (automatically) extract needed values and use a spreadsheet template to make a report.

3.4 General information about field stations

In a data logging system with many data logging stations and several users, it may become of importance to have access to information about each station. This information may include general text, an equipment list, a map, a system drawing and site photographs.

3.5 Data Validation

Incoming data may be validated, e.g. checked for errors, before it is approved for use in graphs and reports. Data Validation has several ways of handling data depending on the nature of the error.

A) One type is sensor error which makes the datalogger log out-of-range and indicating same with a +6999 (or similar value). For example, consider a PT100 with damaged wiring. A momentary disconnected wire to PT100 sensor gives over-range resistance reading, which results in out-of-range error in the datalogger. Clearly, any report for this sensor, like average temperature over a period of time, will be wrong. An out-of-range error can be automatically fixed by replacing one or more sequential out-of-range errors by the last good value.

B) The calibration error of a sensor or instrument will cause logged values to be shifted from the correct value. This error may not show up automatically and has to be manually checked. Sometimes, this error is not discovered until later when the instrument is checked or when it becomes obvious that the value the sensor is returning must be wrong. In either case, the validation process will incorporate re-scaling of measured values from a certain date and time. It will then be run automatically on all new data until the error has been fixed, or until another re-scaling rule must be applied. Clearly, the validation has to incorporate a feature that will work for a certain time period but not on previous data and not on data after a certain date and time.

C) The gap-fault is caused by a datalogger not logging for a period of time due to logger programming, maintenance, shutdown, or because of an unconnected sensor. Gaps in data look bad on graphs and reports based on those values may not be correct. Minor gaps like a missing single period can be bridged automatically by last good value. However if the gap is wider, the user has the option of preparing a best guess estimate for measured values for that period and to import that value to close the gap.

D) Data that is to be forwarded to another system for further handling and will be used in a decision-making algorithm or will be made available to public, often require two-step validation, i.e. data should not be made available to the next step before it has been checked and marked as validated by the Data Supervisor.

3-6 A bridge between real-time data and scheduled data - Manual Collect

One of the characteristics of a data collection center is that data is collected according to a schedule, whether it is collecting once every night, every 8 hours, or even every hour. However, a moment may arise when the user absolutely needs to see latest data available in the datalogger, perhaps because of an impending flood or the widening of certain cracks. A very useful feature is to have a button on the web page to allow authorized users to request the Data Center to immediately collect the latest data from a datalogger. This way, the data will be available on the web graph within minutes.

3.7 Dynamic measurements - Fast Sampling Rate

Dynamic measurements are quite different from the more common, slow moving values of ambient temperature and other environmental variables. Dynamic measurements like those obtained from accelerometers used for vibration measurements, are often made at 100Hz or even higher sampling rate up to 10kHz.

Often the data is only stored if the amplitude (or any other easily defined trigger parameter) is higher than a trigger value. Now, few measurements ahead of the trigger value are stored, as well as measurements for a fixed time or until measured value drops below the trigger value. For example, the sample rate could be 100Hz for 30s, with a total of 3000 values for a single incident. This set of data has to be stored as a separate file in the datalogger and then the Data Center is sent a message to upload the file for further handling and alarming. This dataset is sometimes called Burst-data. Clearly, the burst-data cannot be displayed on graph like periodic data, but rather a special burst data browser is needed.

Another option is to reduce the data by having the datalogger calculate characteristic values from the fast rate sampled data, like peak value, power, total duration, power duration, impact factor etc, and store these along with its timestamp. Now, a graph of the calculated characteristics values may well give a good overview of the situation.

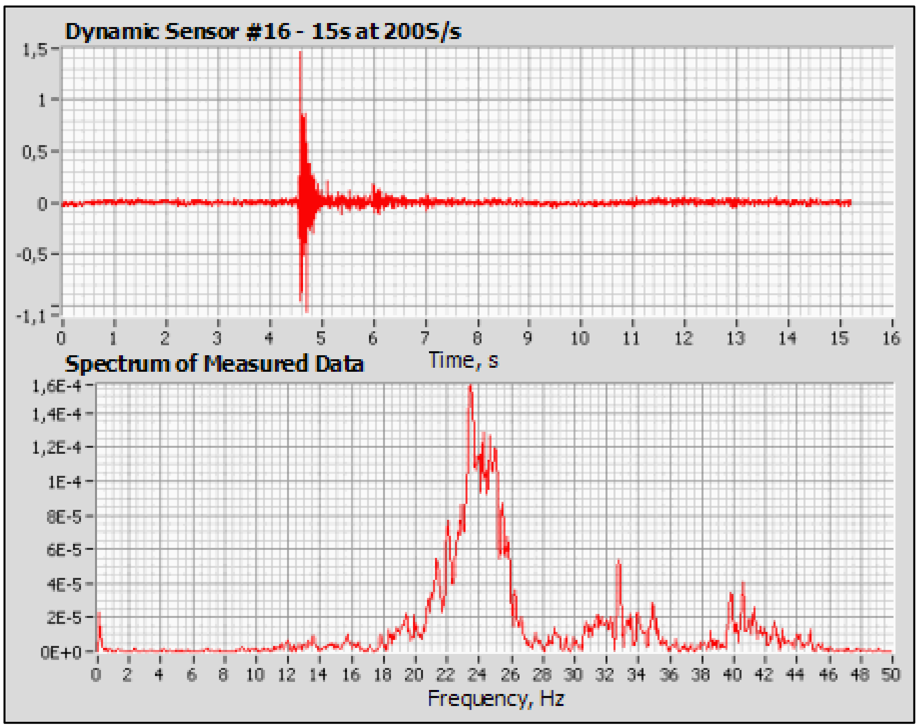

Fig. 12. Screen dump of a view page for a single incident sampled at 200 Samples per second. The upper curve shows the time serie of 15 seconds and the lower curve the corresponding power spectrum. One of the characteristic values is the frequency of the vibration, here clearly seen as 24 Hz.

CONCLUSION

The benefit of a fully fledged database handling system is the short time it takes to implement, the reduced installation and operational costs and the great savings in time and money for all parties involved.

Apart from an organized storage of logged data in a database, easy access is of greatest value. Thereafter, web access to all data, access control, the alarm service and automatic reports becomes the backbone of a streamlined data handling system and which ensures highest quality data service at lowest operational cost for years to come.

REFERENCES

Bjarni BESSASON, Gísli EIRÍKSSON, Ódinn THÓRARINSSON, Andrés THÓRARINSSON and Sigurður EINARSSON. (2007). "Automatic detection of avalanches and debris flows by seismic methods". Paper accepted for publication in *The Journal of Glaciology*.

Elefsen, S. Ó., H. Haraldsson, S. R. Gíslason, H. Kristmannsdóttir (2002). "Real-time monitoring of glacial rivers in Iceland." In: *The Extremes of the Extremes: Extraordinary Floods*, Á. Snorrason, H. P. Finnsdóttir and M. E. Moss (eds.). (Proc. Reykjavík, Iceland Symp. July 2000), 199-204, IAHS Publ. no. 271.

BIBLIOGRAPHY

Andres Thorarinsson (1949). BSc in Electrical Engineering from The University of Iceland 1976. Founded Vista Engineering 1984. Projects in Iceland, China, S-Korea, S-America, USA, Europe. Projects include industrial and utility systems, software development, energy management and communications.