CSCE Annual Conference
Growing with youth – Croître avec les jeunes

Laval (Greater Montreal)
June 12 - 15, 2019

AUTOMATED DATA MANAGEMENT AT SCALE – FLEXIBLE PLATFORM FOR DATA INTEGRATION AND GEOHAZARDS ANALYSIS

Vedala, S.[1,3] and Azarm, R.[2,4]
[1] Canary Systems, USA
[2] Canary Systems, Canada
[3] suchi.vedala@canarysystems.com
[4] rihana.azarm@canarysystems.com

Abstract: Implementation of a complete and dynamic data management system for handling all the types of geotechnical instrumentation for large project is a challenging task. Canary Systems manages data for projects with thousands of concurrent instruments. The types and quantities of data that are available to help clients understand their operational and safety risks have increased tremendously in recent years. Technology advancements have introduced many new types of data beyond traditional geotechnical instrumentation, such as automated total stations, TDR, differential GPS, ground-based radar, satellite InSAR, laser scanners, vibration/seismic recorders, environmental sensors such as rainfall or air quality, among others. While the collection of these data is no longer a technical challenge, the danger presented is the integration and management of all the different sources of data into a single reliable system that can not only store the data but also provide flexible visualization and reporting. Canary System's web-based software allows infrastructure, dam safety and mining industries to import data from various sources and dynamically update all the related charts and reports. The system supports any number of users to be connected simultaneously, allowing clients and external consultants to work together in a single platform and ensure the project safety in unprecedented fashion. The system can be programmed to immediately inform users of data consistency issues and can flag any data that is outside of the expected ranges, either indicating that maintenance might be necessary, or to inform of geotechnical safety considerations. Detailed and complex custom reports can be produced for each site, where data is carefully reviewed and correlated with the other instrument readings or activities in the area to provide a high level of confidence in the geotechnical performance.

1 INTRODUCTION

Geohazards are one of the biggest threats to many forms of infrastructure works. Reliable assessment and mitigation of geohazards thus comprises one of the most critical components of infrastructure design. Technology advancements have introduced many new kinds of instruments to monitor any activity pertaining to geohazards. In addition, geo technicians collect data manually as a part of their analysis. Each of these, generates a pool of data in their daily operations. Each source contributes its part which when taken as a whole can be analyzed to reveal strategically important information that can help mitigate geohazards. While data integration can be a tenuous task, understanding and extracting only vital information from the data, in order to analyze it and use it to drive engineering decisions, it is the most challenging task.

Canary System's web-based software system allows infrastructure, mining and dam safety industries to integrate data from various sources, extract it and analyze the data through detailed custom reports. The system dynamically updates all related charts and reports when new data is processed into the system. Automated anomaly detection can be programmed into the system to indicate any inconsistencies or inaccuracies in data immediately. The automated alarm and threshold configurations greatly help managing risk by alerting the users when any incoming data is not within the expected ranges. This allows users to have a good understanding of the system health and project status and becomes a drive for early-warning systems, routine maintenance and other risk mitigation strategies.

The system enables multiple users to access the software at the same time ensuring data integrity. Interactive data visualization is implemented with 2D and 3D views of data, along with live contouring. Along with static reports that are updated dynamically, interactive dynamic reports can be created for each site that can enhance geotechnical data analysis in a noteworthy fashion. The system is capable of predictive analysis, by implementing curve-fitting and trend line analysis, thus allowing sites to take data-driven decisions with minimal risks. The system above all adheres to the principle of keeping it simple in comprehension and analysis.

2 DATA INTEGRATION

2.1 Imports and Live Connection

The system has the ability of integrating various sources of data on a site to be displayed on one platform thus simplifying the data comprehension for geotechnical engineers. Data warehousing and data modelling is performed in SQL prior to integration of data sources. The system is capable of establishing connections through various methods, from direct connections to remote hardware data loggers, to manually collected data using file imports and direct SQL database connections. Generic data imports can be achieved through custom Python scripts and, allowing integration of virtually any data source, from complex files to Web Services. The system adheres to Extract Transform Load (ETL) data engineering principles, thus minimizing inaccuracies and ensuring robustness of the data. In addition, any data integration failure is alerted by the system logger.

2.2 Quality Management

The software's inherent automated anomaly detection system ensures data quality across the database. Instruments across geotechnical sites can have errors in observations due to physical, environmental and technical issues. Manually collected data is further prone to human errors. Inconsistent or inaccurate data can lead to bad analysis, which in turn can lead to false positives. If data quality is inadequate, improper analysis methods are often used, increasing the risk of missing actual geotechnical hazard warnings.

Ensuring data quality becomes for these reasons of utmost importance in geotechnical data analysis. Using an automated anomaly detection, the system allows to set statistical rules on the incoming data to ensure data uniformity. The system performs a sequence of tests on the data to identify outliers, spikes, and jumps in the recorded values and flags the instruments to be reviewed. The system can automatically invalidate any data which does not pass certain of the tests and allows users to make informed decisions about the health of the instrumentation and any potential safety breaches. This workflow ensures data integrity and allows the users to validate suspicious data for its accuracy.

In some cases, duplicate readings can be made by different data sources. Integrating the same data twice would lead to data duplicity and inaccuracy. The database system is designed however such that for a given timestamp, only one data record can be integrated into the system. For this reason, the system automatically rejects any duplicated data, thus maintaining data consistency. This also prevents users from accidentally overwriting existing data, allowing the process to be fully auditable.

3 DATA PROCESSING

3.1 Advanced Calculation Engine

As part of the process of geotechnical monitoring, many calculations have to be performed with the available raw data in order for it to be analyzed. For some data monitoring, such as rainfall, threshold algorithms have to be implemented to analyze any hazards. These calculations differ from instrument to instrument and site to site. The system's advanced calculation engine allows users to define analytics based on any data in the system. Various statistical and mathematical functions are predefined in the system, allowing users to create any threshold, algorithm or calculation with the data. These calculations are also dynamically updated every time a new data record is observed. The determination of how calculations are configured within an integrated monitoring platform is the most significant and challenging part of a system. Geotechnical teams must set up the best way to transform the raw data by using prior experience and best industry practice. If done incorrectly, outputs from the system will mislead stakeholders on the conditions of a site. Consistency within a monitoring system must also be maintained to provide a simple analysis platform for the geotechnical team and to allow the communication of the results across the teams to be uniform.

Figure 1 demonstrates an example of the calculation configuration wizard of the system, which includes, predefined mathematical and statistical functions including more advanced functions such as the Poisson distribution calculation. The Average Cumulative Resultant Distance per day of a prism is calculated in the figure using the prism's cumulative resultant distance and the system's integrated aggregate functions.

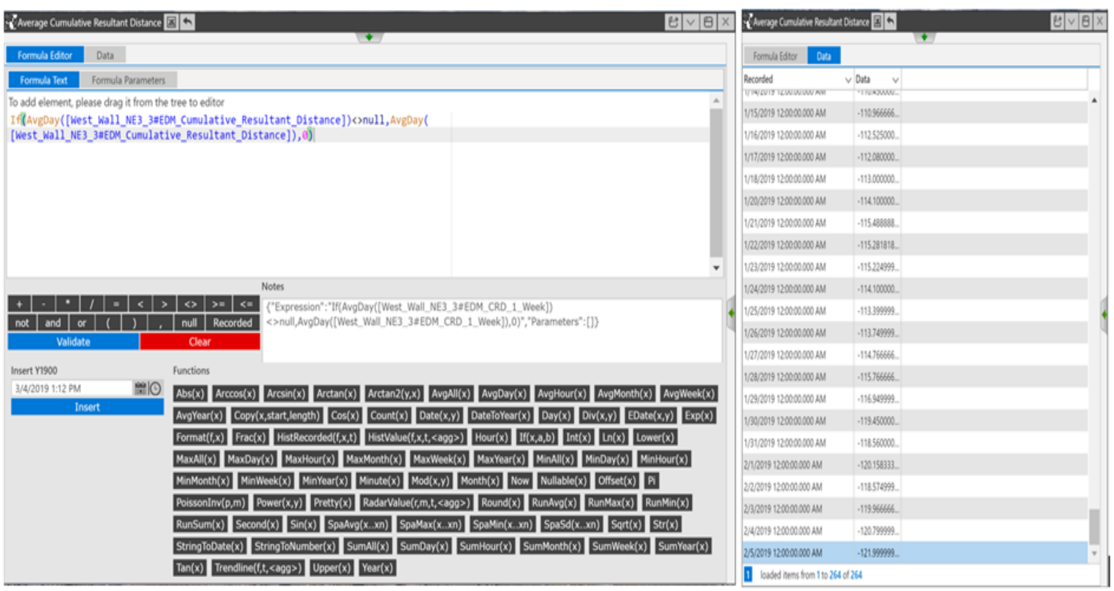

Figure 1: Example of creating a new dynamic calculation using the built-in library of functions.

3.2 Alarm Configuration

In geotechnical engineering, the smallest change in data can lead to hazards causing huge infrastructure, operational, financial and human loss. Monitoring each data record every time a new value is observed and making analysis is a strenuous exercise for geotechnical engineers and will quickly become overwhelming as the quantity and the complexity of the data increases. Automated monitoring is performed by the system by combining the results of complex calculations together with the alarming engine. This module allows setting thresholds on the data and triggers alarms if the observed data is not within the specified range of values. Custom alarm types can be created in the system by the geo technicians depending on the instrument and site, allowing for the alarms to only trigger based on complex conditions, such as if multiple instruments in a given area are simultaneously in alarm or if values are consistently over a threshold for a certain number of readings in a row. The system's alarming engine allows to select from a range of on-alarm/off-alarm actions, including sending email with optional attached reports and charts, thereby alerting users and geo technicians immediate in situations of emergencies.

4 EXTENDED CAPABILITIES TO PREDICTIVE MODELLLING

The goal of predictive modelling analysis is to extrapolate the data into the future and provide predictions and confidence intervals based on the current data. While data monitoring can help understand the current behavior of the site, estimating future trends can change the way sites make decisions. Canary System's software's enhanced features, such as trend lines, allows the system to alarm based on predicted values, providing a proactive approach to risk management and analysis instead of the typical reactive approach widely used in the current geotechnical profession. Trending allows geo technicians to better understand how risks relate to the future and can assist in more accurately made decisions than simply analyzing and understanding historical data.

Figure 2 displays a trend line example from the three-dimensional cumulative resultant distance of a prism with a threshold to notify slope failure. Specific time periods within the data set are used for determining the line. The trend line predicts the slope failure ahead of the actual failure giving geo technicians scope to analyze and mitigate geo hazards. Other examples of trend lines include using them for predicting hydrologic response when piezo metric data is correlated with rainfall predictions, as well as changes in stress if a correlation with draw point extraction is determined in a block cave mine.

Geo technicians can set the trend line type as well as extents and interval of prediction into the future. This dynamic predictive function paired with the system's alarm configuration can alert geo technical engineers and organizations to take measures to mitigate risks in advance, increasing the efficiency of risk management.

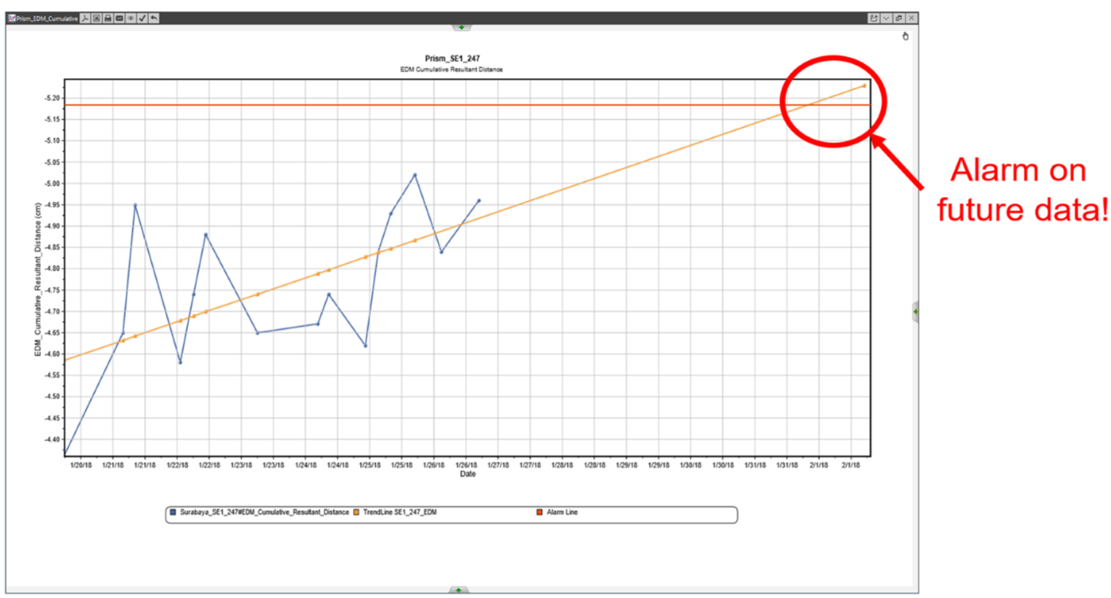

Figure 2: Trend line function created using the three dimensional cumulative distance of prisms.

5 DATA VISUALIZATION

Geotechnical data comes from many different sources and in many different formats. Often the greatest challenge to geotechnical engineers for hazard analysis is simply accessing and viewing geotechnical data in a consistent and convenient format. Visualizing data can be far more effective in making analysis than interpretation through records. The system's web-based software provides 2D and 3D visualization of data along with interactive dynamic features.

5.1 2D Visualization

The system's integrated visualization techniques allow charting any dataset present in the system through simple and line or scatter charts. The charts also allow performing quick statistical analysis by adding features such as aggregate functions and trend lines directly to the chart, improvising the data analysis process. Typically, the best method of understanding correlations between various datasets is by visualizing the trends that affect one another, which is provided by the system's regression lines. Filtering tools available within a chart or image window, provide geo technicians with options to see data pertinent to a specific time range and matching certain data conditions.

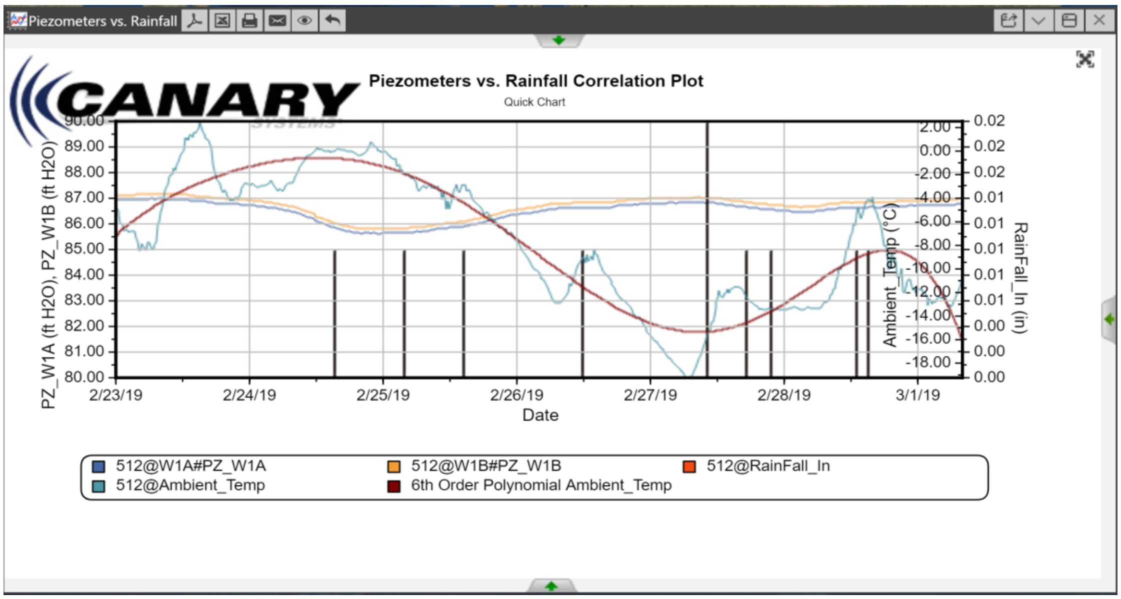

Figure 3: Correlation plot of Piezometer and Rainfall Data.

Figure 3 demonstrates a correlation plot of Piezometers and Rainfall data of an Open pit mine. The multi axes plot implements 6th order polynomial to the ambient temperature which fits the regression line along the data.

Visualizing the layers of a geo technical site and location of the instruments can help in analyzing the overlaying situation. The system's integrated feature-rich GIS component is capable of integrating GIS format files and live map servers at ease. It supports automatic conversions and re projections on the fly between local, global and projected coordinate systems for data and layers. Views and maps can be nested, allowing drilling down to a particular cross-section or plan view. Instruments, reports, charts, documents or can be encapsulated to analyze data right from the map. Dynamic options such as contours, heat maps and vectors based on data values help visualize trends.

Figure 4 visualizes the two dimensional satellite view of an open pit mine embedded with GIS layers and all the instruments located in the mine. Contours are applied to the site to analyze the trends in the data

5.2 3D Visualization

The software's **3D visualization engine** allows for spatial data analysis of any sensor data and display of **3D point clouds and surfaces.** While 2D maps display georeferenced information, instrument locations and system status, visualizing the scene in 3D allows for a much better understanding of the site conditions and the regional effects on the instrumentation. The views, along with the system's filtering options and 3D contouring and scalable 3D vectors provide quick indicators of the instruments data and its behavior with zero or minimal data investigation, reducing time and easing analysis for geo technicians and various users.

Figure 5 below demonstrates 3D visualization of underground mine, cross sectional view of dam, point cloud and slope cross sectional view of INSAR data.

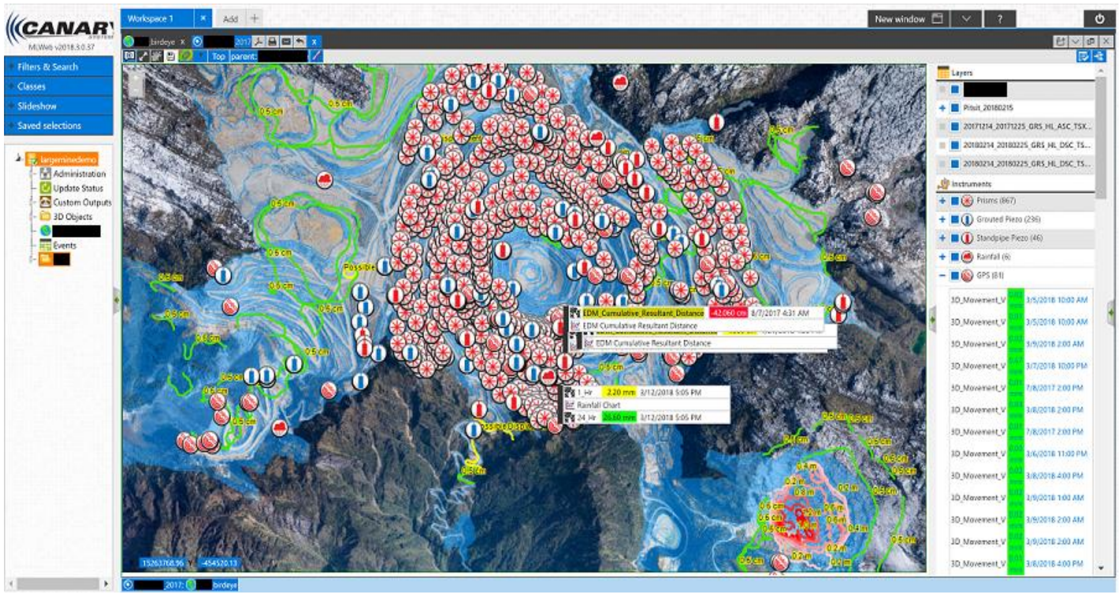

Figure 4: Two Dimensional View of a mine with GIS layers and instruments.

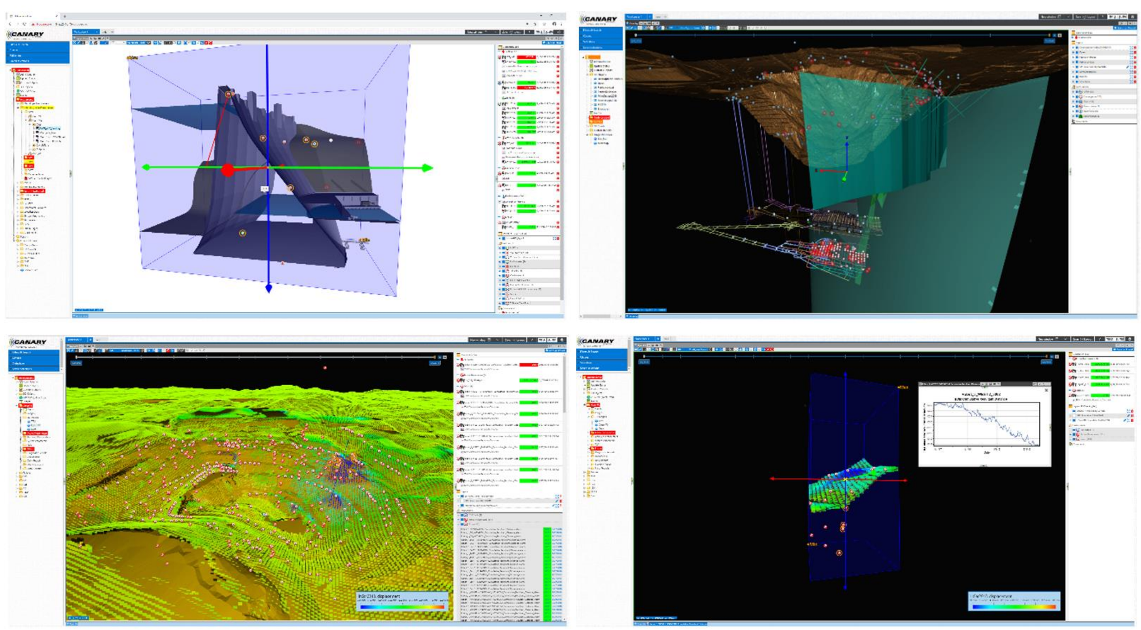

Figure 5: Examples of 3D Visualization.

6 DATA REPORTING

It is imperative for geo-technicians to be able to present the analysis to asset managers and stakeholders. From simple time-series charts, data spreadsheets to specialized outputs for specific data types such as: Inclinometers, SAAs and TDRs; seismic charts including Fourier transformations; wind roses; correlation charts; fully customizable Element Charts, Instrument Reliability Reports and Quick Reports, the system offers plenty of flexible data visualization and output options. Every output in the system can be automatically generated based on a defined schedule, generating customizable PDF reports, which can then be saved on network drives or sent as attachment part of emails.

6.1 Interactive Dashboard

Dynamic and interactive dashboards can be created within the system based on REST API technologies and using the Python scripting language. The ability to setup these dynamic reports directly linked to the monitoring data is unparalleled within the current data analysis process in the geo technical engineering industry. Dashboards are data visualization tools that can be customized to display specific data, however unlike static PDF or paper reports, they are built to visualize and organize data in real-time, allowing users to manipulate and interact with data dynamically. Dashboards can include or exclude any number of live data points, helping geo technicians by giving a broader scope to the data and making them ideal for effective analysis. The system's platform allows embedding these dashboards with ease, making data and hazard analysis efficient along with data collection, automation and visualization.

Figure 6 demonstrates a dynamic dashboard, created in Python using extensometer data of an underground mine. The system's REST API is used to collect the required data. The API ensures data is constantly updated, making the dashboard live and dynamic. The dashboard allows users to filter and extract relevant data to the analysis and sort data based on any given order. Application of powerful functions such as cross correlation, interpolation for performing statistical analysis on the data, along with models to predict data is facilitated with the dashboards as shown in Figure 6. Users can create custom dashboard layouts combining multiple web page elements such as tables and filters to present and visualize the crucial data dynamically in a single view. In addition, the dashboard can help generate dynamic maps, allowing spatial filtering and contouring options.

By summarizing data into powerful BI reports, this visualization and analysis tool helps geo technicians identify the Key Performance Indicators (KPIs) related to any hazard and allows the business units to draw clear conclusions from the system regarding the impact of the data on the operations.

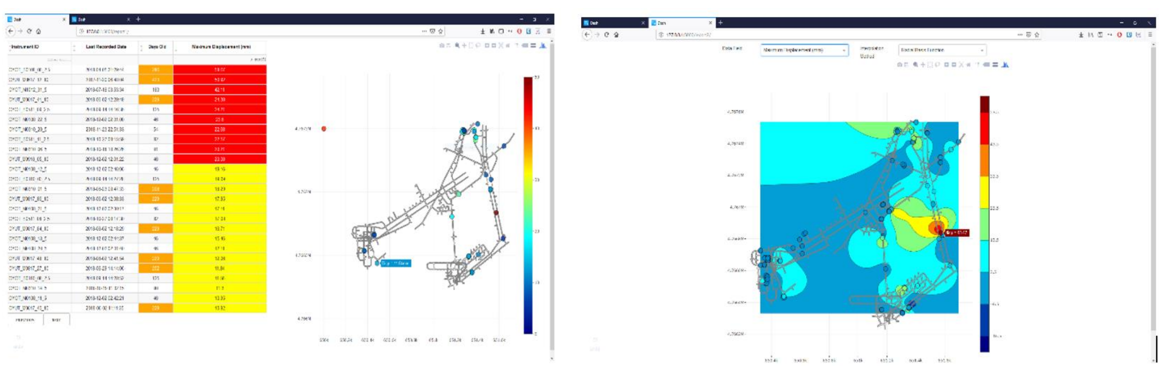

Figure 6: Dynamic Interactive Dashboard created using Python and System's Rest API.

7 SUMMARY

The software system discussed in the paper provides a system that is flexible and efficient in managing large data along with implementing the state-of-the-art data analysis and visualization techniques to facilitate geo hazard analysis. The system enforces the setup of robust data quality checks and ensures data consistency. The system's versatile tools not only help in analyzing the current data, but also predict the future data trends, thus increasing the probability of hazard detection. While the causes of geo technical hazards are often combined with various factors including environmental conditions, a good understanding of the site conditions can help geo technicians take measures to mitigate the hazard risks. The Big Data analysis provided by the system serves this purpose thoroughly.

.

7.1 References

- He C., Huang, J., Ju N. and Xu Q. 2017. Automated Data Processing and Integration of Large Multiple Data Sources in Geohazards Monitoring. *International journal of Georesources and Environment*, IDJE, **3**(1): 9-21.

- Hamman, ECF.,Plooy D. and Seery, J. 2017. Data Management and Geotechnical Models. *Deep Mining 2017,* ACG, Perth, Australia, **1**: 461-488.