# Data analytics for malaria elimination in Zambia

## The challenge

Malaria cases persist in Zambia despite a focus on control and elimination. Health system managers do not have timely visibility into when and where cases occur (or are forecasted to occur) to inform where to allocate scarce prevention and treatment resources.

At the start of the VNM project in Zambia in 2014, health workers supporting the malaria control and elimination program faced several interrelated data challenges —a "problem stack." They were not able to pull as much data as they needed from the existing DHIS2 system. They were also unable to integrate data from other sources; for example, data on medical supply stocks could not be integrated with information on malaria cases. It was not possible to generate the data visualizations needed to answer the complicated analytical questions they needed to answer. From both a technological and public health perspective, the existing data system was insufficient to meet the goal of eliminating malaria-related deaths by 2030.

## The solution

PATH, the Tableau Foundation, and the Government of Zambia launched a partnership in 2015 to integrate new tools and systems for data use to support the Zambian Ministry of Health in its effort to bring the malaria incidence rate down to zero.

The VNM platform drew on a novel approach to tackle these complex analytical questions, which require many different types of data and the tools to turn those data into insights and actionable intelligence; this approach can be applied to many other types of use cases. Operational dashboards were introduced to help district health personnel and frontline health workers track, treat, and report on the disease. Much of the work to inform and design the platform was done by health workers in Zambia to ensure the final system met the country's needs.

In order to scale its early successes, the VNM partnership expanded its technology solutions to support health workers in using data to understand where, when, and how to intervene to control and eliminate malaria. As of 2020, the partnership includes eight technology and service companies that support satellite imagery and location insights, automated workflow for data analytics, cloud-based communication, database solutions, and storage services (Figure 1).

---

**Project dates**
**2015 to present**

**Location**
**Zambia**

### Partners

PATH

Tableau Foundation

Government of Zambia

Alteryx

DataBlick

Mapbox, Slalom

DigitalGlobe

Exasol

Twilio

---

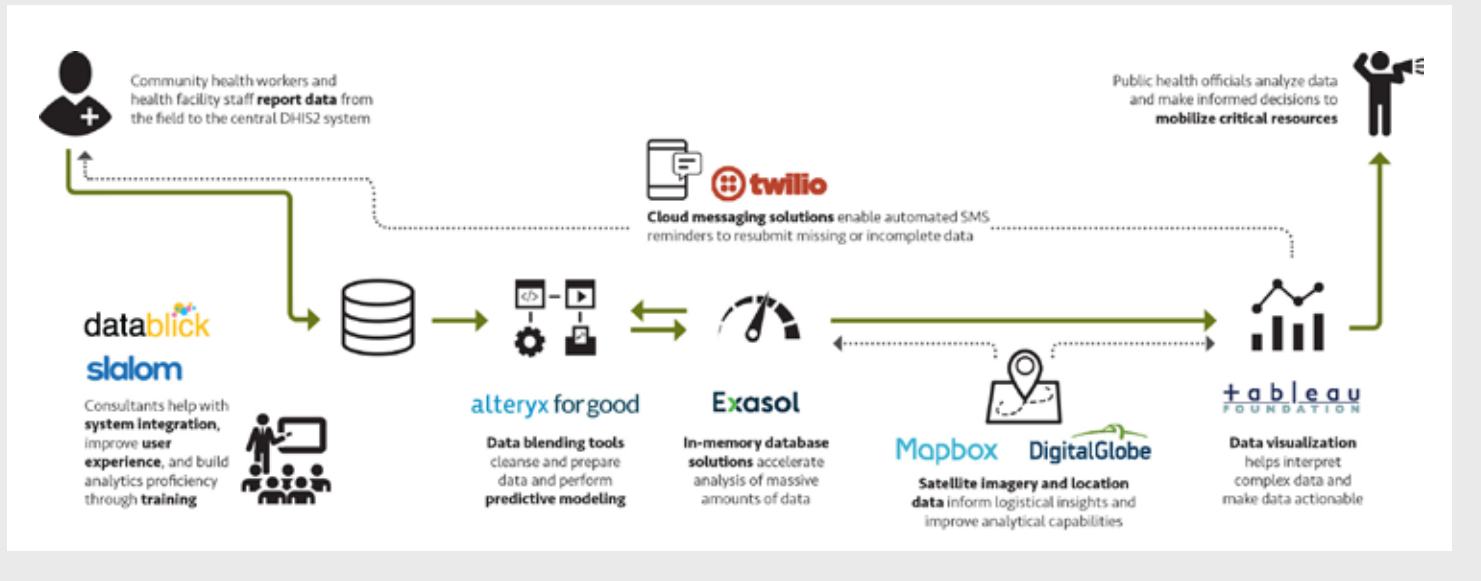### Further information

Visualize No Malaria website @ PATH

Visualize No Malaria – Data for Action Webinar

Using Technology to Advance Global Health: Proceedings of a Workshop. Country-level Digital Health Strategies

An Health Data Science Exchange header

**FIGURE 1.** The Visualize No Malaria (VNM) "stack" of technology solutions and partners
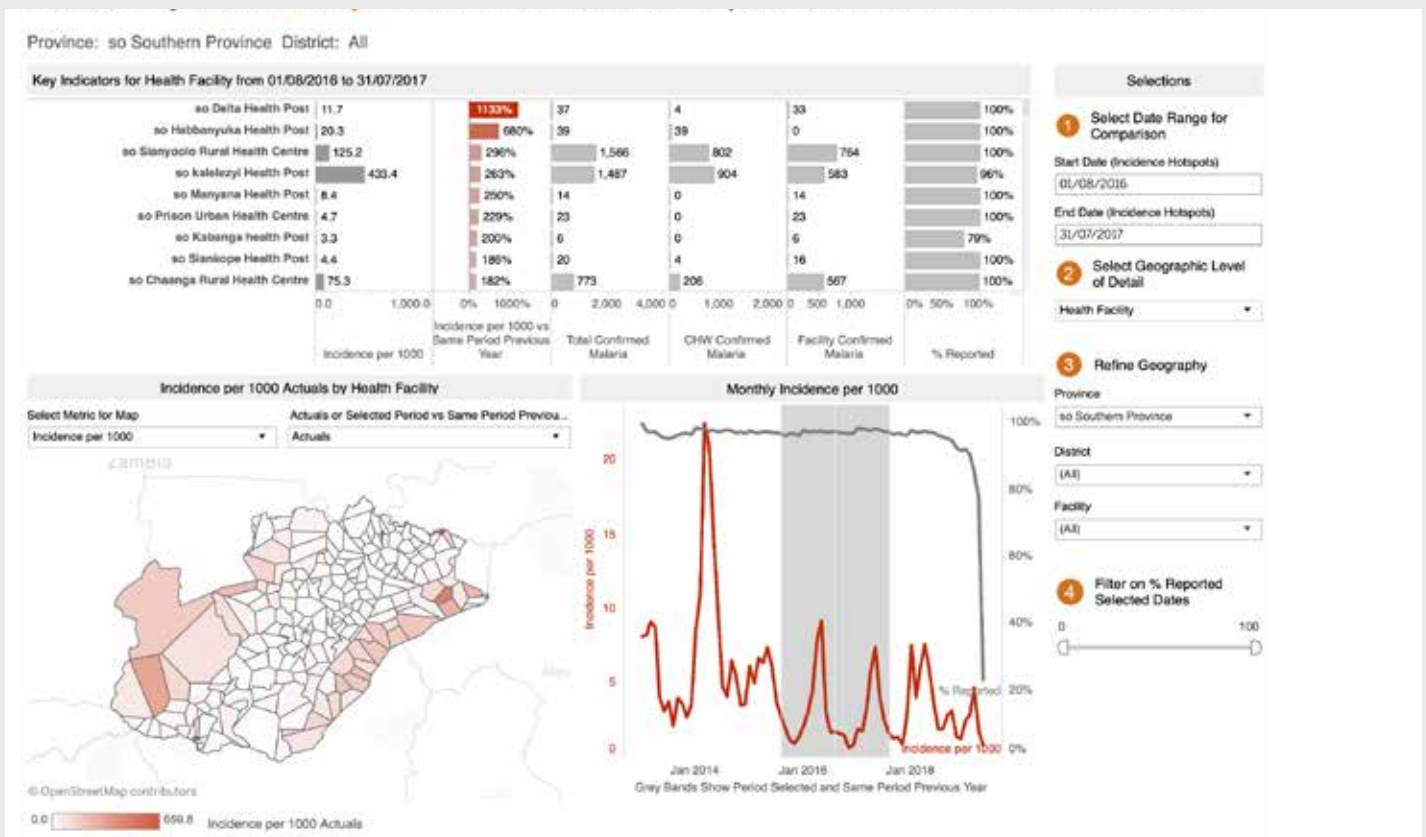
## ⧉ Impact

As part of the VNM project in Zambia, insights from the data visualizations informed how district health managers pursued gaps in surveillance information, deployed health workforce, and adjusted the supplies of malaria drugs and diagnostic tests across Southern Province and eventually across the nation. The VNM platform has contributed to increased malaria surveillance reporting rates and timeliness, improved reporting accuracy, and enabled more data-driven intervention targeting and response.[1] Armed with data literacy and new tools, more than 8,000 frontline health workers now provide high-quality community care and serve as the foundation of Zambia's malaria surveillance system. Users report that the platform has revolutionized the quality and reliability of data and reports and has sped up the accessibility of data and insights to health workers (Figure 2).

The VNM project—together with a complementary set of ongoing interventions, including vector control initiatives like indoor residual spraying and malaria bednet distribution, a mass drug administration program, and intensive CHW training in disease surveillance—contributed to an overall 92 percent decline in malaria-related deaths in Southern Province from 2014 to 2017.[2]

FIGURE 2. Visualize No Malaria incidence dashboard



---

1   https://www.path.org/articles/real-time-malaria-dashboards/

2   https://www.path.org/visualize-no-malaria/

# Scale and future health system applications

Within Zambia, the technology stack used to support the VNM platform is already being adapted to address other health system challenges beyond malaria control and elimination. The platform has also been deployed to support malaria control in Senegal and Ethiopia and has been used to support essential health services in several other countries. In Tanzania, the Alteryx and Tableau assets were used to replace existing data preparation tools used by Tanzania's BID Initiative on childhood immunization. This change improved the data visualizations and significantly accelerated data processing time. Through the US President's Malaria Initiative VectorLink Project in Zambia, Mozambique, Madagascar, Ethiopia, and Mali, the full VNM stack is being used to improve use of malaria case data and improve the supply chain for indoor residual spraying projects.

Each time the VNM platform was adapted for a new environment, the time requirements for scale have significantly decreased, drastically reducing the start-up costs. This acceleration was made possible in part by introducing commercial technology solutions, but also by rigorously prioritizing the use of tools that were fast, were easy to use, and required a minimum of additional training. Further, the VNM platform is adaptable—it is designed for data, not for a particular disease area. Because significant work was done during the Zambia VNM project to test and revise the platform components, less work was required for subsequent applications.

## Applications for emergency response

Several components, or data assets, that are part of the VNM platform were leveraged in the response to the Ebola outbreaks in 2014 and 2015 in Guinea, and in other parts of West Africa in 2016 and 2018.[3] In Guinea, the business intelligence asset from Tableau[4] was used in tandem with a CommCare module from Dimagi to devise a mobile-based patient and contact tracking and tracing system. This digital approach did not replace the existing paper-based system, but added the capability of real-time reporting through the use of smartphones. The assets could be applied to disease surveillance for any infectious disease, as well as many noncommunicable diseases, to provide greater intelligence to local, regional, and national program managers, ministry of health officials, and policymakers about financing and resource needs, to respond quickly to outbreaks, and to help spot small outbreaks before they become large epidemics.

From late 2019 to 2020, data assets that are components of the VNM stack have also been adapted to respond to the spread of COVID-19. As of this publication, Tableau's data visualization software and Mapbox mapping assets have been used to create the COVID-19 Data Hub,[5] which aggregates several vetted datasets and presents a visualization of the disease's spread in every country. The Alteryx data management asset has been used to estimate and forecast hospitalization rates[6] in US counties to predict health workforce and supply needs. Twilio's cloud communications platform is being used to boost telehealth capabilities.

---

3   https://solutionscenter.nethope.org/assets/collaterals/Glob_Health_Sci_Pract-2015-Sacks-GHSP-D-15-00207.pdf

4   http://flipthemedia.com/2015/11/tableau-foundation-building-better-world-data/

5   https://www.tableau.com/covid-19-coronavirus-data-resources

6   https://community.alteryx.com/t5/Alteryx-Use-Cases/Modeling-the-impact-of-COVID-19-on-healthcare-systems-with/ta-p/553437

# ✅ Implementation considerations

Before adopting the health data science assets that are part of the VNM platform, potential users should consider several factors that may determine whether implementation will be successful.

## Health system readiness

The health system should be ready to adopt a sophisticated technology stack; implementation will be faster and more successful if data are captured in electronic systems versus a paper-based system or spreadsheet software such as Microsoft Excel. The optimal environment to successfully implement the VNM assets would already have an existing data pipeline that automatically reports from multiple sources to a central data warehouse. In many countries, health information systems comprise many disparate data sources that require investment of time and skill to clean, combine, and consolidate the data to create dashboards.

Countries should consider data needs and data quality requirements based on high-priority analytical questions. Solutions can start by using the existing data instead of focusing on "perfect" data, as combining data sources and improving their quality can take significant time and resources to achieve. In the case of VNM in Zambia, the Malaria Control and Elimination Partnership in Africa (MACEPA) had worked with the Ministry of Health to develop and establish the community-based reporting tools, systems, and processes of the surveillance system and other necessary inputs to the platform.

When building strong data and reporting pipelines, it is important to remember health workforce requirements. The automatic and manual platform and component updates to run the VNM platform require certain skills, and that will affect what project staff are required to successfully use the platform. Potential users of the VNM platform should consider whether they can employ staff who are experienced with automated reporting systems on the implementation team. There may also be significant data preparation expertise required when the data quality is low, or if data sources are not linked and interoperable.

Additional considerations are required for data visualization, particularly when including maps of geospatial information. Two data characteristics have been important to the success of VNM: (1) whether the source data include coordinates and/or links to geospatial information, and (2) whether the administrative boundaries are up to date. These factors influence how effectively the mapping layers of computation, interactvity, and display will be set up and managed. Potential users of the VNM platform should consider whether their existing data provide datasets with these characteristics.

## User-centered learning culture

VNM put the users—health workers in Zambia's malaria control and elimination program—at the center of the system development process. The VNM team created a learning framework based on four principles (expose, explore, explain, and empower[7]) to create a data culture of learning and informed action. Additionally, VNM intentionally chose tools that were designed with end users in mind; that is, tools that could be adapted to the existing skill levels, fast implementation, and high usability. The data products (dashboards) themselves were also designed and created in partnership with end users. As described in the asset classification below, the user-centered design approach coupled with a willingness to pursue non-integrated data sources and move beyond existing operational systems were both important factors in the success of the VNM implementation in Zambia.

---

7  https://www.path.org/articles/inspiring-data-insight-four-guiding-principles/

## Cost and capability

The VNM platform includes certain data science tools that are proprietary, or not open source, which may or may not fit the needs, policies, and resources of the user's health system. Modular systems like the VNM platform are very flexible and adaptable. However, they can be expensive, especially in terms of upfront costs like purchasing software licenses and training needs. They also require time for training and hiring. However, as a result of the VNM partnership, modular packages, reduced pricing, and licensing grants are now available, which can make the costs more acceptable. Cost must also be balanced with capability. Analytical questions often outpace the capabilities of operational systems used for routine data collection, and having the tools and human resources is necessary to achieve greater impact.

When it comes to data visualization, it may not make sense for a national program to use a software platform like VNM if the goal is to serve a simpler purpose, such as to generate traditional reports. VNM's use of Tableau would be an expensive option if the software is not going to be used to its full capability.

## Partnerships and support

The VNM project was supported by many partners with technical expertise who were able to contribute in-kind donations of staff time to configure the required platforms. The use of volunteers to support the Zambia VNM platform was an integral part of the user-centered design process. Particularly with modular systems that involve several different technology owners, the relationships between key stakeholders are also critical to the success of such large-scale platforms. For VNM, trust between the private sector, NGO partners, and the Government of Zambia was an important facilitator of the project's success.

# 🖧 Data science assets

<table>
<tr><td>**PRODUCT** 🖳</td><td>**PROCESS** 🖧</td><td colspan="2">## Alteryx automated data workflows and data management</td></tr>
</table>

**Data Collection**

**Data Transformation**

**Data for Impact**

Alteryx provides tools that automate workflows to quickly and efficiently process thousands of laboratory results, notify health workers, and track interventions.

The software allows users to rapidly curate, process, and archive large amounts of data to enable near real-time understanding of insights and trends. Instead of waiting months to translate data into action, users can process results daily with automated outputs sent to the end user to inform timely decision-making. Alteryx's data-blending tools lend additional capability to users to cleanse and prepare data and perform predictive modeling.

In Zambia, the Alteryx automated workflow integrates geospatial data and case surveillance data in a data pipeline that informs the data visualization dashboards used by district-level health managers. The data are sourced from the DHIS2 server and local data sets and are integrated using a cloud-based interface. The automated workflows enable faster processing of lab results. This means that positive malaria cases are communicated to health workers on a daily basis instead of monthly; more rapid evaluation, validation, analysis, and action based on survey data as it is collected; and weekly summaries of predicted malaria cases.

**Automated workflow**

**Data blending tools**

**Cloud-based data services**

### *For more information*

Alteryx is a software company headquartered in Irvine, California, with a development center in Broomfield, Colorado. The company's products are used for data blending and advanced data analytics. Alteryx has a stated goal of enabling advanced analytics to be performed by non-specialists. https://www.alteryx.com/

PATH + Alteryx Join Forces to Fight Malaria in Africa. Alteryx Community articles. June 19, 2019. https://community.alteryx.com/t5/Alteryx-Use-Cases/PATH-Alteryx-Join-Forces-to-Fight-Malaria-in-Africa/ta-p/182839

**PRODUCT**  **PEOPLE**

# Mapbox satellite imagery and location data

**Data Collection**

**Data Transformation**

Mapbox provides mapping resources, including satellite imagery and location data. It recruits volunteers around the world to crowdsource data about human settlements and structures that can be overlaid with existing data on infrastructure, waterways, and weather.

Location data

Satellite imagery

Crowdsourcing input approach

In Zambia, the location data were combined with surveillance data to create algorithms that predict where malaria cases emerge. These predictive models enable community health workers to track and treat incidents faster, meaning new cases can be prevented and lives can be saved. Maps and satellite imagery allowed the project team to create more accurate geospatial-based dashboards to visualize case distribution and clustering, and to identify issues with underlying data on the locations of facilities and CHWs.

### *For more information*

Mapbox is the location data platform for mobile and web applications. Mapbox's apps reach more than 300 million people each month. https://www.mapbox.com/

"Supporting malaria elimination for 60 million people" [blog post]. Mapbox Blog. April 25, 2019. https://blog. mapbox.com/supporting-malaria-elimination-for-60-million-people-96053bbd8a0a

---

**PRODUCT**  **PROCESS**

# Supplemental data from nonroutine systems

**Data Collection**

To power the VNM analytics platform, surveillance data are harvested from DHIS2. However, additional datasets are also necessary, such as administrative boundaries, health facility data, additional population data, satellite imagery, and crosswalks and manually generated extension datasets that provide merged and cleansed data for use within automated processes.

Data

Obtaining these data can be a complex and time-consuming effort due to local policies; where there have not been resources put toward data governance, additional effort is required to cleanse and link data so it can be useful.

### *For more information*

Supporting malaria elimination for 60 million people [blog post]. Mapbox Blog. April 25, 2019. https://blog.mapbox.com/supporting-malaria-elimination-for-60-million-people-96053bbd8a0a
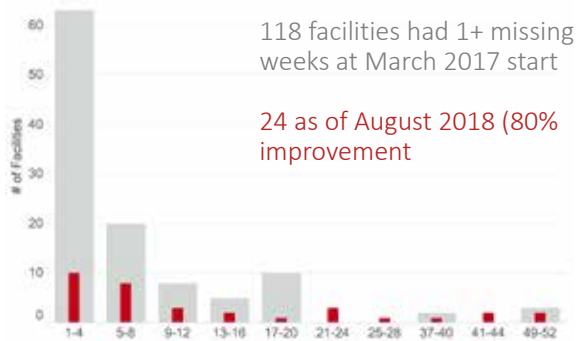
# Twilio cloud communication platform

**Data Collection**

**Twilio provides a cloud communication platform that allows users to exchange valuable insights and enables automated collection of missing and incomplete data from CHWs via text message.**

In Zambia, use of Twilio to communicate with health workers via text resulted in a rapid increase in reporting completeness and timeliness, including back-entry of missing data up to two years prior. Through the VNM partnership, Twilio created a custom cloud-based feedback loop that uses SMS to share data back with frontline health workers. This application has the capability to evolve to include voice and video in addition to text messaging. The bar chart below shows the decrease in the number of facilities with missing reports over the timeline of the implementation of the SMS feedback loop in Southern Province, Zambia.



118 facilities had 1+ missing weeks at March 2017 start

24 as of August 2018 (80% improvement

Viz Alerts, theTwilio communications platform, is an open source Python script designed to identify data gaps, alert the local health workers of those gaps, and integrate missing data when provided by health workers. When the Viz Alerts algorithm detects an incomplete dataset, the script prompts an SMS message via Twilio to the appropriate health worker to submit the missing report; the health worker can quickly reply by submitting the missing data over SMS. Viz Alerts incorporates the missing data automatically, and users can access the updated dataset through Tableau visualizations.

From Marianne Soko, the malaria focal point person in Itezhi-Tezhi district, Zambia: "This is really great because the alert allows me to follow up on the missing reports, and they can be updated both at facility and community level."

**Automated SMS communication**

**Cloud-based data services**

**Health worker communication tools**

### *For more information*

Twilio powers the future of business communications. Enabling phones, VoIP, and messaging to be embedded into web, desktop, and mobile software. https://www.twilio.com/

Talbot J. PATH uses SMS data alerts to help end malaria in Zambia [blog post]. Twilio Blog. June 14, 2017. https://www.twilio.com/blog/2017/06/path-uses-sms-data-alerts-to-help-end-malaria-in-zambia.html.

| PROCESS | POLICY |
|---------|--------|

# Predictive modeling algorithms

| Data Transformation |
| Data for Impact |

### Predictive modeling uses data and stastical methods to predict or forecast outcomes.

Alteryx software is capable of leveraging many different kinds of data to inform these predictions. In the case of malaria, these data could include population mobility, case surveillance, geospatial data, meterological or climate data, and others to create algorithms that inform predictions of where malaria cases may emerge.

Thus far, predictive modeling has not been fully used in support of the VNM platform in Zambia and its application is at the proof-of-concept stage. However, even as Zambia has seen significant progress in its fight against malaria, it will still be challenging to address some of the persistent drivers of the epidemic. Continued innovation and new tools to are essential to reach elimination. For example, malaria parasites that reside in asymptomatic individuals represent one of these epidemic drivers. In the future, predictive models may enable CHWs to determine who might be carrying the malaria parasite and to track and treat cases faster, meaning new cases can be prevented and lives can be saved.

**Custom algorithms**

**Predictive models**

#### *For more information*

Walker A. PATH and Tableau Foundation #VisualizeNoMalaria Project: Development of an Automated Predictive and Forecasting Malaria Cases Capability. Seattle: Tableau; [YEAR]. https://www.tableau.com/sites/default/files/whitepapers/path_and_tableau.pdf.

| PRODUCT | PEOPLE |
|---------|--------|

# User-centered design approach

| Data Collection |
| Data Transformation |
| Data for Impact |

### User-centered design is an iterative process that focuses on the end users and their needs to inform the design process of a tool or solution.

For the VNM platform in Zambia, this methodology focused on understanding who the tools are meant to serve, for what functions, and in service of what goals. This approach also supports the creation or strengthening of a data use culture.

To develop the integrated platform used for VNM, particularly the dashboards used for data visualization and their backend analytics, PATH used a design approach that centered on the users in Zambia at the national, district, and facility levels. PATH and other VNM partners held workshops with local health officials to determine system requirements and data use goals. The resulting platform was iteratively designed to meet the local needs of Zambia's health system.

**Data visualization**

**Collaborative design**

**Data use tools**

#### *For more information*

PATH is an international health organization driving transformative innovation to save lives. Working in partnership with national governments, PATH is leading the way toward a malaria-free world by focusing on new vaccines, treatments, diagnostics, and approaches. https://www.path.org/visualize-no-malaria/

| PRODUCT | PROCESS | PEOPLE | **Tableau data visualization** | Data Transformation |
|---|---|---|---|---|

Data visualization

Data insights

User-centered/
People-centered design

**Tableau provides a powerful data visualization engine that allows end users to create user-friendly dashboards.**

In Zambia, Tableau dashboards were used to analyze the malaria situation in real time. Expert users from DataBlick and Slalom traveled to Zambia and worked closely with district health workers and the Ministry of Health to iteratively develop the dashboards. Other assets from the VNM partnership were able to connect into the data pipeline that feeds the visualizations created using Tableau software.

Before the VNM partnership begn, district health managers generated simple spreadsheets of data in a time-consuming process that could only occur quarterly or annually. Now, with the capabilities of the dashboards, health workers are able to have daily and weekly information at their fingertips to enable them to make decisions on health workforce and supply allocation, track down missing information, and control local outbreaks. Visual data preparation provided a faster method for system design than writing code and also served as a way to work around the lack of data warehouses. In Senegal, local staff were able to quickly learn from the Zambia experience and develop dashboards tailored to local needs.

The designs of a number of the dashboards built as part of VNM are sufficiently useful that they have been re-used for the US President's Malaria Initiative's Malaria Data Integration and Visualization for Eradication (M-DIVE) project that is providing malaria data to countries supported by the US President's Malaria Initiative.

*For more information*

Tableau Software is a software company headquartered in Seattle, Washington, which produces interactive data visualization and analytics products. Tableau's mission is to help people see and understand their data. https://www.tableau.com/

Myrick, Neal. "Visualize No Malaria crosses new frontiers in fight against disease." [blog post]. Tableau Blog. April 25,2019. https://www.tableau.com/about/blog/2019/4/visualize-no-malaria-crosses-new-frontiers-fight-against-disease.

| PEOPLE | PROCESS | **DataBlick and Slalom analytics training** | Data Transformation |
|---|---|---|---|

Analytics training materials

Data use culture approach

Volunteer engagement

**Analytics training for end users can build their capacity to understand and interpret their data, derive meaning from data visualizations, and understand how to use data insights to drive decisions and actions**

In Zambia, volunteer expert users from DataBlick and Slalom helped to build analytics proficiency through training programs for users in Zambia. DataBlick also built a capability to issue reports to district-level managers who need to know when health facilities are not reporting data into the surveillance system. The goal of this effort is to enable the 100+ district-level health information officers and malaria focal point staff to manage and use data effectively in the fight against malaria. Training materials were developed with two aims: to build local capacity to use the data visualization dashboards built through the VNM partnership and to improve overall data literacy.

*For more information*

DataBlick is a boutique consulting firm that transforms data into actionable art. https://www.datablick.com/

Slalom is a modern consulting firm focused on strategy, technology, and business transformation. https://www.slalom.com/home

**PRODUCT**

# Exasol analytics database

Data Collection

Exasol is a high-performance in-memory analytics database that can work with data on premises, in the cloud, or both.

The architecture enables users to perform advanced analytical tasks on large volumes of data within the database itself. It can analyze large volumes of data in real-time to accelerate data insights and reporting.

In Zambia, Exasol is used alongside Alteryx to analyze data from the DHIS2 combined with satellite imagery data. Exasol's in-memory analytics database compresses massive amounts of data and serves them up fast, which is essential in low-resource, lower-bandwidth settings like Zambia. An Exasol database was developed to store the input variables and the output data. It connects both to Alteryx as an in-memory data source for the automated workflow and as a live data source to Tableau for the data visualization dashboards. .

*For more information*

Exasol is an analytic database management software company. Its product EXASolution is an in-memory, column-oriented, relational database management system. https://www.exasol.com/en/products/database-architecture/

**Database management system**

**In-memory analytics**

---

**PRODUCT**

# DHIS2 health management information system

Data Collection

District Health Information Software 2 (DHIS2) is an open source, web-based health management information system platform.

The core DHIS2 software development is managed by the Health Information Systems Programme (HISP) at the University of Oslo. HISP is a global network composed of 11 in-country and regional organizations, providing day-in, day-out direct support to ministries of health and local implementers of DHIS2.

Zambia uses DHIS2 for data collection from health facilities countrywide and as a mobile reporting tool for the country's Malaria Control Program. CHWs and health facility staff report data from the field to the central DHIS2 system. In Zambia, DHIS2 has been integrated directly with Tableau, which allows for embedding Tableau visualizations within DHIS2 and directly pulling data from DHIS2 into Tableau.

*For more information*

https://www.dhis2.org/

**Database**

**Data management**