California Freshwater Harmful Algal Bloom Monitoring Systems Data Ingestion Framework Recommendations

Phase 1: Model Frameworks Leveraging Tribal Government and Community Science Data to Build a Freshwater Harmful Algal Bloom Monitoring Network for California

Prepared and Submitted by Internet of Water
In Partnership with California State Water Quality Control Boards & The Commons

Internet of Water

Contents

EXECUTIVE SUMMARY

Freshwater Harmful Algal Blooms (FHABs) present a growing threat to the state of California, endangering the welfare of wildlife and domestic animals in addition to impairing the recreational and cultural uses of certain bodies of water. In response, a robust system to monitor and report on FHABs is an important step in mitigating the impacts of FHAB events, and to alerting the public to such risks. To increase the capacity of the state to respond to FHABs, monitoring data from Tribal Governments and other non-governmental organizations (NGOs) are needed to integrate with current California monitoring data, creating a robust system for monitoring, reporting, and alert. This report outlines the initial phase of a multi-phase effort to integrate state FHAB monitoring data with non-state FHAB monitoring data to create such a statewide system.

The implementation plan described below is the result of engagements with Tribal Governments and NGOs as well as with California state agencies and other external partners to develop a detailed strategy to incorporate non-State data into the California state data model and alert system. The plan and recommendations described via this document address a need to address data ingestion processes common beyond this specific use case. As such, the information provided here may offer a pathway to beginning this process for other entities looking to modernize their database infrastructure and build programs that accept data from a network of monitoring programs.

Information gathered from engagements and survey results revealed a 1) willingness and ability to support state efforts toward FHAB monitoring; 2) the capacity of non-state entities to enhance state capacity. Based on the information collected via engagements, surveys, and feedback, the report provides 3) a detailed data model outlining the ingestion and data management process; and 4) an explicit set of recommendations for the resulting system to communicate risk to decision-makers and the general public. These recommendations are as follows:

1. Implement a cost-effective, tiered process for gathering data that considers the capacity of the partnering Tribal Governments and community science organizations.
2. Improve how data are accessed and used by external partners and the public through the development of data products (e.g. interactive dashboards) that better communicate the status of reported FHABs in California and the levels of risk associated with associated FHAB data
3. Create an open FHAB data framework that offers the ability to export data in a machine-readable format.
4. Develop alert systems that notify participating members of the public on FHAB events.

This report further details data collection and management requirements, the development of an Application Programming Interface (API), and the use of Artificial Intelligence (AI) and modeling to improve information transfer and delivery. The implementation of such recommendations are based upon funding support for Phase 2 of this project.

PROJECT OVERVIEW

INTRODUCTION

California's water management challenges are complex. One critical challenge is the ability of state, local, and Tribal leaders, and stakeholders to marshal data and information necessary to support sustainable water management decisions, including the unavoidable intricacy of the ecological issues related to water management. Specifically, communities in California face an increasing challenge from freshwater harmful algal blooms (FHABs) in California's lakes and rivers, and the need for timely, actionable information on FHABs is growing.

FHABs occur when algae — simple organisms that live in estuarine and freshwater environments — grow quickly and create toxic or harmful conditions for people and animals. In freshwater, algae are the primary bloom-forming organisms. Blooms cause health problems in humans, pets, livestock and wildlife across the state. Recent data indicate that dog deaths and fish kills are becoming more common when these water bodies become dangerously imparied by FHABs, and their recreational and cultural uses

Figure 1: Clear Lake, CA - Photo by Keith Bouma-Gregson

are becoming more limited. Algal blooms can become toxic in a matter of days or hours, so rapid response is critical.

The Internet of Water (IoW) is partnering with the California State Water Resources Control Board (State Water Boards), Tribal Government representatives, community science groups, and The Commons to develop recommendations for a pilot project designed to improve California's FHAB case management system by 1) defining a data model and developing new tools to receive and handle data submitted by Tribal Governments and community science groups; and 2) engaging with Tribal Governments and community science groups using the improved case management system to provide an early warning system for HABs.

BACKGROUND AND RATIONAL

The California State and Regional Water Resources Control Boards (State Water Boards) work with state and local entities to identify and respond to FHAB incidents throughout California. The State Water Boards first began to formally address this issue in 2005 when it formed the Blue Green Algae Work Group, later renamed the California Cyanobacteria Harmful Algal Bloom Network (CCHAB). An initial product of this group was the Voluntary Guidance Document (original release 2010, updated 2016). Subsequently, the State's Surface Water Ambient Monitoring Program (SWAMP) prepared 2016 California HABs Assessment and Support Strategy[1] to articulate a coordinated program to assess, communicate and manage FHABs in California.

While these efforts represent important progress, and many FHAB incidents are currently reported voluntarily by the public through California's FHAB Program, California currently does not have adequate funding to establish a statewide, routine FHAB monitoring program. Therefore, comprehensive, statewide FHAB monitoring data remains limited.

Fortunately, several third-party groups across California are currently monitoring FHABs, such as Tribal Governments, East Bay Regional Parks, and the Alpine Watershed Group, to name a few. California state agencies are interested in leveraging this third-party science capacity to more comprehensively respond to FHAB events, because incorporating non-state data with state monitoring programs enhances the capacity of state agency monitoring programs and increases data resolution across space and time. Moreover, partnerships with Tribal Governments and non-governmental community science programs (citizen science or volunteer monitoring) highlight the important work of these organizations to address data gaps and speak specifically to their constituents

1 Anderson-Abbs, Beverley A, Meredith Howard, Karen M Taberski, and Karen R Worcester. 2016. "California Freshwater Harmful Algal Blooms Assessment and Support Strategy." Surface Water Ambient Monitoring Program Technical report 925. Sacramento, CA: California State Water Resources Control Board.

or communities to understand and improve their health, economic equity, and social justice. Community Science in particular combines a participatory approach, a flexible and responsive attitude, and a scientifically rigorous process to develop shared ownership of data collected. FHABs represent an area where community science and strengthened engagement with Tribal Governments could play a role in assisting the State's response strategy.

In an effort to support and broaden the CCHAB network, the FHAB program is authoring a Freshwater Ambient Monitoring Strategy for FHABs. This strategy, building on the 2016 strategy document, will lay out a framework for how to build a monitoring program in California and explicitly explores the opportunity to systematically leverage community science capacity more fully to address FHABs.

Undertaken at the same time as the IoW pilot work, the FHAB monitoring strategy deliverable and the recommendations of the IoW pilot will provide critical validation and reference for a fully integrated and functional FHAB monitoring network. Through increased engagement with these stakeholder groups and the integration of third-party data from Tribal Governments as well as community science programs, the monitoring and response network for FHABs can improve the quality of life throughout California.

STAKEHOLDERS

The IoW pilot consists of two participation groups: a Data Model Steering Committee and a Stakeholder Review Committee. Both groups are integral to the success of Phase 1 and will also be important for a successful deployment of the recommendations in Phase 2.

The Data Model Steering Committee consists of representatives from three entities: Internet of Water, State Water Boards, and The Commons. This committee oversaw the execution and implementation of the project including coordination and participation of the second broader stakeholder group. The committee performed all research, analysis, and recommendations, and presented the final recommendations via this document.

The Stakeholder Review Committee consists of members of the California Cyanobacteria and Harmful Algal Bloom (CCHAB) network that sits within the California Water Quality Monitoring Council. The California Water Data Consortium Steering Committee, convened under the auspices of AB 1755—The Open and Transparent Water Data Act provides general guidance and direction to IoW activities in California.[2] These data producers provided guidance, feedback, and organizational insight throughout the information gathering and recommendation development phases.

2 In advance of the formation of the Consortium's Steering Committee, an interim group of California state leaders and Consortium representatives, referred to as the Governing Team, will perform this function.

More details regarding participants and their organizations can be found in Appendix 1.

PROJECT ACTIVITIES

To inform the final recommendations and build stakeholder support for this pilot implementation plan, the Internet of Water led a series of participatory webinars with partners and stakeholders. There were 20 participants in each engagement.

Webinar 1: The objective was to 1) identify the user audience(s) for a FHABs notification dashboard, 2) develop the content for a FHABs notification dashboard, and 3) gather information from stakeholders regarding data practices and variables collected.

Survey: Following this initial engagement, a survey was distributed to gather detailed information regarding data collection practices and variables from each stakeholder organization.

Webinar 2: The second webinar focused on 1) a summary of current California state processes for data ingestion, 2) a discussion about the proposed ingestion process for community science data, and 3) a presentation on current quality assurance/quality control (QAQC) best practices.

Feedback and Review: A draft of the implementation plan was circulated for review and the Steering Committee collected assessments from the stakeholder groups. This feedback was incorporated into the plan, then distributed for a second round of review and feedback.

Webinar 3: After a first round of review and feedback, partners and stakeholders convened to discuss overall impressions of the implementation plan and allow stakeholders to provide feedback and ask questions.

PILOT GOAL

The pilot has been segmented into two phases. In Phase 1, the Data Model Steering Committee produced this CA-FHAB Data Ingestion Framework Recommendation Document. This framework will be the roadmap for the Phase 2 pilot implementation of this project. The implementation of the recommendations presented in this plan as Phase 2 will require additional support.

STATE OF CALIFORNIA WORK TO DATE

DATA USES AND PROGRAM INTENT

The State Water Board's Office of Information Management and Analysis (OIMA) currently manages the FHAB Program. This program focuses on recreational exposures and impacts from FHABs. The program began tracking bloom reports in 2016. Bloom reports can be submitted by the public when they suspect a FHAB is occurring. In addition to bloom reports, OIMA receives monitoring data from partners who voluntarily share their data with the FHAB Program. For example, the Department of Water Resources shares monitoring data they collect at recreational beaches in the State Water Project. These data go into the database and currently have two primary uses: 1) to communicate to the public the status of FHABs across the state and 2) to assist staff in coordinating the response to different blooms.

The FHAB database is linked to an online public interactive map, which displays the location of known blooms, as well as monitoring sites from partners. The map is updated daily with the latest information received by the FHAB program. The data displayed on the map is also publically available on the California Open Data Portal. The data are also used as a repository for relevant details of the follow-up investigation conducted by State Water Board and Regional Board staff to confirm and assess the risks posed by FHABs.

In 2019 the FHAB Program, with additional support from OIMA staff, began a Data Modernization Project. The new data model being developed will improve the Program's ability to:

1. Manage and coordinate bloom investigations by agency staff,
2. Assess status and trends over time, and
3. Incorporate data from more diverse partners.

EXISTING AND PROPOSED DATA MODELS

The current data workflow (Fig. 1) and corresponding data model are focused on collecting information from the public about potential blooms via the FHAB Program bloom report form. All data from the Bloom Reporting form is stored in a Microsoft Access database hosted on Water Board servers. Information collected through the State Water Board and partner staff conducting incident response is stored across emails, personal files, and reports, which are mostly created and curated manually.

The current data workflow and infrastructure used by the State Water Board to monitor and respond to FHABs is cumbersome, compartmentalized, and does not include all the

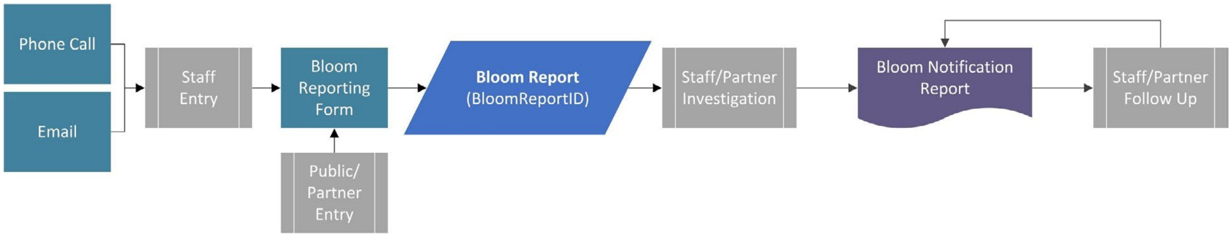


Figure 2: General data workflow currently used by the State Water Board's FHAB Program.

pertinent fields necessary for thorough tracking of and response to FHABs. For example, the current data model is lacking in its ability to collect data to track trends over time, adding new records is constrained to generating new bloom reports, and the database fields are not tailored for data generated by ambient monitoring programs.

To address these shortcomings, the main objectives of the State Water Board's FHAB Data Modernization Project are:

1. Update the data infrastructure to increase automation and efficiency of core data workflows (i.e. FHAB incident response and case management; Fig. 2),
2. Expand and update data infrastructure to improve status and trends assessments,
3. Expand the data infrastructure so it can include Tribal government and community science data, as appropriate, and,
4. Improve how data are accessed and used by external partners and the public through the development of data products (e.g. interactive dashboards) that better communicate the status of reported FHABs in California and the levels of risk associated with their known status.

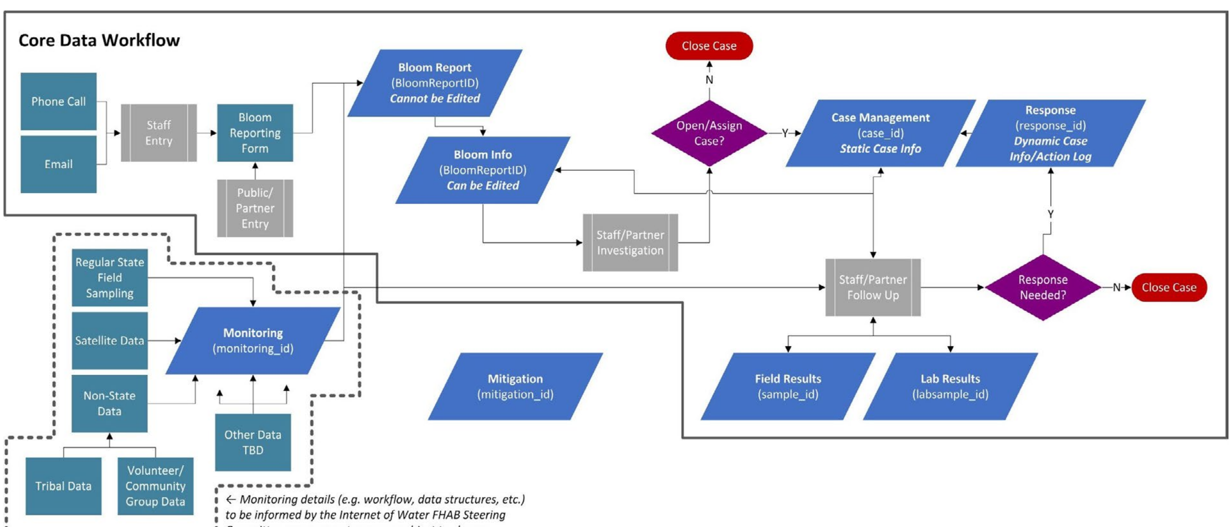


Figure 3: Future data workflow the State Water Board's FHAB Program is in the process of implementation. The core data workflow is surrounded by a solid grey line. The placeholder for data workflows associated with regular monitoring efforts that are outside of the core case management & response efforts, including Tribal Government & community science data, are surrounded by a dashed grey line & are subject to change.

The future data workflow is partially designed to function as a "case management" system to help staff track internal and external communication and record relevant information collected during incident response. While the general workflow (receive data, work with partners, respond) will not change significantly in the future data workflow, the database will be revised to more efficiently receive and better incorporate data produced by agency staff during their bloom incident response into the database (see the Core Data Workflow in Fig. 2). So, instead of the database only containing a single "Bloom Report" table (i.e. blue parallelogram in Fig. 2), the modernized database will contain multiple related tables (i.e. blue parallelograms in Fig. 2 such as Bloom Info, Case Management, Response, etc.). Additional improvements include the ability to store analytical lab results and photographs from field investigations, which cannot be easily stored in the current data model.

DATA MANAGEMENT OF MODEL

Currently, the State Water Board FHAB team has been focused on updating the data model to reflect the improved core data workflows (Fig. 2, solid grey box). Building the Tribal Government and community science monitoring aspects of the data model (Fig. 2, dashed grey box) requires input from Tribal Government and community partners, to ensure their data can be incorporated into the FHAB program data infrastructure. This pilot is powering the input of Tribal Government and community partners and will serve as the foundation upon which the both the Tribal Government and community science aspects of the data model will be built.

Just as with the Data Model, the State Water Board FHAB team has focused on updating the data management process for the core data workflows thus far. For example, data will be accepted through a modernized Bloom Reporting Form (in development) and managed by the FHAB program via their modernized database. As the data workflow and model associated with Tribal Governments and community science data is developed, the Program also needs to determine how they will collectively manage that data so that all parties have access to data and data products they need. It is important to note that any data that is included in the State Water Boards FHAB Data Modernization Project will be made publicly available, as appropriate. Some example questions that should be answered include:

- What will the minimum QAQC requirements be?
- Who will store the original data and where?
- Who determines what data is made public?
- How will the data be submitted to the modernized FHAB database?

DATA DISSEMINATION WORKFLOW

The data dissemination workflow for the core data processes that currently exist will be updated and be made more efficient through the State Water Boards FHAB Data Modernization Project (e.g. the online public interactive map will be re-linked and updated). The State Water Boards FHAB Data Modernization Project would like to put into place workflows that effectively utilize, and disseminate Tribal Governments and community science data in a way that is meaningful for all parties. Doing so will require all parties' commitment to discussion, collaboration, and coordination to determine which dissemination workflows and products work best for everyone.

CURRENT STATUS OF UPDATES

The State Water Board FHAB team is working on finalizing the core data infrastructure and workflows. The data model is currently being built in SQL and Microsoft Access, and the FHAB program plans on launching the revised core workflow in the summer of 2020. Tribal Government and community science workflow and data models are not under current development. Their creation will rely on input from and data gathered via this pilot and the Freshwater Ambient Monitoring Strategy. Both efforts provide integral details about how to best design the workflow and data models to efficiently and effectively receive data from participating Tribal Governments and community entities.

COMMON PARAMETERS AND CONSIDERATIONS OF FHAB MONITORING PROGRAMS

As the data models and ingestion methods for data inputs are considered, the parameters of importance play an integral role in any effective updates to the current process. As mentioned above, a concurrent effort is underway to build monitoring protocols and methodologies for application by entities looking to adopt a new program or revise an existing program.

The types of measurements made and parameters collected depend on the goals of the individual FHABs monitoring program. Considering FHABs are inherently a biological phenomenon, the parameter options for monitoring FHABs are many[3]. For

3 Graham, Jennifer L., Keith A. Loftin, Andrew C. Ziegler, and Michael T. Meyer. 2008. "Guidelines for Design and Sampling for Cyanobacterial Toxin and Taste-and-Odor Studies in Lakes and Reservoirs." U.S. Geological Survey Scientific Investigations Report 2008–5038. U.S. Geological Survey. https://doi.org/10.3133/sir20085038.

public and environmental health risk assessment, measurements usually focus on the density of cyanobacteria or concentrations of cyanotoxins in the water or mat samples. Cyanobacterial density or relative abundance are often measured with microscopic cell counts, chlorophyll-a concentrations, or remote sensing algorithms to estimate abundance from satellite data. Additionally, secchi depth is often included in many lake monitoring programs and measures water clarity, but does not discriminate cyanobacteria from other factors that also affect clarity. Not all blooms are toxic, however, and so these parameters may not always correlate with cyanotoxin concentrations. Counting the number of toxin producing genes in a sample using qPCR provides information about the potential for toxin production is often used as a screening tool prior to cyanotoxins analyses.

Measuring cyanotoxin concentrations and/or counting toxin producing genes involves collecting a sample for laboratory analysis or using a field-test kit. Analytical results vary in terms of units and data types for the various toxin types, sample media, and methods.

Lastly, additional water quality parameters can be measured that do not directly address public health risks, but relate to water body health and function. These may include nutrients (nitrogen and phosphorus), temperature, dissolved oxygen, conductivity, etc. Parameters such as these may help identify drivers of FHABs as well as other non-toxic impacts of a FHAB on a water body. Some of these parameters can be collected as discrete samples or with deployable sondes, which generate time-series data. Time-series data require specific data workflow considerations to be ingested by a database. An effective database structure will need to be able to accept all these different types of FHABs related parameters.

Clear Lake, CA - Photo by Keith Bouma-Gregson

SURVEY SUMMARY

INTENT/GOAL OF SURVEY

Following the initial engagement with CA-FHAB partners and stakeholders, IoW distributed a survey to gather more detailed information about stakeholder organizations, types of data collected, methods of collection, frequency of collection, format of data collected, data management practices, data processing, presentation of data to the public or identified audiences, and types of information and format of information desired for a FHAB Notification Dashboard. The intent of this survey was to quantify and categorically assess the current state of monitoring programs in terms of data management and desired assets for the selected participants.

This survey covers key components of the data management cycle familiar to all monitoring programs. In a nutshell, when looking at these results we can ask the following: why do parameters matter? Why does an entity's data storage strategy matter? And finally, why do we care where else the data is sent? In short, the answer to all of the questions is that data management is an indicator of data quality and data quality is integral to determining how data can be applied and used. This survey allowed our team to get a handle of the current climate for data management across all entities - the current strengths and weaknesses as well as individual entity's roadmaps for future use of their data.

This information will be used to better understand challenges and capacities of stakeholder organizations, to identify areas of potential support, and to form the basis of recommendations in this plan.

RESPONDENTS AND METHODOLOGY FOR COLLECTION

Eleven of the fourteen recipients completed the survey (n=11). Participants of the survey represented state agencies, Tribal Governments, and non-governmental organizations (NGOs). Of these organizations, 100% (n=9/9) use employees to collect data while 55% (n=5/9) use volunteers, 33% use interns (n=3/9), and 22% use community members not affiliated with the organization (n=2/9).

The geographies represented among these organizations was diverse, ranging from organizations that accept information about any water body in the state of California to specified water bodies such as:

- Barner Slough
- Castaic Lake
- Clifton Court Forebay
- Klamath River

- Lake Oroville
- Lake Perris
- Lake Tahoe
- O'Neill Forebay
- Pyramid Lake
- Sacramento-San Joaquin Delta
- Salmon River
- San Luis Reservoir
- Scott River
- Shasta River
- Silverwood Lake
- Tahoe Basin
- Thermalito Afterbay
- Thermalito Forebay

SUMMARY RESULTS

By engaging state, Tribal Governments, and NGO representatives we were able to draw parallels between strategies and isolate areas where minimal overlap exists. Keeping in mind the groundwork that the State Water Board has already laid down in building a data ingestion system, the goal of this information was to inform how to design data ingestion, sharing, and communication methods for the state's FHAB data modernization plan from non-California State entities.

DATA COLLECTORS

The data collection components are of critical interest to building out both Tribal Government and community science component to the FHAB data ingestion structure. First and foremost, who collects the field sample matters. Staff, trained volunteers, and untrained volunteers can all offer valuable monitoring data, but the data that the 100% of entities that rely on staff collections is likely of higher caliber than those collected by untrained volunteers. The high level of trained monitors indicates that the FHAB system can provide the state with a robust opportunity to monitor and respond with confidence to possible blooms. That being said, untrained volunteers can still provide invaluable observations that may trigger more thorough analysis of a possible FHAB occurrence, resulting in a more efficient use of scarce resources.

WHO COLLECTS DATA IN THE ORGANIZATION?	
Type of Data collector	Percent of Respondents (n=9)
Employees	100% (n=9/9)
Volunteers	55% (n=5/9)
Interns	33% (n=3/9)
Community Members	22% (n=2/9)

Table 1:Respondents responses to the survey question: Who collects data in the organization? Note the percent of respondents may not total 100% because respondents could select more than one option in their response.

DATA COLLECTION

The survey asked respondents to identify the frequency of data collection (n=9). In developing an ingestion model and subsequent portfolio of uses for the data, the frequency of collection represents an integral component to understanding the quality of the data and what can be done with the data after it is ingested.

The frequency of data collection varied widely, with most respondents (55%; n=5/9) collecting samples at irregular intervals. However, many other organizations collect samples weekly (44%; n=4/9) or monthly (44%; n=4/9), with a significant number listing "other" as their collection frequency (44%; n=4/9). A small number collected data annually (11%; n=1/9), 2-3 times per month (11%; n=1/9), or daily (11%; n=1/9). Participants defined "other" as, in response to bloom reporting, based on bloom status statewide, or as needed.

Half of respondents selected 'irregular intervals'. Based on the information available, we can only make inferences as to the intermittent frequency of monitoring. The program could have developed a Quality Assurance Project Plan (QAPP)that relies on visual monitoring where data is only collected when monitors identify a possible FHAB sighting and that triggers an actual sampling event. Conversely, the frequency may be the result of lack of a QAPP or standard operating procedures (SOP) for the monitoring program. Finally, sometimes minimal requisite resources (i.e. human, hardware, or financial) lead to an infrequent or sporadic monitoring schedule.

While any data is better than no data, collection that occurs at a known interval will be better suited for tracking baseline data and noticing spikes in key parameters that indicate the presence or imminent arrival of FHABs and better inform all stakeholders of the state of FHABs in their water bodies.

AT WHAT FREQUENCY DOES YOUR ORGANIZATION COLLECT DATA?	
Frequency of Data Collected	Percent of Respondents (n=9)
Irregular Intervals	55% (n=5/9)
Weekly	44% (n=4/9)
Monthly	44% (n=4/9)
Other	44% (n=4/9)

Table 2:Respondents responses to the survey question: At what frequency does your organization collect data?

PARAMETERS COLLECTED

Participants selected from a list of 72 potential types of data (parameters) for collection. Parameter redundancy indicates that an ingestion model can successfully bin data arriving from distinct sources and use that aggregated data for statistical analysis as well as communicating the state of water bodies to general audiences. Of these parameters turbidity, pH, dissolved oxygen saturation, dissolved oxygen concentration, and cyanotoxin concentration were the most commonly collected along with photographs, written text notifications, GPS coordinates, and other location data being the most common media of data. This aligns closely with the parameters of value for FHAB monitoring programs as presented previously in Section 2.

We are confident that the survey results give enough indicators and guidance to the State's FHAB team and Data Model Steering Committee to recommend a universally adoptable ingestion structure for the data model. Furthermore, the received responses on parameters collected indicate that collectors will have the desired data points ready as-is to port into the data ingestor once built.

LABORATORY TESTING LOCATION

For sample testing, 88% (n=8/9) of participants use an external laboratory while 33% (n=3/9) use an internal laboratory. Approximately 44% (n=4/9) use field test kits and others use outside consultants (22%; n=2/9). A small percentage of participants (11%; n=1/9) do not conduct water quality testing, and an equal number report "other" means of testing (11%; n=1/9).

The timeline for receiving results and the quality of data are often inversely related. When testing is conducted in an external laboratory, the results can take upwards of a week to receive but the level of confidence in those results is very high. Alternatively, field test kits allow for a quick turnaround of results but with a lower level of confidence in the result.

WHERE DOES YOUR ORGANIZATION CONDUCT TESTING?	
Places Where Testing is Conducted	Percent of Respondents (n=9)
External laboratory	88% (n=8/9)
Internal laboratory	33% (n=3/9)
Field test kits	44% (n=4/9)

Table 3:Respondents indicated that they conducted testing of their water samples at either an external labor, internal lab, or via field test kits.

DATA FORMAT

The format of the data widely varied from unstructured text or media to more structured data forms. Data that is already or can easily be restructured into machine-readable formats provide more functional value not only for the data collectors but also for data ingestors. That being said, unstructured text and media such as text descriptions and photos provide invaluable supporting information that contextualizes the quantitative data. As an interesting aside, advancements in text recognition and media conversion software applications may open new opportunities to convert unstructured data to a machine-readable format in the medium- to long-term.

WHAT FORMAT IS YOUR ORGANIZATION'S DATA IN?	
Format of Data Collected	Percent of Respondents (n=9)
Unstructured text (Word doc, PDF)	55% (n=5/9)
Unstructured media (images, video)	77% (n=7/9)
Structured tabular (Excel, Access)	77% (n=7/9)
Flat file (CSV) or Tab delimited file	66% (n=6/9)
Other	22% (n=2/9)

Table 4:Format of data collected categorized by type.

DATA MANAGEMENT

Data management focuses on data storage practices after the sample results are in hand. A majority of the data are managed on a local hard drive (66%; n=6/9)) while only 22% (n=2/9) are managed in a cloud storage provider. Although 33% (n=3/9) of participants indicate their data are managed in a custom database, no participants use an enterprise software package to manage data. Approximately 55% (n=5/9) of participants listed "other" as their data management process. Note: respondents were able to choose more than one option.

Respondents recognized clear deficiencies in data storage strategies. More than half of respondents reported that their data was stored in a collection of files on a local hard drive. While this may have been the status quo for the past few decades, advancements in software storage solutions have pushed this precarious system into obsolescence. Widespread availability of cloud-based, multi-user data management options can secure data storage systems in stable structures. Single-user access to unstructured data increases the chance of losing data and also reduces the ability to nimbly port data into external sharing opportunities or regional collaborations. Redundancy and back-ups are your data and data manager's best friend.

WHERE DOES YOUR ORGANIZATION MANAGE DATA?	
Places Where Data is Managed	Percent of Respondents (n=9)
Local hard drive	66% (n=6/9)
Cloud storage provider	22% (n=2/9)
Other	55% (n=5/9)

Table 5:Where is data managed?

PUBLIC DATA SHARING

When sharing data, all parties want to know the quality of those data in order to assess the level of confidence in what information they are sharing or consuming. Understanding how data are publicly shared and the desired uses of the collected data has a two-fold purpose for this effort. First, it helps determine the desired base tier of quality required, or filtered, for any data in order to continue to inform these public uses. Second, it assists the State in creating a short-list of communication tools it can build and share to support the desired communication use cases of the data collectors, bringing that data life cycle full circle.

All participants indicated that they make their data available to the public. The most common forms of this process were to push the data to a public repository (55%; n=5/9)) or through interactive web maps or applications (55%; n=5/9). A majority of participants (66%; n=6/9) indicated they send data directly to parties who communicate with their organization, while 33% (n=3/9) indicate they provide links to folders or collections of files for the public to access.

HOW DOES YOUR ORGANIZATION USE TO MAKE DATA PUBLIC?	
Processes to Make Data Public	Percent of Respondents (n=9)
Public repository	55% (n=5/9)
Interactive web maps or applications	55% (n=5/9)
Send data directly via request	66% (n=6/9)
Links provided to data files	33% (n=3/9)

Table 6: Process to make data public, as identified by respondents.

The audiences for participant data also varied (Table 7).

WHAT AUDIENCES USE YOUR DATA?	
Audiences for Collected Data	Percent of Respondents (n=9)
General Public	88% (n=8/9)
Recreational Users	88% (n=8/9)
Occupational Users	66% (n=6/9)
Cultural and/or ceremonial users	44% (n=4/9)
Public Health Officials	66% (n=6/9)
Decision-Makers	77% (n=7/9)
Policy Makers	66% (n=6/9)
Scientists	77% (n=7/9)

WHAT AUDIENCES USE YOUR DATA?	
Audiences for Collected Data	Percent of Respondents (n=9)
Other	33% (n=3/9)

Table 7: Breakdown of audiences for participant-collected data. Note that respondents could select all audiences that applied..

PREDICTIVE PUBLIC FHAB DATA USE

Following up on how participants share data publicly, participants were asked to describe how they would use FHAB data if it were made publicly available. These responses focused on two main objectives, 1) on getting information to decision-makers in a timely manner and with greater confidence and 2) educating the public to inform individual decision-making around potential for FHABs at water bodies they wish to visit. The types of decisions or questions that need to be addressed with such data were divided into ten possible categories for participants. In ranking order, participants prioritized:

HOW WOULD YOU USE HAB DATA IF IT WERE AVAILABLE?	
Perfered use of HAB Data	Ranking Order
Human Health	1
Water Contact Recreation	2
Non-Contact Water Recreation	3
Domestic Animals	4
Aquatic Life	5
Other Wildlife	6
Other Beneficial Uses	7
Organoleptic Qualities (taste and odor)	8
Other	9

Table 8: Ranks the preferred use of HAB data if a larger dataset was made available.

PROPOSED DATA COMMUNICATION STRATEGIES

In order to communicate risks associated with FHABs, participants suggested tools such as applications, simple graphics, interactive maps, videos, and support for organizations to better educate and communicate risks. The ideal approach to convey this information varied by audience, according to participants. Below is a heat map summary of the audience by type of communication or data representation, with darker colors indicating a higher level of interest (Fig. 3).

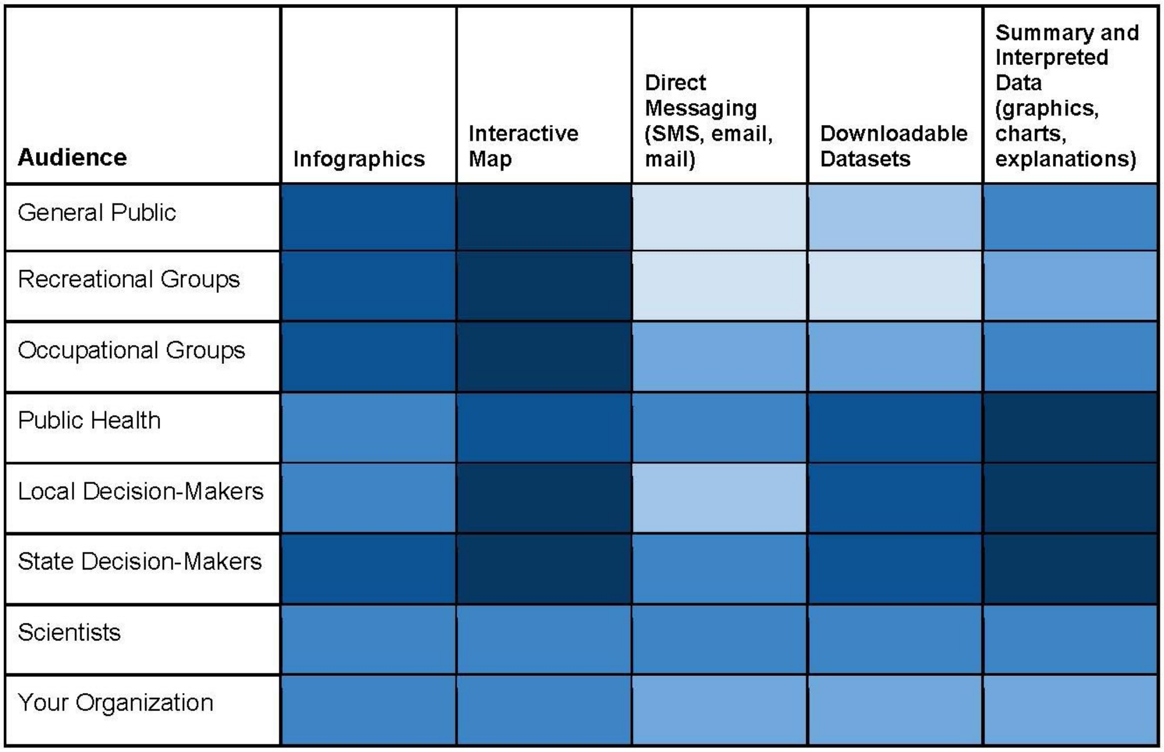


Figure 4: Heatmap showing data communication method and potential audience. Darker colors indicate a higher level of interest by the survey respondents in that communication medium for a specific audience type.

SURVEY RESULTS DISCUSSION

Previous work undertaken with collaborative efforts for data management at The Commons and through the IoW has verified countless times that programs with digitized, machine-readable data are better prepared to participate in and benefit from collaborations and external (i.e. state level) data-driven programs. While this respondent sample size may be small and representatives from each contributing sector minimal, the message is clear and universal: there is a need that any participation in an FHAB response network offer a simple intake process that can utilize machine-readable data that meets basic qualifications without sacrificing the existing applications of the data.

General goals for existing monitoring programs focused mainly on data use, especially via analysis and sharing. Goals included a desire to use monitoring results for edification and strategic planning. Some respondents detailed multiple places that they already share their monitoring data - a clear reminder that any FHAB data ingestion system will be one more destination for this highly in-demand information. Considering that the mandate or impetus for monitoring changes across the respondents depending on authority, funding sources, stakeholder demands, and other motivations, the common thread of pushing data analysis public bodes well for the importance of continuing this current effort beyond planning and moving to implementation.

Before any data can enter a storage structure, however, it must be collected. As the quality of the data will be determined in part by the level of training of the individual, the frequency of monitoring, and the reliability of the collection methods hold weight as well. All data holds value to monitoring FHABs, from geo-located photos to laboratory analyzed bacteria samples, and building an early response system. As such, the State Water Board will weigh the significance of the data ported to their system based on the tiered level of data type - as will be discussed in the recommendations section of this document.

Similarly, when identifying parameters collected, the responses confirm that entities build monitoring programs that capture the parameters of critical need to their waterbody. We can infer from the parameters collected that identifying FHABs may be one component of a broader monitoring program for each of these entities. While common parameters are collected, such as dissolved oxygen concentration and saturation, pH, turbidity, and cyanotoxin concentrations, specific programs value the continued monitoring of additional parameters too. Diversity in programs does not doom this initiative, it actually allows us to look holistically at the content being delivered and identify critical criteria for participation in a broader effort. By acknowledging the value of both the geo-located photos and device collected data points in building a response network, the State is leveling the playing field in contributions to water quality monitoring efforts.

By sharing their existing data management structures in their survey responses, respondents highlighted that our recommendations will be most successful if they acknowledge existing methodologies rather than require adoption of standard QAPPs in order to participate. Most likely, the State Water Board will structure a filter feature in their model to allow for the isolation of certain types of data to inform different agency decisions or actions. Despite differences, the commonalities across existing monitoring efforts are quietly present and offer a pathway for any desired participants to enter their data into a central repository that functions both statewide and locally. In offering up the existence of this central hub there is an underlying goal of the individual entities retaining or even expanding the impact and reach of their existing monitoring efforts. Furthermore, by structuring an ingestion pathway, the State Water Board may also be able to provide previously unattainable portable analysis that can double as communications for use by the individual entities, a need mentioned by multiple respondents in their descriptions of extraction and manipulation efforts.

Any recommendations via this pilot on how to structure data and offer software solutions to support the participation of third-party entities in the FHAB monitoring network should recognize that participation must have an added benefit to the individual entity that chooses to participate. We expect that entities will voluntarily elect to participate in the data sharing program, bringing along their monitoring strategies, contributors, and existing goals for their activities. The State Water Board is well positioned to offer opportunities to address some of the legacy issues that these groups have identified via the survey around data management by offering tools for best practices for data collection and storage as well as for data sharing, analysis, and publication. By offering tools that address the needs and management deficiencies represented throughout the survey, the State is creating more incentive for participation in this collective monitoring effort.

The distributed survey, in short, validates that an effective way for the State of California to expand the FHAB monitoring network as well as institute a reliable and referenced monitoring system is through collaboration with a network of Tribal Governments and community science groups. By surveying respondents on data management in terms of collection, storage, and analysis, we were able to confirm shared goals, uncover commonalities, and expose common deficiencies that, if not properly planned for, may cause issues for ingestion. All recommendations to the items addressed in this discussion appear in the final, Data Management and Model Recommendations section of this document.

Clear Lake, CA - Photo by Keith Bouma-Gregson

DATA MANAGEMENT AND MODEL RECOMMENDATIONS

Based on the survey results and detailed input from stakeholders via meeting minutes, the requirements solidifying a successful CA FHAB Monitoring Data Framework will include the following:

1. Update the data infrastructure to increase automation and efficiency of core data workflows (i.e. FHAB incident response and case management);
2. Expand the data infrastructure so it can easily accommodate Tribal Governments and community science data, as appropriate; and
3. Improve how data are accessed and used by external partners and the public through the development of data products (e.g. interactive dashboards) that better communicate the status of reported FHABs in California and the levels of risk associated with associated FHAB data.

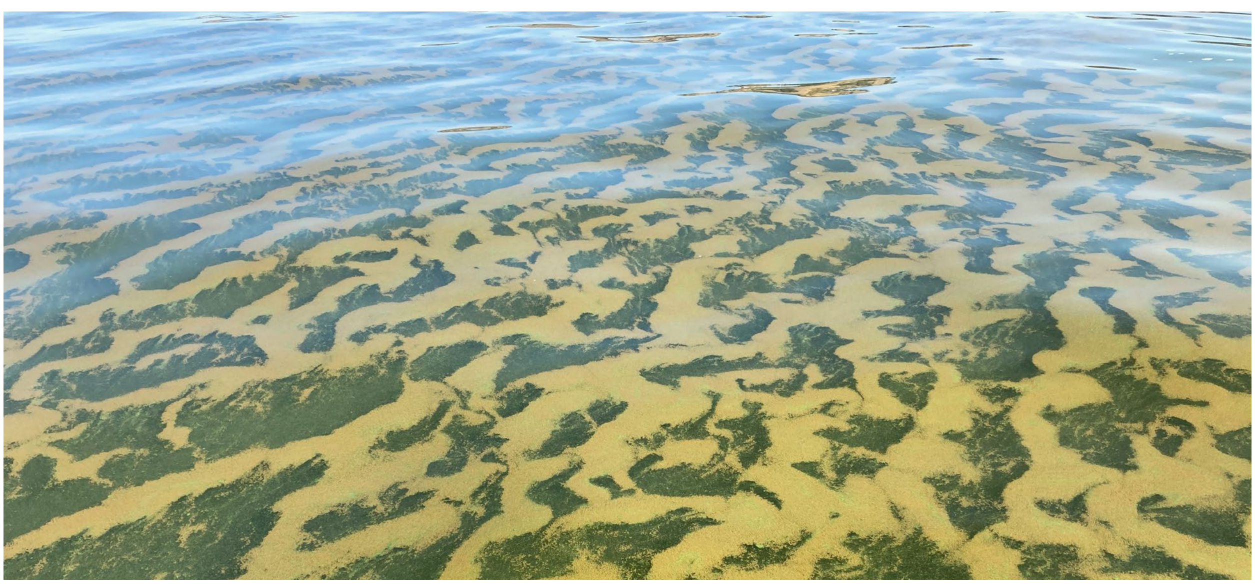
The core objective of our recommendations are to provide an easily adopted framework and approach for integrating Tribal Governments and community science data into the State of California's Core Data Workflow as structured in Fig. 2. Data producers and program administrators of Tribal Governments and community monitoring programs are the front line of assessing the health of California's water bodies. These programs provide a level of detail that represents one of the most effective early screening systems for FHABs because of the expanse of the monitoring network and frequency of sampling. Tribal Governments and community data equate to hundreds of thousands of hours of monitoring time and are an invaluable asset to be leveraged but supported in the same vein. Our recommendations are designed to build upon existing workflows and limit any technical disruption that could cause interruption in data collection activities.

Successful implementation of the new vision of the CA-FHAB Monitoring Data Framework will require an iterative development approach and an application architecture that ensures data and workflows can function according to the following core principals where all data systems are:

Organized around business capabilities and need. Under this principle, data management, visualization, and publishing platforms integrated into the CA-FHAB Data Ingestion Framework must be targeted to serve the business practices of Tribal Government and community science programs as well as the State of California. A monolithic platform such as a data portal or clearing house can be incredibly effective and part of the framework; however, to incentivize contribution and standardization of data, it is imperative that any software deployed support end users in repeatable workflows that will generate structured, machine-readable data that can be leveraged in day to day business functions.

Cost-effective to maintain. Whether a data management, visualization, or publishing platform is a custom build, Configured Off the Shelf (COtS), or Software as a Service (SaaS), the system must be affordable and offer a low cost of overhead to maintain and operate. It is also important that any solution considered balance business processes with constraints in the stakeholder's annual budget and technical expertise.

Can be purposefully integrated. A fundamental component of including voluntary monitoring data into the CA-FHAB Monitoring Framework requires formatting and posting of data from Non-California State programs to the State of California's Modernized FHAB Database. Many Tribal Governments and some volunteer monitoring programs are mandated to upload water quality monitoring results to state and or federal systems such as The US Environmental Protection Agency's (EPA) Water Quality Exchange (WQX). WQX serves as a national standard that facilitates interoperability between highly variable data sets by ensuring any information uploaded conforms to data of known quality. Any technical solution(s) proposed must adhere to a structured data standard and possess a means to export data at minimum. Ideally any third party system would include a public or semi-public application programming interface (API) that would enable a series of GET requests that could facilitate pushing data dynamically to the CA-FHAB Monitoring Framework applications and other third party data sharing platforms. It is important to note that all of this work hinges on meeting organizations and users where they are. Any forward movement in transposing data is acceptable as long as it is providing value in supporting the goals of the State Water Board in ingesting Tribal Governments and community collected data for the early warning, detection, and monitoring of FHABs.



Clear Lake, CA - Photo by Keith Bouma-Gregson

TRIBAL GOVERNMENTS AND COMMUNITY SCIENCE DATA MODEL INGESTION RECOMMENDATIONS

Given the broad opportunities for engagement with Tribal Governments and community science volunteer data collection programs, we recommend implementing three tiered functions for gathering the greatest amount of data of known quality, and requiring the least possible amount of time given the Tribal Governments or volunteer's bandwidth. Developed by The Commons' project team, the Water Monitoring Data Elements and Business Functions diagram (Fig. 4) provides a high level blueprint focused on marrying organizational workflow with early stage feature requirements.

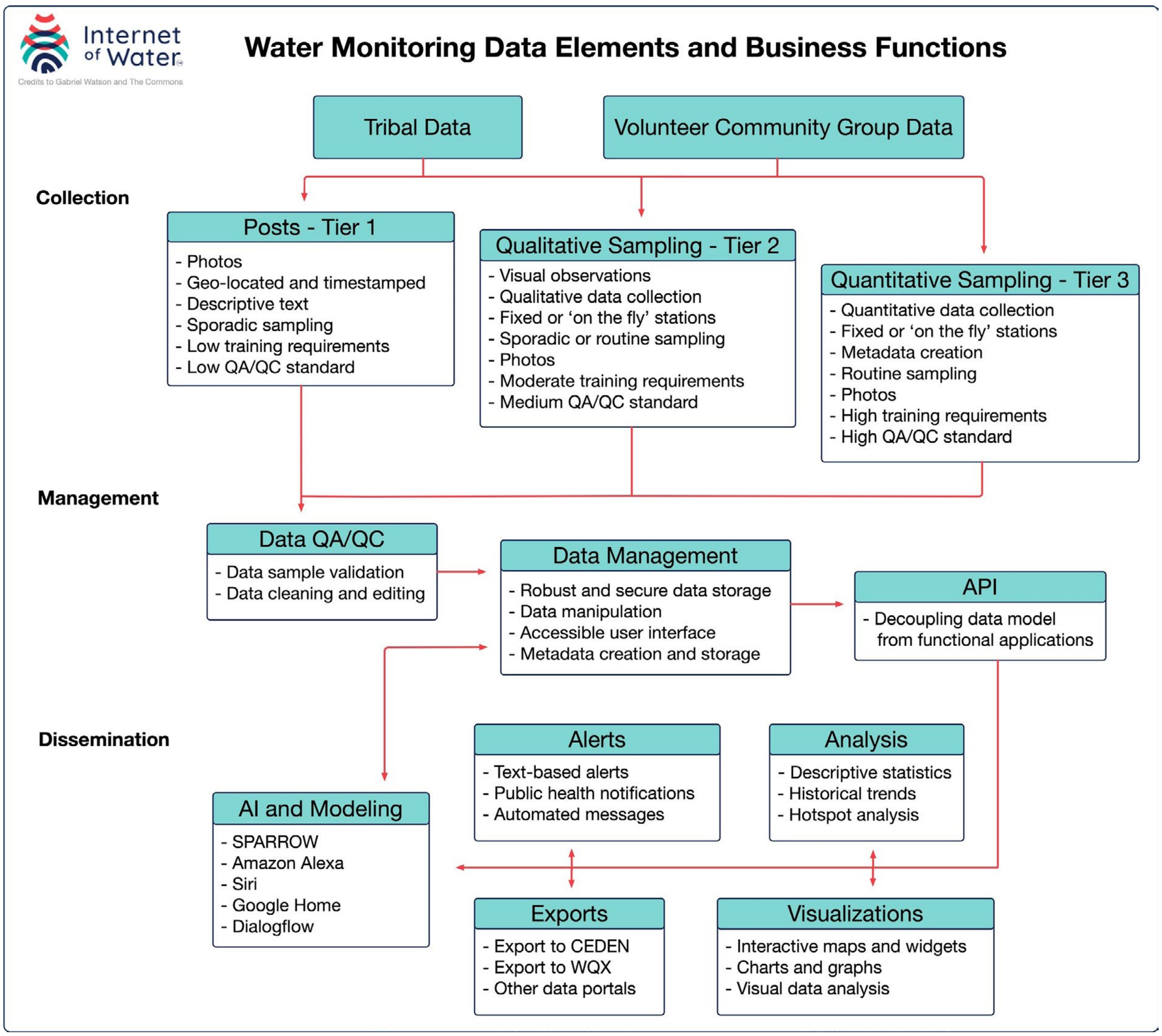


Figure 5: Water Monitoring Data Elements and Business Functions display a three tiered decision-tree for structured collection, management, and dissemination based on posts, qualitative, or quantitative sampling decisions.

COLLECTION

Starting with Collection (Fig. 4), survey results indicated a range of data ingestion methods used by our project stakeholders, with 26% of respondents working to get these results into structured, machine-readable formats that are then maintained by software programs such as Microsoft Excel and Microsoft Access. At the stage of collection, Tribal Governments and volunteer or community science monitoring programs should work to segment their data collection efforts based on the following three tiers:

- Posts - Tier 1
- Qualitative Sampling - Tier 2
- Quantitative Sampling - Tier 3

Ideally Tribal Governments and community science monitoring programs would integrate all three of these data models into their monitoring programs and target each to the appropriate demographic of monitor. Data models and uses are outlined as follows.

POSTS - TIER 1

The following data model can be leveraged when volume of contributions and a clear call to action is required for those volunteer collectors that have a limited amount of time for training but still wish to participate in being the eyes and ears of their local water body. For example, a member of a volunteer organization would have likely attended a training at their local watershed organization event. Program managers adopting this data model and use case would share information on visual observations a volunteer identified on a given water body that would trigger an event to collect these data points. While largely unstructured, the following data model provides a means of balancing ease of collection with helpful information that can be extended to other workflows.

For example, if volunteer monitors are leveraging a smart phone application such as Water Reporter, they would possess the ability to easily rely on the application's geo-location services to gather latitude and longitude. Utilizing their smartphone's camera or photo library, an image can be captured (e.g. BloomWatch app). Comments can be added as a general means of describing the particular observation or issue. Lastly, with location described as point (Latitude and Longitude), most modern databases such as PostgreSQL can be configured to automatically relate the observation's location, via Post's powerful point in poly function to another data set such as a hydrologic unit code twelve (HUC 12) watershed which has national coverage. Finally, with an established location, other ancillary functions can be performed such as triggering an email with turn by turn directions to the new observation.

Under this function users are tasked with submitting observations of the following data fields and types (Table 9):

EXAMPLE USE CASE 1: THE WATERSHED MANAGEMENT GROUP

Consider the Flow365 Monitoring program conceived and managed by The Watershed Management Group (WMG) in Tucson Arizona. This program's purpose is to monitor the status of flow in Tucson area creeks and rivers, check groundwater levels in accessible wells, and observe the plants and animals in local riparian habitats. The WMG Flow365 study trains volunteers to pick a fixed monitoring location that they visit on a weekly or by-weekly basis to collect visual and qualitative data. The data, collected via the Water Reporter mobile app, supports the long-term goal of restoring Tucson's heritage of flowing creeks and rivers. The mix of observational and qualitative data that is collected by trained volunteers informs long-term strategic planning for their River Run Network. This network more broadly offers educational and advocacy opportunities to restore free-flowing rivers in Tucson based on tangible research and monitoring work.

Along with the data fields listed in Table 9, volunteers collect photos and observational estimates of flow states: Flood Flow, High Flow, Moderate Flow, Low Flow, Trickle, Ponding, No flow. Data is visualized on an interactive map. By collecting this qualitative data at fixed sites and intervals the effort is building a collection of flow rates over time to use throughout their ongoing effort.

TIER 1: DATA FIELDS & TYPES		
FieldAlias	DataType	CollectionMethod
record_id	Integer	Machine Generated
latitude	Decimal	Pulled from device
longitude	c	Pulled from device
comments	String	User entered
image	String	Captured via Camera
watershed_huc	Integer	Automated based on location
watershed-name	String	Automated based on location
report-owner-fn	Text	User entered
report-owner-ln	Text	User entered
collection-date	Date	User entered

Table 9: Identifies FieldAlias, DataType, and CollectionMethod columns display possible data fields and the associated qualifying information that the data type and method.

It's important to note that each sample collected is represented with an explicit latitude and longitude point and is flat in its format. This differs from a traditional time series data model where a relationship of one fixed station relates to many samples and readings. See Quantitative Sampling - Tier 3 below.

QUALITATIVE SAMPLING - TIER 2

Under this data model, Tribal Governments and volunteer monitoring programs are able to extend the utility of Tier 1 by adding additional, categorical, data fields and field types to their collection methodology. For example in Tier 1, a single "comment" field is the only area to input qualifying information about the observation being collected. While the simplicity of this approach leads to more data contribution, the lack of structure makes leveraging the information very challenging. Qualitative Sampling - Tier 2 enables users to establish their own schema for data collection by adding custom form fields and data types to facilitate structured data entry by Tribals Government and volunteer monitors. The general data structure in Tier 2 is similar to Tier 1 in that it is flat in nature. Each row represents a new record where additional attributes are appended as columns.

Example data fields and types can be observed in Table 10.

TIER 2: DATA FIELDS & TYPES		
FieldAlias	DataType	CollectionMethod
reading_id	Integer	Machine Generated
latitude	Decimal	Pulled from device
longitude	Decimal	Pulled from device
watershed-huc	Integer	Automated based on location
watershed-name	String	Automated based on location
report-owner-fn	Text	User entered
report-owner-ln	Text	User entered
collection-date	Date	User entered
{Append n number of Parameter Names & Parameter Data Types}	Text, Date, Enumeration, Integer, Decimal, Document, Image	User entered

Table 10: Identifies FieldAlias, DataType, and Collection methods for all data fields associated with Tier 2 sampling data models.

Many modern day SaaS platforms support this functionality and workflow. For example, Google Forms, ArcGIS Online/ArcGIS Collector, and Water Reporter all provide an efficient means for enabling program administrators to intake structured data of all types including

EXAMPLE USE CASE 2: CALUSA RIVERKEEPER

Calusa Riverkeeper, Southwest Florida water quality juggernauts have developed a volunteer monitoring program where participants, known as Water Rangers, monitor for Harmful Algal Blooms (HABs). During self-defined monitoring outings, Rangers track for the presence/absence of new bloom locations. Their observations trigger a scientific analysis of water quality at a perceived impacted location. The collected data from the volunteers is used both as standalone credible data to inform the general public of repercussions of water quality as well as the gateway to building a rigorous scientific data set used to pursue legislative changes that will address the underlying triggers of the blooms.

Rangers are trained to act as community scientists to conduct surveys, report conditions, and collect observations while on local waters. Volunteers set their own schedules and commit to monitoring in project area zones of their choice. By funneling all reports through Water Reporter, the incoming data set is seamlessly consolidated and standardized in machine-readable and structured format. The geo-located samples and instantaneous shift in chain of custody allows the scientists to assess the need to respond with minimal time lag to conduct in-situ sampling to confirm the presence of a bloom. The only equipment that volunteers need in order to participate in this Tier 2 monitoring program is a secchi disk which monitors may elect to dip to share a water transparency reading.

While monitors focus on monitoring for the presence or absence of HABs, their observations also tend to capture the impacts that water quality health can have on the surrounding ecosystem - sightings of everything from robust sawfish to dead manatees. This circumstantial monitoring data may be outside the original scope of the Calusa Water Ranger Harmful Algal Bloom program yet its existence in the data set provides valuable educational data at no expense to the integrity of the monitoring data.

photos and document storage. A key factor in establishing a successful Tier 2 Qualitative Sampling workflow consists of ensuring data contributors are effectively trained and are comfortable with the data entry process. Further, in some cases an entity's QAPP may require an update to ensure it is aligned and references the necessary form fields being collected by participants.

Example use cases showcase how entities leverage the above Tier 2 data model to engage volunteers in monitoring programs using Water Reporter.

QUANTITATIVE SAMPLING - TIER 3

The third tier of the data collection is characterized by fixed monitoring points and a data model that is capable of accomodating time series information. Quantitative Sampling - Tier 3 is relational in that many readings generated by participants over time, can belong to a single monitoring site thus enabling users to gain better insight on how conditions change due to seasonality and other environmental factors such as flow and tide. This data model requires the greatest amount of effort from a data management side because managers must ensure monitoring sites all possess unique identifiers and that samples are carefully added to the fixed station. Quantitative sampling at routine monitoring sites are generally conducted with organizations that have a QAPP and have established an index of protocols and methods for their work in the field. Metadata is also established that describe the monitoring program and appropriate uses of the data being collected.

Example data fields, Table 11, follow a very similar structure as the Qualitative Sampling - Tier 2 model however, the primary difference is that a 'site_id' is necessary in order to generate a one to many relationship.

TIER 3: DATA FIELDS & TYPES		
FieldAlias	DataType	CollectionMethod
site_id	Integer	Machine Generated
latitude	Decimal	Pulled from Device
longitude	Decimal	Pulled from Device
watershed-huc	Integer	Automated Based on Location
watershed-name	String	Automated Based on Location
site-name	Text	User Entered
site-description	Text	User Entered
site_id	Integer	Machine Generated
reading_id	Integer	Machine Generated
collection-date	Date	User Entered
{Append n number of Parameter Names & Parameter Data Types}	Text, Date, Enumeration, Integer, Decimal, Document, Image	User Entered

Table 11: Display field names, data types, and collection methods for Tier 3 sampling data models across the relational data model.

MANAGEMENT

QAQC AND DATA MANAGEMENT

QAQC and Data Management is generally one of the most complicated components of the Water Monitoring Data Elements and Business Functions workflow. While critical, these components can lengthen the lag time between data acquisition and sharing if too complex. Complexities exist to ensure quality and control but can lengthen the time between data acquisition to analysis of the data. Furthermore, QAQC processes themselves require a level of in-staff training and inevitable transitions in staff can lead to a loss in knowledge in what methods are conducted when reviewing and validating samples for inclusion in the database. Limited access to funding and technically proficient staff can result in data loss or database corruption creating massive amounts of sample re-entry work. As a result, any solution considered for use by Tribal Governments and community science groups connecting their data to the CA FHAB Data Workflow must allow for easy and reliable management of new and existing data without compromising the integrity of the incoming data.

The Data Model Steering Committee recommends identifying and adopting solutions for each tier of data that fit the needs and skill sets of the individual monitoring programs. It's important to note that there are many solutions that can aid in effective data management. Based on the survey results, it is important that users select a data management option that meets the following criteria:

- Provides a data model that can be easily expanded and match the vocabulary of its user base while offering a high degree of standardization.
- Maintains services and data management applications in a universally accessible environment such as the cloud.
- Possesses an active and well-funded developer community that helps to align diverse data management needs into feature sets that are representative of Tribal Governments and community science monitoring programs.
- Maintains or is working to establish a publicly facing API that enables third parties to innovate and build applications on top of a common data management infrastructure.
- Ensures that all entities adopting data management services retain ownership of their own data.

Additional structures should be put in place depending on the data type being ingested by the model. For Tier 1 data, QAQC may look different than its more rigorous counterparts. Nonetheless the QAQC process for observational should allow for the rapid ingestion of data that is self-selected by the data collector without causing a backlog from displaying results due to required expert review. The QAQC process can include training of volunteers in how to categorize their observations. While this requires a small investment in training the general public, the volunteer can self-select a category that a trained expert can validate at a later time. That being said, the public-facing version of the post should always display a status of review of the post as determined by a qualified individual. Status can show an additional data field that displays confirmation of the post report, a status update, or a comment by an expert.

APPLICATION PROGRAMMING INTERFACE (API)

An API is a critical element that allows for the decoupling of an organization's database and data model from any functional applications that are either planned or in use. For example, if a Tribal Government or a watershed organization needed to convert their data to a different format, instead of having to change the field names and relationships in their own database to meet a given standard, an API could serve as a bridge to map the organization's internal schema to the desired one. APIs can also be used to connect different software applications and functionality. For example a Post Tier 1 record of a volunteer observing a bloom could be collected in the field using a smartphone. Upon receipt of that record, an email trigger could occur to send an email or text alert to a list of users subscribed to the watershed (e.g. BloomWatch app). This would rely squarely on an API to handle the event and trigger the necessary functionality.

In the case of water quality monitoring, an API can play a significant role in the structuring and dynamic transfer of data from one system to another. Any system being developed or adopted for integration into the CA FHAB Workflow should possess a public or semi-public API that enables users to programmatically request data at a practical organizational unit. A simple instance would consist of a series of documented GET endpoints that expose useful pieces of data for analysis or integration into third party applications.

If deemed feasible, we recommend that the API support authenticated requests and responses in JavaScript Object Notation (JSON), for it's readability and widespread adoption in application development. An example of a JSON output for a monitoring site and its respective readings could be structured as follows (Fig. 6):

- User posts a get request for stations at the following endpoint:
- Authentication: {User would register and input their API Key}
- GET: https://api.mywaterdata.org/datasourceid/stations/

Alternatively a user may wish to return readings and other ancillary data that is made available through a series of documented endpoints. APIs will only become useful if they are structured around serving out data, tasks, and functions that are requested by a community. We recommend that all members start small by exposing a few useful elements of the data being collected to gauge use and adoption. After the initial

```
},
    "graph_cache": null,
    "hibernate": false,
    "id": 8845,
    "image_id": null,
    "indicator_id": null,
    "is_active": true,
    "is_public": true,
    "is_scored": true,
    "json_url": null,
    "key": "69c57ae0d29e049f",
    "name": "JAMES RIVER AT HOLCOMB ROCK, VA",
    "organization_id": 130,
    "owner_id": 1216,
    "processing": false,
    "raw_id": "02025500",
    "score": null,
    "statistical_profile": null,
    "territory": {
      "huc_10_code": 208020303,
      "huc_10_name": "Harris Creek-James River",
      "huc_12_code": 20802030301,
      "huc_12_name": "Judith Creek-James River",
      "huc_6_code": 20802,
      "huc_6_name": "James",
      "huc_8_code": 2080203,
      "huc_8_name": "Middle James-Buffalo",
      "id": 78329
    },
    "territory_id": 78329,
    "updated": "2019-10-31T19:17:32.184495"
  },
```


Figure 6: An example of a GET request.

configuration or release, gather feedback from the users and adjust endpoints and your roadmap accordingly.

Some example data elements to start with could consist of the following:

- Monitoring Sites
    - Ancillary metadata such as station name, description, and location
- Readings
    - Parameter Name
    - Parameter method
    - Sample result and unit of measure

DISSEMINATION

VISUALIZATIONS

Our recommendation for visualizations to explore desired outputs after the completion of the data model and ingestion process for the Tribal Governments and community science data.

AI AND MODELING

To date, many of the technologies developed to serve the environmental movement have been created explicitly to support the information needs of the environmental practitioner. The current assumption our community mistakenly makes is that the mission of restoring our natural world is a strong enough case to convert a member of the general public to dig into complex data, use our systems, and begin getting involved in tackling local water quality threats. While this is far from the reality of how individuals outside the choir leverage technology, it presents a massive blank slate to begin exploring new approaches for making water quality data pervasive in daily decision making of the general public. Building on Google's dominant JSON Linked Data (JSON-LD) standard and highly robust AI framework for Natural Language Understanding (NLU), the Data Model Steering Committee recommends leveraging forward looking platforms that will enable monitoring data and watershed organization programs to be searchable and discoverable online while also exposing the necessary functionality allowing Google Home users to ask their device a variety of questions that can be answered based on data supplied by a given user base. Current frameworks exist to deploy assistant AIs that have the capabilities to help watershed residents conduct the following basic and complex tasks:

- Identify what watershed they are in
- Gain information on who the local watershed organizations are in their area and what they are doing to monitor and improve their local waterway.
  - Donate to said organization
  - Sign up for events related to local cleanups and volunteer monitoring training.
- Based on relevant thresholds and data, ask if it's safe to swim at a beach nearest to them.

While this functionality is less utilitarian as perceived from the lens of an environmental practitioner, the user patterns emulate what a stunning majority of internet users are accustomed to experiencing. This recommendation could provide the necessary bridge that brings localized efforts to restore water quality, one step closer to the 7.2 million U.S. residents that currently own a Google Home device. Further, integration of the JSON-LD data standard is a step that will allow samples and information managed by Tribal

Governments and community science groups to occupy significant real estate when returning search results from Google. By focusing on these simple, but agreed upon definitions for each organization's monitoring data elements and interpretations, our cumulative efforts can be indexed and standardized for easy public dissemination. This means there is a greater chance of this information being discovered in a context that is of utility to an interested third party. Looking ahead, our work to integrate data managed effectively with JSON-LD standards and new AI frameworks, will unlock new and more efficient ways for the general public to get involved with their local Tribal Government or volunteer monitoring program, joining the ranks to help improve our natural world.

ALERTS

Alert frameworks are deceivingly simple to implement. While they do not require a significant amount of technical proficiency to implement, a clear direction on content and context for framing how data informs a user's action is of utmost importance. For example, alerts should only be triggered to the public until after authorities such as the State, Tribal Governments, and community science leads align on what actions should be taken by the public. For example, if a Tribal Government or community science monitor shares an observation for what they feel is a water quality threat that may require additional assessment, these data should serve as the common thread of analysis that crafts direction to the general public. Once all parties are aligned on the call to action the alert should contain the necessary actions members of the public should take, based on the data observed.

In an effort to meet members of the public where they are, information delivery could occur in the form of an email or push notification, as well as alerts via social media from authoritative State, Local, or Tribal Government sources. It's important to note that no matter what the channel, State and Tribal Governments, and community science authorities should share the same data interpretation and common messaging when recommending actions to either take or not take to the public. These results should be rapidly interpreted in the appropriate context of public health, common actions should be crafted, and these next steps should be packaged in the form of an alert to the general public as quickly as possible. Useful frameworks or technical services that can support this effort consist of Twilio, a modern SMS, Push, and Interactive voice response (IVR) API that provides easy data integration across any platform, as well as Mandril, a transactional email API that helps automate send triggers and content.

OPEN FHAB DATA: EXPORTS AND ANALYSIS

The ability to export data in a machine-readable format provides an almost infallible framework in ensuring organizations can maintain ownership of their data and not be reliant on development cycles of their service provider. The ability to export metadata will also be paramount in order to share information relevant as an open data source. In this day and age, software provides its greatest value based on context and integration. It is not cost effective for a system or product developer to duplicate the functionality of a platform that has already been created, but rather provide a means for interoperability. For example, out of the box, systems like ArcGIS, Tableau, and Water Reporter offer base functionality, an API, and data export via software as a service. This puts the initiative on the user to identify what their analytical workflow needs are. The ability to export or leverage an API for one or all of these systems means that they can all be used together or independently to meet the requirements of the institutions they serve. It's imperative for this effort that organizations focus internally on what their business functions are and identify the suite of software packages that can help them achieve rather than leaving it up to software systems to serve as a silver bullet fallacy.

Good analysis and the utility of data export stems from having a core plan or need for what you anticipate to use structured and machine-readable data for outside its existing system or original intent. This could include matching the schema of a state or federally adopted data format such as CEDEN or WQX, or it could also include accessing a raw format for your organization's own internal or external information products. At the core of any of these organizational business functions, authorized parties must be able to access raw data from the data management system they institutionalize.

Clear Lake, CA - Photo by Keith Bouma-Gregson

APPENDIX 1: MEMBERS AND ORGANIZATIONS

DATA MODEL STEERING COMMITTEE

The **Internet of Water**, a project based at Duke University's Nicholas Institute for Environmental Policy Solutions, seeks to fundamentally change how water is managed by improving access to water data for real-time decision-making. Its mission is to build a dynamic and voluntary network of communities and institutions to facilitate the opening, sharing, and integration of water data and information. This network will connect data producers, hubs, and users to enable the discovery, accessibility, and usability of water data and information.

The Commons is a registered 501(c3) organization with a mission to aid individuals and organizations to access, organize, and share data that inspires action that restores our environment. The Commons aids environmentally focused NGOs and government agencies to access robust software development services that facilitate water quality improvement at a national scale. They have a proven track record for solving problems at a local scale while still ensuring data and products can be utilized to inform and drive national efforts. They design their software as a service (SaaS) products with an approach that ensures their technology supports a variety of overlapping needs exhibited by their user base allowing them to build systems one time, that serve the greatest number of organizations at the least possible cost. They are dedicated to working exclusively on projects that facilitate environmental improvement with a unified philosophy to create the code for change.

The **California State and Regional Water Quality Control Boards** (California Water Boards) works to preserve, enhance, and restore the quality of California's water resources and drinking water for the protection of the environment, public health, and all beneficial uses, and to ensure proper water resource allocation and efficient use, for the benefit of present and future generations. Within the Water Boards, the **Surface Water Ambient Monitoring Program (SWAMP)** oversees the **Freshwater Harmful Algal Bloom (FHAB) Program**, which is part of a statewide initiative to address cyanobacteria blooms and cyanotoxins in California's freshwater systems and help ensure that the public is protected and informed.

The **California Water Quality Monitoring Council** (Monitoring Council) serves to integrate and coordinate water quality and related ecosystem monitoring assessment and reporting between the boards, departments and offices within California. The Monitoring Council oversees the **California Cyanobacteria and Harmful Algal Bloom (CCHAB) Network** which works towards the development and maintenance of a comprehensive, coordinated program to identify and address the causes and impacts of FHABs in California.

Internet of water

Peter Colohan, *Executive Director*
Kyle Onda, *Data Architect*
Lauren Patterson, *Data Analyst*
Ashley Ward, *Engagement and Outreach Associate*

The Commons

R. John Dawes, Jr, *Executive Director*
Erin Hofmann, *Strategy and Application Lead*
Gabriel Watson, *Data Analyst*

California State and Regional Water Quality Control Boards

Erick Burres, *Statewide Citizen Monitoring Coordinator*
Keith Bouma-Gregson, *Co-Lead of the FHAB program for the Water Boards*
Marisa Van Dyke, *Co-Lead of the FHAB program for the Water Boards*
Greg Gearheart, *Deputy Director, Office of Information Management and Analysis (OIMA)*
Anna Holder, *Environmental Scientist in OIMA/ SWAMP*
Meredith Howard, *Environmental Program Manager at Central Valley Regional Water Quality Control Board*

California Office of Environmental Health Hazard and Assessment

Beckye Stanton, Jr, *Environmental Scientist, Division of Scientific Programs: Pesticide and Environmental Toxicology Branch and CCHAB Network co-chair*

STAKEHOLDER REVIEW COMMITTEE

Several groups are currently monitoring FHABs in California, including stakeholders within the CCHAB Network, Tribal Governments, and community science groups.

NATIVE AMERICAN TRIBAL GOVERNMENTS

Tribal Governments have a long history of involvement in FHABs monitoring and policy development in California. In 2017, the State Water Boards approved new Tribal Beneficial Uses to address the traditional interactions between tribes and aquatic ecosystems and input from Tribal Governments provide valuable information to ensure adequate water quality protection.

Big Valley Band of Pomo Indians

Sarah Ryan, *Deputy Tribal Administrator/ EPA Director and CCHAB Network co-chair*

Karuk Tribe Department of Natural Resources

Susan Fricke, *Water Quality Manager*

COMMUNITY SCIENCE GROUPS: NON-PROFIT AND NON-GOVERNMENTAL ORGANIZATIONS

Restore the Delta

Barbara Barrigan-Parrilla, *Executive Director*

The Watershed Project

Helen Fitandes, *Project Coordinator and Water Quality Monitoring Council member*

League to Save Lake Tahoe

Jesse Patterson, *Chief Strategy Officer*

Emily Frey, *Citizen Science Program Coordinator*