Designing, Implementing, and Evaluating a Municipal Water Quality Dashboard

Anonymous
Anonymous Department
Anonymous University
Anonymous Place
Email

Anonymous
Anonymous Department
Anonymous University
Anonymous Place
Email

***Abstract*— At the Cary/Apex Water Treatment Facility in North Carolina, laboratory staff and sampling technicians perform over 100,000 water tests every year. This abundance of data is validated manually and analyzed with spreadsheets. This paper describes the process of designing, implementing, and evaluating a data visualization dashboard in PowerBI to help this water treatment facility effectively monitor the quality of the drinking water in their distribution system. The effectiveness of the dashboard was evaluated through surveys given to the project stakeholders. The feedback from these surveys showed that, overall, the users found the dashboard useful but would like to have real-time data updates.**

Keywords—dashboard, data visualization, PowerBI, water treatment monitoring

I. Introduction

The Cary/Apex Water Treatment Facility (CAWTF) provides drinking water to approximately 290,000 customers in central North Carolina [1]. Laboratory and sampling technicians need to:

- Collect ~2,600 samples from required coliform sites each year,
- Perform and track over 20,000 field tests and 80,000 lab tests per year on water samples,
- Verify that water quality measurements, including total chlorine and nitrite, meet regulatory requirements.

It is important to review and trend distribution results regularly to verify compliance and protect public health.

The CAWTF store test result data in the SampleMaster® LIMS [2], an online system hosted by Accelerated Technology Laboratories, Inc. (ATL). Field data is entered into the LIMS using the web-based SampleMaster® iMobile portal [3]. The quantity of tests and disparate software systems have created challenges for CAWTF staff. Test result data have been overwritten and required monthly samples have nearly been omitted. Aggregating sample and test data from multiple systems is an effort-intensive, manual process, and out-of-bounds test results have been missed among the volume of data.

It is imperative for water plant personnel to notice if the drinking water does not meet regulatory standards or if there are concerns with nitrification in a timely manner to protect public health. While these challenges have not impacted water safety, an opportunity exists to improve these processes with modern software tooling.

This article describes the design, implementation, evaluation, and deployment of a PowerBI [4] dashboard to improve the collection and analysis of water samples. The field water quality test results were manipulated in PowerBI by pivoting tables, grouping columns, filtering rows, applying conditional formatting, and applying custom Data Analysis Expressions (DAX), resulting in several data visualization elements including tables, gauges, maps, and slicers. The pipeline and end-user dashboard enable the CAWTF to monitor multiple statistics, analyze the samples for trends, and receive alerts for out-of-range values. The new system helps personnel to ensure the drinking water meets regulatory standards more efficiently and accurately.

The remainder of this paper is organized as follows: Section II provides additional background on the facility's challenges and current technologies; Section III reviews literature on data dashboard design; Section IV describes the dashboard implementation; Section V provides evaluation and discussion, and Section VI discusses takeaways from this effort and opportunities for continued development.

II. Background and Challenges

Requirements for our software system were derived by observing the current processes and interviewing stakeholders using methods prescribed by Robertson and Robertson [5]. The current water sampling and analysis process was documented through interviews with subject matter experts, documentation review, and on-site apprenticeship.

A. Current Water Sampling and Analysis Process

The CAWTF is a surface water treatment facility jointly owned by the Town of Cary and the Town of Apex in North Carolina. Water from Jordan Lake is treated using ozonation, mixing, clarification, and filtration. The drinking water is disinfected with chloramines before being pumped into water towers, homes, and businesses through a network of over 1,000 miles of drinking water pipes [1].

State regulations require the CAWTF to collect at least 200 coliform samples per month from the drinking water distribution network. The sampling technicians also measure chlorine levels at the sample sites. Technicians also collect weekly measurements for pH, monochloramine, and free ammonia from specific long residence time locations. Technicians occasionally measure the water temperature, conductivity, and turbidity for process control purposes. Importantly, technicians measure nitrification at locations known to have stagnant water, which can lead to bacterial growth [6]. The technicians record around 1,550 field test results per month while at the sample locations using the SampleMaster® iMobile software platform [3] on their iPads.

At the start of the month, sampling technicians are given labels for the sampling sites they are to visit. The technicians drive to the sampling sites to collect water samples and perform field measurements. The technicians track their progress by reviewing which labels are left for the month. Labels have been lost causing confusion and risk of missing samples or inefficient route planning once the omission is discovered.

The technicians bring the samples to the laboratory, sign the chain of custody paperwork, and give the samples and paperwork to laboratory staff. The laboratory staff review the results for abnormalities (typos) and concerns (e.g., low chlorine, high nitrite) and document receipt of the samples. If chlorine or nitrite exceed thresholds, the sampling technicians are asked to flush the hydrant to improve water quality.

The laboratory staff perform coliform and other bacteriological tests on the water samples. In total, the laboratory staff run approximately 7,000 tests per month on the samples collected from the distribution system and from other points within the treatment process. Laboratory staff enter the laboratory test results into the SampleMaster® LIMS program [2]. The laboratory and field test results are collated in LIMS. Within 10 days of the end of the month, the CAWTF laboratory supervisor exports the data required for regulatory reporting (all coliform and total chlorine results) from SampleMaster® LIMS to an Excel file. The supervisor reviews the data to verify the required monthly samples were collected and that there are no empty fields or typos. The file is uploaded to the North Carolina Public Water Supply's online portal, where compliance with the chlorine and coliform limits is tracked.

B. *Stakeholder Requirements*

A focus group meeting was held in June 2023 with seven stakeholders from the CAWTF to understand their needs for the proposed data. Following introductions and a brief review of the project, the stakeholders were asked to explain the limitations of their current systems for data management. The consensus was that a user-friendly data dashboard for aggregating and visualizing the sample collection and analysis process would be of most benefit. Specifically, the stakeholders prioritized four challenges for the dashboard to address.

1) Challenge 1: Support efficient route planning by showing geographical positions of sampling locations and provide indicators of collected and uncollected samples.

The first issue raised was that sampling technicians could not efficiently see which water samples needed to be collected. At the beginning of every month, each technician is provided with a printout of their assigned sample labels (between 40 and 68) containing the addresses of each sampling station they need to visit. The technicians have the autonomy to decide which stations they visit each day, but all labels must be collected before the end of the month. However, there is no map representing the geographic locations of the sampling stations. It can be challenging, especially for less experienced technicians, to keep track of which samples still need collecting and to plan efficient routes. To support efficient route planning, the dashboard should be equipped with a map of sampling locations that shows a distinction between sites that have and have not been collected.

2) Challenge 2: Prevent data loss due to errors with the user interface.

The second problem discussed was that, within the current system, there is the potential for data to be lost entirely. The SampleMaster® iMobile software is reportedly cumbersome to use and data can be overwritten or lost if buttons are pressed incorrectly. Sampling technicians have developed ways to prevent loss of data, such as taking screenshots, entering test results into a second portal, or writing results on paper. The dashboard should address the issue of potential data loss by displaying field data captured in iMobile in real time to increase confidence that the data were uploaded successfully.

3) Challenge 3: Make concerning test results more visible to alert staff faster.

At present, the laboratory staff manually review field test results to identify any out-of-range values or typos. The staff examine results printed on paper with no ability to filter or search. If a suspected typo is found, the lab staff consults the sampling technician and makes corrections if necessary. If an out-of-range value is found, then precautionary measures, like flushing hydrants, are scheduled. The dashboard should display data in a format that makes out-of-range results more apparent, helping to make the process of reviewing results more efficient. The stakeholders suggested adding color-coded icons to the map and highlighting concerning results in the table display.

4) Challenge 4: Automate data processing of LIMS data to improve analysis efficiency.

Currently, CAWTF staff manually export data from the LIMS to analyze trends in nitrification and other values of interest. They prepare this trended data for use in monthly meetings during the summer where they discuss the results and strategies for mitigation. This is time consuming and labor intensive. To aid in the team's data analysis process, the dashboard should show how the nitrite and total chlorine values have changed over time. It should update automatically (or when manually refreshed) so that the most up-to-date data are available when needed.

After the focus group, hi-fidelity prototypes were created for three dashboard pages: (1) Cary bacteriological and sampling statistics, (2) Apex bacteriological and sampling statistics, and (3) nitrification analysis. The prototypes were provided to CAWTF stakeholders, who provided feedback prior to implementation.

III. Related Work

A. Data Visualization and Performance

Research shows that visualizing data promotes better data cognition than text-only numeric displays [7]. Properly interpreting the data is an integral step in the process because misinterpreting the data and justifying a decision on a false truth could lead to a disastrous outcome, especially when a company's customer base is fully dependent on them. As evidenced by the studies listed below, graphically representing data improves the analyst's understanding of it.

Ballard sought to discover whether US Public Health Managers prefer to use data presented in tables or graphs [7]. Half of the participants were provided performance information using a bar chart while the other half were presented with the same data in a table. Then they were asked to rate their intention to use the data from 1 (least likely) to 10 (most likely). In all cases, graphically displayed data were rated more likely to be utilized [7]. Similar findings were obtained from a study involving 32 intensive care nurses. The study found that nurses using a visual dashboard to monitor patient health were able to notice and act upon changes in the patient's condition faster than the nurses who used tabular displays [8].

Bradley et al. report on the benefits of a data management program at the Philadelphia Water Department (PWD) [9]. The PWD uses a data dashboard to track distribution system nitrification by displaying trends in drinking water storage tank chlorine residuals and water temperatures. Their dashboard also monitors water system pressure readings, water main breaks, and total chlorine residuals. Bradley et al. report that the use of dashboards helps the PWD share data faster and easier by reducing the need to prepare reports, reducing human errors by removing the need for manual data editing and has kept the statistical analysis methods consistent over time.

B. Data Dashboard Design Principles

Dashboards are versatile and can be equipped with charts, displays, and maps to show both numerical and categorical information. First, and most importantly, a dashboard should be simple, concise, and free of cluttering information. It should allow the user to see everything they need in one place all at once, and preferably on a single page. Few observes that "this single-screen display of the most important information people need to do a job, presented in a way that allows them to monitor what's going on in an instant, is a powerful new medium of communication. At least it can be, but only when properly designed" [10]. Additionally, Few states [10] that "Well-designed dashboards deliver information that is:

- Exceptionally well organized,
- Condensed, primarily in the form of summaries and exceptions,
- Specific to and customized for the dashboard's audience and objectives, and
- Displayed using concise and often small media that communicate the data and its message in the clearest and most direct way possible."

Malkani et al. state that an expertly designed performance dashboard should have the following qualities: "1) easy navigation, 2) high usability, 3) use of adjustable thresholds, 4) use of diverse chart selection, 5) compliance with the Americans with Disabilities Act, 6) use of charts with tabulated data, 7) incorporated user feedback, 8) simplicity of design, 9) adding clear descriptions for charts, and 10) comparison data with other entities" [11].

Out of the principles discussed above, I primarily focused on incorporating the following into the dashboard I designed:

- Determine the information most important to the end user and put it in front.
- Incorporate simple, helpful, and intuitive displays.
- Only display the most current and relevant data.
- Incorporate as much useful data into the dashboard to improve user efficiency without making it overcomplicated.

IV. Dashboard Product

Two major activities were required to address the CAWTF stakeholders' challenges. First, the existing data pipeline was revised to support modern software tooling. Second, a data dashboard was implemented in Microsoft's PowerBI [4]. PowerBI was chosen due to the product's maturity, community support, and availability within the CAWTF's Microsoft Office365 package.

A. Data Pipeline Revisions

Fig. 1 shows the data pipeline that delivers data from the sampling technicians' devices to the data dashboard. The final deployment of the dashboard required modifications and enhancements to this pipeline by multiple parties.

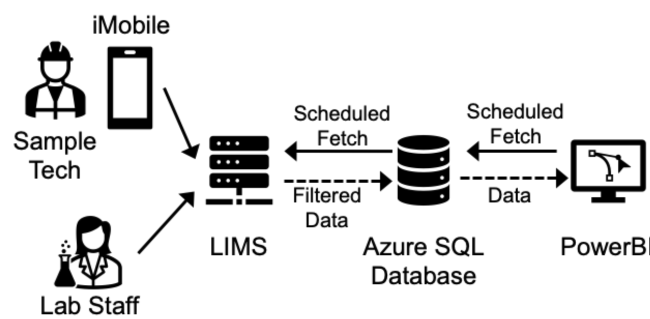

Fig. 1. Diagram depicting data flow from field to dashboard.

Sampling technicians enter field water quality test results into iMobile, which stores the data in the cloud-based SampleMaster® LIMS. LIMS provides over 25,000 data fields for each sample, most of which are empty and irrelevant for CAWTF samples. The stakeholders identified 19 relevant fields, and Cary's IT team created a JSON file to filter the LIMS data to those fields, reducing the transmission size of a single sample from approximately 1300KB (mostly field labels with empty values) to approximately 12 KB – a 100x reduction.

An Azure SQL Database operated by the Town of Cary retrieves the sample data and filters it using the JSON file several times per day through the LIMS application programming interface (API). The filtered data is pushed automatically to an Open Data Portal hosted by Opendatasoft [12], where it is accessed via a URL with the appropriate API key. The Town of Cary (the municipal body that oversees the CAWTF in-part) uses this "Open Data Portal" to publish various municipal data for public use. PowerBI then imports the updated data from the Open Data Portal. This data transfer cycle occurs at 6 designated times per day.

The revised data pipeline helps to address Challenge 2 – prevent data loss through an improved user interface. The Azure SQL queries LIMS using collect date and order ID/sample number, resulting in a large dataset with all values, including those previously received. Then duplicates and tests pending analysis (where no results are available) are filtered out. The overall process takes about 5 to 15 minutes and results most times in no (or very few) updates. Based on carefully analyzing the data with regards to volume of data, nature of delays in publishing and on availability of results, it was determined that scheduling the querying of LIMS at times when data was expected to be published was the proper way to overcome the inefficiencies. Although the data are not updated in real time or on demand, the LIMS data is updated in the PowerBI dashboard six times per day, five of which are during prominent sampling hours. This enables sampling technicians to regularly verify that field data were saved in LIMS throughout the workday.

B. Water Sample Monitoring Pages

Figures 2-3 show pages of the implemented PowerBI dashboard. Data transformations and DAX were implemented to perform calculations and transform data underpinning the visualization widgets. The PowerBI dashboard pages implement the hi-fidelity prototype designs reviewed by CAWTF stakeholders. The process of designing various iterations of the dashboard spanned five calendar months. The dashboard and its elements that address the stakeholder challenges are described below.

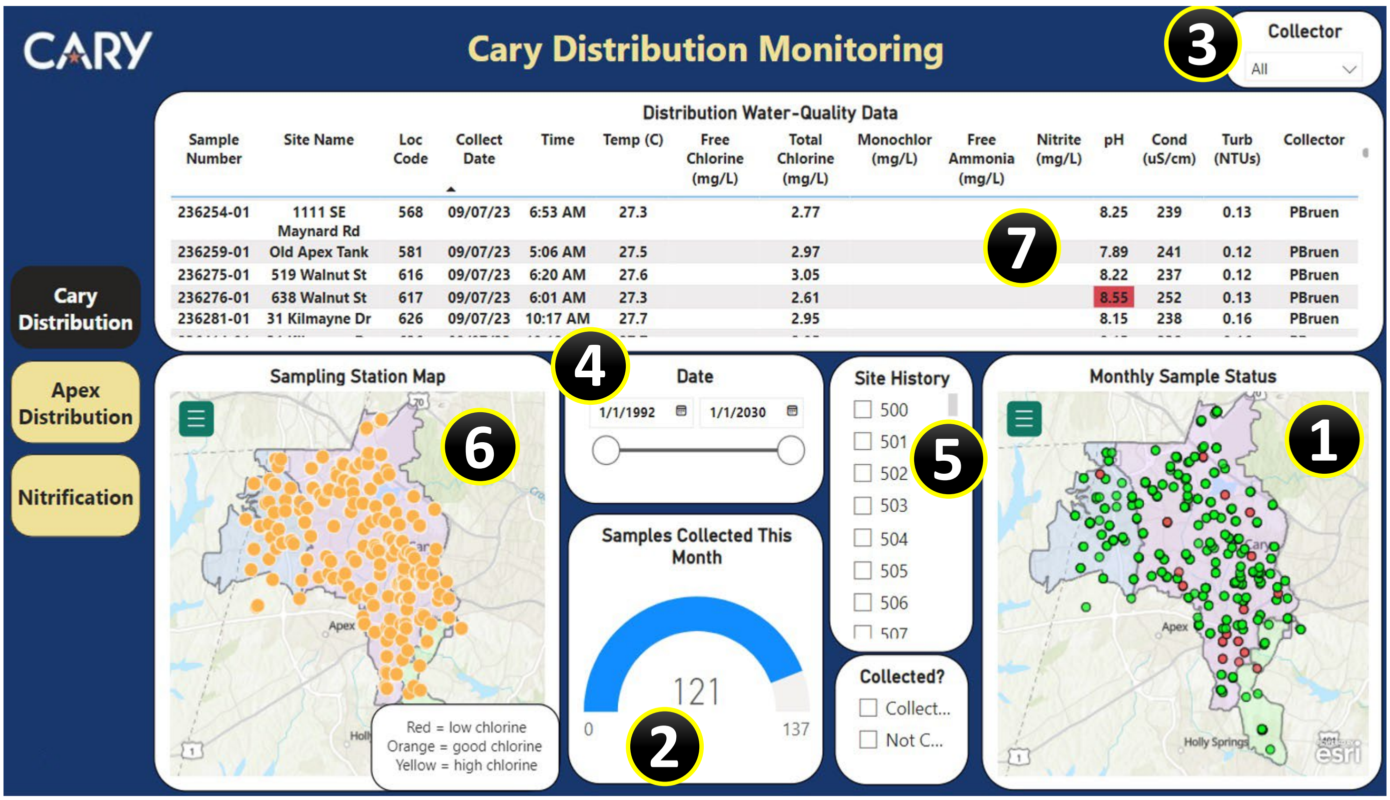

Fig. 2. Town of Cary Distribution Monitoring Page

Challenge 1 is to better support route planning and sample collection tracking. The Distribution Monitoring page (Fig. 2) contains a sample collection map (1) shows where samples have and have not been collected in the current month. The page also shows a gauge (2) with the number of samples collected out of the total required for the month to help CAWTF staff track overall sampling progress. The user can further refine the test result table and chlorine map by using filters for sampling technician (3), date (4), and location code (5), which assists individual technicians in planning their work.

Challenge 2 is to prevent data loss through an improved user interface. This issue is partially addressed by the infrastructure of the data pipeline. Although the data are not updated in real time or on demand, they are set to update six times a day, 5 of which are during prominent sampling hours.

Challenge 3 is to make concerning test results more visible to the analyst. Test results and icons on maps are programmed to change colors to alert viewers of concerns. The Distribution Monitoring page (Fig. 2) shows a map of chlorine levels at different sampling sites (6) with color-coded dots that reflect the test results. The page shows a table (7) of recent test results and highlights any that are out of range. On the Nitrification page (Fig. 3), results for nitrites are plotted on a graph (8) and total chlorine results are plotted on graph (9) that shows whether the values are above or below the limit line. This page also features a heat map (10) that highlights areas of the distribution system on a map where nitrification levels are of interest and how they have changed over time.

Challenge 4 is to automate data processing of LIMS data to improve analysis efficiency. The data pipeline addresses this issue by updating the dashboard automatically several times a day. The combined features of the data dashboard enable CAWTF staff to view trends and quickly scan and filter erroneous values, avoiding the need to export and manually analyze the data from LIMS. The most up to date data is ready for review at any time, saving staff many hours of manual manipulation every month.

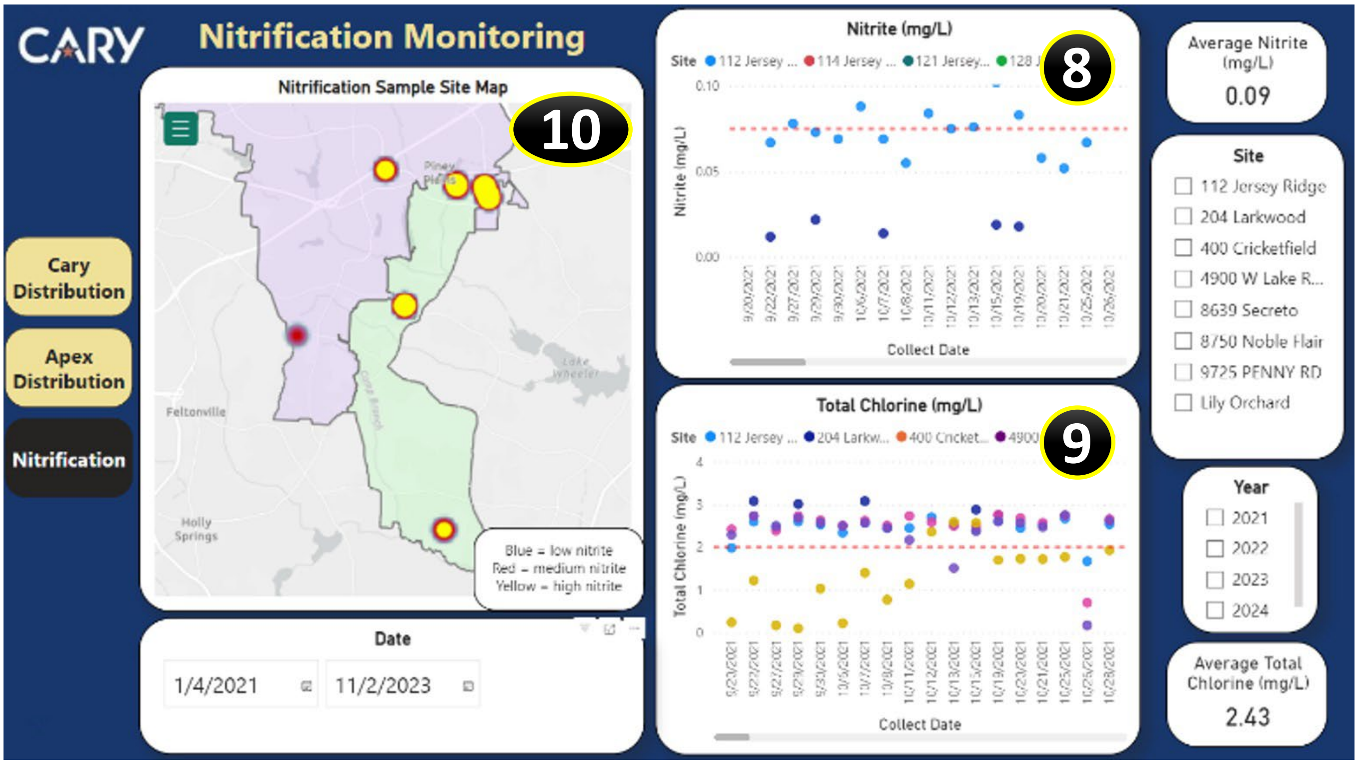

Fig. 3. Nitrification Monitoring Dashboard

V. Evaluation and Discussion

The PowerBI dashboard was deployed on a private municipal network accessible to CAWTF personnel via a web browser. After deployment, a review session was conducted with seven CAWTF stakeholders (a mix of sampling technicians, laboratory staff and their supervisors). The most important features of the dashboard were demonstrated, and the attendees were asked to provide feedback. After the meeting, all attendees were emailed an anonymous survey to learn what met their standards and what areas needed further improvement. The survey contained the following questions:

- Did you find the dashboard easy and intuitive to use?
- Does the dashboard allow for you to see quality concerns in a timely manner?
- Does it allow you to see what distribution samples still need to be collected for the month?
- Does this tool allow you to review the data faster?
- Can you get useful information from the dashboard?
- Please rate the usefulness of the maps.
- Are there any must-have features missing?
- What current features can be enhanced to make it more useful?

Five of the seven stakeholders (71%) at the review meeting responded to the survey. The feedback was positive. All five respondents stated that the dashboard:

1. was easy and intuitive to use,
2. allowed for them to see quality concerns in a timely manner,
3. allowed them to see what distribution samples still needed to be collected for the month,
4. allowed them to review the data faster, and
5. gave them useful information.

The respondents all indicated that they would probably use it daily and that it would be useful. One respondent stated, “I feel like this is a tool that I will look at on a regular basis and that it will be useful for water plant staff and sampling staff." However, three respondents expressed their desire for the data to update in real time: "would be great if it was 'real-time' so samplers can double check data entry." This feature would help samplers avoid accidental loss of data because they would have an indication that the data they entered were captured.

Two stakeholders also volunteered to complete a usability test while their screens were recorded. They completed three typical analysis tasks: (1) determine of average total chlorine has increased or decreased from 2022 to 2023 at locations where nitrification is monitored, (2) identify when a sample was last collected from sample station 504, and (3) determine how many samples have been collected this month. Both volunteers completed each task within seconds without help or training. However, one participant inadvertently checked a filter selecting a particular sample location. So, when looking for whether the chlorine increased or decreased from 2022 to 2023, he was looking at the average chlorine levels for that specific site and not for all the locations. This suggests the dashboard could be improved with better indicators of the system status, i.e., which filters are applied, as suggested by Nielsen's usability heuristics [13]. The volunteers evaluated the dashboard using the System Usability Scale [14]. The average SUS score was 99, indicating the two volunteers found the dashboard to be highly usable and highly learnable.

VI. Takeaways and Future Work

Overall, the CAWTF stakeholders praised the data dashboard and planned to use it daily. We offer the following recommendations, learned through this project, specifically for data dashboard developers.

1) Prototype extensively with stakeholders. The dashboard must use visualization techniques familiar to and preferred by the users [6]. In our project, the initial prototypes were revised multiple times. Implementation of the dashboard widgets can proceed with confidence with this target reference available and reduces the need for rework on the final product.

2) Design and implement the data pipeline as early as possible. Any delay in data pipeline development will delay the dashboard's delivery. Further, the data schema provided by the pipeline constrains the capabilities of the dashboard. In our project, technical challenges prevented the data from being updated in real-time as desired by the users. Ultimately, the data pipeline constrains the dashboard solution, and so it is essential to discover these constraints as soon as possible.

In the future the dashboard could be improved by implementing features the stakeholders mentioned they would like to have but were not considered essential. These include a map with total coliform/*e. coli* positives; a map with sensitive water customers (such as manufacturing and pharmaceutical companies) and critical customers (such as schools and hospitals); and a map showing sampling station details such as date installed. The dashboard can also be expanded to display many more test results captured in the LIMS, such as customer complaints, disinfection byproducts, perfluorinated compounds, as well as the number of tests performed and the performance of each analyst with respect to quality control analyses. We will continue to enhance the application as its usage grows with CAWTF staff.

Acknowledgment

The authors thank the Town of Cary and Accelerated Technology Laboratory, Inc (ATL) staff members that collaborated on this project. Thank you to the Water System Laboratory Supervisor, Rachel Monschein, who served as the initial contact and provided details on the water treatment facility's sampling routines and current data-management methods. Thank you to Damian Bucovsky, Cary's IoT Solutions Architect and Gregory Power, Cary's Data Scientist who assisted with establishing the data pipeline. Thank you to Cary Engineers Sarah Braman and Corrie Bondar for their Power BI expertise and all of Cary's water treatment staff including David Brooks, Patrick Bruen, Erin Lee, John Conley, Davis Reynolds, and Seth Burleson, who gave their input on how the project could best suit their needs.

References

[1] Town of Cary, "2022 Annual Water Quality Report," Cary, Aug. 2023. Accessed: Jan. 14, 2024. [Online]. Available: www.townofcary.org/waterreport

[2] Accelerated Technology Laboratories, "Sample Master LIMS." Accessed: Dec. 30, 2023. [Online]. Available: https://atlab.com/products/sample-master/

[3] Accelerated Technology Laboratories, "Mobile Data Collection for your Laboratory." Accessed: Dec. 30, 2023. [Online]. Available: https://atlab.com/products/sample-master-imobile/

[4] Microsoft, "Power BI - Data Visualization | Microsoft Power Platform." Accessed: Dec. 30, 2023. [Online]. Available: https://www.microsoft.com/en-us/power-platform/products/power-bi

[5] S. Robertson and J. Robertson, Mastering the Requirements Process, 2nd ed. Upper Saddle River, NJ, USA: Addison Wesley, 2009.

[6] Town of Cary, "Cary/Apex Water Treatment Facility Nitrification Action Plan," Cary, Jan. 2021.

[7] A. Ballard, "Promoting Performance Information Use Through Data Visualization: Evidence from an Experiment," *Public Performance & Management Review*, vol. 43, no. 1, pp. 109–128, Jan. 2020, doi: 10.1080/15309576.2019.1592763.

[8] S. Anders *et al.*, "Evaluation of an integrated graphical display to promote acute change detection in ICU patients," *Int J Med Inform*, vol. 81, no. 12, pp. 842–851, Dec. 2012, doi: 10.1016/J.IJMEDINF.2012.04.004.

[9] T. C. Bradley, N. Horscroft, K. Mattson, and G. A. Burlingame, "Data Management Programs Offer Multiple Benefits," *Journal AWWA (American Water Works Association)*, vol. 113, no. 10, pp. 22–32, Dec. 2021, doi: 10.1002/AWWA.1824.

[10] S. Few, *Information Dashboard Design: The Effective Visual Communication of Data*, 1st ed. Sebastapol, CA: O'Reilly Media, Inc., 2006.

[11] D. Malkani, M. Malkani, N. Singh, and E. Madan, "Best Practices for the Design of COVID-19 Dashboards," *Perspect Health Inf Manag*, vol. 20, no. 1, p. 1b, 2023, Accessed: Dec. 30, 2023. [Online]. Available: /pmc/articles/PMC9860470/

[12] Opendatasoft, "Create the best data experiences - Opendatasoft." Accessed: Jan. 14, 2024. [Online]. Available: https://www.opendatasoft.com/en/

[13] J. Nielsen, *Usability Engineering*, 1st ed. San Diego, CA, USA: Academic Press, 1993. doi: 10.1524/itit.2002.44.1.003.

[14] J. Brooke, "SUS: A 'Quick and Dirty' Usability Scale," in *Usability Evaluation In Industry*, CRC Press, 2020, pp. 207–212. doi: 10.1201/9781498710411-35.