AQ-SPEC
Air Quality Sensor Performance Evaluation Center

Air Sensors International Conference 2022
Session 2B: Swimming in Data: The Current and Future State of Data Management Platforms

Lessons Learned in Designing, Developing, and Implementing the South Coast AQMD AQPortal Environmental Data Management Solution

Vasileios Papapostolou, Sc.D.
Program Supervisor, AQ-SPEC
(Presented by Ashley Collier-Oxandale, Ph.D.)

May 11, 2022
Pasadena, CA

AQPortal Overview

(Structure and the use of tools such as RStudio Team offers to ability to revise and add to Dashboards as needed)

Internal Databases (data produced by South Coast AQMD)

- Low-cost sensor networks
- Air monitoring station data (regulatory)
- Laboratory samples
- AB 617 Community Air Monitoring
- Rule 1180 Refinery Emissions Monitoring
- Air Quality Assessment (AQI data with associated health messaging)
- Special projects (e.g., MATES V)
- Air quality advisories
- Facilities information

(note, includes continuous/time integrated data as well as stationary/mobile data)

External Databases

- Traffic count data
- NOAA meteorological data
- Fence line data collected by refineries
- NASA satellite data (OMI, TROPOMI)

(note, ability to add more in the future)

Azure Cloud Infrastructure (database for storage, organization, and processing)

Data Export

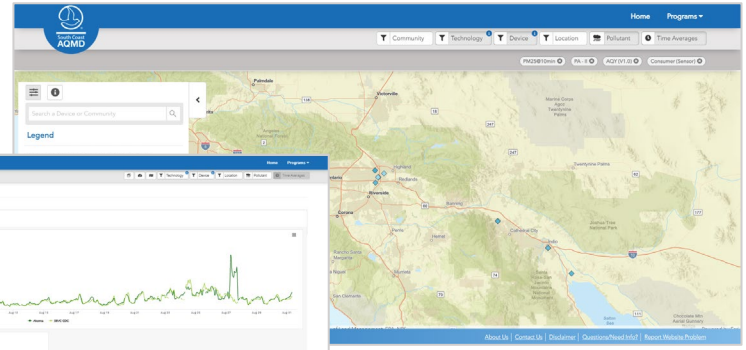
- Easier access to data
- Ability to customize downloaded data

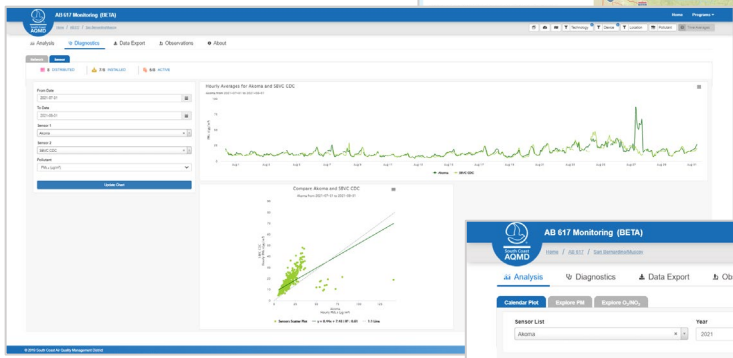
External Dashboards (for the public and staff)

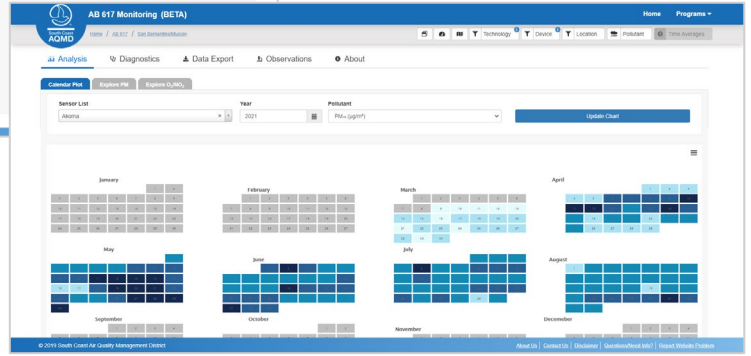
- Designed for intuitive use and engaging interaction with the data for all types of users
- Web-based
- For the public to better understand their local air quality AND explore special projects
- Wide range of visualization types
- Dashboards ranging from the "All Programs" or global/district level to the program-, project-, and even community-specific levels

Internal Dashboards (for district staff)

- For diagnostics and more complex analysis







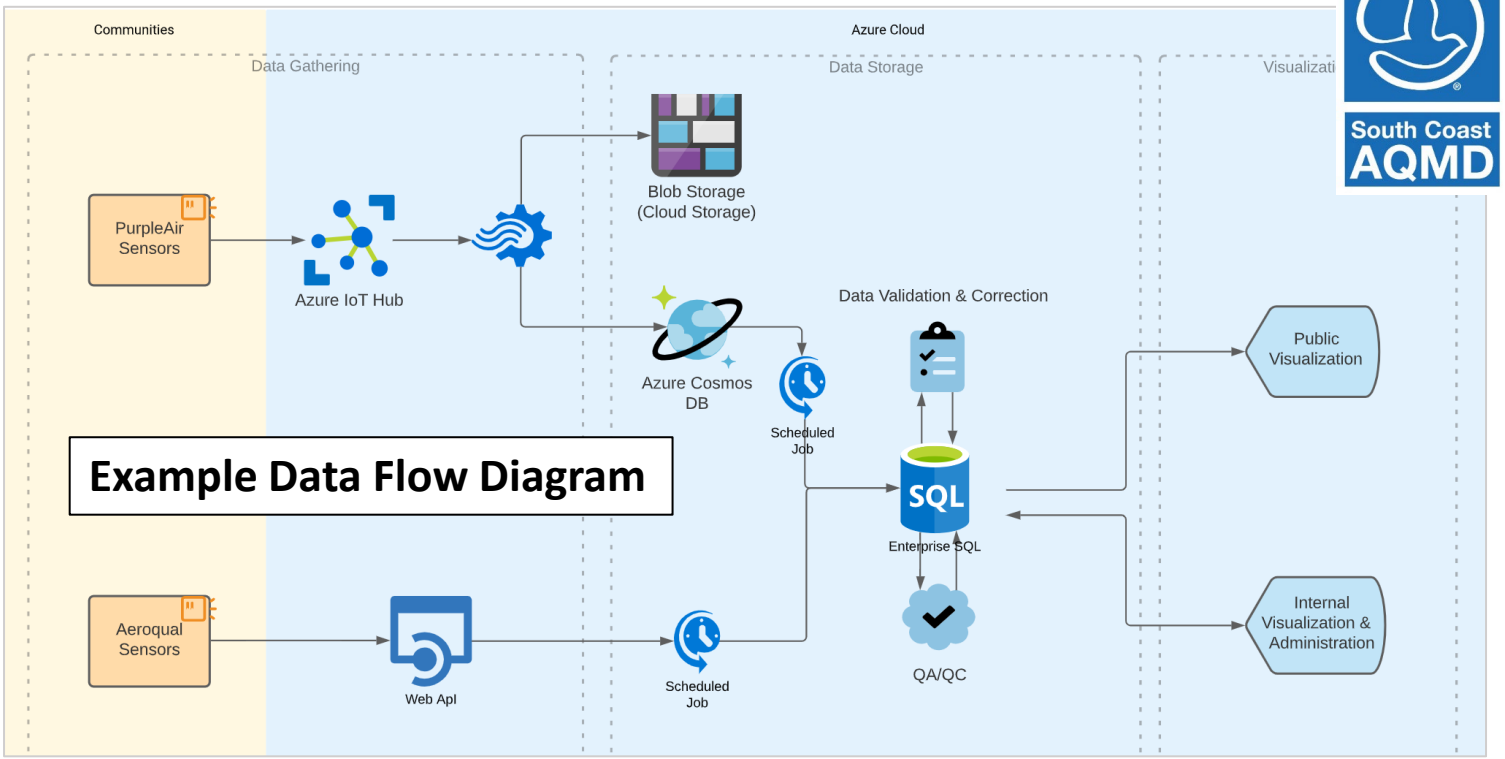
This solution harmonizes different types of air monitoring data, collected using different types of instrumentation, offering a single platform where district staff and the public can more easily access and analyze the data using advanced tools

Back-end System Development – Cost Considerations

Cloud Platforms typically utilize **pay-as-you-go** business models, meaning cost is incurred when data is **ingested**, data is **processed**, and data is **stored** in a database

- Design must consider both performance requirements and cost considerations
- Resources should be turned off or scaled down when not in use, to minimize costs
- Infrastructure cost must account for three almost identical environments: developmental environment, stage/test environment, production environment
- Use short-term data storage (hot-storage) that is quickly available and long-term data storage (cold-storage) for older data that is not needed as much
- Regularly reevaluate processes and usage to optimize efficiency

Databases – Our Approach



Phase I – database for sensor data
(begin with a single database)

1. Program staff outline data access and transfer pathways
2. Contractor develops data flow diagram
3. Review/revise the flow diagram
4. Implement flow diagram resulting in sensor data being pushed or pulled into the cloud database

Phase II - Repeat Phase I procedures with other programs to bring in other types of data **(14 databases total)**

Other Recommendations

- Split the processing and final database resources to ensure adequate resources available on VM to provide data quickly to UI platform for data visualizations
- Sensor data access via vendor hosted APIs may be a cheaper option than accessing sensor data directly from IoT device through Azure IoT hub

Data Processing and Quality Checks

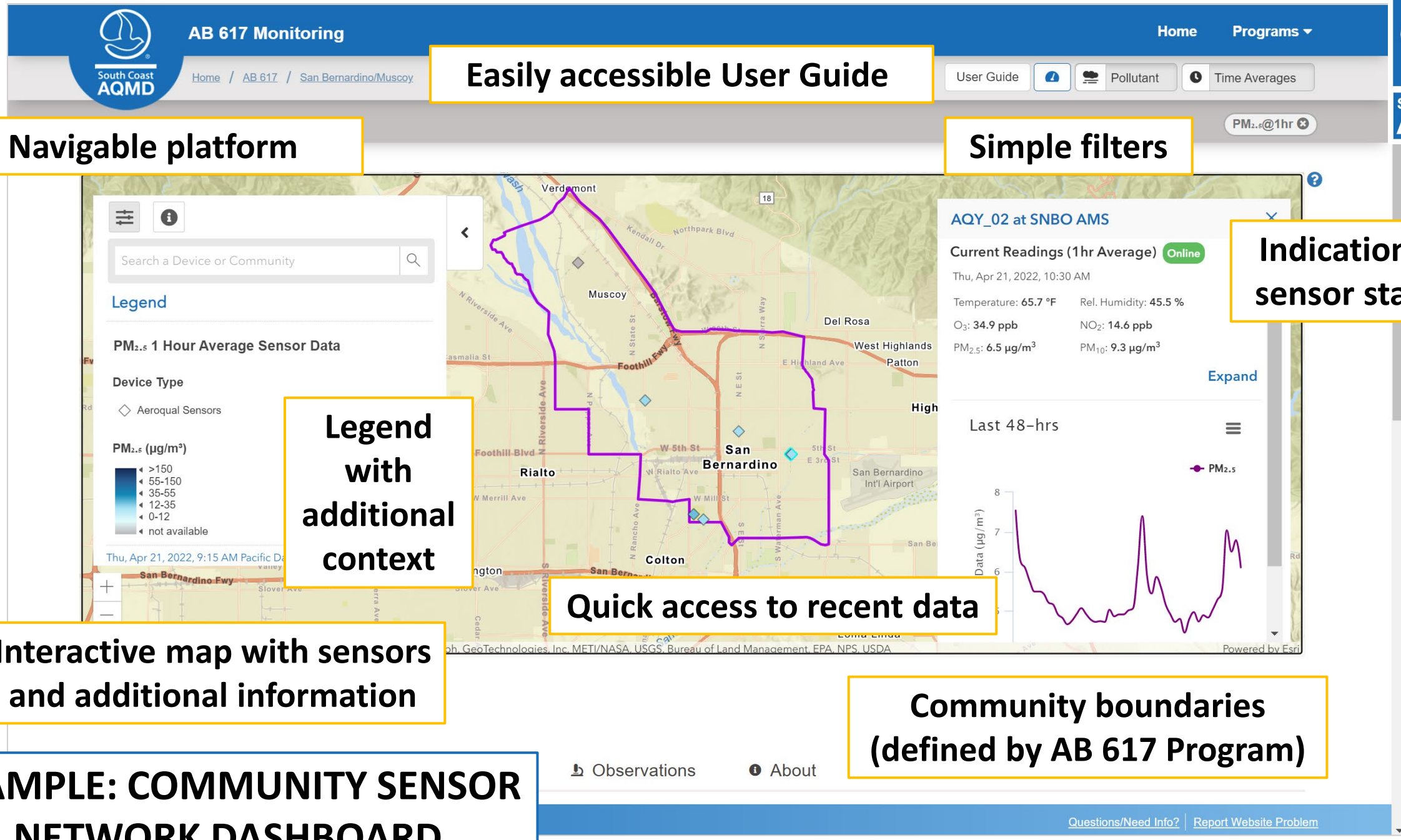
- Data processing includes developing and applying quality checks (QC), time averaging the data, and/or modifying data in some other way
- To ensure data is readily accessible for the data visualizations, **raw and processed data are both saved in separate tables**
- Recommendation, at the outset of the development
- Identify **a small dataset for quickly testing** and validating all QC rules
  - Identify **all the data that will need to be processed** (esp. if there is a substantial amount of historical data)
  - Identify the **ideal formats** for the final data, for example, specific time averages or time-matched data

Case Study

- Contractor may not be familiar with typical approaches to processing sensor or air quality data
- One QC rule was developed to identify data that was “flatlining”, however, it was not clarified how missing values (NaN’s or Null values) should be accounted for
- Thus, an earlier version of the rule resulted in data passing the QC rule that should not have
- Our team worked with the contractor to revise the rule during User Acceptance Testing

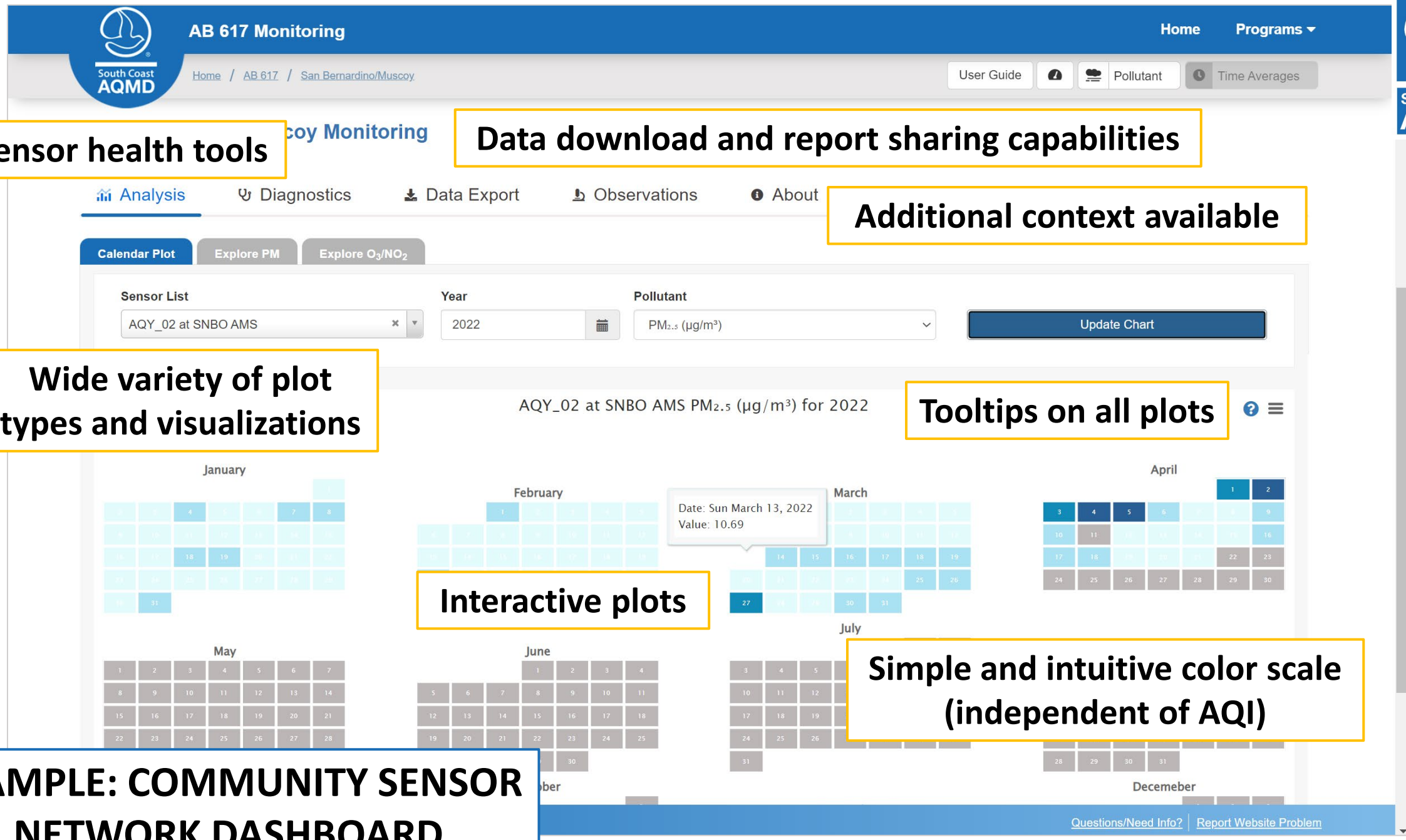
The Process of Developing Dashboards

- Early input from **a wide range of stakeholders** is key
  - Lessons learned WRT data communication and visualization for the public were shared by staff from different programs throughout planning
- Involve **both ground staff as well as management** in the planning stages
  - Management can provide appropriate framing for the tools, and ground staff can define key details
- Provide Contractor with examples such as **flow charts and mock-ups**
  - Our team used PowerPoint to develop wireframes
- **Wireframes** (requirements) → **Mock-ups** (organization, look & feel) → **Development** (functional dashboards)
  - Develop consistent workflow for identifying and tracking changes requested of the Contractor to ensure no requested changes are missed, iterative process
- Leverage **different technologies for different purposes**
  - For example, High Charts for standard (contractor-developed) dashboards, RStudio Team for custom/specialized dashboards and visualizations (staff-developed)
- Form **review groups**
  - External (within the Agency and outside) non-air quality expert (e.g., legal, public affairs, media, communication, social sciences background) review group



EXAMPLE: COMMUNITY SENSOR NETWORK DASHBOARD

https://aqspec.aqmd.gov/public/community/1002



EXAMPLE: COMMUNITY SENSOR NETWORK DASHBOARD

https://aqspec.aqmd.gov/public/community/1002

Using this Solution to Communicate and Share Data with the Public

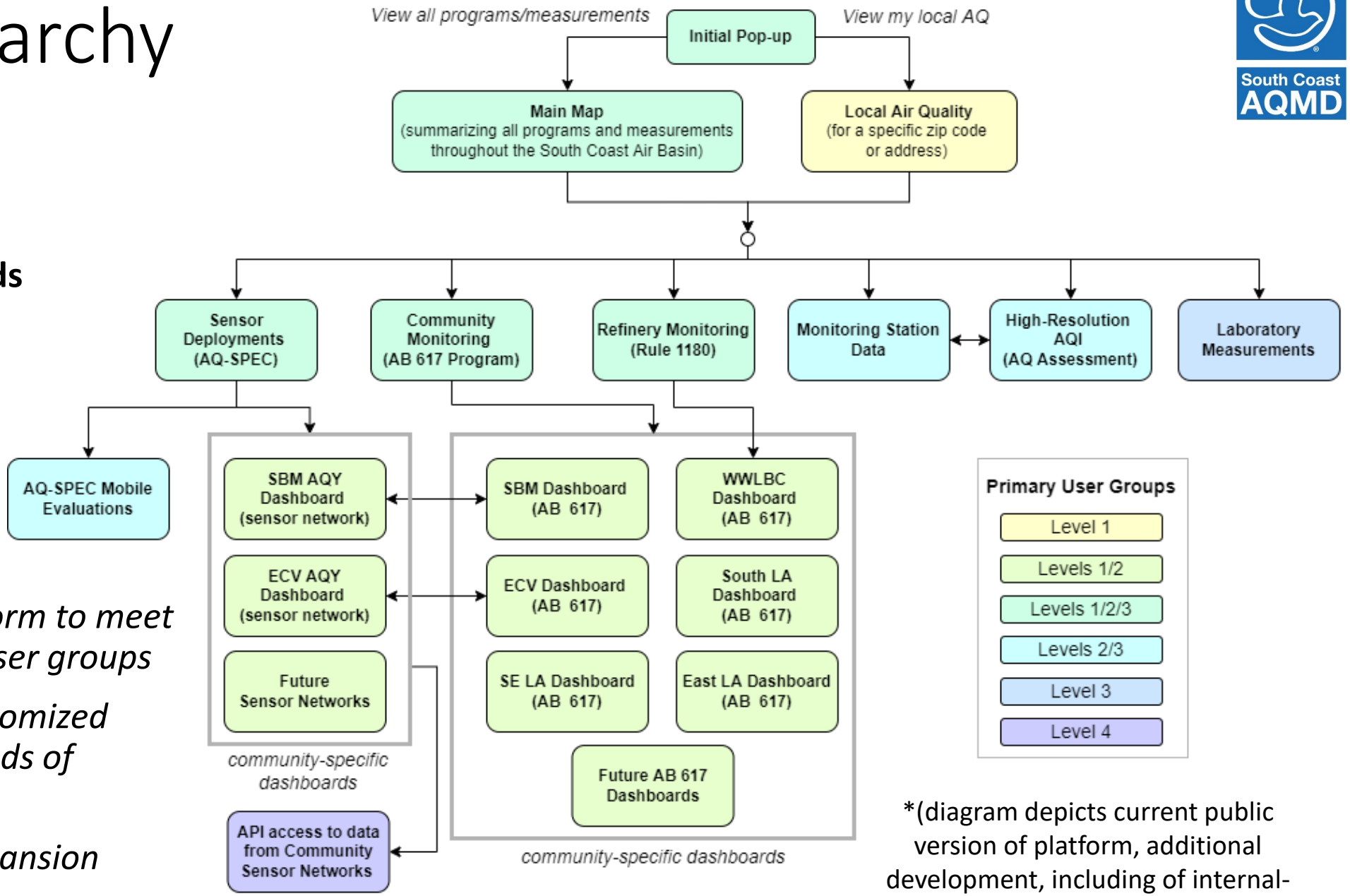
- Share **"beta" version** with the public (or appropriate end users) to collect feedback
- Supplemental resources that our team found useful to provide alongside the Dashboards include:
  - **User Guides** with clearly labeled screenshots, explanations of different plot types, and tips for users
  - **Video tutorials**, it is generally recommended to keep these short (e.g., less than 2 minutes each)
  - **Surveys** to collect feedback from users (in particular, from the beta user group)
  - **Tutorials** built into the final dashboards are also recommended (esp. for first-time users)
  - **Tooltips** built into every plot to provide background and context for the results displayed
- If your end users include users from a wide range of backgrounds and technical abilities, it may be helpful to define **expected user groups**, for example
  - Level 1 Users: members of the general public
  - Level 2 Users: citizen/community scientists, environmental/public health non-profits
  - Level 3 Users: researchers, students, community group leads / air quality experts, agency personnel (internal and external to South Coast AQMD)
  - Level 4 Users: API access users (e.g., students, researchers, app developers)

Overall Hierarchy

Global Dashboards

Program Level Dashboards

Community-specific and Specialized Dashboards



- *Structure enables platform to meet the needs of different user groups*
- *Dashboards can be customized to meet the specific needs of individual programs*
- *Structure allows for expansion and scalability*

*(diagram depicts current public version of platform, additional development, including of internal-only dashboards is ongoing)

Acknowledgements

- Co-authors: Brandon Feenstra, Ashley Collier-Oxandale, Aneeta Dev, Brian Roche, Xin Chen, Ron Moskowitz, Jason Low, Vasileios Papapostolou, Andrea Polidori
- Additional South Coast AQMD staff in AQ-SPEC, Information Management (IM), and our internal contractors (Varsun eTechnologies)
- Additional collaborators in the AB 617 Program, Rule 1180 Program, Air Monitoring, Air Quality Assessment, Laboratory, and Quality Assurance

Contact AQ-SPEC

www.aqmd.gov/aq-spec
info.aq-spec@aqmd.gov

Contact the Speaker

Vasileios Papapostolou, Sc.D.
Program Supervisor, AQ-SPEC
vpapapostolou@aqmd.gov
(909) 396-2254