E-Enterprise
IoT/Sensor Sprint
Summary Report
V 1.3

August 27, 2020

Table of Contents

Executive Summary

The E-Enterprise Digital Strategy conducted the IOT and Sensor sprint to obtain an understanding of the current status of Federal and State sensor activities. This paper outlines the key findings from the interviews, including the challenges they identified. EPA, States and Tribes are doing a lot of work in the sensor space, and the team started their work by reviewing existing information (see Appendix A for examples of existing sensor research and projects). The team then interviewed Federal, State and Tribal Representatives, including:

- Environmental Protection Agency: Dwane Young (Office of Water (OW)), Phil Dickerson (Office of Air and Radiation (OAR)), Andrea Clement (Office of Research and Development (ORD)), Leah Ettema (Water, Region 3), Dave Smith (OMS), Ethan McMahon (OMS)
- Tribes: Kari Hedin (Fond du Lac Bank of Superior Chippewa), Brandy Toft (Ojibwe Nation)
- State: Andy Putnam (Colorado)

Key Findings

The following are the key findings and challenges identified in the interviews.

1. There is a need for more accessible data management, integration, analytical, and visualization platforms
    - EPA's Data Management Analytics Platform (DMAP) concept has the potential for becoming more than a pilot and could provide a key platform for collecting, analyzing, and making accessible sensor data from a wide variety of initiatives, but it needs further development and investment.
    - States and Tribes may want to share platforms or utilize EPA platforms to realize economies of scale. Some States and Tribes lack the resources to develop their own platforms. Investment, cost sharing and governance would need to be established to provide partners with greater access to visualization and analytical tools (e.g., geospatial, COTS such as Tableau, CliqSense)
2. There is a need for standards and tools:
    - **Metadata** is essential for data use, but metadata standards vary depending on subtypes of networks (e.g., surface water is different than particulates).
    - **QA/QC Tools and Approaches** are needed, including normalization tools to detect anomalies in large data sets. States and Tribes would benefit from an E-Enterprise approach vs. developing them independently (some States and Tribes do not have the resources to develop the tools and approaches independently).
    - **Catalogs and Aggregation (Mashup) Tools** are needed as indices of what data are being collected from where and by whom so that aggregation and querying can occur.

3. Guidance is needed to correlate how sensor data can be used.
    - There needs to be guidance and information on what each set of sensor-collected data can do to assist in environmental decision making. This is a function of the quality of the sensors, their readings, collected metadata and analytical techniques.
4. There is a need for subject matter expert (SME) assistance to partners who want to use sensors but do not have the expertise. Partners would benefit from a clearinghouse of architecture examples and a central repository of best practices.
5. There is a need for capacity building among the regulators to process large amounts of sensor data and the corresponding “demand” once it is collected.
6. E-Enterprise needs to support moving sensor programs from “pilot” deployments to more systematic programs to support environmental decision making.

Recommendations

Based on the findings, the following are the recommendations for moving forward.

1. Evaluate the correlation between the notional sensor architecture and the draft E-Enterprise target architecture that the Digital Strategy Team is developing to ensure consistency of concepts.
2. Build on existing available resources and identify further areas where EPA and E-Enterprise partners need to develop standards or guidance on when and how to use IoT sensor data including communications of such data with the public (whether regulatory agency or publicly generated).

1 Sprint Purpose and Scope

The E-Enterprise Digital Strategy Team's Internet of Things (IoT) sensors sprint goal was to obtain a better understanding of federal and state and tribal sensor activities. Through this sprint, the team identified common needs and issues and developed an architecture of related technologies that offer significant opportunities for faster, better, cheaper and less burdensome environmental data gathering and analysis. In addition, the team identified new, different and innovative sensor capabilities.

1.1 Sprint Activities

In the course of this sprint, the team conducted the following activities:

1. Reviewed a large amount of existing information and resources on sensors from ongoing EPA, State, and Tribal initiatives. See Appendix A for examples of existing sensor initiatives.
2. Interviewed EPA, State, and Tribal staff on a range of sensor applications in the air (e.g., particulates and hazardous air pollutants) and surface water environmental areas. The following people were interviewed as part of this sprint:
    - Dwane Young, EPA Office of Water (OW) (continuous monitoring data sharing)
    - Phil Dickerson, EPA Office of Air (OAR) (particulate sensors)
    - Andrea Clement, Ingrid George, Rachelle Duvall, EPA Office of Research and Development (ORD) (particulate and hazardous air pollutant sensors)
    - Leah Ettema, EPA , Region 3 OW (water quality sensors)
    - Kari Hedin, Fond du Lac Band of Lake Superior Chippewa (water quality sensors)
    - Brandy Toft, Leech Lake Band of Ojibwe (air monitoring sensors)
    - Dave Smith, EPA Office of Mission Support-Environmental Information (OMS-EI) (data storage and analysis)
    - Andy Putnam, State of Colorado (multiple air monitoring programs)
    - Ethan McMahon, OMS-EI (citizen science program)
3. Established a common set of architectural components used by sensor applications to understand how sensor data are collected, integrated, stored, analyzed and visualized;
4. Identified related initiatives (e.g., EPA's Data Management Analytics Platform (DMAP)) that could be important components of the architecture to support sensor data collection and management;
5. Evaluated similar examples relevant to environmental missions from another Federal agency, the Department of Homeland Security (DHS).
6. Identified issues and challenges, such as data quality, use of IoT data, and technology limitations (including sourcing considerations), that affect IoT development and data integrity/assurance.

1.2 Sprint Outcomes

This initial sprint summary provides the outcomes from the sprint, including:

- Basic components of an IoT-Based Sensor Network (Section 2)
- Use cases that demonstrate how sensor data are currently being collected and managed (Section 3)
- Cross-cutting challenges and gaps in sensor architectures (Section 4)
- A Notional Interim Target Architecture (Section 5)
- Recommended next steps for future consideration and action (Section 6)
- List of some of existing resources and initiatives underway with published tools and guidance (Appendix A)
- Three use cases on sensor architectures (Appendices B-D)

2 Basic Components of an IoT-Based Sensor Network

The ideal goal of a sensor architecture is to store, process, and analyze large volumes of data quickly. This includes the ability to collect data from multiple sensor sources (sensor networks) in real time, share data in numerous ways among co-regulators and the public, and enhance the ability to use these data to make faster and more comprehensive environmental decisions. Sensor network architecture needs to support metadata that includes information about sources and inform an assessment of quality. They use open-standard technologies, APIs and open data techniques. Sensor networks also must have a governance model that addresses data ownership, management and sharing permissions.

Any device connected to the internet can be considered part of IoT; however, in the case of environmental sensing, this discussion is limited to the small, relatively inexpensive devices with which EPA, states and tribes are currently experimenting. The assumptions for this sprint include:

- IoT sensors are typically physically smaller, lighter and significantly cheaper than conventional devices, lending them to use cases that need high densities, low costs and easy deployability.
- IoT sensors are invariably COTS products, developed and marketed for multiple, both industrial and other purposes. The pace of development is intense. EPA and the environmental community can exert only minor influence on the direction of the technology and product mix.
- IoT sensors currently do not meet EPA reference accuracy standards and are therefore not suitable for most regulatory-level work. It may be many years, if ever, before these sensors can begin to replace traditional data collection methods; however, the data from the sensors may inform regulatory evaluations and considerations where allowed by law or regulation. Data from IoT sensors can be used to identify trends and elevated pollution levels as well as to characterize ecosystems.
- IoT sensors operate autonomously with QA/QC largely limited to inspections of operating conditions or what can be deduced from the data stream since laboratories are not involved and sites are generally unsupervised during operation.
- Sensors are typically connected wirelessly, usually with low power and uncertain reliability.

2.1 Continuous Monitoring vs. IOT Sensing

Continuous monitoring is defined as continuous or regularly-repeated monitoring transmissions. For example, an air or water sensor might transmit an observed value continuously, or at intervals every day, every hour, or more frequently. IoT sensing generally produces continuous data sets, but it can also produce event-driven reporting, such as when a pollutant is detected. IoT sensor platforms, whether they are event-driven or continuous, may also transmit a regular ‘heartbeat’ signal to confirm to the network that they are operational. The network may also query the sensor and/or perform service (e.g., firmware updates, recalibration or command/control).

The following are potential concerns with IoT continuous monitoring:

- **The resulting data sets are relatively large.** Those interviewed for this sprint raised the large data sets as a primary concern as the quantity of data generated by continuous monitoring is typically far larger than that produced by traditional regulatory networks. Storage, processing, metadata, QA/QC parameters and analysis are significantly different from the usual data collection methods; however, because the data streams are typically “thin” (low volume), the total amount of data generated is not unmanageable. Interviewees reported potential storage requirements for their applications in gigabytes or terabytes, but not larger, and storage costs are decreasing.
- **Data standards and metadata management must be greatly streamlined.** Because wireless transmissions are typically low-power and low-bandwidth, metadata must be kept to a minimum but be adequate to allow the data’s use. For example, descriptors need to be applied to the overall transmission rather than individual readings. IoT sensing therefore may require dedicated standards custom-developed for this data environment.
- **Archiving will be difficult.** Federal records management implies the need to archive potentially large or compressed data sets. Although compression is possible (e.g., storage in binary form), it may not always be necessary to archive the original granular readings. Data summarization may be acceptable or storing only anomalous data. Archiving is thus both an IT and a policy issue, both of which will affect costs.
- **Comprehensive access and data sharing is, at least at present, a major problem.** Today’s experimental applications are distributed (i.e., data are stored in multiple locations). Even where shared software is available (as it is for EPA’s Office of Water’s Continuous Monitoring Strategy), archiving strategies are not yet defined.

2.2 Conceptual Flow of Sensor Data

Figure 1 shows the conceptual flow of sensor data from its point of collection through evaluation and analysis.

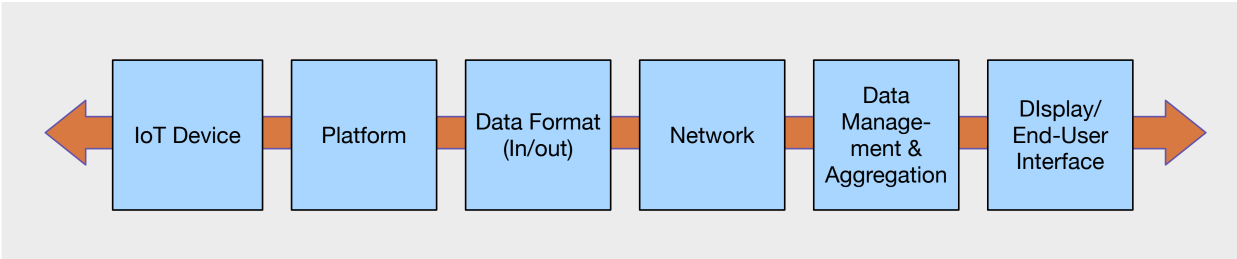


Figure 1: Sensor Network Basic Components

The following are definitions for the components outlined in the figure above.

- **IoT Device.** Some form of internet-connected sensor, generally assumed to be small, relatively inexpensive, but robust enough for environmental deployment (temperature and moisture resistant, suitable for underwater or adverse weather conditions, etc.). Operational field life is a major issue.
- **Platform.** (optional) The platform on which the device is mounted. Considered part of the architecture if (1) the platform itself is an IoT device, such as a unmanned aerial vehicle (UAV) or (2) if the platform serves as a distributed data manager (i.e., for batching, QA/QC, averaging or otherwise condensing or managing the data to be transmitted within the network, etc.).
- **Data Format.** Input/output data formats apply to the IoT endpoint device as well as any enabling platform. These formats are often customized to the needs of continuous monitoring using lightweight data protocols.
- **Network.** The (typically wireless) network transmitting the internet protocol signals. These can include options such as local area networks (LANs), WI-FI, radio (including cellular) and satellite transmissions. IoT networks may often operate intermittently, with variable reliability and/or at low bandwidth and signal strength.
- **Data Management and Aggregation Platform.** The data integration, processing and management platform (including storage) needs to be customized for continuous monitoring uses.
- **Display/User Interface.** The platform from which data can be displayed or accessed. These are most often existing COTS tools vs. unique software programs for sensors specifically.

For the purposes of an E-Enterprise sensor architecture, the focus of development is on all but the last elements (Display and End-User Interface). There were no data storage, analysis and visualization needs encountered that could not be handled by existing COTS tools (e.g., Tableau, ClikSense).

3 Current IoT Sensor Use Cases

The IOT sensing applications in this sprint covered both a water application (Integrated Water Network, http://54.210.62.171) and air application Purple Air Particulate Sensors (https://www2.purpleair.com). These two use cases were selected because they have existing pilot programs underway. The third use case is an example state to illustrate the type of future sensor program needs that will likely exist. Tribal and EPA Regional interviews helped identify issues that programs have faced with implementing sensor programs.

The programs interviewed are experimental, evaluating the quality and reliability of potential results and demonstrating architectural proofs of concept. The programs use commercial sensors and their data management involves a mix of custom data management software and proprietary software and storage available from vendors.

There are two models for sensor architectures:

- Centralized – All data from sensors are transmitted directly to a central platform, and analysis, access and business functions are called from this platform (though analysis can also be conducted remotely);
- Distributed – Data may be collected and stored in one location, then cataloged and called when needed using APIs for specific purposes. Tools for analysis and visualization may also be in the same or a different location than where storage occurs. This is the most prevalent model described from those interviewed. Issues of archiving to meet Federal records requirements will need to be addressed as a part of this model.

Table 1 below summarizes the three use cases identified from the interviews.

Use Case C	Type of Network	Purpose	Architecture and Example Gaps
EPA Office of Water Integrated Water Network (pilot deployment)	Distributed (storage of data at point of collection appliance and catalog; tools to register and access data from storage as needed)	Sharing of multi-purpose data developed by multiple agencies within a single watershed. Uses can be for both regulatory and non-regulatory purposes depending on legal and regulatory requirements.	Architecture is evolving from a decentralized system of agency-owned and operated data stores to a common system of stores operated within DMAP on AWS. A separate catalog, probably to be housed within WQX, will provide access and data sharing. Archiving might be done in collaboration with NOAA. **Gap:** Adopting a new common data standard (SensorThings) that allows the scale up of the volume of data; finish building the appliance and catalog and get new uptake on users for the platform.

Use Case C	Type of Network	Purpose	Architecture and Example Gaps
EPA Office of Air and ORD Purple Air Particulate Sensors to Augment AIRNOW Health Alerts	Distributed collection; commercial vendor storage using Purple Air cloud.	Augmenting limited AIRNOW particulate data with denser network of IoT sensors to detect smoke associated with wildfires. Collect particulate data from wide variety of private sensors; correlate with regulatory reference values to provide daily health and safety alerts to the public.	Purple Air sensor data is input to a centralized storage platform operated by the commercial vendor. It is then available to any party through APIs. AIRNOW may bin extracts temporally and spatially in DMAP to align with AIRNOW's 12-hour averaging. **Gaps:** (1) Concern about the quality of Purple Air sensor data; ORD is developing correction factors to compensate for temperature and humidity variation across regions; (2) need for archiving; OAQPS is discussing archiving summarized data in binary form using tools developed by NASA; EPA and Federal records officers need to be consulted.
State Hazardous Air Pollutant Analytical Support (example future use case)	Distributed (State storage, potential for shared analytic needs)	Example State is required to collect and use large volume of sensor data from private parties to monitor hazardous air pollutants under a State law.	Example State may implement a cloud storage and intake network but may want to use shared EPA tools to conduct analysis and visualization of data to avoid individual license purchase costs. **Gap:** How to access tools in a shared environment (authentication, access, and borrowing of shared tool capacity); analysis of software licensing requirements will be part of this evaluation.

4 Main Challenges

IoT sensors are evolving so rapidly that new challenges for their full exploitation will continue to emerge. This section summarizes the challenges that are apparent based on interviews at the federal, state and local levels. Some of them are unique to IoT sensors; others are similar to other types of input devices, but sufficiently different in degree that they require custom responses.

4.1 Data Storage and Integration Platforms

In the three use cases identified for this sprint, distributed models are used with data storage located separately from where the data analysis and visualization occur. For the OW Continuous Monitoring Strategy, the plan is to centralize data appliances within DMAP and provide common tools and common access through a catalog, thus enabling the sharing of data among contributing organizations for watershed-scale decision-making. For the Purple Air sensors, EPA's Office of Air Quality Planning and Standards (OAQPS) is discussing staging data from separate Purple Air data sets into DMAP for processing and integration and then flowing that data into AIRNOW.

There may be a number of reasons why shared environments make sense for IoT monitoring data, including:

- **Financial efficiency:** As small-scale experiments scale up to practical implementation, the large volumes of data collected mean that cloud storage, shared tools and shared licenses could lower everyone's costs. Many practitioners are operating on very small R&D budgets, and therefore systems like DMAP could expand their opportunities substantially.
- **Large scale analytics:** The volume of data available offers opportunities for data reuse and analysis across data sets. As has been seen using DMAP to analyze the ACE data set, the tools available in systems like DMAP may enable deeper analysis than the traditional end-use environments. Rich data sets developed by continuous monitoring offer new analytic opportunities not anticipated in traditional data management architectures. They can best be analyzed if housed together.
- **Consistent Full Service/Access Reliability:** Centralized systems offer the full range of services to customers, including ingestion, QA/QC, standards applications, analytics and visualization, while avoiding dependence on individual system maintenance by separate organizations.
- **Standards Convergence:** Users may be able to more quickly evolve common data formats, metadata standards, QA/QC tools, shared analytics, archiving technology and policy and governance structures.
- **QA/QC Evolution:** As sensors develop, it is possible that longitudinal analysis across data sets in shared environments might produce insights into such phenomena as sensor useful life, long-term data drift patterns, insight into reliability and accuracy performance by vendor and other conclusions valuable to EPA and its stakeholders, both public and private.
- **Integration with Conventional Data Sets:** DMAP is not conceived primarily for IoT-based monitoring; it was designed as a general cloud-based architecture for heterogeneous data management. Experience gained in such applications of the Purple Air → AIRNOW might be valuable in learning how to integrate new streams of monitoring information with traditional data sets.

The notional IoT Data Sharing Architecture presented in the last section of this report suggests an environment that combines the planned features of a centralized platform (e.g., DMAP) with a Catalog function like that under development for the OW Continuous Monitoring Strategy, all within a holistic data management framework that integrates IoT and traditional data sets.

4.2 Visualization and Analytic Tools

Visualization and analytics were assumed to be a likely challenge when this sprint was scoped, but as noted above, conventional COTS tools seem adequate for processing sensor data sets. There were requests to explore sharing of these tools so that each individual State, Tribe or EPA Region did not have to establish their own instance due to costs and lack of expertise in standing these up. Methods for sharing costs among users will be needed as well as other governance processes.

4.3 Metadata Standards

The IoT sensing market is large and very diverse. Environmental monitoring is only one business application. Vendors are moving quickly to serve many markets with diverse needs for size, cost, product lifecycles and measurement accuracy. Federal agencies are customers and onlookers—they invest in sensor R&D only in the most specialized markets. The current marketplace was described in our interviews as the 'wild west.'

Because vendors have every incentive to push proprietary rather than standards-based products, metadata standards are difficult to develop and enforce. IoT sensing has unique constraints that call for customized standards. IoT data structures must support data streams, accommodate the limited accuracy and discrimination of many sensors, be tolerant of interruptions, support sensor side computations (such as aggregations, summaries and averages) and be able to accommodate batch transmissions. The challenge is to work broadly across organizations to evolve metadata management within this very heterogeneous marketplace. An equal challenge is to develop processes that result in stakeholders agreeing on the standards that will be used with sensor technologies.

Metadata Standards Example

For water monitoring, the OW Continuous Monitoring Strategy (2016) settled on Sensor Markup language (SensorML) and Water Markup Language (WaterML); as of today, OW is additionally looking at the SensorThings standard, which is similar to SensorML but more attuned to the limitations of IoT. These standards are available from different international groups that EPA has been evaluating to determine the best standard for the EPA sensor program to adopt based on scalability and flexibility.

4.4 QA/QC Tool Analytical Platform

Conventional monitoring methods implement QA/QC at several points in the reporting chain. For example, lab sample analysis requires insertion of duplicates, spikes and blanks in the same workflow as the environmental samples. Chain of custody documentation may be another requirement.

To gain adequate QA/QC for data flows from IoT sensors, new methods may be required. For example, networks can be designed with redundant sensors or with co-location with reference monitors. Detailed metadata may also provide critical information about the QA/QC of networks. These methods will rely on regular communications between sensor manufacturers and network operators to develop appropriate quality approaches and developing and sharing lessons learned.

Individual sensor checks might be useful to weed out bad sensors prior to deployment, but the main type of QA/QC available is through analysis of the data stream itself for detection of range errors, stuck/frozen sensors, sensor drift over time and other phenomena. Users must understand the factory settings related to calibration requirements and sensitivity standards that are unique to each sensor type.

This implies that QA/QC may need to be integrated into the architecture for each sensor network type, and potentially for specific products and product deployments. Customized QA/QC tools may need to be embedded at the data ingestion, data management/analytics and even the data visualization layers of an architecture.

QA/QC Example:

ORD is close to releasing correction factors for adjusting readings from the Purple Air sensors for temperature and humidity variations. The hope is that this will compensate systematically for observed variation of Purple Air readings in different settings. On the other hand, other users do not use these sensors because they believe the variations are random and sensor specific—not correctable based on ambient conditions. The challenge is not just to develop QA/QC tools for known applications, but to define the actual nature and scope of QA/QC required along the full data development chain.

4.5 Distributed Appliances and Data Sets Catalog

Experimentation with IoT sensing is being done by independent groups at different levels of government, with different use cases, hardware and data management systems. This may continue, especially since with current technology most use cases would involve short-term applications rather than long-term installations supporting national data sets.

In the absence of a large-scale fully deployed sharing environment, data storage and management are likely to remain distributed and institutionally fragmented at least for the foreseeable future. Publicity about, and access to, these data sets requires some sort of catalog with a link to data points (sensors) that are available and appropriate access controls to each set.

Exceptions: Some sensor vendors, notably Purple Air, are maintaining data sharing warehouses of their own. Commercial data storage is an attractive idea because of costs and potential access to other users' data sets, but it may not be a desirable solution for the long term. Commercial data storage seems likely to reinforce proprietary specialization and discourage convergence on common standards. In addition, costs for commercial data storage are not predictable or controllable, and data sets in this model would not be organized by use case, application or program use, but arbitrarily by vendor.

Data Catalog Example:

OW's Continuous Monitoring Data Strategy recommends the development of a data catalog (exposed directly or indirectly through OW's Water Quality Exchange (WQX) that points back to service endpoints exposed by data appliances tied to the contributing organizations. OW is collaborating with the Consortium of Universities for the Advancement of Hydrologic Science (CUASHI) to develop a pilot catalog (Discovery Tool) that can link to existing appliances. Figure 1 provides an overview of the proposed OW Continuous Monitoring Data Sharing Architecture.

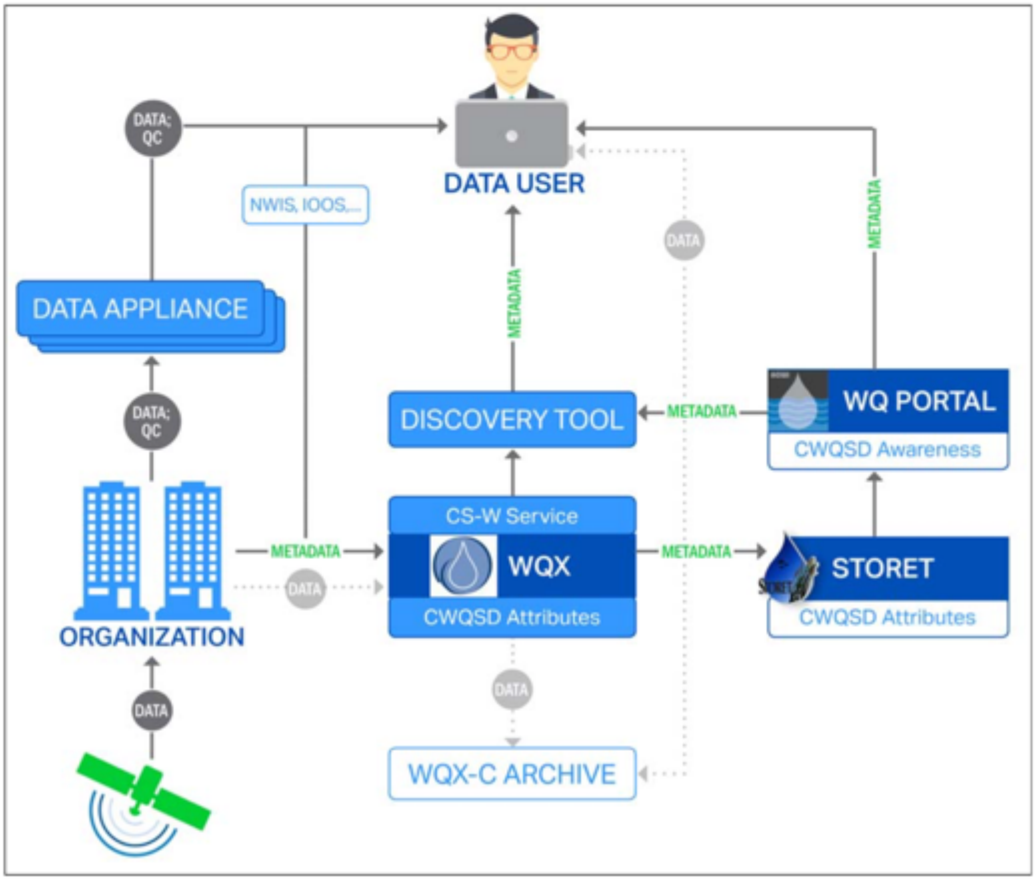


Figure 1: The Proposed OW Continuous Monitoring Data Sharing Architecture

OW's sensor architecture distributes copies of a common data appliance to participating state, local and tribal agencies. A central Catalog (discovery tool) housed within the Water Quality Exchange (WQX) is under development by the CUASHI to provide data sharing and user access to the distributed data sets.

OW recently received authorization to operate (ATO) to implement the architecture on EPA's Data Management and Analytics Platform (DMAP) in the Amazon Web Services (AWS) cloud, allowing central access to the Data Appliance. The Catalog will still be used to provide visibility into the separate data sets, and presumably will still be located within WQX.

The recommended system architecture calls for the standards-based implementation of a centralized catalog, data appliances and archive as shown above, where (1) site and deployment metadata are submitted to an extended WQX serving as a catalog; (2) data are available through service endpoints exposed by data appliances tied to the organization; (3) data users discover data of interest with a discovery tool by querying metadata in the catalog, and then retrieve the data from the data appliances; and (4) data are archived for backup, redundancy and/or regulatory reasons in a modified WQX.

4.6 Correlating Sensor Technology and IoT Methods with Business Purposes

Most of today's IoT sensors are far less accurate and consistent than EPA's reference monitoring methods. How much less depends on the parameter measured. For simple parameters like temperature, they may be close to reference levels. For others, like measuring ambient levels of hazardous air pollutants, they may be useful for detection at best.

Sensor accuracy limits the business purposes for which the current sensor generation is suitable. None of the interviewees anticipated using IoT sensors for regulatory-level analysis or policy purposes in the foreseeable future. They did, however, see opportunities to use IoT sensing to supplement other systems or be used where reference-level accuracy is not necessary. Examples types of decision making include source detection, trends analysis, public awareness, public safety alerts and warnings, research, citizen science and incident response/responder safety. Appropriate business purposes may often involve limited deployments over set periods of time, rather than permanent installations reporting continuously year over year.

Sensor accuracy has implications for data management as well. Reporting methods should fit business purposes, including considerations for reducing data volume (not reporting more data than is necessary for the purpose), metadata specifications and data retention/archiving strategies

Air Now Example

The OAR AIRNOW system wants to provide local-level alerts and warnings of wildfires, mainly in the western states. They are working on ways to extract smoke particulate signatures from background particulates using inexpensive Purple Air sensors. This would enable them to field the necessary number of sensors to provide adequate reporting granularity in high-risk areas, or perhaps to leverage Purple Air data developed by other agencies and users.

4.7 Clearinghouses, Best Practices, Publicity and Other Non-Tech Issues

The sensor knowledge uncovered in this sprint from EPA, States, and Tribes demonstrates how compartmentalized and independent current sensor initiatives are. Managers at the higher levels may be generally aware of activity in other programs and organizations, but individual researchers and mission staff were less aware and rely on ad hoc or informal networks to discover what colleagues are doing. In the Department of Homeland Security's (DHS) Science and Technology (S&T) example below, there seems to be no communication between EPA and DHS S&T, even though OAQPS and the S&T are working in very similar areas for air sensing.

E-Enterprise might evaluate various options for setting up clearinghouses, supporting workshops and conferences, and publicizing best practices and data sharing catalogs. This relatively early stage of IoT development seems the ideal time to encourage cross-fertilization and accelerate interactions.

Non-Technical Challenge Example

EPA and DHS's S&T are working in two IoT areas independently. (1) S&T has a new Wildland Urban Interface (WUI) program supporting the Federal Emergency Management Agency (FEMA). One project area is using IoT particulate sensors to provide local alerts and warnings. EPA's AIRNOW program is essentially doing the same thing. Both are trying to differentiate smoke signatures from background particulates. (2) S&T has worked for years on chemical and biological detectors as part of its Countering Weapons of Mass Destruction (CWMD) research. Their results and investments with a range of vendors should be useful to EPA's interest in fence line monitoring and Hazardous Air Pollutants applications, but there does not seem to be interaction between the agencies on this.

5 Notional Interim Target Architecture

A notional architecture for managing and using data from IoT sensors is shown below.

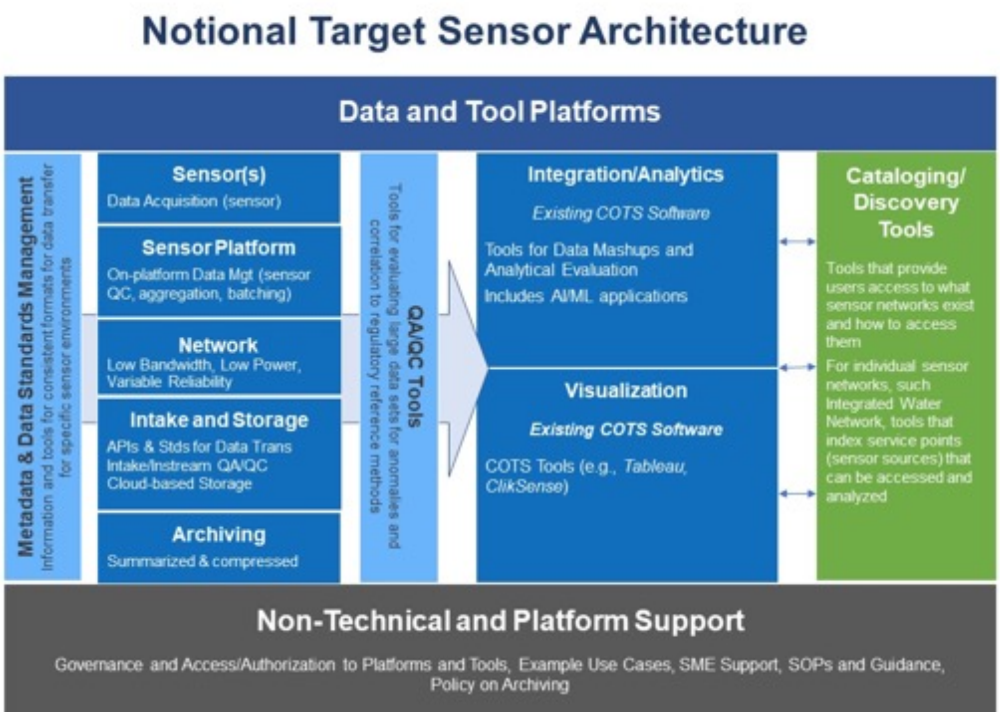


Figure 2: Notional Target Sensor Architecture

The left column shows the basic data ingestion and management track, with parallel functions for Metadata and Standards Management and QA/QC. Data developed on the left is then analyzed and visualized by the tools in the center column and accessed through discovery tools on the right.

- **Metadata and Data Standards Management:** Modernization of standards such as the Open Geospatial Consortium WaterML, SensorML and SensorThings directly affects most steps in data flows between sensor and eventual archiving
- **Sensors:** The commercial devices collecting the data.
- **Sensor Platforms:** (1) Optionally included in the architecture as IoT-controlled collection platforms such as UAVs (drones) and (2) necessarily included in the architecture if they perform distributed data management functions such as aggregation, summarization and QA/QC.
- **Network:** A consideration for the architecture for handling issues such as interrupted data flows, batch flows, mesh networking and satellite communications.
- **Intake and Storage:** The core function of the data management flow.
- **Archiving:** As yet unaddressed for IoT sensing but required for data management for federal programs.

- **QA/QC Tools:** Necessarily part of the architecture because the great majority of IoT sensor and data QA/QC is handled within the data stream itself.
- **Integration and Analytics:** No special requirements related to IoT sensors.
- **Visualization:** No special considerations related to IoT sensors.
- **Cataloging/Discovery Tools:** Necessary at this phase of development to allow stakeholders to discover each other's data and to tie distributed data stores together for analysis and reporting.
- **Non-Technical and Platform Support:** Supporting the whole IT enterprise are non-technical and platform support functions which may be handled separately from the architectural components. Non-technical issues also include the ability of a receiving agency to have the capacity to manage the influx of large amounts of data, particularly in the area of citizen science where many public entities may have an interest in both providing data as well as receiving and interpreting data collected. This issue was raised in several interviews and relates to the organizational and resource capacity of regulatory agencies.

Discovery tools may be an interim phase of this architecture. Once IoT data development is more regularized and integrated into related EPA and state data flows, the need for a separate catalog and accessing strategy may fade.

6 Recommendations and Further Considerations

Based on the findings in this paper, the following are the recommended next steps.

1. Evaluate the correlation between the notional sensor architecture and the draft E-Enterprise target architecture that the Digital Strategy Team is developing to ensure consistency of concepts.
2. Build on existing resources and programs and identify further areas where EPA and E-Enterprise partners need to develop standards or guidance on when and how to use IoT sensor data including communications of such data with the public (whether regulatory agency or publicly generated).

The following are potential topics of interest related to the research presented in this paper.

- Review additional challenges raised from the review group. As this work has progressed, there has been considerable discussion about where the boundaries lie between IT, sensor R&D and policy. As has been laid out in this paper, QA/QC and metadata and standards management appear to blur into network and sensor hardware issues. A discussion of E-Enterprise partner's highest priority in this area would be valuable.
- Conduct additional policy and technical research into Unmanned Aerial Systems (UAS) and other platforms being used (e.g., weather sensors, unmanned land or water/underwater vehicles, etc.). Sensors on drones was within the original scope of this sprint but has not yet been addressed.
- Involvement in ongoing pilot programs is a possibility. The continuation of OW's Continuous Monitoring Strategy, the evolution of the 2016 architecture into DMAP and current Open Geospatial Consortium standards (such as SensorThings) might be another avenue to pursue.

- Plans for a longer-term investment in DMAP, including perhaps its use as a staging level for AIRNOW, would explore the utility of IoT sensors in major ongoing EPA programs.
- Further review the California effort on Hazardous Air Pollutants fence line monitoring at refineries to evaluate that architecture model and add any relevant findings to this paper. This would directly involve state, local and industry participation in an IoT architecture.
- Evaluate the correlation between the notional sensor architecture and the target architecture being simultaneously developed to ensure consistency of concepts.
- Identify areas where EPA and E-Enterprise partners need to develop policies or guidance on when and how to use IoT sensor data including communications of such data with the public (whether regulatory agency or publicly generated).

Appendix A: Examples of Existing Sensor Resources and Programs

The following are examples of existing sensor initiatives' reports, guidance documents, tools and other resources.

E-Enterprise Resources

- Advanced Monitoring Technology: Opportunities and Challenges. A Path Forward for EPA States and Tribes (November 2016)

Air Sensor Resources and Programs

EPA Resources

- Air Sensor Toolbox
- Air Sensor Guidebook
- Air Quality Data Collected at Outdoor Monitors Across the US
- AirNow
- Air Aware 2019
- Deliberating Performance Targets for Air Quality Sensors Workshops
- Air Sensor Guidebook
- 2019 Regional/State/Tribal Innovation Projects
- Inter-Tribal Environmental Council's Clean Air Program
- Region 4 and 6: Collocated Air Sensor Shelters for Tribes and Citizen Science

State Resources

- California South Coast Air Quality Management District Air Quality Sensor Performance Evaluation Center (AQ-SPEC)
- Los Angeles Community Air Monitoring Network

Tribal Resources

- Coeur d'Alene Tribe UAS Program

Academic Resources

- European Commission Joint Research Centre Review of sensors for air quality monitoring
- Development and Implementation of a Platform for Public Information on Air Quality, Sensor Measurements, and Citizen Science
- Environmental Defense Fund Air Quality Maps

Water Sensor Resources and Programs

EPA Resources

- Water Quality Surveillance and Response
- Village Blue Research Project
- Low Cost Sensors for Real-time Continuous Water Quality Monitoring
- Region 5: Real-Time Sensors to Screen for Water Quality Impacts from Wastewater Discharges
- Monitoring and Remote Sensing
- A Review of Emerging Sensor Technologies for Facilitating Long-Term Ground Water Monitoring of Volatile Organic Compounds
- Online Water Quality Monitoring Resources
- Nutrient Sensor Action Challenge Stage II

Academic Work

- Sustainable Environmental Research: A system for monitoring water quality in a large aquatic area using wireless sensor network technology
- Low-Cost, Open Source Wireless Sensor Network for Real-Time, Scalable Groundwater Monitoring

USGS Resources

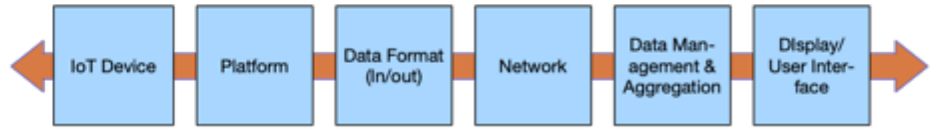
- USGS Water Data for the Nation
- Cloud-optimized USGS Time Series Data

Appendix B: IoT Architecture Model: Fond du Lac Band of Lake Superior Chippewa – Water Flow Sensing for Climate Change Impacts

The Fond Du Lac Bank of Lake Superior Chippewa has a pilot program in place to collect water level data from five streams and one lake. The following explains the components of the sensor architecture in place.

Spatial Hydro-Ecological Decision System (SHEDS) Flow Picture Explorer

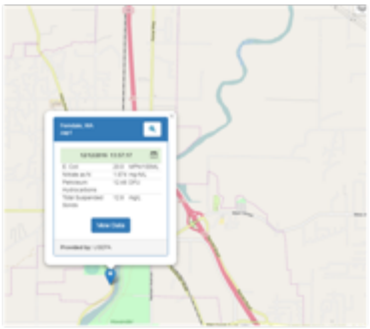
Hobo Logger and Outside staff gauge



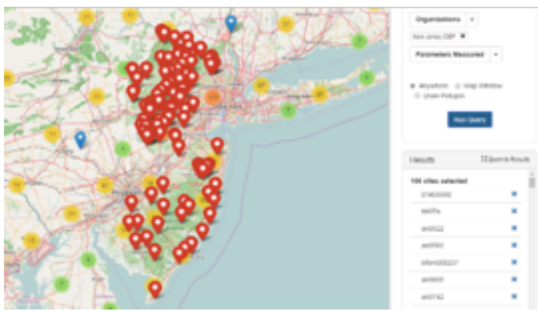
- **IoT Device.** U20 HoboWater Level Water Level Logger. Accurate to .1 feet. Five streams and one lake are monitored. In one location, USGS and R5 are running an innovation grant experiment with an installed camera and using ML application to test whether photos can provide same information as the logger. Camera in place for two seasons now and algorithms being developed for testing.
- **Platform.** Vented standpipe placed in the stream holding the logger, data read manually once each season using handheld device (sensor plugged into to device and data transferred). Paired with outside staff gauge.
- **Data Format (In/Out).** HoboWare data standard measuring water flow and temperature. Data format also accounts for barometric pressure to pair with water level readings. Outside staff gauge to calibrate water level. Rating curve generated to output water discharge.
- **Network.** Manual data reading once each season.
- **Data Mgt &Aggregation**. 30-minute readings taken throughout data collection. Downloaded data are stored on Tribe central server. Server also hosts a QA/QC tool developed by R5 for this sensor (manually applied) to account for drift, spikes and other anomalies. Two staff currently have data access; Tribe wants to make data more publicly available eventually. Data are transferred to WQX through the Ambient Water Monitoring System.
- **Displays/User Interface.** Data are downloaded to HoboWare software platform (proprietary platform part of sensor package), and graphing and display capabilities are available through this product; also transferred to Excel for additional analysis. Data are manually shared with EPA, other Federal agencies (FEMA for floodplain mapping, USGS) as requested.

Appendix C: IoT Architecture Model: Integrated Watersheds Network

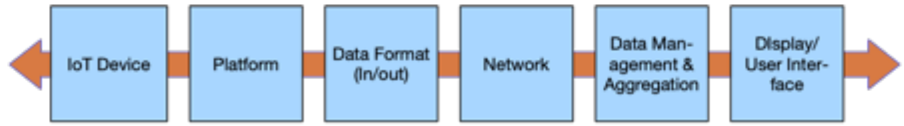
The Integrated Watersheds Network sensor pilot integrates sensor data from a wide variety of State and local partners as well as other Federal agencies collecting water quality data.



Sensor data can be reached through the IWN from registered partners



Example sensor locations available in IWN pilot



- **IoT Devices.** Multiple partners with existing water quality sensors across the country register their sensor platforms into a catalog that allows users of the network to reach out and gather data from selected sensors. Data can be used for a wide range of water monitoring and analysis purposes consistent with Clean Water Act data collection practices.
- **Platform.** Determined by each partner who is responsible for hosting the sensor. Purpose of the IWN is to reach to the data repository where the resulting data is stored and pull these data into a common format that is capable of being analyzed and displayed.
- **Data Format (In/Out).** IWN project adopted existing common metadata standard (OGCWater ML2) established by the Open Geospatial Consortium. Because of scaling issues, now migrating to SensorThings.
- **Network.** EPA pilot developed a common data appliance that enables partners to register their sensor data points, ingest data, and have it published in a common data standard. A search index and catalog of available network participants is underway by CUASHI. This catalog will incorporate data from any data appliance registered, USGS, and NOAA.
- **Data Mgt &Aggregation.** Data from large or subsets of sensors can be selected through the catalog and queried by the IWN users.
- **Displays/User Interface.** Accessed data can be downloaded and displayed through standard analysis tools.

Appendix D: AirNow Sensor Data Pilot: Enhancing Wildfire Smoke Detection

The AirNow Sensor Data Pilot is a partnership between EPA and the U.S. Forest Service. The Forest Service has a Web page showing Purple Air readings; EPA's version will add reference sites. AirNow has conducted extensive consultation with State/Local air quality agencies. At the initiation of this pilot, EPA and the U.S. Forest Service were unsure if agencies would participate. Ultimately 148 of 150 of the air quality agencies opted in (DE and MS opted out). EPA's Office of Research and Development (ORD) manages the relationship with Purple Air.

Existing AirNow Reference Monitors

Reference Monitors Augmented by Existing Purple Air Sensors

IoT Device: Purple Air sensor units: PM2.5 micron particulate measurements, dual internal sensor chips (made in China, not by Purple Air; estimated cost @$25 each). Also report temperature and humidity. $229 for typical indoor/outdoor unit. Connects to network via WIFI. Effective range 0 to 500 µg/m3, max 1000 µg/m3. Purple Air updates firmware remotely.

Platform: No special platforms. Data includes all readings in the Purple Air Database for outdoor measurements.

Data Format: Purple Air native data formats. Pilot uses all data reported: A & B Channels (particulates) plus Temperature, Humidity, Time.

Network: API from Purple Air database.

Data Mgt & Aggregation: QA Processing by ORD

- Use only outdoor measurements
- Screen to ignore stuck sensors, non-reporting sensors; require close agreement between twin sensors in unit
- Apply ORD Correction equation to correct for regional/weather variations (sensors tend to read high; correction brings values down)

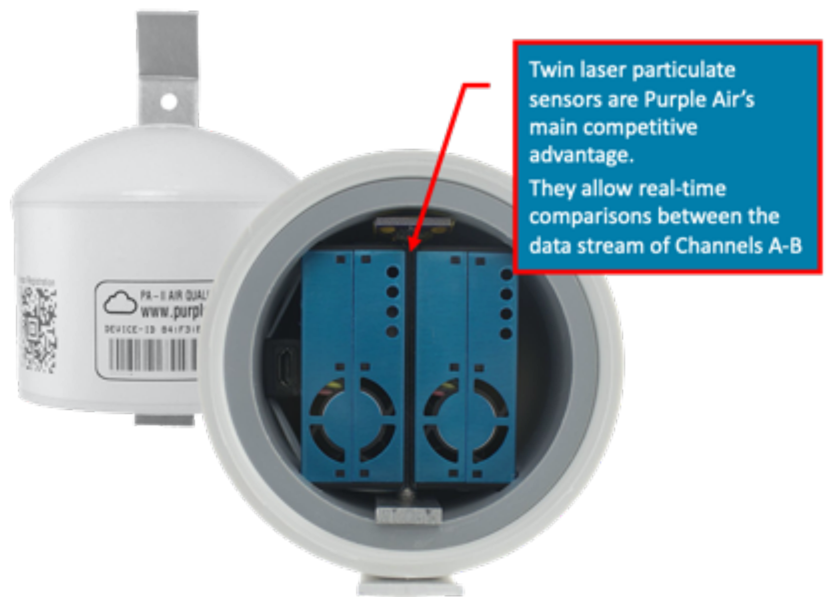
Displays/User Interface:

- Only pilot output will be a visual display
- Processed data will become a new data layer on the AirNow Wildfire and Smoke Map on AirNow.gov. Allows direct comparisons with reference sensors.

Air Now: Data Management

All of the quality assurance/quality control (QA/QC) for Air Now is done within data stream.

Purple Air basic sensor unit: $226



View from underneath

Purple Air Database

- Aggregates ALL customers: many AQ agencies, but also general public
  - Pilot is a de facto citizen science application within a formal agency program
- Database covers entire country, but AirNow wildfire interest is mainly in western states
  - Total number of additional sites not calculated; California alone, however, adds about 10,000 sites.
- Purple Air separates outdoor from indoor readings only
  - Provides raw data only -- no QA, good sensors and bad

EPA Data Processing

- Use outdoor readings only, filter non-reporting sensors, stuck sensors
- A/B channel comparison: must agree within 5μg/m3 during low readings, 70% during high readings
- For pilot, two channels must be within 5% to accept; values are averaged
- 2-minute sensor readings aggregated to 1-hour average
- Apply EPA national correction equation (PM2.5 corrected = 0.534*[PA_cf1(avgAB)] - 0.0844*RH + 5.604
- Apply Nowcast algorithm to relate hourly readings to AQI categories

Comparisons with Reference Sensors

- AQI category correctly reported 93% of time; within 1 category 100% of time
- Most disagreement at the boundaries between AQI categories
- Preliminary tests indicate good agreement with wildfire particulates