Presenting Air Quality Data Exchange (AQDx) - A Standardized Format for Data Sharing and Harmonization

Ashley Collier-Oxandale, Erik Joplin, Leah Gibson, Jason Schroder, & Michael Ogletree

National Ambient Air Monitoring Conference, August 2024

COLORADO
Air Pollution Control Division
Department of Public Health & Environment

Overview

I. Project Motivation & Development
II. Overview of the Standard and Metadata
III. Next Steps & How to Get Involved

Objective

Photos of an air toxics monitoring site and instrumentation

Develop a standard data format suitable for...

- Different types of technology
  - Sensors -> research-grade -> regulatory-grade
- Different temporal scales
  - High-time resolution and time integrated measurements
- Different pollutants
  - Criteria & air toxics (established and novel monitoring methods)

SOLUTION: AQDx - Air Quality Data Exchange

Std Development

Project Partners

TDEnviro

Step 1: Compared existing data formats and standards to identify a recommended model for the format

Step 2: Interviews w/ SMEs to gain input on key considerations for draft data standard

Step 3: Broader survey to collect feedback on draft standard from key stakeholders

Step 4: Finalized draft standard for review/ dissemination/ modification

Modelled on AQCSV

- AQCSV - Air Quality Comma Separated Values

Rationale:

- Input/output format for AirNow -> *meets needs of regulatory instruments*
- Self-describing -> *user friendly for a wide audience*
- Leverages AQS coding -> *builds on an existing system*
- Backed by an authoritative organization -> *EPA*

OVERALL: well-established in the air quality monitoring community

An Overview of Air Quality Exchange (AQDx)

Two versions offered, each includes the same parameters, intended for different applications

(1) CSV
(Comma-Separated Values)

- For finite datasets, such as historical data
- For relatively smaller datasets
- Human-readable
- Recommended format for general public

(2) JSON
(JavaScript Object Notation)

- For real-time or streaming data
- May support direct submission to a database, enabling automated processing (e.g., automated QA and reformatting to save needed information in preferred format)

The AQDx Standard - CSV Version

Each record contains the following fields:

data_steward_name, device_id, device_manufacturer_name*, *datetime, lat, lon, duration, parameter_code, method_code, value, unit_code*, *autoqc_check, corr_code, review_level_code, qc_code, qualifier_codes, data_license_code, elev

For each record, the following apply:

- **The field delimiter**: , (a comma, which is ASCII character 44)
- **Data field definitions and specifications**: See **Table 1, Table 2** and the supporting guidelines.

- **IDENTIFIERS: where did the data come from?**
- **DATA: what is the complete data record?**
- **DATA QUALITY INDICATORS: what is the data quality?**
- **& ADDITIONAL DATA: some additional information about the data**

AQDx (CSV version) - an Example

Example 1:

The community-based organization CleanAirVision (data steward) operates and manages a PM air sensor in New Orleans, LA that outputs $PM_{2.5}$ and PM_{10} data . Air Sensing, LLC manufactured the device with device id 11402. Data processing has involved automated pollutant range and sticking checks as well as an internal review of the data by a research partner. The first three full data records describe $PM_{2.5}$ data and the last three describe PM_{10} for three minutes on May 23, 2023 between 11:50 PM and 11:52 PM CST.

```
CleanAirVision,11402,AirSensingLLC,2023-05-23T23:50:00-06:00,30.06111,-89.94084,3600.,88501,,3.1,105,1,0,1,0,,,WGS84,3.2
CleanAirVision,11402,AirSensingLLC,2023-05-23T23:51:00-06:00,30.06111,-89.94084,3600.,88501,,2.8,105,1,0,1,0,,,WGS84,3.2
CleanAirVision,11402,AirSensingLLC,2023-05-23T23:52:00-06:00,30.06111,-89.94084,3600.,88501,,4.2,105,1,0,1,0,,,WGS84,3.2
CleanAirVision,11402,AirSensingLLC,2023-05-23T23:50:00-06:00,30.06111,-89.94084,3600.,81102,,10.2,105,1,0,1,0,,,WGS84,3.2
CleanAirVision,11402,AirSensingLLC,2023-05-23T23:51:00-06:00,30.06111,-89.94084,3600.,81102,,9.7,105,1,0,1,0,,,WGS84,3.2
CleanAirVision,11402,AirSensingLLC,2023-05-23T23:52:00-06:00,30.06111,-89.94084,3600.,81102,,12.3,105,1,0,1,0,,,WGS84,3.2
```

JSON for Real-Time Data

- JSON is often used for communicating data in web applications, where information exchange is less structured and frequent
- Offers a more flexible structure than CSV, with the ability to be backwards compatible with newer AQDx versions
- Can be stored in a noSQL database such as MongoDB for efficient queries

CSV

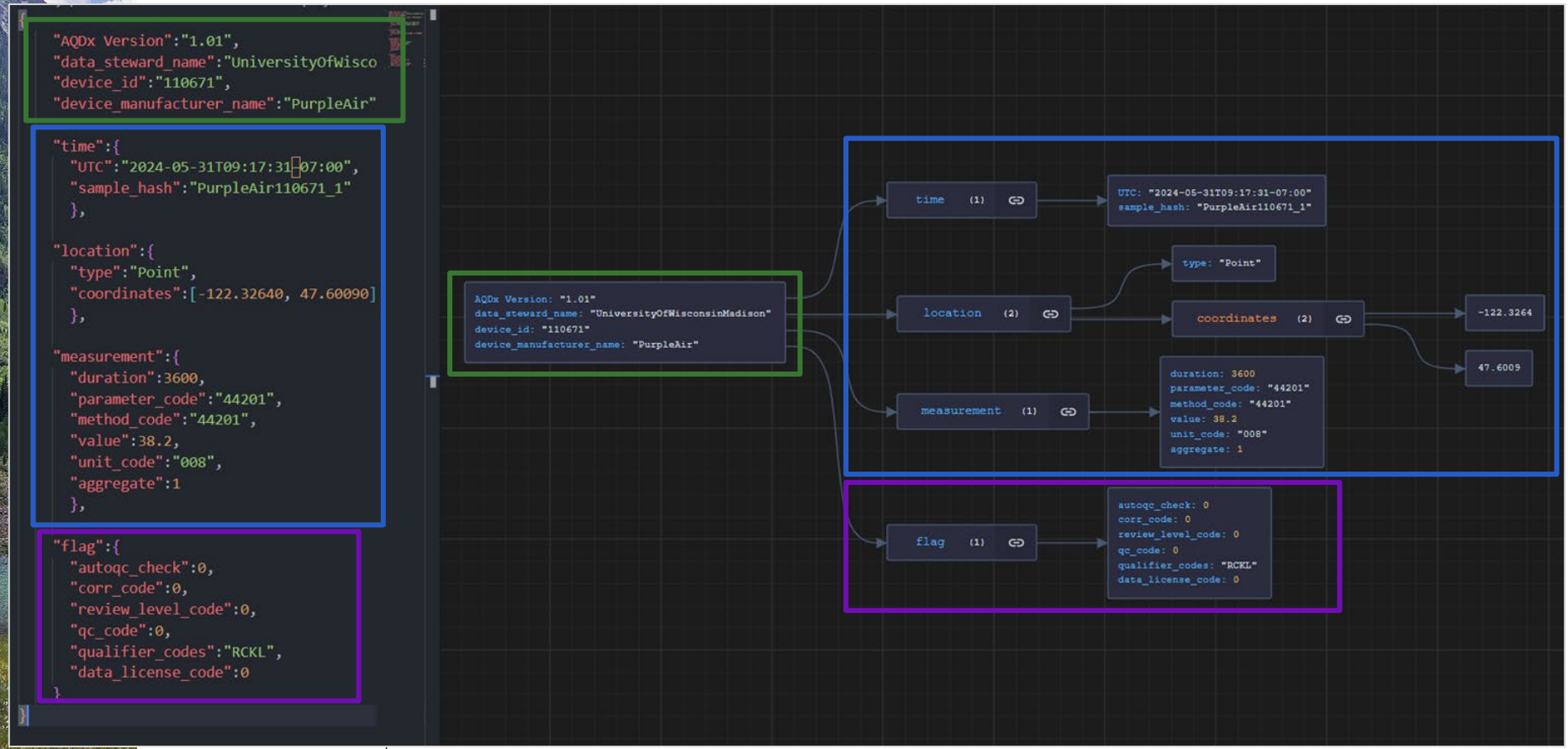
```
one,two,three
1,2,3
4,3,2
```

JSON

```
[{"one":"1","two":"2","three":"3"},
{"one":"4","two":"3","three":"2"}]
```

JSON allows for a more flexible format ideal for streaming data, however total file sizes will be larger than CSV

The AQDx Standard - JSON Version



Example 1: CSV compared to JSON

Example 1:

The community-based organization CleanAirVision (data steward) operates and manages a PM air sensor in New Orleans, LA that outputs $PM_{2.5}$ and PM_{10} data . Air Sensing, LLC manufactured the device with device id 11402. Data processing has involved automated pollutant range and sticking checks as well as an internal review of the data by a research partner. The first three full data records describe $PM_{2.5}$ data and the last three describe PM_{10} for three minutes on May 23, 2023 between 11:50 PM and 11:52 PM CST.

CleanAirVision,11402,AirSensingLLC,2023-05-23T23:50:00-06:00,30.06111,-89.94084,3600.,88501,,3.1,105,1,0,1,0,,,WGS84,3.2

JSON:


```
{"data_steward_name":"CleanAirVision","device_id":"11402","device_manufacturer_name":"AirSensingLLC","time":{"UTC":"2023-05-23T23:50:00-06:00","sample_hash":"aee0123affeg"},"location":{"type":"Point","coordinates":[-89.94084,30.06111]},"measurement":{"duration":"3600","parameter_code":88501,"value":3.1,"unit_code":105,"autoqc_check":1,"corr_code":0,"review_level_code":1,"qc_code":0,"elev":3.2}}
```

Data Quality Level Indicators (all versions)

The following three fields provide a more complete picture of the level of processing, correction, and review for data -> facilitating appropriate use of data

- **autoqc_check**: has the data been processed using automated tools, algorithms, or software? [inputs: 0 = no, 1 = yes]
- **corr_code**: has the data been calibrated against a known standard, or has the data have been adjusted using an established method?
  [inputs: 0 = no, 1 = yes]
- **review_level_code**: what level of review has the data received?
  [inputs: 0 = none, 1 = internal, 2 = external, 3 = certified]

The Standard - MetaData Form (2 of 4 tabs)

A	B	C	D	E	F	
Type	**Field Name**	**Response**	**Required (Y/N)**	**Format**	**Description**	**Options**
	Site Name		Y	string (64)	Name of site	
	Latitude (degrees)		Y	decimal (9,5)	Latitude in decimal degrees of the device. Latitudes north of the Equator are positive and latitudes south of the Equator are negative; for example, 45.25325 N = 45.25325 and 45.25325 S = -45.25325. Report values to the 5th decimal point.	
	Longitude (degrees)		Y	decimal (9,5)	Longitude in decimal degrees of the device. Longitudes west of the Prime Meridian are negative and longitudes east of the Prime Meridian are positive; for example, 100.36124 W = -100.36124 and 100.36124 E = 100.36124. Report values to the 5th decimal point.	
	GIS datum		Y	string (10)	Code indicating the datum associated with the latitude and longitude measurements used to plot the data in geographic information systems (GIS)	
	Address		N	string (64)	Address of site	
	State code		Y	integer (2)	Two digit AQS state code	see AQS state codes
	County code		Y	integer (3)	Three digit AQS county code	see AQS county codes
	Site owner		Y	string (64)	Name of site owner	
	Link to site photo(s)		N	string (85)	Link (url) to photo(s) of site	
	Description of site surroundings		Y	integer (1)	Category describing the surroundings of the site	1 - urban, 2 - rural, 3 - suburban, 4 - in
	Description of nearby pollution sources		N	string (100)	Text describing pollution sources near the site	
	AQS ID		Y	integer (9)	Nine digit monitoring location code	AQS site IDs
	Monitoring Scale (Spatial Scale Represented)		Y	integer (1)	Spatial scale of representativeness is described in terms of the physical dimensions of the air parcel nearest to a monitoring site throughout which actual pollutant concentrations are reasonably similar	1 - micro, 2 - middle, 3 - neighborhood
	Site type		Y	string (40)	General monitoring objective	1 - highest concentration, 2 - populatic
	Groundcover		Y	string (64)	Type of groundcover at the site (select dominant type)	1 - grass, 2 - shub, 3 - trees, 4 - flower

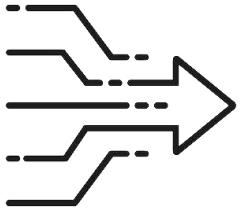
Metadata applicable to:
All
Regulatory

Data Steward Details | Site Details | Instrument Details | Data Quality Level Details

Trade-Offs

Advantages

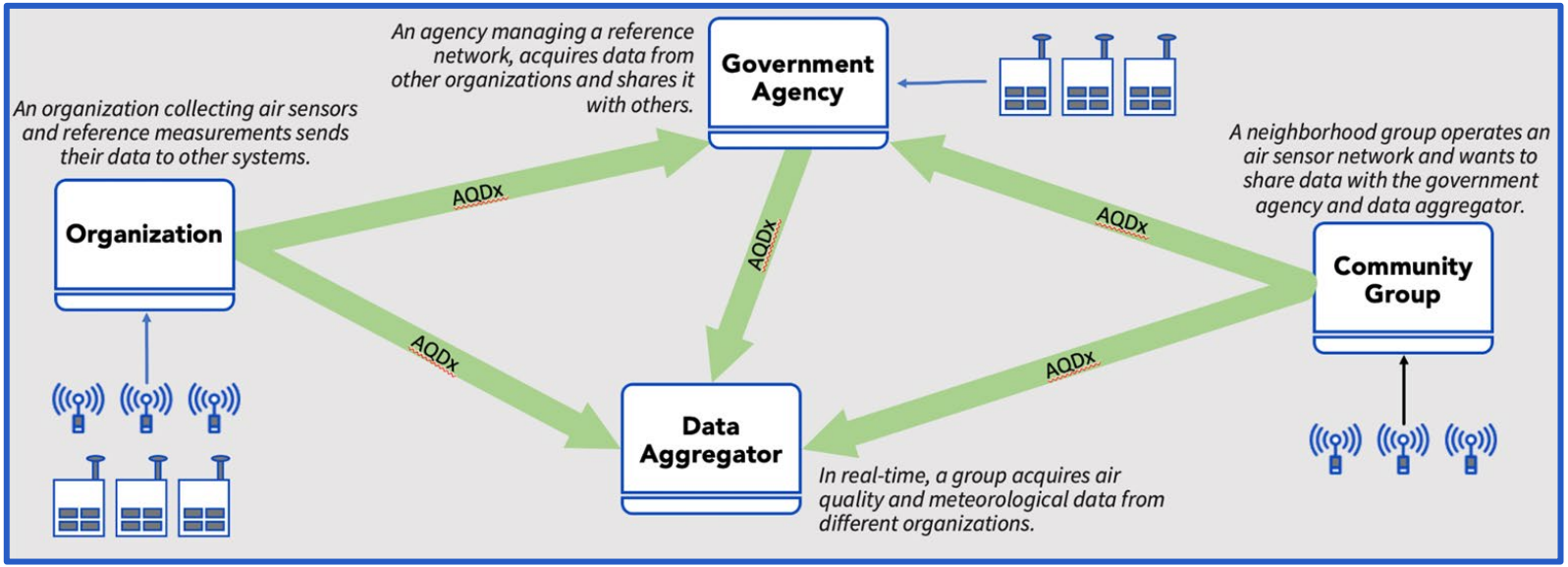
- Comprehensive
- Supports data exchange
- Support collaboration
- Supports the integration of different data sources



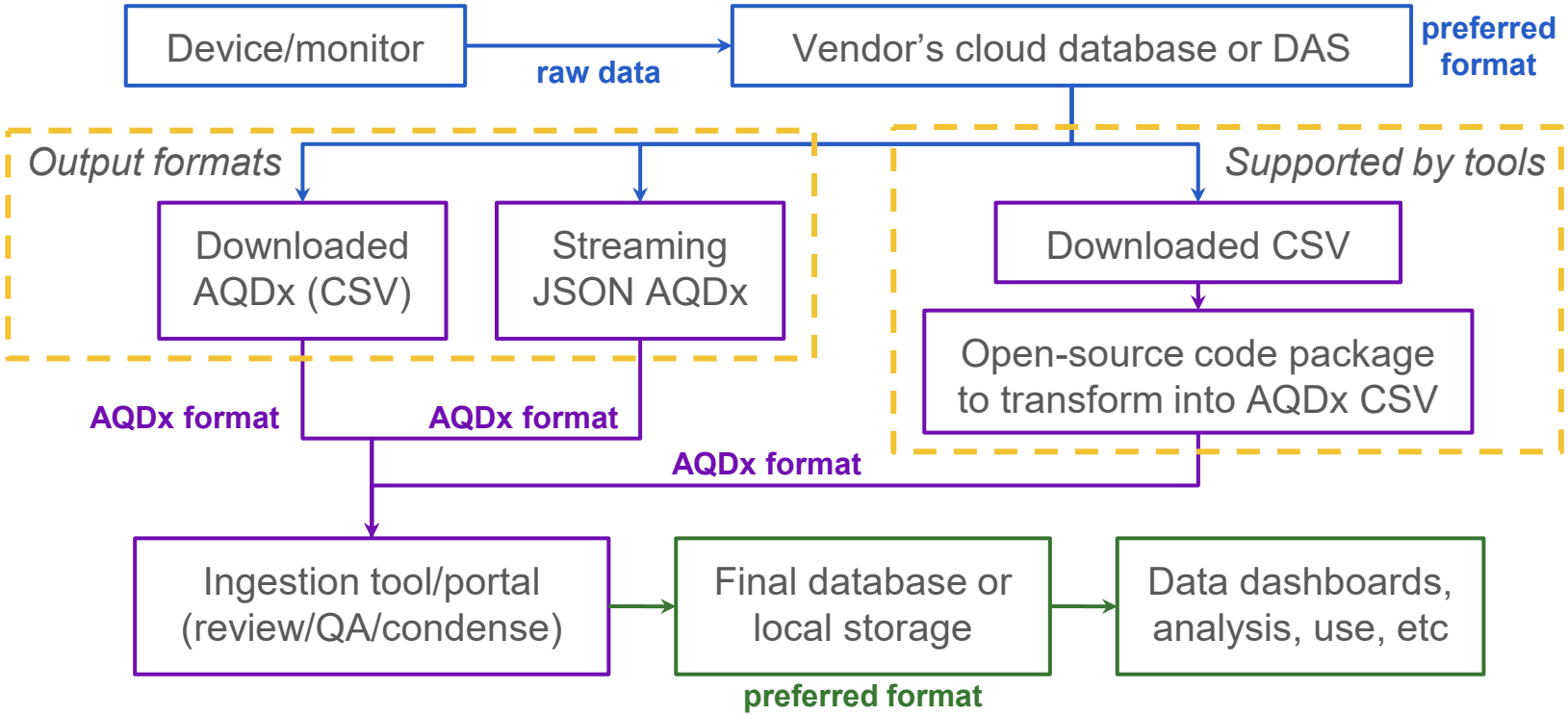
Challenges

- Verbose
- Not ideal for bulk data transfers
- Not a data/instrument acquisition format
- Not ideal for data storage

Purpose of AQDx = Data Exchange



How/Where Data is Transformed into AQDx?



Exploring this Question through Data Pilots

Data Pilot 1 - a community sensor network

- Collaboration between state gov + CBO + manufacturer
- Objective: build a real-time AQDx output for different sensors that may be used as an ingestion format for data visualizations

Data Pilot 2 - real-time FRM data

- Collaboration between state gov + local gov + contractor
- Objective: build a real-time AQDx output for different sensors that could be used to share data on a visualization platform

Data Pilot 3 - historical data

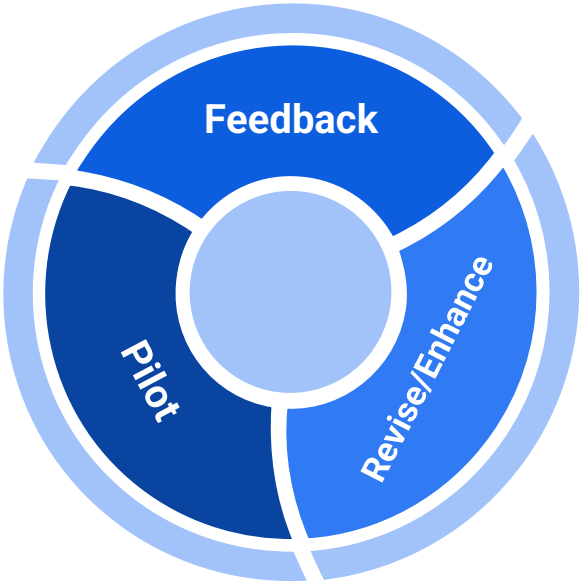
- Collaboration between state gov + local gov + contractor
- Objective: pilot transforming historical data into AQDx to evaluate as delivery format to share data with state gov for use

Data Pilot Approach - Guiding Questions

- Was the CSV or JSON tested?
- Describe how the format was integrated into existing data storage and management systems
- What is the lift to transform the data format?
    - How much time does it take to either transform the data or build tools to output the data in the new format for exchange? (tracking sheet to record tasks and time)
- How does the size of the data record compare to the previous or default format?
- Are there any fields that should be changed?
- Are there any new issues you foresee with this standard format?
- Do you have recommendations for other groups in terms or where or how they transform data into AQDx, or make the format available?

Status & Next Steps

- First official version is available: https://cdphe.colorado.gov/air-quality-data-exchange
- Continued collaboration with Working Group (EPA, South Coast AQMD, and Bay Area AQMD) to revise and enhance
- Continued Stakeholder engagement
- Integrate into rules and regulations
  - Reg. 7 - requires the submission of fenceline data from oil and gas facilities during certain phases of operation
  - AQDx will harmonize data from different sources



Final Discussion & Questions

- Questions?
- Thoughts, feedback, or recommendations?
- Interest in testing/piloting the AQDx?

[AQDx (Version 2.0)]
[Metadata Form]
[Metadata Form Guidance]
[AQDx Webpage]

- *Ways to get involved*
  - Join the Stakeholder Group
  - Review and provide feedback
  - Email: ashley.collier-oxandale@state.co.us

Acknowledgments

- *Thanks to the AQDx Working Group*: Ron Evans, Andrea Clements, Vasu Kilaru, Richard Wayland, Phil Dickerson, Corey Mocka, Robb Wildermann, Brandon Feenstra, Wilton Mui, Michael Flagg, Michael Ogletree, Jason Schroder, Amber Eglund, Madelyn Percy, Loren Speer, Erick Mattson, Erik Joplin, Leah Gibson, Nathan Bradley, Steve Szocik, William Vicars, Steve Mccannon, & Jim Reasor
- *Thanks to the additional partners at CDPHE*: Jenna Channel, Stephen Johnson, & Angelika Selviyan
- *Thanks to project partners at TDEnviro*: Tim Dye & Helena Pliszka

Thank you!

For questions or additional information, please contact:
ashley.collier-oxandale@state.co.us

COLORADO
Air Pollution Control Division
Department of Public Health & Environment