



Issue/gap identified

Multi-Drug Resistant Organisms (MDROs) pose a major challenge to long-term healthcare facilities (LTHFs) in Alexandria. To monitor the burden, Epidemiologists at Alexandria Health Department (AHD) conduct regular Point Prevalence Surveys (PPS). However, this process is highly manual and time-consuming due to:

Inconsistent census formats

LTHFs submit patient and current infection data using inconsistent formats and varying data fields.

Unreliable current infection status

LTHFs report outdated or incorrect infection status, requiring manual verification in the state surveillance system.

Unstandardized lab data

Lab results use inconsistent terminology and are stored as separate entries, complicating patient-level analysis.

Multiple lab systems

Switching lab providers (ALRN to DCLS) introduced structural differences between historical and current data, preventing easy consolidation.

These gaps place a significant manual burden on epidemiologists during each PPS—one they are unable to address given their existing workload. Investing in data modernization to automate the cleaning and standardization of lab data would streamline this process, reduce the burden on the epidemiology team, and strengthen MDRO surveillance capacity.

Uncleaned DCLS C.Auris Data Example

| Firstname | Lastname | Dob | CollectionDate | TestResult |
|-----------|----------|----------|----------------|---------------------------|
| John | Doe | 9/9/1999 | 1/11/2025 | Candida auris DETECTED |
| John | Doe | 9/9/1999 | 1/11/2025 | No Candida auris isolated |



Uncleaned DCLS CPO Data Example

| Firstname | Lastname | Dob | CollectionDate | TestResult |
|-----------|----------|----------|----------------|---|
| John | Doe | 9/9/1999 | 1/11/2025 | NDM positive Klebsiella pneumoniae group Isolated |
| John | Doe | 9/9/1999 | 1/11/2025 | KPC gene not detected |



Cleaned Merge Data Example

| FirstName | LastName | Dob | CollectionDate | CAuris | KPC | NDM | NDMOrganisms |
|-----------|----------|----------|----------------|----------|----------|----------|-----------------------------|
| John | Doe | 9/9/1999 | 1/11/2025 | Positive | Negative | Positive | Klebsiella pneumoniae group |



Completing an MDRO Escape Room with the Epidemiology Team

Outputs

Python script that cleans, standardizes, and merges lab data from DCLS and ALRN, transforming test-level rows into one row per patient per collection date.

Test script to thoroughly test the cleaning, standardization, and merging logic.

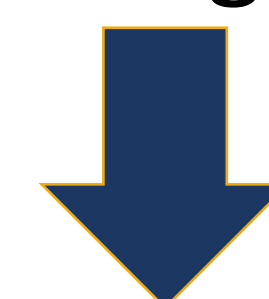
Generated datasets include:

- Cleaned C. auris** lab data
- Cleaned CPO** (Carbapenemase-Producing Organisms) lab data
- Merged C. auris + CPO** lab data
- Master MDRO dataset** combining historical and newly pulled records

Documentation describing each dataset's structure, content, and fields.

Outcomes

- Consolidated, organized, and simplified** lab data, streamlining patient-level analysis
- Improved standardization and consistency** within and across lab data sources, facilitating aggregation and analysis by test result
- Increased adaptability** to future lab changes and data format variations



- Reduced manual workload** for Epidemiologists when conducting PPS
- Faster turnaround** for MDRO analysis, reporting, and summaries
- Reusable tool** to support future PPS using DCLS and ALRN lab data in streamlining MDRO surveillance efforts



Image source: Centers for Disease Control and Prevention. Candida auris.
<https://www.cdc.gov/antimicrobial-resistance/media/pdfs/candida-auris-508.pdf>

Acknowledgments

This project was made possible through the guidance and support of the Epidemiology and Informatics teams at AHD. Special thanks to Sean Kolaskar for his mentorship and feedback throughout the project.

Challenges

- Anticipating future variability:** Accounting for potential new data fields, formats, and test result variations that may be introduced by labs in the future.
- Interpreting lab logic:** Understanding when different types of tests (e.g., PCR or colonization) are conducted and how their results should be interpreted—particularly when they have conflicting results for a single sample.
- Result summarization:** Developing rules to accurately consolidate multiple test-level results that use inconsistent terminology into a single, meaningful patient-level result.
- Automating workflows:** Learning how to automate the execution of scripts and data pipelines.