Operational Network of Air quality Impact Resources

Best Practice Guide

BP404 | Manage and analyse data

Data interpretation: overview

Introduction

Air quality data from smart low-cost air quality sensing devices has the potential to support a wide variety of high-impact activities. However, raw data (i.e. data that has not been processed or corrected) on its own has limited use. To extract value from raw data, it must be interpreted to make it useable and able to produce insights or functional outputs that users can trust as being accurate. The potential value of data therefore depends on the capacity of an organisation (or its partners) to undertake effective data interpretation, and to act on the insights that emerge. This may require specialist knowledge from a data analyst; however, these skills can be learnt with appropriate training.

Effective data interpretation matters because it can:

- help to build **trust** in the data that is collected
- support the distillation of data into a **useable** format for your organisation
- ensure the **quality** and **accuracy** of results and insights
- improve understanding of how you can use and share data **appropriately** and **effectively**
- support the leveraging of data to drive **impact**.

Who is this resource for?

This resource is for local governments and other organisations undertaking similar projects. It is intended for staff engaged with the design and delivery of air quality monitoring projects, including project managers, environmental officers, smart city leads, and planners. It is also a useful reference for senior management who wish to understand the complexities and challenges related to this kind of project.

How to use this resource

This Best Practice Guide chapter is the first in a series of four chapters on the topic of data interpretation. It provides a high-level overview of the topic, as well as presenting some foundational concepts (such as the OPENAIR *framework for categorising air quality sensing devices*). For more detailed guidance on the three main steps of data interpretation, please refer to these Best Practice Guide chapters in the order outlined in Figure 1.

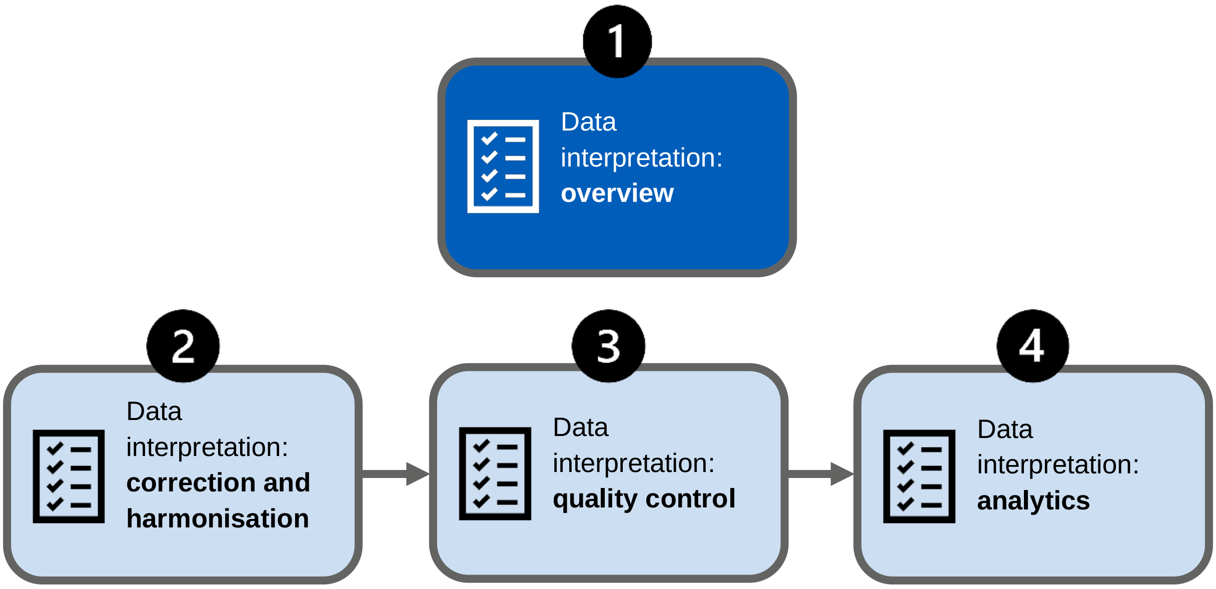

Figure 1. OPENAIR data interpretation Best Practice Guide chapters

Key messages

The key messages in this chapter are:

- Make sure you have a clear understanding of the problem your project aims to solve, or the question you want your data to answer
- Familiarise yourself with the terminology and definitions relating to air quality and air quality data interpretation
- Consider how your project relates to the OPENAIR tiered data use framework (as presented in Figure 4), as this will guide your choices regarding data collection and processing
- Allocate sufficient resources (in terms of time, funding, and expertise) for data management, interpretation, and analysis throughout the lifetime of your project.

RESOURCING DATA INTERPRETATION: Data interpretation requires sufficient resourcing. You will need to ensure that you clearly define a technical solution capable of supporting data interpretation that meets your impact objectives. You will also need to assign responsibility for data interpretation tasks, and ensure they are allocated time and budget as part of your project plan. You may also need to engage external expertise for more complex data interpretation.

Data interpretation process and definitions

Increasing levels of data interpretation

Raw sensor data	
Data correction and harmonisation	The correction and harmonisation of raw (unprocessed) data into a useable and readable standard format. This process may include: • *calibration corrections*, such as correction for all devices of one type; drift correction (adjusting correction over time to account for sensor degradation); corrections for inter-device measurement variability; calibration for different types of particulate pollution • *environmental interference factor corrections*, where sensor response is significantly influenced by environmental conditions (especially when there are extremes in either relative humidity (RH) or temperature). Interference from RH is a known problem associated with optical sensors. *Please note that in many instances corrections are automatically applied to the data, though not always. It is worth checking with your equipment provider.* • *data harmonisation*, which involves converting data expressed in different formats into a single, harmonised format (defined by a data schema). For details on how to develop a data schema for your project see OPENAIR Best Practice Guide chapter *Data labelling for smart air quality monitoring*.
Data quality control	The process by which interpreted data is checked to ensure it can be trusted, and is useable for analysis. This may include: • *field testing*, where the ‘normal’ operation of a device (and the quality of the data it produces) is verified prior to main data collection activities • *data cleaning*, where data anomalies, outliers, and other bits of unusable data are detected and removed • *verification of data quality during standard operations*, where data is verified against external references (or through internal cross-verification) to improve user trust in its ability to support impact.
Data analysis	The interrogation and utilisation of data to produce insights or functional outputs. Data analysis may include: • statistical analysis (e.g. that shows locations where air pollution is higher over a 1 h or 24h period) • temporal interpolation • spatial aggregation and interpolation • complex geospatial system modelling • AI and machine learning applications.
Impact	

Figure 2. An overview of the three stages of data interpretation

Data interpretation for impact

To create impact with your air quality sensing project, it is critical that you process data appropriately. You should not assume that a commercial supplier of sensing devices has thought through all the important data interpretation considerations that relate to your particular situation. Each project is unique, and it is your responsibility to understand your data interpretation needs to achieve the desired project impact.

Data from low-cost devices has the potential to support a wide variety of activities and impact areas*, including:

- education and community engagement
- advocacy
- pollution hotspot identification
- improved decision-making and policy development
- improved service delivery
- improved air quality alerts.

Develop a clear understanding of how your data can support impact

Figure 3 shows the steps needed to interpret raw sensor data to create impact.

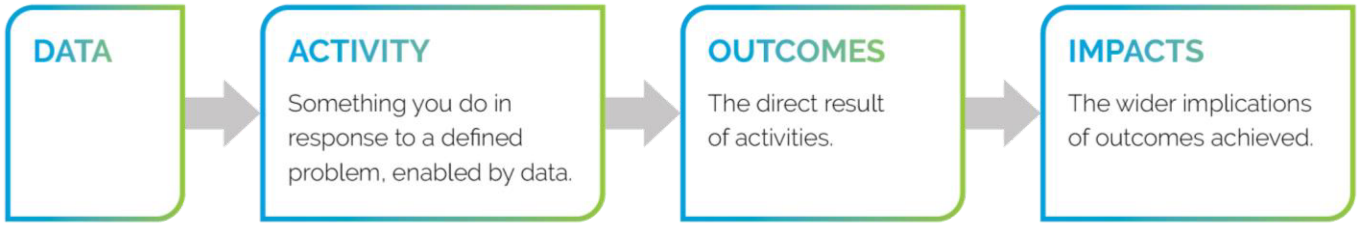

Figure 3. Data interpretation framework

The effective and appropriate interpretation of raw data can produce insights or functional outputs (such as automated alerts), which can then drive activities for impact creation.

At the start of your project, you should identify clear project objectives that connect a problem statement with fit-for-purpose data, data users, and an impact plan. This will be captured by a Data Use Action Statement (see the OPENAIR supplementary resource *Identify Template* for help in developing this).

The value of data relates to its potential to support a given impact objective. This can be explored by considering the limits to impact creation that are imposed by the context and design of a project.

* For additional information on potential impact areas, refer to the OPENAIR supplementary resource *Activities for impact compendium*.

Contextual limitations to impact creation

- **The specific context of your monitoring location**. Climate and weather, topology, vegetation, and relative position of pollution sources will all impact the type and quality of data you are able to collect, and your data interpretation requirements. Engage with your technology vendor and/or regulatory authorities to help you understand these factors in your location.
- **The air quality guidelines/standards that apply in your region**. In New South Wales, there are several state-specific regulations and guidelines. A good starting point is the Protection of the Environment Operations Act 1997, which sets the statutory framework for managing air quality in NSW. This legislation is supported by the Protection of the Environment Operations (General) Regulation 2022, The NSW Government is also a signatory to the National Environment Protection Measure (NEPM) (Ambient Air Quality). The NEPM provides standards for maximum concentrations and averaging periods for several key pollutants, including $PM_{2.5}$ [†]. Understanding these (and local standards) can have implications for how you should interpret your data relative to a particular use case.
- **Your capacity as an organisation to support data interpretation**. This includes the skills and expertise of your staff, either to design and implement data interpretation in-house, or to ensure that appropriate data interpretation outcomes are achieved by third parties on your behalf. It includes the availability of staff to engage with data interpretation during the project.
- **Your capacity as an organisation to act on data to create impact**. This includes your capacity to make data available to prospective users across your organisation, and the capacity of those users to undertake activities for impact that make use of data. Please refer to the OPENAIR Best Practice Guide chapter *Engaging your organisation with air quality data* for more in-depth guidance.

Design limitations to impact creation

- **The attributes and limitations of your chosen sensing devices.** Attributes include sensor performance metrics relating to data quality, device functionality, and configurable settings for things like sampling rate and reporting interval. These attributes will place limitations on the type and quality of data you can produce, and how you can appropriately use it. Understanding this allows you to make use of the supplementary resource *A framework for categorising air quality sensing devices*. Refer to the OPENAIR supplementary resources *Technical requirements template* and *A guide to developing technical requirements* for further details.
- **The design of your monitoring network.** The number of devices, the way that you configure devices, the density of device deployments, and device micro-siting arrangements within the local landscape all impact the type and quality of data that you produce, and your data interpretation requirements. For more on this topic, refer to the OPENAIR Best Practice Guide chapters *Sensing device deployment planning: high-level design* and *Sensing device deployment planning: detailed design*.

[†] PM (particulate matter) refers to airborne solids or liquids. Its size is measured in micrometres and is indicated by the subscript. E.g. $PM_{2.5}$ has a diameter of 2.5 micrometres or less. (NSW Health, 2020)

- **The data interpretation capacity of your IoT solution.** This includes the functionality and limitations of your IoT architecture. Data correction and harmonisation may be applied as a series of automated actions that are programmed into devices, or into the IoT platform that manages your devices. Quality control and analysis can occur at various levels in your architecture. Aim to understand your data interpretation needs when you make technology procurement decisions.

A framework for understanding appropriate data use relative to technology and methodology choices

The OPENAIR supplementary resource *A framework for categorising air quality sensing devices* determines how data from smart low-cost sensors might be appropriately used. It is recommended that you refer to this framework to guide your overall approach to data interpretation.

The framework (summarised in Figure 4) is designed as a simple tool to help you connect your intended data impact with project decisions, such as the configuration and calibration of your sensing devices, the design of your monitoring network, and the approach you take to data interpretation. Each tier of the framework describes a relationship between a type of data and possible data uses. The way that data can be used varies between levels or tiers.

Data from smart low-cost sensors is applicable to Tiers 1 to 3. Start by identifying which tier of the framework your project aligns with, then use the following guidelines to develop your data interpretation strategy:

- In general, Tier 1 applications require less sophisticated and rigorous data interpretation. The emphasis may be on supporting community members to understand and do basic interpretation themselves, as part of a project's educational content.
- Tiers 2 and 3 can often have unclear boundaries, so your project may relate to aspects of both (e.g. some hotspot monitoring is done to a high enough standard that it can supplement data in levels above Tier 2).
- Tier 2 applications may require less sophisticated interpretation than those in Tier 3. Often, all you need to do is identify trends and variations. However, correction of raw data remains critical. If you are running a hybrid network (with more than one type of sensing device), then internal harmonisation of data will also be critical to allow direct comparison of data from different types of device. Quality control (particularly to identify and remove outliers) is also essential for Tier 2 applications.
- For Tier 3 applications, data harmonisation and quality control are particularly critical, as these are vital precursors to effective data sharing, which is the foundation of supplemental monitoring (where local air quality data is shared with the appropriate government authority and compared with Tier 4 data). In some cases, analytics may be outsourced to the regulatory authority.

For additional details, refer to the OPENAIR supplementary resource *A framework for categorising air quality sensing devices*.

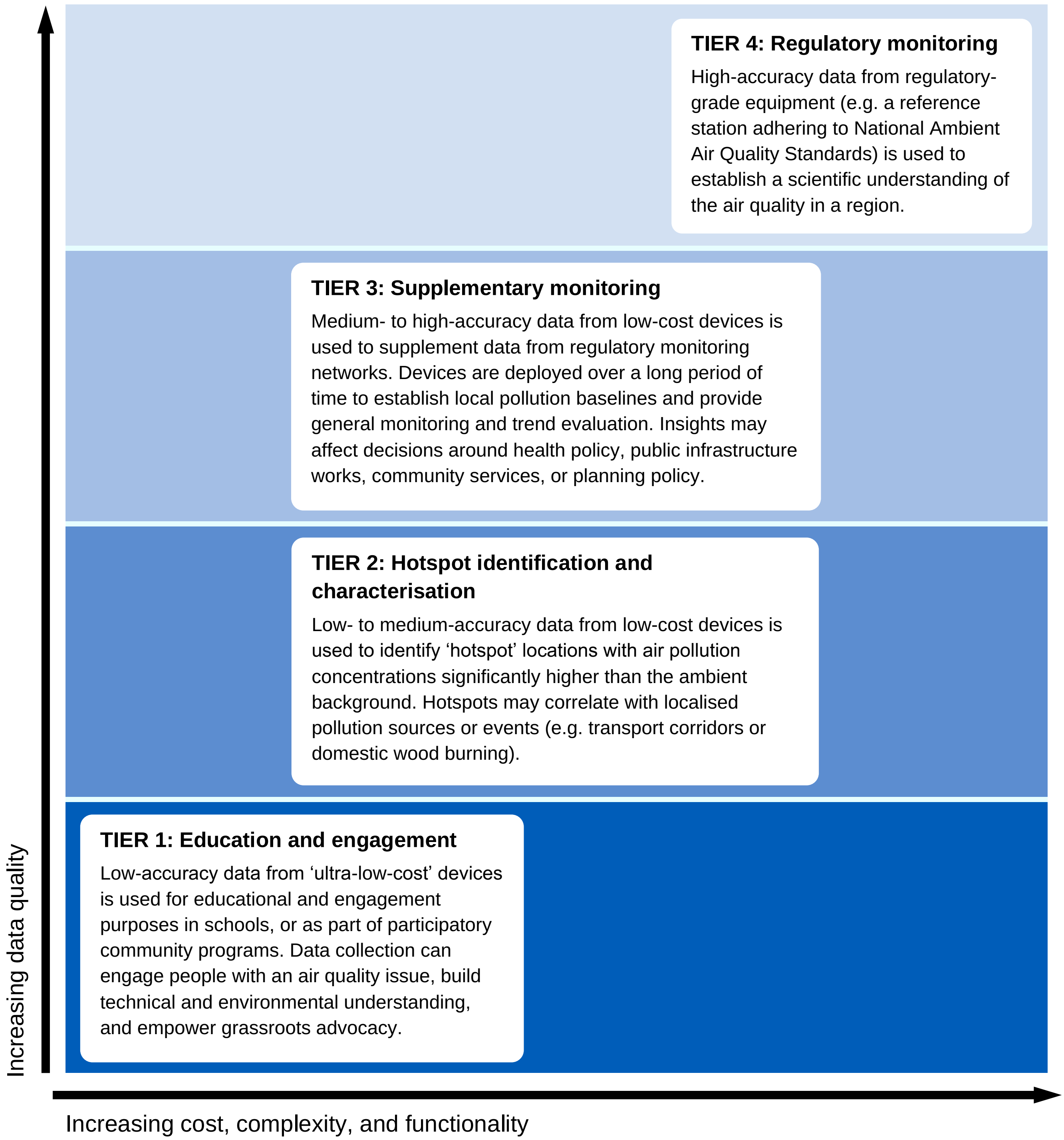

Figure 4. A framework for categorising air quality sensing devices (basic overview)

References

NSW Health. (2020). *Particulate matter (PM10 and PM2.5)*. NSW Government. https://www.health.nsw.gov.au/environment/air/Pages/particulate-matter.aspx

Additional resources

U.S. Environmental Protection Agency* | *Air Sensor Toolbox

The U.S. Environmental Protection Agency (EPA) has developed extensive guidance on the use of low-cost sensor data, including a range of information on how to understand your sensor data readings.

Associated OPENAIR resources

Best Practice Guide chapters

Data interpretation: correction and harmonisation

This Best Practice Guide chapter provides guidance for correction and harmonisation of data produced by smart low-cost air quality sensors. It introduces several types of correction factor that may need to be applied to raw sensor data, and explores how data formatting and labelling should be harmonised with a project data schema to support effective data management and sharing.

Data interpretation: quality control

This Best Practice Guide chapter provides guidance for the quality control of data produced by smart low-cost air quality sensors. Data quality control helps to isolate trusted data that can then be used to support chosen activities. This chapter explores approaches for cleaning static data sets to prepare them for analysis, and approaches for operational verification and quality control of live data streams.

Data interpretation: analytics

This Best Practice Guide chapter introduces common analytical approaches that can be applied to data produced by smart low-cost air quality sensors. These include statistical analysis; temporal interpolation; spatial aggregation and interpolation; complex geospatial system modelling; and machine learning.

Data labelling for smart air quality monitoring

This chapter is a high-level introduction to data schemas and data labelling.

Activities for impact

This Best Practice Guide chapter introduces a range of activities that can be undertaken by a local government to create impact relating to an air quality issue. Activities are categorised into four impact areas: transport; built environment; green infrastructure; and community engagement.

Engaging your organisation with air quality data

This Best Practice Guide chapter discuss organisational capabilities and the concept of smart city maturity models.

Sensing device deployment planning: high-level design

This Best Practice Guide chapter explores the high-level design of a smart air quality monitoring network. It provides general guidance for selecting where to deploy devices, what to mount them on, how to mount them, and how to support their operation.

Sensing device deployment planning: detailed design

This Best Practice Guide chapter explores the detailed design of a smart air quality monitoring network. It builds upon high-level design activities, and provides guidance for planning and documenting the details of specific device deployments.

Supplementary resources

Activities for impact compendium

This resource presents an extended, detailed compendium of activities that can be undertaken by a local government to create impact relating to an air quality issue. Activities are categorised into four impact areas: transport; built environment; green infrastructure; and community engagement.

A framework for categorising air quality sensing devices

This resource presents a new framework for categorising air quality sensing devices in an Australian context. It identifies four tiers of device types, separated in terms of functionality and the quality and usability of their data output. It is designed to assist with the selection of devices that are appropriate for meeting the needs of a project and an intended data use case.

Identify template

This template supports creation of a business plan and 'data use action statement' as strategic foundations for a smart low-cost sensing project.

Technical requirements template

This template is an extended, step-by-step tool that supports the development of technical requirements for a smart air quality monitoring project. These requirements define the details of technologies (sensing devices, platforms, and services) that can meet the specific needs of a project, and are intended to support procurement decision-making.

A guide to developing technical requirements

This resource is a companion guide to the technical requirements template.

Further information

For more information about this project, please contact:

Peter Runcie

Project Lead, NSW Smart Sensing Network (NSSN)
Email: peter@natirar.org.au

This Best Practice Guide section is part of a suite of resources designed to support local government action on air quality through the use of smart low-cost sensing technologies. It is the first Australian project of its kind. Visit www.openair.org.au for more information.

OPENAIR is made possible by the NSW Government's Smart Places Acceleration Program.

Document No: 20231211 BP404 Data interpretation: overview Version 1 Final