AusEnHealth

An Online Environmental Health Decision Support Platform

AusEnHealth Initiative Deliverable 5

Methods Report

April 2022

Document Control
Authors: A. Price, K. Mengersen, F. Barar, R. Gupta, C. Davis.
Reviewers: P. Fievez, J. Keysers, AusEnHealth Steering Committee
Date: 27-04-22
Version: 1,0

AusEnHealth Methods Report

A. Price[1], K. Mengersen[1], F. Barar[2], R. Gupta[2], C. Davis[1].

1. Queensland University of Technology; 2. Australian Urban Research Infrastructure Network

April 2022

Acknowledgements

The AusEnHealth Initiative has been funded through a collaborative research agreement between FrontierSI, Western Australian Department of Health, Environmental Protection Agency Victoria, Queensland University of Technology (QUT), NGIS, Australian Urban Research Infrastructure Network (AURIN) and the NCRIS Terrestrial Ecosystem Research Network (TERN). We are grateful for the support of these agencies and individuals who provide additional advisory through the AusEnHealth's Steering Committee.

Team Member	Role	Organisation
Mike Lindsay	Executive Director, Environmental Health Directorate	Western Australia Department of Health
Andrea Hinwood (Jun20-Feb21), Martine Dennekamp (Feb20-Present)	Chief Environmental Scientist, Deputy Chief Environmental Scientist	Environmental Protection Agency, Victoria
Phil Delaney	Chief Innovation and Delivery Officer	FrontierSI
Stuart Barr	Professor and Director	Australian Urban Research Infrastructure Network
Beryl Morris	Director	NCRIS Terrestrial Ecosystem Research Network
Claire Sparke	Unit Head, Population Health	Australian Institute of Health and Welfare

We would also like to acknowledge the guidance and advice provided by AusEnHealth Advisory Board. This group comprises of experts in multidisciplinary fields and are a significant source of knowledge to the project team. These individuals and organisations are provided in Appendix B.

Table of Contents

1. Introduction 1
2. Data Audit 2
   2.1. Use Case Workshop 3
   2.2. Related Projects 4
   2.3. Indicator Development Frameworks 5
      Indicators and Parameters in the Literature 5
   2.4. Approach to Data Audit 5
      Data Search 6
      Data Source Auditing 6
   2.5. Data Catalogue 6
      Data Custodian Information 6
   2.6. Key Data Properties 7
      Secondary Data Sources 7
      Other Data Collections 7
      Ongoing Data Collection and Collaboration 8
   2.7. Summary Findings (Data Audit) 9
      Health Outcome Data 9
      Climate Data 9
      Air Quality Data 9
      Water Quality Data 9
      Built Environment Data 10
      All Data 10
3. Data Analysis Scope 10
   3.1. Data Selection 10
      Climate Data 11
      Air Quality Data 12
      Built Environment Data 12
      Demographics Data 13
      Health Outcome Data 14
   3.2. Data Handling 16
      Data Integration Techniques 16
      Automated Integration 17
      Data Acquisition and Formatting Processes 18
      Metadata 19
   3.3. Data Analysis Methods 19

Exploring Environmental Health Variables, Indicators, and Indices 19
Variable and Indicator Creation Methods 20
3.4. Population Vulnerability Indices 27
Social Vulnerability 27
Utilising exposure, sensitivity, and adaptive capacity framework 29
AusEnHealth Vulnerability Index Components 32
3.4.1. Methodology 34
3.5. Summary Findings (Data Analysis Scope) 40
Data Handling 40
Analysis Methodology 40
4. Visualisation and Communication Tools 41
4.1. Overview of Visualisation Techniques 41
4.2. Spatial Visualisation Techniques 42
4.3. Visual Variables 43
4.4. Summary of Existing Applications Review 43
4.5. Summary Findings (Visualisation and Communication Tools) 46
Messaging and Communication 46
5. Platform Development Scope 46
Technology Readiness 47
5.1. User Journey Mapping 47
Climate and Health Outcomes Related to Heat 48
Air Quality and Respiratory Health Outcomes 49
5.2. Functional Requirements 49
Business Requirements 49
Administrative Functions 49
User Requirements 50
System Requirements 50
5.3. Non-Functional Requirements 50
Usability 50
Availability 50
Scalability 50
Performance 50
Supportability 51
Security 51
5.4. Solution Design 51
Landing Page 51

5.5. Heat | Risk Assessment 52
5.6. Heat | Climate Change Assessment 53
5.7. Heat | Multi-Cloud Health Data Analysis 55
5.8. Air Quality | Impact Analysis 57
5.9. Air Quality | Public Awareness 59
5.10. Air Quality | Campaign Awareness 61
5.11. Architecture 63
Front-End Architecture 63
Back-End Architecture 64
Data Analysis Architecture 65
5.12. Schedule 65
5.13. Summary Findings (Platform Development Scope) 66
Data Sharing and Integration 66
6. Conclusion - Moving Towards a Cohesive Environmental Health Ecosystem 66
6.1. Environmental Health Ecosystem Recommendations 67
AusEnHealth Integration 67
Collaborative Ecosystem 67
Appendices 68
A. AusEnHealth Recommendations 68
Data Audit Recommendations 68
Data Analysis Recommendations 68
Visualisation and Communication Recommendations 68
Platform Development Recommendations 68
Environmental Health Ecosystem Recommendations 69
B. AusEnHealth Advisory Board 69
C. Use Case Ranking 69
D. Data Catalogue Tables 70
Parameters from Key Data Sources 70
Additional Data Collections 76
E. Data Quality and Accessibility Heatmaps 77
Air Quality 77
Climate 77
Water Quality 77
Built Environment 78
Health Outcomes 79
F. International Classification of Diseases Code Summary 80

Respiratory Health 80
Heart Health 80
Brain Health 81
Heat Health 81
Other 81
G. Proposed Data Integration Architecture 82
H. Formatted Data Used in AusEnHealth Pilot 83
I. Vulnerability Index Variable Definitions 84
J. Demographics Imputation Assessment 85
Bibliography 86

1. Introduction

Humans depend on the environment for energy and the materials necessary to sustain life [1]. In 2016, the World Health Organisation (WHO) stated that *"13.7 million deaths per year in 2016, amounting to 24% of the global deaths, are due to modifiable environmental risks"* [2]. The ongoing consideration of environmental factors is crucial for maintaining and improving the lives of humans [3]. The consideration of public health with specific respect to the environment is referred to as *environmental health*. The United States of America's National Environmental Health Association (NEHA) defines environmental health as *"the science and practice of preventing human injury and illness and promoting well-being by identifying and evaluating environmental sources and hazardous agents and limiting exposures to hazardous physical, chemical, and biological agents in air, water, soil, food, and other environmental media or settings that may adversely affect human health"* [4]. Similarly, the Australian Government Department of Health (DOH) states that environmental health *"involves those aspects of public health concerned with the factors, circumstances, and conditions in the environment or surroundings of humans that can exert an influence on health and well-being"* [5].

The Australian Environmental Health (AusEnHealth) initiative aims to build a national, spatially enabled, ongoing data infrastructure resource that will enable users to access, visualise and analyse environmental health data, reports, and models. Such a resource will allow the general public, media, policy analysts, decision-makers, researchers, and scientists to engage with environmental health in ways not previously possible in Australia. Enabling an understanding of Australian environmental factors in different regions across Australia can save lives through improved intervention targeting, prevention initiatives, and community awareness. While there is decades-long nation-wide interest in improving environmental health [6] understanding in Australia, existing approaches have either been limited in success or scope. This is largely due to differences in state-based policies for recording and sharing environmental health data, as well as a lack of long-term funding opportunities for projects working to produce improved data infrastructure.

The first phase of AusEnHealth is focused on the production of an interactive demonstrator which utilises existing environmental health data from around Australia. These data will be used to create indicators and indices, which are of high interest to policy analysts, decision-makers, scientists, and researchers working in the field. Environmental health indicators *"describe the link between the environment and health"* [1] and usually describe one aspect of the environment-health relationship. Environmental health indicators are mainly classified as either exposure indicators (possible environmental determinants of health), and health indicators (health effects from exposure to environmental hazards and risks). These indicators are then grouped by exposure domain; categories which contain a number of related environmental determinants of health. Such domains include air quality, climate, and built environment, among others.

This report details the process of building a digital environmental health decision support platform from the ground up. First, the report details the identification of high value use cases, relevant data, and metadata in the Data Audit Section. The report then outlines the selected data custodians, data formatting and processing methods, as well as indicator and index methodology in the Data Analysis Scope Section. A review of visualisation techniques and existing digital platforms is provided in the Visualisation and Communication Tools Section, followed by the list of users, scenarios, and user requirements for AusEnHealth application in the Platform Development Scope Section. Finally, the Conclusion - Moving Towards a Cohesive Environmental Health Ecosystem Section outlines the direction of AusEnHealth within the environmental health community. An AusEnHealth activity

pathway is shown below in Figure 1, with only the *"Demonstration Session & Revised Pilot"* and *"Assessment & Road Mapping"* activities remaining at the time of writing.

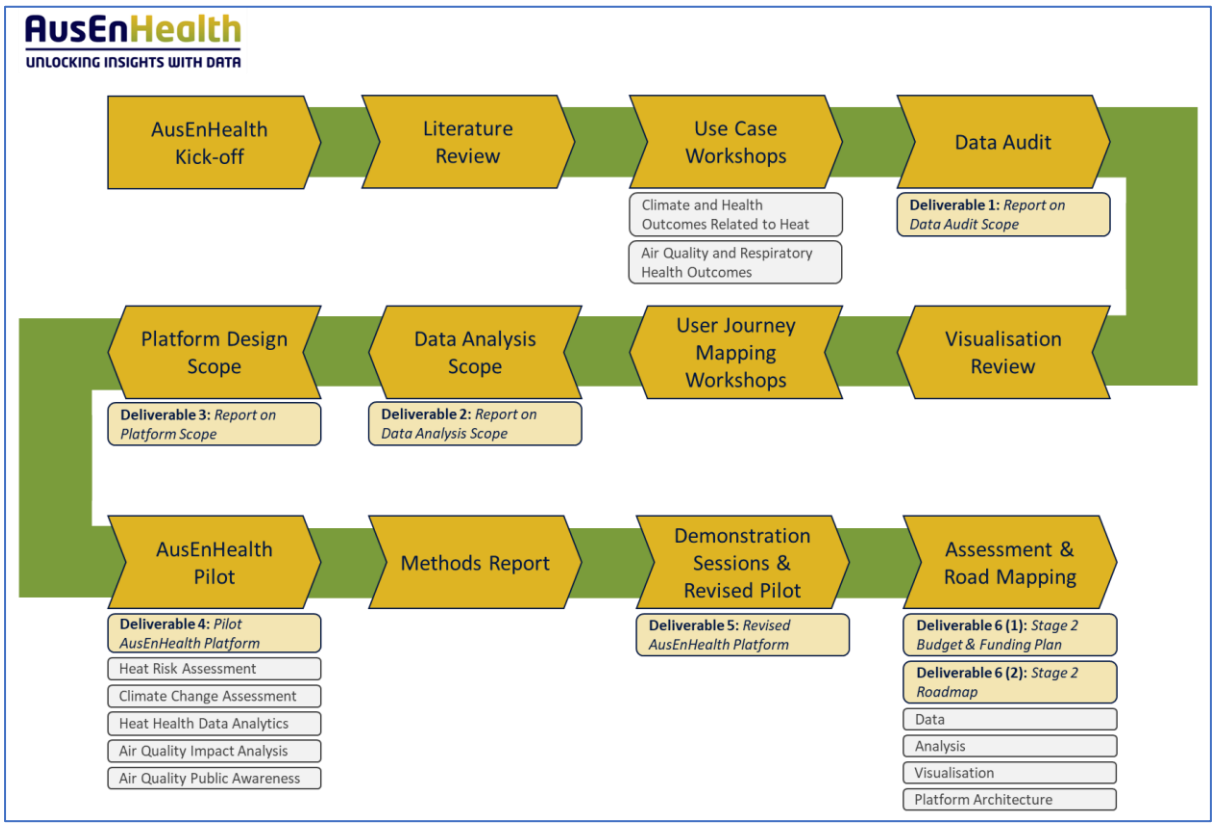

Figure 1: Activity Pathway for AusEnHealth Stage 1.

2. Data Audit

It is critical that the first phase output is capable of demonstrating high value to ensure future development is supported. To achieve this, one of the first activities of AusEnHealth was to determine the scope of the pilot, including identifying and selecting the highest priority areas of environmental health (use cases) upon which to build the pilot and identifying target data sources. A use case is defined in this report as an exposure domain and at least one of its associated health outcomes, described by health indicators. Some examples of use cases include:

- Air quality and asthma,
- Air quality and respiratory health outcomes,
- Water quality and gastrointestinal disease, and
- Land contamination and health outcomes related to lead exposure.

The grouping of exposure and health indicators is used to split the broad topic of environmental health into tangible sections, which can be prioritised for implementation into the AusEnHealth application. The use cases mentioned in this report have been determined through the review of similar projects and research papers in environmental health both in Australia and internationally. This section of the report covers the approach taken to identify the pilot's two key environmental health use cases, outlines the processes taken to identify data sources, and notes the key discussions held with relevant data custodians and environmental health experts to decide on the datasets pursued in the creation of the AusEnHealth pilot.

2.1. Use Case Workshop

AusEnHealth's *Use Case Workshop* was held on the 4th of December 2020 and was the first use case prioritisation task. For this workshop, AusEnHealth's Advisory Group, Delivery Team, and other contacted experts in environmental health met to decide on five key use cases. Initially, the workshop attendees were asked to rank exposure domains for further consideration. The results of which are displayed below:

Exposure Domain	Rank
Air Quality	**2**
Climate, Weather & Extreme Weather	**1**
Water Quality & Quantity	**3**
Food Safety & Supply	6
Vector Distribution & Ecology	**4**
Built Environment	**5**
Land & Land Contamination	7

From this ranking activity, Food Safety & Supply and Land & Land Contamination were excluded for this phase of AusEnHealth. The remaining exposure domains were then expanded to potential use cases: a combination of exposure domain and selected health outcome(s) based on environmental health literature (with some added through discussion during the workshop). The use cases were then discussed based on identified data sources at the time of the workshop and were further considered using the following criterion:

- Can data related to the use case be collected on a national scale?
  - *Availability:* Is data available and accessible?
  - *Reach:* Can data be attained from multiple jurisdictions?
- How important is the use case?
  - *Clarity:* Are there clear links between environmental indicators and health indicators?
  - *Urgency:* Does the use case address a significant or emerging health hazard?
- How feasible is the use case?
  - *Timeliness:* Is this use case feasible to explore in this first phase (12 months)?
  - *New Knowledge:* Would including this use case enable the ability to discover new knowledge that might otherwise go unchecked?

This criterion is the result of consideration of the literature [7], [8] and how it applies to AusEnHealth. Figure 2 shows an example slide used to facilitate discussion on potential air quality use cases during the use case workshop.

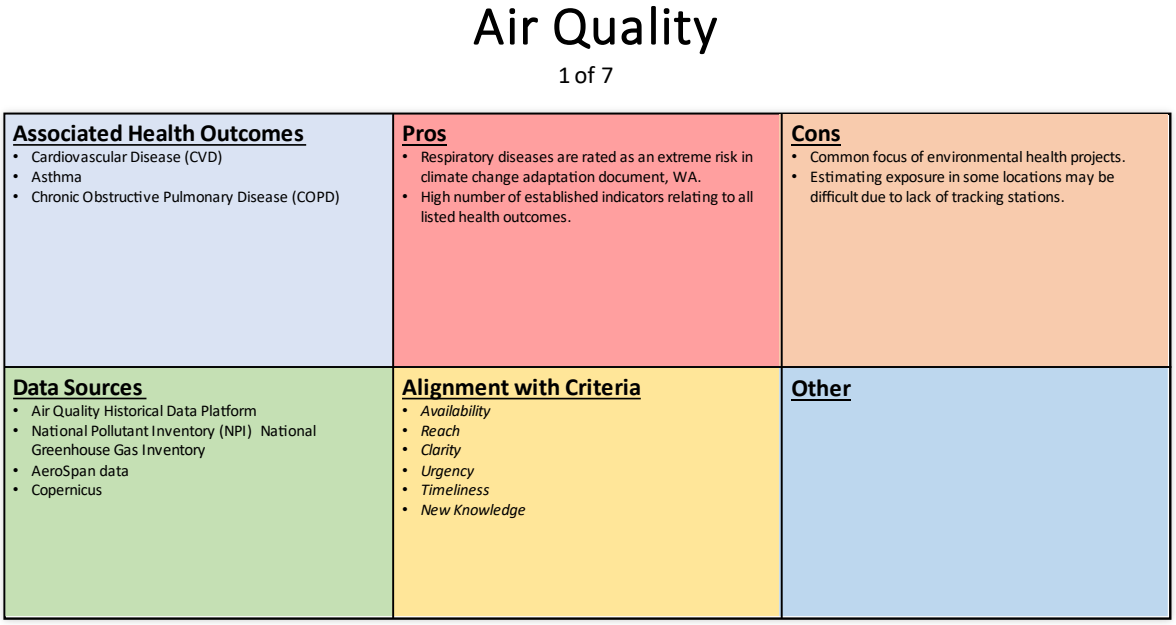

Figure 2: Example slide from use case workshop discussion.

Post discussion, each attendee was asked whether each use case should be considered further, with *yes*, *maybe* and *no* response options. To summarise these responses, a *score* was attributed to each response, with a *yes* response given a score of two, a *maybe* response a one, and a *no* response a zero. The results were used to create three ranking methods (a sum of the *maybe* and *yes* responses, a sum of the *yes* responses and an average score), which were used to arrive at the following five key use cases:

1. Climate and Health Outcomes Related to Heat
2. Built Environment and Health Outcomes Related to Heat
3. Water Quality and Water Borne Disease
4. Air Quality and Asthma
5. Climate and Mosquito Borne Disease

All considered use cases are listed in Appendix C, Table 16, as well as their rankings using the various ranking methods. The national data audit activity began directly following this workshop, with a focus on the above five key use cases.

2.2. Related Projects

There are two important objectives of AusEnHealth. The first objective is to enable improved access to national data relating to environmental health. The second objective is to provide models and visualisations, which enable the interpretation of data for decision making, essentially making raw data meaningful. Data platform development, in a general sense, is by no means a new concept. In fact, there have been many successful national and international projects which focus on both data provision and visualisation [9, 10, 11], including projects with an environmental health scope [12, 13, 14, 15].

AusEnHealth is not the first Australian effort to provide a resource aimed at improving access to environmental health information on a national scale. In 2010, a report titled *Investigation, Scoping and Planning of a National Environmental Health Tracking System* was produced by the WA Department of Health **Environmental Health Directorate** (commissioned by enHealth) [7]. This scoping project shares a large number of goals with AusEnHealth and provided insight into the

difficulties that would be faced in the pursuit of a national environmental health platform, including data custodian and stakeholder perception, as well as data access and interpretation. The report recommended the development of a tool which focuses on the collection, analysis and reporting of "*core information relative to the practice of environmental health in Australia*". To ensure the success of projects in the future, the report notes that "*negotiation with data custodians will have to occur around specific datasets and will have to be negotiated around individual topics and end uses*". The findings and recommendations from the work in [7] were carefully considered throughout the course of AusEnHealth to ensure that the resulting AusEnHealth pilot has the potential to be maintained and expanded in the future.

2.3. Indicator Development Frameworks

The aforementioned enHealth report [7] also identifies the need for consistent standards for the recording and reporting of information on a national scale. To help achieve this, the creation of nationally recognised environmental health indicators is also recommended. Indicator development literature has been considered in AusEnHealth, which referred to several frameworks. These include:

- Environmental public health conceptual framework (EPHCF) [16],
- Driving force-pressure-state exposure-effect-action (DPSEEA) [17],
- Driving force-pressure-state-impact-response (DPSIR) [18],
- Pressure-state-response (PSR) [19], and
- Driver-pressure-environmental condition-health impact-action (DPEHA) [20].

Indicators and Parameters in the Literature

As the development of indicators is a large task, AusEnHealth instead relies on existing literature to determine indicators and parameters of interest. A number of these sources rely on the above indicator development frameworks [1, 21, 22, 9, 23], while other sources refer to research literature [24, 25] and government-backed reports [26, 27], which state indicators and parameters without reference to existing development frameworks.

Information has been collected from the above sources to produce an indicator summary document, which can be found at the AusEnHealth website (https://ausenhealth.com/home). This document lists the exposure domains, indicators, and variables noted in each of the sources discovered prior to AusEnHealth's use case workshop activity. For details of the indicators used in the AusEnHealth pilot, see Section 3.3.

2.4. Approach to Data Audit

The goal of this data audit is to identify valuable data sources, as well as to better understand the landscape of Australian environmental health data collection at both state and national levels. The indicators and parameters from the literature have been used as a starting point in the identification of Australian sources of environmental health data in AusEnHealth's first phase. The process undertaken involved (a) searching for terms similar to those identified in the indicator summary document, (b) auditing the results based on data quality and accessibility, and (c) consultation with data custodians to discuss potential data gaps.

This process of identification and recording of environmental health metadata is limited by its online nature. Attempts have been made to contact a number of data custodians to discuss relevant data that may not be noted online. Resulting discussions have revealed that some custodians sell data services which include temporal and spatial resolutions and time series data not otherwise available. Over time, these additional options will be noted in the live data catalogue as knowledge of existing data continues to develop.

Data Search

Each of the known parameters and indicators were sought via four main data repositories for Australian environmental health data:

1. **Data.gov.au**: A national aggregation of data sources used by governmental organisations. For each data source shown on the platform, the source location of that data was searched until the access point could be found. Many of the Australian national data repositories were identified through this method, including BOM, ABS, and Geoscience Australia.
2. **Google**: A standard google search was undertaken for each indicator to locate sources outside the Australian data repositories. International organisations with global datasets were found through this method, including LANCE-MODIS, CRU.TS, and TerraClim.
3. **National Map**: National Map is an aggregator of publicly available data repositories and is especially useful for identifying potential sources of built environment data.
4. **Research Articles**: A short search using Google Scholar for data repositories was undertaken to discover recently developed datasets for any of the parameters in the indicator document.

Each of these searches were aimed at determining the access point for the dataset specific to each parameter. Each of the sources identified in this step were then audited.

Data Source Auditing

The following criteria were used to determine the quality of the data source:

1. National Coverage: The data sets must cover at least the Australian mainland (including Tasmania) or be able to be merged from state level datasets.
2. Current: The data must fall within the year range of 2000-Present. Additionally, there must be sufficient data in each year.
3. Spatially Precise: Preference was given to point data or rasters with high resolution (0.07° or better).
4. Accessible: The data must be freely accessible (except in the case of BOM derived variables) and publicly available (or available through agreement).

Of over 100 data sources identified, roughly 50 have been listed as suitable sources for the AusEnHealth pilot. Once a data source was identified for a parameter, a metadata listing was then created in the data catalogue with a range of attributes, including data custodian, access website link, timescale, update frequency, spatial resolution, and data file format, among others.

2.5. Data Catalogue

Throughout the data audit activity, information (metadata) relating to data and custodians were recorded. This section displays this information, separated into data custodian information, key data identified by AusEnHealth's project team and other data which may be relevant to the five key use cases discussed in Section 2.1 of this report.

As it is likely that data sources may change in the future, a live version of the data catalogue has been created and can be found at https://frontiersi.com.au/ausenhealth-digital-twin. This live catalogue accepts contributions from the public and will be updated on a regular basis (see the Ongoing Data Collection and Collaboration Section of this report).

Data Custodian Information

There are a large number of environmental health data custodians in Australia, each with different practices and goals. Throughout the search process, our understanding of key data and data sources for visualisation and modelling expanded. As a result, a list of key data custodians who provide a

significant portion of environmental health data at a national level was established. These are presented alongside their motivation for use in AusEnHealth in Table 1.

Data Custodian	Motivation for Use
Australian Bureau of Statistics (ABS)	High resolution point data for demographics.
Bureau of Meteorology (BOM)	High resolution temporally and spatially.
Geoscience Australia (GA)	Provides high quality, built environment data and analysis-ready satellite data through their Digital Earth Australia platform.
Centre for Air Pollution, Energy and Health Research (CAR)	Combines data from various air quality measuring platforms to provide modelled results that minimises the weaknesses of individual platforms.
Australian Urban Research Infrastructure Network (AURIN)	Provides high quality, built environment data.
Australian Institute of Health and Welfare (AIHW)	Provides useable data at SA3 resolution for health outcomes.
National Notifiable Diseases Surveillance System (NNDSS)	Provides useable data at postcode resolution for health outcomes.
NASA Earth Observations (NEO)	Provides medium resolution global rasters that can be used as a backup for modelling.
Copernicus Atmosphere Monitoring Service	Provides medium resolution global rasters for a large variety of air pollutants.

Table 1: List of Key data custodians.

A number of other sources that contain similar data have been considered and are included in this report's data catalogue in the Secondary Data Sources Section. These sources either provide data at a lower spatial or temporal resolution or provide less data overall than those listed in Table 1.

2.6. Key Data Properties

The data catalogue is split into three sections. The first section contains specific parameters from the key data custodians listed in the previous section, which are displayed both online and in Appendix D. In addition, a heatmap is provided in Appendix E to highlight data availability. Included in the list of key data are health outcomes, some of which are separated by International Statistical Classification of Diseases and Related Health Problems, Tenth Revision, Australian Modification (ICD-10-AM). These codes are valuable for use in modelling health outcomes for particular use cases and should be utilised where possible.

Secondary Data Sources

The second section of the data catalogue is a list of secondary data sources and the data they provide. In the event that a key data source is no longer able to provide access to certain parameters, it is important to have secondary sources on record. Some of these sources provide additional parameters relevant to environmental health, all of which are recorded and provided both online and in Appendix D, Table 18.

Other Data Collections

A number of additional data collections have been identified through the data audit, which contain environmental health datasets. Some of these datasets may be useful for other projects relating to environmental health, as well as future phases of AusEnHealth. Information on these data collections is also provided both online and in Appendix D, Table 19.

Ongoing Data Collection and Collaboration

It should be stated again that the collection of data and resources within AusEnHealth's project team does not end with this report. As previously mentioned, the data catalogue will continue to be updated online, and contains names of data providers, data categories and lists of parameters. This resource (screen capture displayed in Figure 3) also displays website links where available, functioning in essence as an environmental health focused data repository. Note that the online data catalogue may differ significantly from the catalogue listed in this report in future, both in datasets listed as well as format.

Category	Parameter	Access	Metadata Link	Source	Curator	Timescale
Climate	Temperature	WEBSITE	LINK	BOM Climate Maps	Bureau of Meteo	1910 -
Climate	Rainfall	WEBSITE	LINK	BOM Climate Maps	Bureau of Meteo	1900 -
Climate	Extreme Temperature	Upon Request	LINK	ACORN-SAT	Bureau of Meteo	2000-
Climate	Extreme Rainfall	Upon Request	LINK	ACORN-SAT	Bureau of Meteo	2000-
Climate	Solar Exposure	WEBSITE	LINK	BOM Climate Maps	Bureau of Meteo	1990 -
Climate	UV Index	WEBSITE	LINK	BOM Climate Maps	Bureau of Meteo	2000-
Climate	Humidity	WEBSITE	LINK	BOM Climate Maps	Bureau of Meteo	1971 -
Climate	Southern Oscillation Index	WEBSITE	LINK	BOM Climate Maps	Bureau of Meteo	1870 -
Climate	Temperature	WEBSITE	LINK	Air Temperature from 21 Weather Stations	UNSW CFRC	2016-2019
Climate	Floods	WEBSITE	LINK	Landsat (Water Observations from Space)	Geoscience Austr	1986 -

Figure 3: Online data catalogue example layout.

In addition to ongoing maintenance of the online resource by AusEnHealth's project team, readers are invited to contribute environmental health data sources, reports, research papers, etc. A contribution can be made using an online form (see Figure 4) available within the data catalogue.

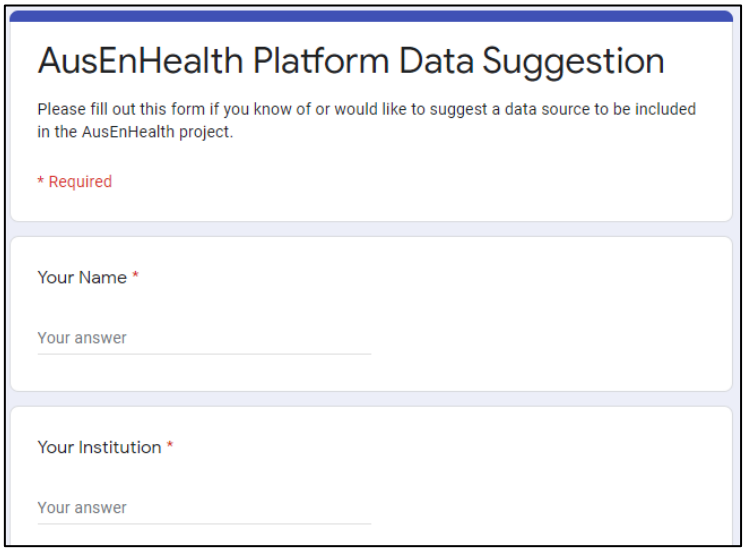

Figure 4: Public contribution Google Form.

Submissions made using this form will be checked regularly and utilised by AusEnHealth's project team once the contributions themselves are audited for quality. Any contributions made in this way are greatly appreciated, as new information will help to keep the AusEnHealth pilot's data up to date in the future.

2.7. Summary Findings (Data Audit)

The data audit activity successfully identified a range of data custodians which allow open access to a range of environmental health data. This activity also identified a significant number of gaps and opportunities. A summary of findings, gaps, opportunities, and recommendations for the data audit are presented below. Recommendations arising from this summary can be found alongside all other project recommendations in Appendix A.

Health Outcome Data

In an ideal environment, the pursuit of health outcome data would involve a searchable catalogue which shares data custodians and related metadata both nationally and by jurisdiction and allows for open access to various temporal and spatial resolution data. Health outcome data access in its current state is heavily restricted by privacy concerns and governance. In the AusEnHealth Project, this has led to limited insight into the metadata recorded at state level, with understanding gained through either project partners, such as WA Health or AIHW, or through national data custodians, such as NNDSS and ABS. One example of an existing metadata repository is AIHW's METeOR [28], which holds metadata for a range of AIHW datasets and is publicly available. However, the minimum datasets held by AIHW are available only by request and may require special approval or is available only through tables prepared by AIHW, for which charges may apply. The development of a public Australian health data catalogue with a supporting metadata catalogue would greatly improve the flow of information to decision makers, as well as improving the visibility of the state of Australian health outcome data.

Additionally, where data requests are a necessity due to privacy concerns at either the state or national level, the creation of a system which clearly lays out the data access process, including data availability and data request requirements, would be of significant benefit to future health outcome related projects.

Climate Data

As a broad finding, climate data is the most straightforward and complete data identified in this audit. While features such as extreme temperature and extreme rainfall are listed as paid options, the temperature and rainfall parameters themselves can be used to derive such features. However, it is acknowledged that climate includes a much larger range of variables than the parameters included in this audit.

Air Quality Data

For the first phase AusEnHealth pilot, there are a range of custodians to consider when pursuing national air quality data. The air quality data recorded in this data audit include modelled results from the Centre for Air pollution, energy, and health Research (CAR). This modelled data is produced through research funding and is not ongoing. Using this data source in the short term is acceptable, however, alternatives (such as Copernicus air quality data) should be considered in the event that CAR stops producing modelled data.

While the modelling aspect of the air quality data can be repeated, there are potential issues with fine temporal modelling. Air quality data is recorded by many states using rolling averages, which protects from equipment malfunction but inhibits real-time analysis. If this data were released without smoothing, more complex models could be achieved and at a more sensitive time scale.

Water Quality Data

While the parameters identified in this audit are at a fine scale and are accessible, there are many absent parameters which, if accessible, would enable modelling of water borne disease. Some examples of desirable parameters for modelling include pH, dissolved oxygen levels and chemical

compound amounts such as copper, lead, zinc, or iron. We recommend future work to bring water quality monitoring to the same standards as climate data in preparation of potential spikes in water borne disease incidence.

Built Environment Data

There are a range of built environment parameters which exist on a national scale. These parameters share a number of limitations: a large number of the national parameters have only been recently developed and consist of data from 2020 onwards, the parameters are mostly one point in time compared to the time series data recorded in other exposure domains, and the update frequency for many parameters is yearly. It is clear that built environment indicator development is ongoing. The main recommendation relating to built environment data is to continue recording time series data to allow for more complex analysis to be performed.

All Data

When considering national data in Australia, a common issue communicated by national data custodians is differences in formatting and collection across states. If each state adapted to a single agreed framework for the collection or recording of state-based data, repeated data collation and cleaning efforts would no longer be a strain on analysis groups and comparison of data between states would become more straightforward. On a similar note, metadata collected for various environmental health categories (down to indicators and their parameters) differs by state, which makes preparing national data infeasible for some parameters. National data standards enable seamless cross-jurisdictional data comparisons and would allow for increased collaboration between researchers and policy makers in the future. Consideration of international data standards is a clear extension of this opportunity, which would lead to increased global collaboration.

3. Data Analysis Scope

The AusEnHealth Pilot relies on data to fulfill its design capabilities as a strategic planning digital twin. As data is a crucial aspect of the platform, the way in which data is integrated, analysed, and presented is of critical importance. Development of these processes have been shaped in many ways throughout AusEnHealth's development. The data audit process has identified Australian variables related to environmental health and the two user journey mapping workshops have identified end users and scenarios to guide development. In addition, an evidence-based approach to data selection and visualisation has been carried out through careful consideration of literature related to environmental health, as well as through consultation with various experts in the areas of climate and air quality.

This section begins by noting the data chosen from the data audit for implementation into the AusEnHealth pilot. This section then covers the various data handling aspects of the project, including data integration methods, formatting choices (e.g., imputation and aggregation methods employed), and metadata recorded for each data source. Finally, this section provides a comprehensive breakdown of environmental health variables and indicators, and vulnerability indices created for use in the AusEnHealth pilot.

3.1. Data Selection

A large volume of both national and international data has been identified through the project's data audit process. Data has since been considered for implementation into the AusEnHealth pilot based on relevance, quality, and impact relating to user scenarios identified through the user journey mapping workshops (covered in Section 5.1). In this section, all datasets incorporated into the pilot are listed by broad category, alongside a brief justification for inclusion.

Climate Data

The highest priority use case in the AusEnHealth pilot was determined to be climate and health outcomes related to heat. Over the 20th century, global surface temperature has risen by 0.6°C, with the warming trend from 1950-2000 being nearly twice that of the first half of the century [29]. This increase in overall temperature is mainly due to a slow increase in minimum daily temperatures, which will continue to increase in the 21st century, with an estimated further increase of at least 2-3°C by the year 2100 [30]. Australia does not avoid the global trend of increasing temperature, with all climatic models showing an Australian-wide warming trend [31]. It is becoming extremely important to adjust to changes in climate, due to the impact of prolonged extreme heat exposure on human health [32] [33]. These impacts include heat exhaustion, heat stroke, exacerbation of various respiratory, cerebral, and cardiovascular diseases, and death [34]. One of the most important issues facing Australia is understanding how to adapt to a changing climate [31]. Extreme heat is the focus of the majority of research investigating the relationship between climate and human health. However, cold-related mortality is shown to exceed heat related mortality both internationally [35] [36] and in Australia [37], especially when considered alongside increased respiratory conditions [33] [38]. As a result, both extreme heat and extreme cold are considered as part of the AusEnHealth Project.

Prior to AusEnHealth's user journey mapping workshops, various climate data were collected, including southern oscillation index [39], daily precipitation (mm), maximum daily temperature (°C), minimum daily temperature (°C), daily mean vapour pressure (hPa), solar radiation, and UV index. These variables were all sourced from the Bureau of Meteorology (BOM) [40], the official Australian data custodian for weather forecasts and observations. AusEnHealth's user journey mapping workshop revealed a priority in the area of climate risk assessment, with a specific focus on identifying communities most at risk from extreme heat. Through consultation with AusEnHealth advisory group members and relevant literature linking climate exposure data with health outcome data, this list was refined to the following:

Variable	Temporal Properties	Geographical Properties	Data Source
Maximum Temperature (°C)	2000-2020 (daily)	Australia (0.05° raster)	BOM
Minimum Temperature (°C)	2000-2020 (daily)	Australia (0.05° raster)	BOM

Table 2: Climate variables collected for use in the AusEnHealth pilot.

Due to rising trends in temperature and widespread knowledge of heat-related health outcomes, both minimum and maximum temperature are valuable variables to include in the platform to indicate heat exposure across Australia. Vapour pressure was initially included for its role in the calculation of apparent temperature [41] [42], which is demonstrated to have an impact on human health [43] [44]. However, wind speed is also required for the calculation of apparent temperature. While Copernicus [45] offers wind speed data, its geographical resolution is not consistent with climate data, and values are estimated. Out of concern for potentially inaccurate apparent temperature calculations, both vapour pressure and wind speed were not included in the pilot.

Researchers from the Australian Cancer Council advised against the consideration of cancers in the first phase of the AusEnHealth Project due to the complexity of evaluating the lag between exposure and cancer incidence. As a result, both solar radiation and UV index (two variables associated with cancer incidence) have been excluded from the pilot. Southern oscillation index, which indicates heating and cooling cycles of the Pacific Ocean and the atmosphere above it, has also been excluded from the pilot as the project team did not identify evidence in the literature linking this variable to climate related health outcomes. While consecutive time periods with low or high precipitation can be indicators of drought or flooding events, respectively, this climate variable was not identified by the project team as having a direct link to other health outcomes related to heat. Precipitation is a

valid environmental health variable [46], however, it does not suit the *climate and health outcomes related to heat/cold* use case, nor the scenarios outlined in the user journey mapping workshop.

Air Quality Data

From the AusEnHealth use case workshop, the air quality and respiratory health outcomes use case was selected by vote as the second highest priority use case. Air quality is determined by the measurement of several air pollutants, which are introduced into the air in a variety of ways, including road traffic, industrial processes, and bushfires. In fact, while Australian air is generally among the cleanest in the world [47], pockets of poor air quality are said to be responsible for the deaths of roughly 3000 Australians per year [48].

Australia has its own set of standards for air quality, dictated by the National Environment Protection Measure for Ambient Air Quality (Air NEPM). The two pollutants which most frequently exceed these standards are ozone (O3) and particulate matter less than 2.5µm in diameter (PM2.5). While there is a consistent effort made to monitor air quality across Australia [49], there is a renewed interest in the impact of air quality on human health stemming from the devastating 2019-20 bushfires. A resource such as AusEnHealth would enable policy makers to make timely, evidence-based decisions and better assess population vulnerability utilising changes in air quality trends.

After careful consideration of geographical and temporal resolutions for national air quality data, Copernicus [45] was chosen as the primary data source for nitrogen monoxide, nitrogen dioxide, sulphur dioxide, ozone, carbon monoxide and particulate matters <10µm, <2.5µm, and <1µm. These variables are provided by and were created by combining model data with global air quality observations. The user journey mapping workshop which focused on air quality led to the design of scenarios focused on impact analysis, public awareness, and campaign awareness for air quality. Consultation with researchers from CAR, coupled with information directly from the Copernicus data custodian and other relevant literature led our team to include all of the air quality parameters in our pilot. The air quality variables included are as follows:

Variable	Temporal Properties	Geographical Properties	Data Source
Nitrogen Monoxide, NO (kg kg^{-1})	2003-2020 (3-hourly)	Australia (0.75° raster)	Copernicus
Nitrogen Dioxide, NO2 (kg kg^{-1})	2003-2020 (3-hourly)	Australia (0.75° raster)	Copernicus
Sulphur Dioxide, SO2 (kg kg^{-1})	2003-2020 (3-hourly)	Australia (0.75° raster)	Copernicus
Carbon Monoxide, CO (kg kg^{-1})	2003-2020 (3-hourly)	Australia (0.75° raster)	Copernicus
Ozone, O3 (kg kg^{-1})	2003-2020 (3-hourly)	Australia (0.75° raster)	Copernicus
PM10 (kg m^{-3})	2003-2020 (3-hourly)	Australia (0.75° raster)	Copernicus
PM2.5 (kg m^{-3})	2003-2020 (3-hourly)	Australia (0.75° raster)	Copernicus
PM1 (kg m^{-3})	2003-2020 (3-hourly)	Australia (0.75° raster)	Copernicus

Table 3: Air quality variables selected for use in the AusEnHealth pilot.

Built Environment Data

The way in which cities are designed and built can enable or restrict human activity and can lead to populations being exposed to or protected from extremes in the natural environment [50]. It is important to consider the difference that urban environment makes to population vulnerability. Recently, technologies have enabled an increased capability to measure and collect data which improve the understanding of urban and regional population health. With both motivation and capability, projects such as AusUrbHI [51] and Digital Twin Victoria [52] aim to improve the link between built environment and human health through the creation of new data and identifying how to best present data to inform government, industry, education, and the general public.

A wide range of data has been collated in the built environment data category through collaboration between the AusEnHealth Project and the Australian Urban Research Infrastructure Network (AURIN). The user journey mapping workshop on *climate and health outcomes related to heat/cold* revealed interest in utilising urban variables in the AusEnHealth Project. Through ongoing consideration of research literature for both climate and urban impacts on heat risk, and through ongoing discussion with AURIN and other project teams currently working on heat risk assessment, a range of urban variables were collected from the Australia Bureau of Statistics (ABS) [53], Geoscience Australia's Digital Earth Australia (DEA) [54], NASA's Moderate Resolution Imaging Spectroradiometer (MODIS) [55], and the Terrestrial Ecosystem Research Network (TERN) [56]. The variables collected are given in the following table:

Variable	Temporal Properties	Geographical Properties	Data Source
Hospital Counts (per population)	2021 (yearly)	Australia (SA2, LGA)	ABS
Green Space (% parklands)	2016 (yearly)	Australia (SA2, LGA)	ABS
Water Bodies (%)	2019 (yearly)	Australia (shapefiles)	DEA
Vegetation Index (NDVI)	2000-2019 (monthly)	Australia (0.05° raster)	MODIS
Canopy Cover (%)	2000-2021 (seasonal)	Australia (25m grid)	TERN

Table 4: Built environment variables selected for use in the AusEnHealth pilot.

A similar set of built environment variables is listed in a recent BOM report which presents a heat vulnerability index [57]. Notable differences between the list provided in [57] and the above list of variables are cool places and health services data. These data were not available from the data custodian (AURIN) at the time of writing and are otherwise available only through individual jurisdictions. Due to the short time-frame of the AusEnHealth Project, this data was unable to be acquired.

Demographics Data

Understanding population characteristics around Australia enables government and communities to determine areas which may be more vulnerable to natural hazards than others both socially and economically [58]. Demographics data has been utilised in numerous studies which demonstrate associations between vulnerable populations and increased mortality and morbidity. Every five years, the ABS holds a Census [53] which records a wide range of demographic variables. The following table represents the Census datasets utilised in the AusEnHealth pilot. For the specific data used in the platform, see Section 3.3 (Data Analysis).

Dataset	Temporal Properties	Geographical Properties	Data Source
ABS_Pop_2014-19	2011-2019 (yearly)	Australia (SA2, LGA)	ABS
Income_2014-19	2014-2017 (yearly)	Australia (SA2, LGA)	ABS
EducationEmployment_2014-19	2011, 2016 (yearly)	Australia (SA2, LGA)	ABS
FamilyCommunity_2014-19	2011, 2016 (yearly)	Australia (SA2, LGA)	ABS
HealthDisability_2014-19	2011-2018 (yearly)	Australia (SA2, LGA)	ABS
EconomyIndustry_2014-19	2014-2019 (yearly)	Australia (SA2, LGA)	ABS
Motor Vehicles by Dwellings	2011, 2016 (yearly)	Australia (SA2, LGA)	ABS
Social Health Atlas of Australia	'11, '14, '17 (yearly)	Australia (LGA, PHA)	ABS, PHIDU

Table 5: Demographics datasets utilised in the AusEnHealth pilot.

In addition to the ABS Census data, existing health condition data is utilised from the Public Health Information Development Unit (PHIDU) social health atlas [59] [60] to highlight vulnerability present in communities through existing conditions such as asthma or risk factors such as high blood pressure.

Health Outcome Data

The fundamental objective of the AusEnHealth Project is to provide insight into the impacts of the changing environment on human health. It is important, therefore, to consider health outcome data as part of this project. It is well understood that air quality and climate extremes (the two key use cases considered in the first phase of AusEnHealth) have significant impacts on human health [61] [62] [63] [64]. These impacts are identified by applying carefully considered statistical models to environmental health data alongside health outcome data. Relevant health outcome data is identified using a list of disease classification codes, known as the International Statistical Classification of Diseases and Related Health Problems (ICD), adopted by the World Health Organisation in 1948 [65].

ICD codes are revised roughly every 10 years. In 1992, the 10th edition of this list was released for public use. In 2019, an 11th edition was accepted, with implementation planned for 2022 [65]. The ICD codes used in Australia currently use ICD-10-AM codes (the Australian modification), which was implemented in Australia in 1999 [66]. A review of ICD code usage in the literature was undertaken to identify relevant health outcomes to the AusEnHealth Project's climate and air quality use cases. Key health outcomes identified in this review include:

- Asthma,
- Chronic obstructive pulmonary disease (COPD),
- Cardiovascular disease,
- Ischemic heart disease, and
- Cerebrovascular disease.

For a full list of identified health outcomes, see Appendix F.

Process for Health Data Access

AusEnHealth's data audit activity revealed that health outcome data (e.g., hospitalisations, emergency department presentations, and mortality) are either not available on a national level or are available at geographical or temporal resolutions which do not enable straightforward statistical analyses or comparisons to be made due to privacy restrictions. Some examples of public health outcome data include AIHW mortality over regions and time (MORT) books (5-yearly mortality data at SA3 and LGA geographical resolution) [67] and AIHW's 2019-2020 Australian bushfire report data (weekly emergency department presentation and mortality data at SA4 geographical resolution) [68]. To enable the assessment of environment-related health outcomes at improved temporal and geographical resolutions, it is necessary to request health outcome data from one of the Australian health data custodians.

As jurisdiction-specific data requests were infeasible due to the timing of the AusEnHealth Project, data was requested through AIHW. This process involved initial correspondence to understand which data is potentially available and the processes required to acquire data from AIHW. The ICD code summary provided in Appendix F have been created in part thanks to these discussions with AIHW. Requests for hospitalisation data and mortality data are handled by different teams within AIHW, each with different requirements. Prior to a formal request for hospitalisation and mortality data, it is required that certain ICD code lists are supplied to AIHW, with hospitalisations requiring ICD-10-AM codes, mortality requiring ICD-10 codes. The process for arranging these code conversions was lengthy, and involved correspondence with AIHW's population health unit, QUT's school of public health and social work, and AIHW's classification team. Two AIHW interns handled the ICD code conversions over a period of 6 weeks and provided additional information relating to external cause codes to support the data requests.

Formal submissions to AIHW for hospitalisation and death data were for 6-monthly (2010-2020) SA3 (across Australia) datasets, split into three age groups (<19, 19-64, 65+). The ICD conversion process revealed a number of other relevant ICD codes to include in the request. A complete set of ICD-10-AM, ICD-10, and external cause codes are provided in the external document "*EX. AusEnHealth ICD Breakdown*". The ICD-10-AM and ICD-10 codes were grouped into categories, which are provided in the following table.

ICD Code Grouping	Mortality	Hospitalisations
Respiratory Health	x	x
Heart Health		x
Brain Health		x
Heat Health	x	x
Cold Health		x
Burns		x
Deprivation	x	x
External Causes, Environmental	x	x
External Causes, Social		x

Table 6: Categories or groupings of ICD codes used in the AusEnHealth pilot.

A request for the same health outcome data has also been submitted to the Western Australia Department of Health (WADOH) for LGA geographic resolution data. With the health outcome data requests from AIHW and WADOH, the list of health outcome data sought throughout the AusEnHealth Project are as follows.

Variable	Temporal Properties	Geographical Properties	Data Source
MORT Books: Deaths	2014-2018 (single)	Australia (SA3)	AIHW
2019-20 ED Visits, Deaths	2017-2019 (weekly)	Australia (SA4)	AIHW
AIHW Hospitalisations, Deaths	2011-2019 (6-monthly)	Australia (SA3)	AIHW
WADOH Hospitalisations, ED Visits, Deaths	2000-2019 (6-monthly)	Australia (LGA)	WADOH

Table 7: Health outcome datasets selected for use in the AusEnHealth pilot.

Environmental Exposure and Health Outcome Associations

Care must be taken when drawing associations between environment and health variables due to; (a) the complexity and enormity of environmental health interactions, and (b) the presence of ecological biases, such as the modifiable areal unit problem (MAUP) [69]. The analysis of health outcome data is approached in two main ways in environmental research literature. The first approach is a typical statistical analysis, where environmental variables are assessed against a health outcome or health outcomes of interest [70] [71] [72]. The second approach is the development of indicators or indices, which are based on relevant literature, rather than being directly assessed against one or more health outcomes [73] [74] [75] [76] [77] [78]. Occasionally, indicators developed in this way are then compared against health outcomes [57] [79] [80] to assess how effectively said indicators reflect risk.

AusEnHealth has taken the second approach in the analysis of health outcome data, where important information is drawn from air quality and climate data and presented as an indication of risk separately from health outcome data. The AusEnHealth Project aims to also create, and present vulnerability indices related to air quality and climate. For these indices, a middle-ground approach is taken. First, an extensive literature review, combined with the data audit, determined which variables are suitable for use in the development of vulnerability indices (i.e., known links to health outcomes or population

vulnerability). Then, the variables were combined utilising weights based on correlations between acquired index variables and available health outcome data.

3.2. Data Handling

Once data was identified and selected for inclusion, it then needed to be collected, collated, and formatted for use in both the analysis component of the project and for visualisation in the AusEnHealth pilot. This process is known as data integration. There are a range of options available for data integration, the selection of which is based on end user requirements. This section notes the considered data integration approaches, as well as the formatting processes carried out to enable the creation of indicators and indices.

Data Integration Techniques

Data integration is a process of acquiring data from different sources and combining them to make them compatible with one another to facilitate tasks, such as analytics and visualisation, for decision making [81]. Data integration typically involves multiple steps, such as pulling data from various sources, followed by data cleaning, and finally merging them into a database. The following factors may be considered when deciding on an appropriate data integration process for a research project.

a) **Extract, Transform, Load** [82]: These tasks collect, process, and upload data from different sources and ensure a clean and consistent database. The AusEnHealth project requires data in multiple formats and varying complexities that need to be uploaded and processed. Hence, there is a need to design and develop methodologies to extract individual datasets into a local data frame. These local data frames will be then used to clean the incoming data using imputation techniques. The next step will be to transform the columns or features in the datasets as per the modelling requirements to create indicators/indices. Some of the datasets are updated frequently, and hence an automated ETL workflow will be the best solution in the long term. Without an ETL process, data preparation for analytics may be ad hoc and consume significant time during each run. In addition, a good ETL process can ensure scalability as the data sources increase.
b) **Data Wrangling** [83] [84]**:** Apart from the traditional ETL processes, data wrangling deals with aspects such as maintaining the data quality, merging data from multiple sources, ensuring reproducible processes, and managing provenance. For a large volume of data, manual ETL processes may be expensive and labour-intensive. Data wrangling recommends flexible architecture and adaptive pay as you go solutions, which will be suitable to AusEnHealth platform.
c) **Change Data Capture** [85]**:** Currently, data is captured by various sources and manually extracted by the data officer/platform administrator when needed. For a dynamic ETL workflow, triggers can be introduced to check and pull data at regular intervals. This scheduling will depend on the frequency of data captured by the data source, e.g., temperature or air quality. Any extraction schedule might range from hourly to yearly, depending on how this is made available. These triggers can be added as part of the end-to-end data collection workflow, and additional processes may be needed to extract other data releases that may be ad hoc or bespoke.
d) **Data Replication and Storage** [86] [87]: The platform needs to consider some key design aspects that are necessary for the high performance of data-intensive applications. Data replication and storage are categorised under costs, availability, security, reliability, bandwidth, and response time. Platform capabilities need to be evaluated in regard to these various data replication characteristics to achieve the best performance possible under constraints such as project budget, skills, or infrastructure dependencies. Overall, data storage should be located as close as possible to the data analysis tools, thereby ensuring that data access/transfer can be as quick as possible for tasks such as processing or analytics.

e) **Data Virtualisation** [88] [89]: It helps to provide separation between data preparation and data analytics. Users can access the data through the portal, maps, dashboard, or reports without being concerned about data accumulation or pre-processing steps. As a result, data integrity, quality and availability are virtually separated from the end-users.

An automated ETL process is valuable in an operational product to eliminate manual tasks required for upkeep. A manual ETL process has been selected for AusEnHealth's first phase, as the initial AusEnHealth pilot will not frequently update data.

Automated Integration

This subsection outlines advanced technical implementation processes that may be applied in future work phases to improve the efficiency of data integration tasks and support analytics for real-time data. Even though the manual process takes less time to design, it is not scalable and impossible to manage in the longer run.

Data orchestration is an automated process that uses software solutions to collect, clean, transform, and merge data from multiple sources [90]. For AusEnHealth's first phase, two data orchestration tools were investigated to develop end-to-end data pipelines for use in the future: AWS Glue [91] and Apache Airflow [92]. Due to their advanced nature, both approaches require experts to build and manage their solutions and will have costs associated with resources and infrastructure.

In terms of development time, both systems consume time for designing the data preparation workflow. However, AWS Glue has many inbuilt features and is tightly integrated with other tools such as Amazon S3 buckets that were utilized in the pilot phase for data storage. Overall, this mix means that Glue has a steeper learning curve but is easier to develop. In contrast, Airflow has a flat learning curve being more or less standalone from other tools. It requires a developer to design and code the ETL processes using a series of steps, some of which are represented in DAGs (Directed Acyclic Graphs). Since the project deals with data from multiple sources and formats, there can be a significant increase in the development cost and time.

Overall, Airflow provides considerable flexibility, however, this flexibility comes with the cost of setting up the server and infrastructure around the Airflow ecosystem. The AusEnHealth pilot is currently utilising an existing AWS server, provided by the Office of eResearch at the Queensland University of Technology (QUT), for data storage. By leveraging the expertise and support available at QUT, implementing and managing AWS Glue is a feasible option for future data pipelines. However, longer term costs (e.g., supporting servers, data transfers, etc.) will need to be considered in future.

Since the AusEnHealth pilot project deals with data from multiple sources and formats, there could be a significant increase in the development cost and time if Airflow were adopted. Given that QUT has an existing relationship with Amazon with some project team members having direct experience with its functionality, AWS Glue is recommended over Apache Airflow as long as the project's budget can incorporate funds to maintain the service, both throughout the application's development, through its maintenance, and support activities. More details on the proposed architecture can be explored in Appendix A.

We have currently investigated two orchestration tools, however, there are many feasible data integration options (e.g., Luigi, Metaflow, Stitch, K2View, and Prefect). Should automated integration be required in future phases of the AusEnHealth Project, these tools can be explored to select the best approach. A proposed future data integration architecture is included in Appendix G.

Data Acquisition and Formatting Processes

The creation of indicators and indices on a wide set of data works best if the data is uniformly formatted. From Section 3.1, there are a small number of important data custodians, including BOM, ABS, Copernicus, PHIDU, and AIHW. This subsection describes the processes used to acquire and format the data ahead of the indicator and index creation methodology.

Collection Methods

The most common method of data collection among datasets is a manual download from the source using file transfer protocol (FTP) to download direct to local disk, e.g., publicly available ABS demographics data. This usually occurs when data is prepared (aggregated, formatted, etc.) and presented publicly by a data custodian. In some cases, such as BOM climate data, datasets are only available in a series of files, e.g., yearly, monthly, or even daily datasets. Manually downloading daily data files over a number of years is not feasible and must rely on automation. This is achieved using the statistical software R. More specifically, the `curl` package, which works as a HTTP client to send a HTTP request to the online data storage server.

Formatting Inconsistencies

The data collected throughout AusEnHealth's first phase is not consistently formatted, as each data custodian has different ways of collating, storing, and presenting data. Some immediate differences in data formats are geographical and temporal resolutions. Some data is provided according to the Australian Statistical Geography Standard (ASGS) standard resolutions (SA2, LGA, etc.), where others are provided at a much finer grid level. Some data is formatted for public use and has deliberate empty columns or rows. In some cases, more variables are provided in a dataset than necessary. There are also minor discrepancies in the use of rows and columns to represent changes in time and over space. All of these differences must be considered and resolved in order to efficiently apply indicator and index methods.

Data Formatting

Data alignment takes place in two separate scenarios. The first is linking spatial aggregated datasets using a consistent set of ABS ASGS standard geography polygons and the second scenario is the extraction of ASGS aggregated regions from gridded (raster) data.

For the first scenario, ABS correspondence tables provide linking between identifiers of each statistical and administrative region. The data is linked using the `dplyr` and `sf` packages in R, specifically utilising the `merge` function. In all cases possible, data is aligned with the base LGA shapefiles, as well as the base SA2 shapefiles (which can be aggregated up to SA3, SA4). In the second scenario, the `exactextractr` package uses a zonal extraction method to extract a function of data from each pixel of a raster within a given polygon of a shapefile. The function used in this project was the mean value of each pixel value. Grid data in all cases was extracted to both LGA and SA2 shapefiles to match other prepared data.

GIS was used in place of R when the grid data was high enough resolution that computing with R was not feasible. Using ArcMap, the shapefile is joined with an LGA and SA2 shapefile. The count of polygons and total polygon surface area from the data's shapefile within each SA2 or LGA polygon is computed and added to the administrative region shapefile's data.

In addition to the above aligning processes, a number of general data cleaning processes were applied to arrive at the final set of formatted data. Steps were taken to remove missing rows and columns, rename variables, and join datasets where appropriate. In some instances, steps were taken to change how time is represented in datasets, e.g., changing entries such as "*1-7 September 2019*" to a more

consistent 3-column year, month, and date system. A complete table of formatted data used in the AusEnHealth pilot is provided in Appendix H.

Metadata

Metadata stores can be differentiated by their coverage, data granularity and the detailing of their descriptions [93]. There can be local, institutional, national, and discipline-specific metadata. According to "*Best Practices for Publishing, Retrieving, and Using Spatial Data on the Web*" [94] some of the essential best practices regarding metadata management deal with discoverability, accessibility, portrayal, information modelling, data exchange, and spatial reference systems. FAIR (Findable, Accessible, Interoperable, Reusable) [95] data policies need to be considered to facilitate knowledge discovery and reusability. Metadata from spatial datasets require some extra details over a regular dataset. These details include the size of spatial datasets, coordinate reference systems, projections, temporal dimensions, thematic layering, and spatial semantics, scaling and quality [94]. Metadata good practices [96] are followed as part of the AusEnHealth Project.

Since the AusEnHealth project deals with data from multiple sources, it requires the design of a metadata schema that adheres to commonly available metadata standards such as DataCite [97], ISO/CD 19115-3 [98] and Data Catalogue Vocabulary (DCAT) [99]. Further developments on the AusEnHealth pilot will need a robust metadata solution that can be sustained with minimum manual inputs while keeping the service disruption to a minimum. The project team has decided to adopt a metadata solution for the AusEnHealth pilot that uses a combination of dynamic spreadsheets and static records stored in Amazon S3 Buckets maintained at QUT servers. These records contain relevant information regarding the original datasets utilised and notes describing the logic applied for each indicator. This approach is fit for the pilot and scope exists to move towards a more extensible and sustainable solution in the future that can draw on a range of other infrastructures such as data.gov.au or AURIN. Existing metadata tools such as GeoNetwork [100] or CKAN [101] can also be explored.

3.3. Data Analysis Methods

As previously noted in Section 3.1, there are two dominant approaches used in the literature to study the link between the environment and human health. The first approach is a typical statistical analysis, where environmental variables are assessed against a health outcome or health outcomes of interest [70] [71] [72]. The second approach is the development of indicators or indices, which are based on relevant literature, rather than being directly assessed against one or more health outcomes [73] [74] [75] [76] [77] [78]. The first phase of AusEnHealth takes the second approach, which is further outlined in this section.

Exploring Environmental Health Variables, Indicators, and Indices

Environmental health indicators describe the link between the environment and human health [1]. "*Indicators have become widely used in many different fields and play as useful role in highlighting problems, identifying trends, and contributing to the process of priority-setting, policy formulation and evaluation and monitoring of progress*" [102]. Indicators are typically used to draw information from a variable that is difficult to otherwise infer by comparing to thresholds related to adverse health risks. The broad indicator definition highlights the challenge of creating indicators in the AusEnHealth POC which can be used most effectively to examine increases/decreases in population vulnerability. For example, consider the following two different ways to utilise temperature data. Firstly, average monthly temperature, which is calculated using average daily temperature. This variable can be used as a relative comparison of heat risk, noting that spikes in temperature might be missed due to the averaging method. Secondly, the number of region-specific high heat days, an indicator calculated using average daily temperature. There is an improved understanding in this indicator of heat risk, as

spikes in temperature are included in the count. However, this indicator lacks information on temperature values. The methodology used to create variables and indicators can either highlight or inhibit the presentation of data and must be carefully considered.

An index, or composite indicator, offers an additional lens through which an observer gains insight(s). This is achieved by gathering information from multiple variables or indicators and combining this information into a single value which represents combined information from each of its components. Indices have seen increased use in recent years, especially in the lens of vulnerability [103] [104]. Indices offer a way to condense the growing number of valuable indicators into a focal point of insight, eliminating the need to view numerous indicators and compile insight.

With reference to key vulnerability index literature [105] [57], vulnerability indices were created as part of the AusEnHealth Project for the agreed upon key uses cases: climate (both heat and cold) and air quality. This section of the report describes variable, indicator, and index creation methodology, utilising the formatted variables detailed in Section 3.2 and displayed in Appendix H.

Variable and Indicator Creation Methods

As part of the AusEnHealth Project, acquired data have been utilised to create environmental health variables and indicators, relating to both climate and air quality. Indicators have become widely used in both environment and health planning because of their ability to provide a simple and digestible window into sometimes complex data. Indicators are "*important for informing the public and decision-makers about key health and environmental problems, and actions required for their management*" [102]. This subsection details the variables and indicators chosen for visualisation in the AusEnHealth Project's proof-of-concept.

Extreme Temperatures

While information about raw temperature values can allow a direct and informative comparison between regions, the indicators prepared for the AusEnHealth Project aim to provide focused insight into both extreme cold and extreme heat historically across Australia. Extreme temperature is noted in most research literature as a single value representing a region-specific threshold [76] [106] [75]. For example, a climate change-related health vulnerability index created for British Columbia, Canada [75] counts the average days per year above 25°C as a comparative measure of extreme heat. Research works considering heat across Australia have instead focused on region-based historical temperature percentiles of daily mean temperature [107] [108], noting the significantly different climates in the regions which make up Australia and their populations' acclimatisation to these climates. Daily mean temperature (in place of daily maximum or minimum) is used in these measures to capture daily temperature fluctuations and overnight heat dissipation. Daily maximum and minimum temperature data for day i, region j, is converted to daily mean temperature (DMT) using:

$$DMT_{i,j} = \frac{T_{max,i,j} + T_{min,i,j}}{2} \quad (1)$$

The classification of high temperature percentiles in the literature is not consistent, with some stating the 85th percentile [109] [110] [111], and others the 90th percentile [107] [112]. Likewise, extreme temperature classification varies also, noted in some research literature as above the 95th percentile [113] [108], and above the 98th and 99th percentiles in others [114] [110].

Based on the above literature, the AusEnHealth pilot classifies high and extreme region-based temperatures as above and below the 85th and 95th percentiles, respectively. These conservative classifications are aimed at providing a complete picture of climate highs and lows across Australia with respect to the significant global research efforts attributed to extreme temperature events. While

daily temperature data is available, there is limited insight gained from national comparisons of a single day's data. Hence, the extreme temperature indicators available in the AusEnHealth pilot will be provided at weekly, monthly, and yearly scales. There are a number of ways to process temporal aggregation for daily temperature data, which are detailed as follows, alongside other processes used to create the set of extreme temperature indicators used in the AusEnHealth Project.

Number of "at risk" and "high risk" temperature days

This indicator offers an easy comparison between regions by simply counting the number of days (in a week, month, or year) in which a region's temperature is above the high and extreme percentile thresholds outlined above. These thresholds are named in the platform for a simplified user experience:

Percentile	Category Name
<85	Negligible risk
85-95	At risk
>95	High risk

1. Rank $DMT_{i,j}$ (in ascending order for abnormal heat, descending for abnormal cold) across all days and divide by the total number of days to create region-specific historical percentiles.
2. Count the number of days in the week/month/year above the 85th and 95th percentiles for high and extreme heat, respectively.

The following figures are example visualisations of these indicators.

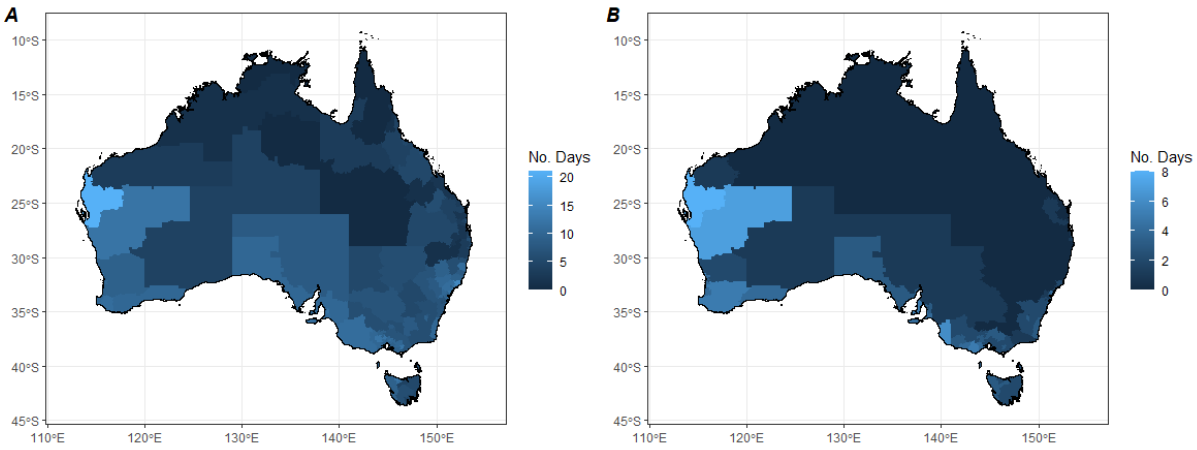

Figure 5: Number of (A) at risk, and (B) high risk heat days, March 2006 (SA3).

Nationally ranked region-specific temperature percentiles across Australia

National comparisons of raw temperature values will always be heavily influenced by the latitude of the Australian regions and will not adequately represent risk due to a populations capability to acclimatise to region temperatures over time. A better indication of high heat, notably referenced in heatwave literature [108] is to compare regions to their own historical data to produce a relative measure of extreme temperature. To understand which regions in Australia are facing the most extreme heat, these historical comparisons are then ranked spatially. For the purpose of user interpretation, this variable has been calculated using two different approaches. The first uses aggregated mean temperature to create percentiles that can be displayed in the AusEnHealth pilot.

1. Calculate weekly, monthly, or yearly mean $DMT_{i,j}$ (depending on the aggregation period).

2. Rank the mean $DMT_{i,j}$ values (in ascending order for abnormal heat, descending for abnormal cold) historically and divide by the total number of aggregation periods to create region-specific historical percentiles. **These region-specific aggregated historical percentiles can be displayed in the AusEnHealth platform.**
3. For each aggregation period, rank the historical percentiles spatially, and divide by the total number of regions to create spatial percentiles.
4. Convert percentiles into a 5-tier categorical indicator by splitting the data into 20th percentile groups (1 = lowest extreme, 5 = highest extreme).

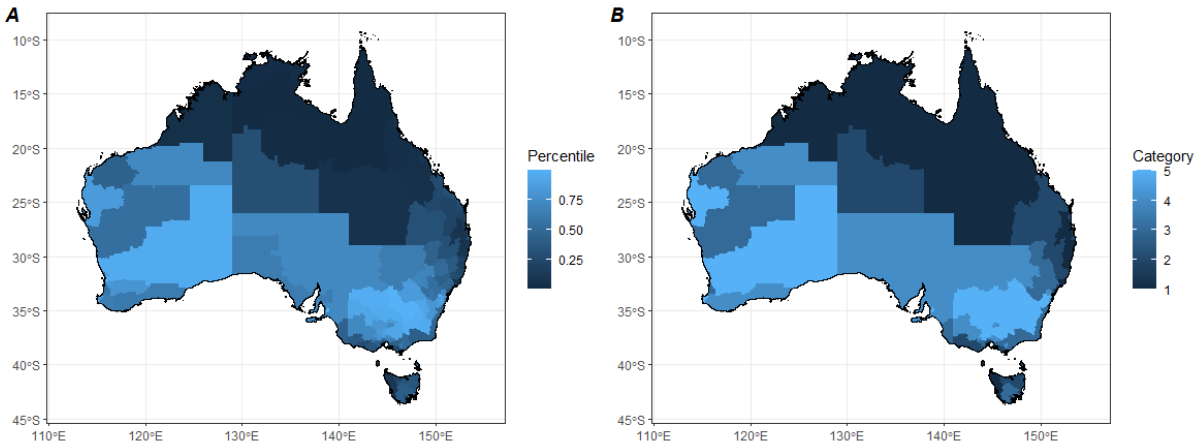

Figure 6: (A): Spatial percentiles of region-specific aggregated historical average temperature percentiles, January 2009 (SA3). (B): Categorical version of data presented in (A).

The second method adds up the daily historical percentiles, which captures additional information about the variation of temperature within each aggregation period.

1. Rank $DMT_{i,j}$ (in ascending order for abnormal heat, descending for abnormal cold) across all days and divide by the total number of days to create region-specific historical percentiles.
2. Sum the historical percentiles for each week, month, or year (depending on the aggregation period). **These summed percentiles will not be displayed in the AusEnHealth platform, as they do not have a straightforward interpretation.**
3. For each aggregation period, rank the historical percentiles spatially, and divide by the total number of regions to create spatial percentiles.
4. Convert percentiles into a 5-tier categorical indicator by splitting the data into 20th percentile groups (1 = lowest extreme, 5 = highest extreme).

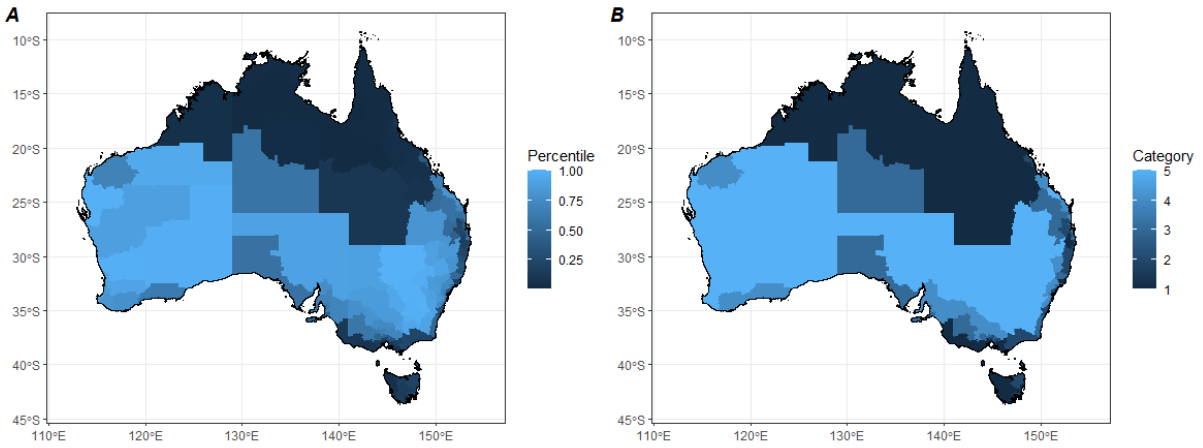

Figure 7: (A): Spatial percentiles of region-specific aggregated historical summed temperature percentiles, January 2009 (SA3). (B): Categorical version of data presented in (A).

The spatial rankings for this approach are more informative relative to the aggregation periods, as they include information regarding variation of the temperature within each aggregation period. However, as the raw values cannot be meaningfully interpreted on their own, this method is not used for visualisation. Instead, this method is used in the creation of the AusEnHealth Project' vulnerability indices (see Section 3.4.1).

Heatwave and Coldwaves

Heatwaves impact human health, agriculture, the economy, natural hazards, and ecosystems, and are among the most common natural hazards experienced across the Australian community [115]. Heatwaves have increased in intensity, frequency, and duration across Australia over the past 67 years [116]. In the 2000s, heat-health warning systems, such as southeast Queensland's "*apparent temperature*" service or Melbourne's service using a one-day daily temperature, were incomparable and made heatwave events difficult to visualise or compare across time or jurisdiction. An objective definition for heatwaves and heatwave severity, the Excess Heat Factor (EHF), was proposed in 2013 [115]. EHF has since been assessed as a valuable tool in the identification of vulnerable populations [117] [118] and has been utilised in heat health research and in the creation of heat vulnerability indices [57] [119]. While there are other definitions of heatwaves in the literature, EHF is used in AusEnHealth due to its prevalence in Australian heatwave literature. The EHF is calculated by considering both *excess heat* and *heat stress*.

Excess heat is unusually high heat arising from a high daytime temperature that is not sufficiently discharged overnight due to unusually high overnight temperature. Maximum and subsequent minimum temperatures averaged over a three-day period are compared against a climate reference value to characterise this unusually high heat in an excess heat index (or significance index), given by

$$\mathrm{EHI}_{\mathrm{sig}} = \frac{\mathrm{DMT}_{\mathrm{i}} + \mathrm{DMT}_{\mathrm{i+1}} + \mathrm{DMT}_{\mathrm{i+2}}}{3} - \mathrm{DMT}_{95}, \qquad (2)$$

Where DMT_{95} is the 95$^{\mathrm{th}}$ percentile of daily temperature ($\mathrm{DMT}_{\mathrm{i}}$) for the climate reference period (originally 1971-2000) calculated using all days of the year. Across the 24-hour period, DMT on day i is in °C. The units of $\mathrm{EHI}_{\mathrm{sig}}$ are also °C. From the equation, the excess heat index measures anomalies in three-day DMT with respect to the 95$^{\mathrm{th}}$ percentile of historical daily mean temperature.

Heat stress arises from a period where temperature is warmer, on average, than the recent past. Maximum and subsequent minimum temperatures averaged over a three-day period and the previous

30 days are compared to characterise this heat stress in a second index. This is expressed as a short-term (acclimatisation) temperature anomaly. The heat stress index (or acclimatisation index) is given by

$$\mathrm{EHI}_{\mathrm{accl}} = \frac{\mathrm{DMT}_{\mathrm{i}} + \mathrm{DMT}_{\mathrm{i}-1} + \mathrm{DMT}_{\mathrm{i}-2}}{3} - \frac{\mathrm{DMT}_{\mathrm{i}-3} + \cdots + \mathrm{DMT}_{\mathrm{i}-32}}{30}, \tag{3}$$

The units of $\mathrm{EHI}_{\mathrm{accl}}$ are also °C. From this equation, the heat stress index measures anomalies in three-day daily mean temperature with respect to the previous 30 days.

Excess heat factor combines the effect of excess heat and heat stress, calculated as an index to provide a comparative measure of intensity, load, duration, and spatial distribution of a heatwave event. Heatwave conditions exist when the EHF is positive. There are multiple ways of calculating EHF [115] [120]. For the AusEnHealth Project, EHF is calculated using

$$\mathrm{EHF} = \max\left(0, \mathrm{EHI}_{\mathrm{sig}}\right) \times \max(1, \mathrm{EHI}_{\mathrm{accl}}). \tag{4}$$

This variant of the definition sets all negative values of EHF to 0. A severe heatwave occurs when a positive EHF value exceeds the 85$^{\mathrm{th}}$ percentile of the distribution of positive EHF values at a given location (denoted EHF_{85}). Extreme cold can also be assessed using the above calculations with respect to cold and is termed the Excess Cold Factor (ECF).

EHF and ECF are presented in the AusEnHealth Project in a similar way to high percentile temperature days. Using aggregation periods leads to count indicators representing both the number of regular heatwave and coldwave days, as well as separate indicators representing the number of severe heatwave and coldwave days. Spatial rankings of historical EHF and ECF percentiles are also provided using the same methodology as for extreme temperature.

Number of "all" and "severe" heatwave and coldwave days

This indicator offers an easy comparison between regions by simply counting the number of days (in a week, month, or year) in which a region experiences a regular or severe heatwave or coldwave. The derivation for this indicator is as follows.

1. Calculate EHF/ECF using equations (2)-(4).
2. Count the number of days in the week/month/year for which EHF/ECF is above 0, and above the 85$^{\mathrm{th}}$ percentile for all and severe heatwaves, respectively.

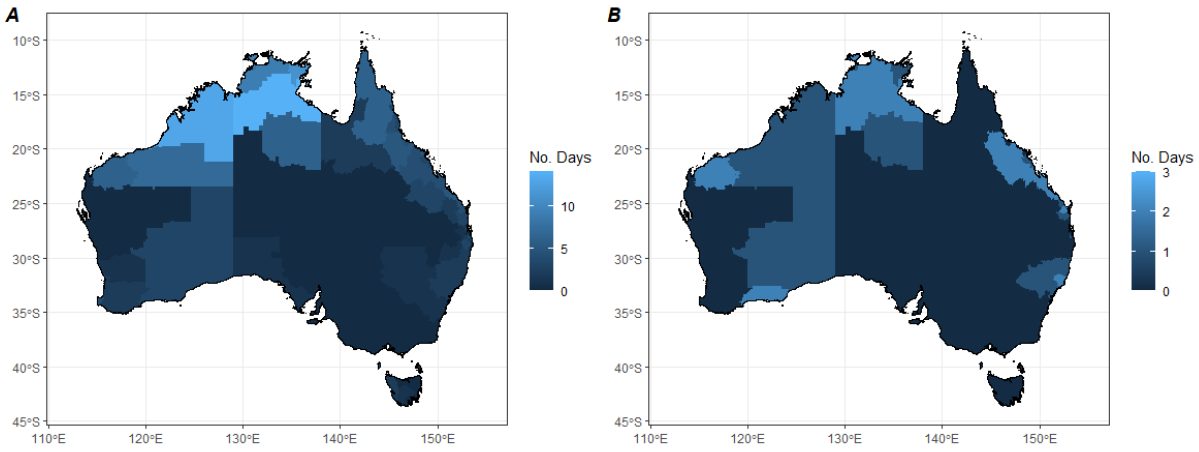

Figure 8: Number of (A) all, and (B) severe heatwave days, December 2004 (SA3).

Nationally ranked region-specific EHF/ECF averages across Australia

An alternative assessment of heatwaves across Australia is a region-specific historical comparison, i.e., how severe are the current heatwaves compared nationally over a given aggregation period? This variable takes the average EHF/ECF values over the aggregation period to create spatial percentiles which are then categorised to simplify interpretation.

1. Calculate EHF/ECF using equations (2)-(4).
2. Calculate weekly, monthly, or yearly mean EHF/ECF (depending on the aggregation period).
3. For each aggregation period, rank the mean values spatially, and divide by the total number of regions to create spatial percentiles.
4. Convert percentiles into a 5-tier categorical indicator by splitting the data into 20th percentile groups (1 = lowest extreme, 5 = highest extreme). "*NA*" values in this indicator represent no heatwave/coldwave.

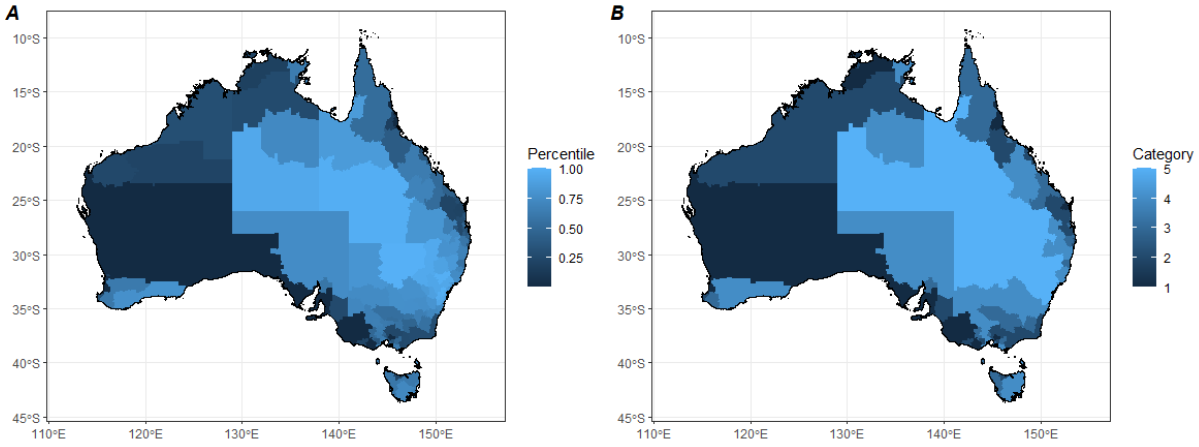

Figure 9:Spatial percentiles (A) and categories (B) of heatwave days, December 2017 (SA3).

Air Quality - Pollutants

Variables and indicators based on historical and spatial percentiles are also planned for the acquired air quality variables, which are PM10, PM2.5, PM1, CO, NO NO2, SO3, and O3. While there are national guidelines suggesting risk concentrations for air pollutants [121], these vary internationally [122] and have been challenged by research literature [123]. Due to this uncertainty, created indicators will be presented alongside the raw (albeit aggregated) air quality variable data. Due to uncertainty surrounding the identification of "*risk*" relating to air quality, indicators instead compare to region averages.

Number of "above average" pollutant days

This indicator offers an easy comparison between regions by simply counting the number of days (in a week, month, or year) in which a region experiences air pollution either above average, or above a risk threshold. This indicator is produced for all air quality variables. Due to the uncertainty surrounding the national guidelines for air quality pollution, a potential risk threshold is set as the 85th percentile across all data. This threshold has been selected as a marker to distinguish between levels of air pollution, where increased levels relate to increased potential risk. As there are no identified spatial trends in air quality data, percentiles are calculated across both space and time. These thresholds are named in the platform for a simplified user experience:

Percentile	Category Name
<50	Below average
50-85	Above average
>85	Potential risk

The derivation for this indicator is as follows.

1. Rank daily air pollutant (in ascending order) across all days and regions and divide by the total number of data points to create region-specific historical percentiles.
2. Count the number of days in the week/month/year above the 50th and 85th percentiles for above average and potential risk categories, respectively.

The following figures are example visualisations of these indicators.

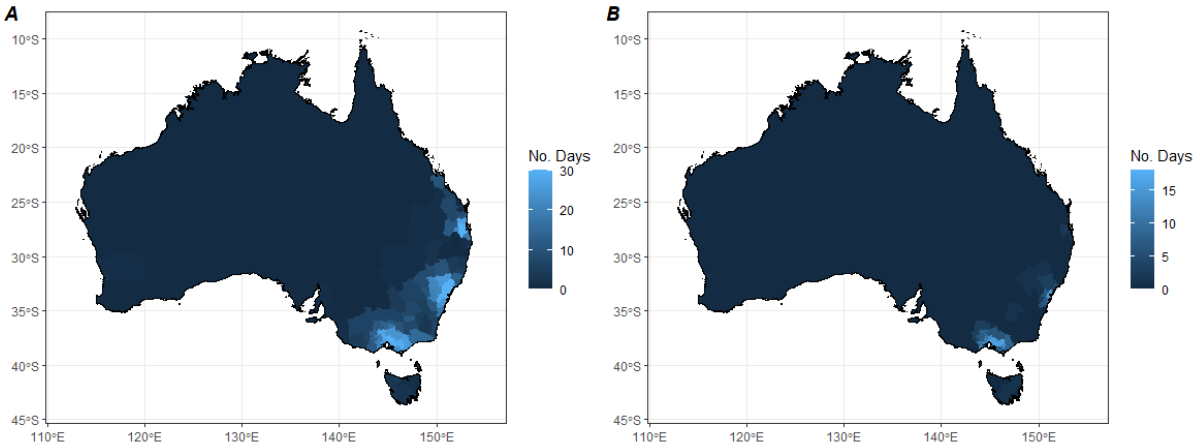

Figure 10: Air quality threshold counts, April 2014 (SA3). Counts of above average days (A) and potential risk days (B) of Nitrogen Dioxide (NO2).

Nationally ranked region-specific air quality percentiles across Australia:

Again, as there are no identified spatial patterns for the air pollutant data acquired in the AusEnHealth Project, the spatial and temporal patterns in air quality can be combined. To understand which regions in Australia are facing the most extreme air pollution, air pollution levels are ranked over all time and space. This variable has been calculated using two different approaches. The first is a spatial and historical percentile calculation that can easily be displayed for user interpretation.

1. Calculate weekly, monthly, or yearly mean air pollutant levels (depending on the aggregation period).
2. Rank the pollutant values (in ascending order) across all days and regions and divide by the total number of data points to create region-specific historical percentiles.
3. Convert percentiles into a 5-tier categorical indicator by splitting the data into 20th percentile groups (1 = lowest extreme, 5 = highest extreme).

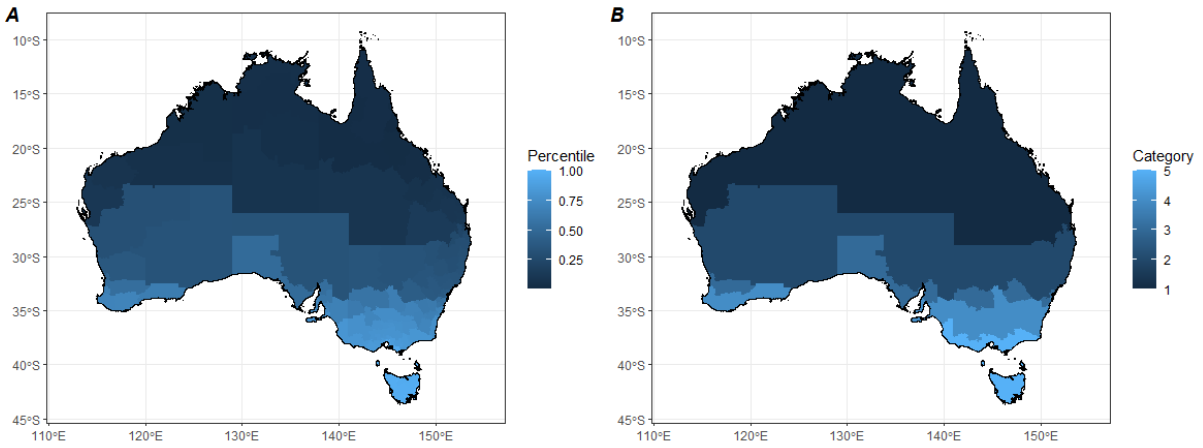

Figure 11: Historical and spatial percentiles (A) and categories (B) of Ozone (O3), February 2019 (SA3).

The second method is a spatial ranking only, for use in the creation of air quality vulnerability indices.

3.4. Population Vulnerability Indices

The dynamics, frequency, and extremes of many natural hazards are intensifying [124]. The changing nature of many hazards, coupled with growing and ageing populations and infrastructure in exposed areas is leading to increased vulnerability across Australia and internationally. Effective adaptation and risk management strategies and practices require an in-depth understanding of population vulnerability, as well as the various exposures facing populations. In addition, it is important to properly assess changes in these areas [125]. "*Hazards only lead to disaster if they intersect with an exposed and vulnerable society and when the consequences exceed its capacity to cope*" [124]. Among the many goals of the AusEnHealth Project is the aim to provide tools to support the assessment of population vulnerability through an environmental health lens. This section presents a condensed summary of vulnerability assessment and vulnerability index methodology identified in existing research literature, as well as the methods implemented to create the various vulnerability indices presented in the AusEnHealth Project's proof-of-concept strategic planning digital twin.

Social Vulnerability

Historically, significant efforts have been made to characterise the human and financial risks related to natural hazards. More recently, these risks have become attributed more to the vulnerability of exposed human populations rather than the exposure itself [126]. The term vulnerability, in social science research, generally describes a state of people and populations, and can significantly vary geographically [127]. There are many works providing insight into population vulnerability through demography: the statistical study of human populations. Due to the complex relationships between the environment and humans, no single investigation into vulnerability will provide a holistic and comprehensive answer to general environmental health vulnerability [128]. Despite this, local government agencies need information to support vulnerability assessment in order to effectively develop policy, apply intervention, provide emergency disaster management, and aid their respective community's ability to adapt to changes in environment.

Local government agencies have limited resources with which to carry out an effective assessment of population vulnerability related to the wide variety of potential environmental hazards [103]. An effective solution to this problem is the creation of social vulnerability indicators, which expedite the insight gained when exploring certain population demographics. In 2010, a nation-wide social vulnerability index (SVI) was created in the United States by combining a wide range of population

variables [103]. This alternative to vulnerability indicators was introduced to assist state and local personnel to identify the location of socially vulnerable communities in order to more effectively target and support community-based efforts to mitigate and prepare for disaster events. Social vulnerability indices, as well as resilience and risk indices, have since become prominent tools which provide insight into disaster response, recovery, readiness, adaptability, resilience, and risk assessment [129].

Table 8 notes succinctly the variables included in three prominent social vulnerability indices: the above noted social vulnerability index (SVI) [103], the social vulnerability index for the United States (SoVI) [130], and Australia's index of relative socio-economic disadvantage (IRSD) [104].

Variable	SVI	SoVI	IRSD
Low Income	x	x	x
Jobless with Children			x
Highschool Education	x	x	x
Unemployed	x	x	x
Labourers			x
Low Rent		x	x
Single Parent House	x		x
Disability	x		x
Divorced			x
Machinery Operators			x
Low Skill Workers		x	x
No Vehicle	x	x	x
Crowded Dwelling	x		x
No Education			x
Poor English Language	x	x	x
Income	x	x	
Age >65	x	x	
Age <17	x	x	
Minority	x	x	
Multi-Unit Structures	x		
Mobile Homes	x	x	
Group Quarters	x		
Median Age		x	
Households Receiving Government Pay		x	
High Income		x	
Female		x	
Female Headed Households		x	
Nursing Home Residents		x	
No Health Insurance		x	
People Per Housing Unit		x	
House Value		x	
Female Participation in Labour Force		x	
Unoccupied Houses		x	
Hospitals Per Capita		x	

Table 8: Comparison of variables utilised in social vulnerability indices.

The selected variables can then be reduced to a single value representing vulnerability. There are two prominent methods for index creation. The first is a percentile rank-sum approach [103], which transforms variables into percentiles ranked from highest to lowest inferred vulnerability across all

assessed regions. These percentiles are then added to produce the SVI. The second prominent approach is a modification to the first approach, where variables are weighted using principal component analysis (PCA) before being combined [130] [104].

While these indices are shown in their respective sources as effective measures of vulnerability, other approaches have expanded such indices to capture other aspects of population vulnerability, including readiness, adaptability, and exposure. Heat vulnerability indices in research literature take components from social vulnerability assessments and pair them with exposures relating to climate [131]. A 2005 heat vulnerability index created in Southern Quebec [106] included high heat days and consecutive high heat days alongside population demographics. A 2006 heat vulnerability index created in Vermont [80] included canopy cover, high heat days, mortality, and heat-related hospitalisations in addition to socio-economic population data. Additional research literature in the United States include adaptability components such as data on air conditioning in the home [74], alternative vulnerability variables such as existing chronic conditions [73], and additional variables concerning land use [72].

Air quality vulnerability indices have also been considered in the literature, to a much lesser extent. A 2015 air pollution index applied in China [132] incorporated sulphur dioxide (SO_2), nitrogen dioxide (NO_2), and particulate matter (<10µm) with population data to create a vulnerability index. In 2021, an air pollution index was considered in Africa [133] which combined socio-demographic data with nitrogen dioxide (NO_2). Other air vulnerability index literature, including those extending into epidemiology and COVID-19 [134] have incorporated various mortality and morbidity into vulnerability indices, as well as simply implementing entire existing social vulnerability indices.

There are two main conclusions drawn from literature of social vulnerability indices and their environmental health related expansions. Firstly, population vulnerability with an environmental health focus is difficult to capture (and thus indices vary significantly between source), and secondly, the variables included in a vulnerability index heavily rely on data availability in the focus area.

Utilising exposure, sensitivity, and adaptive capacity framework

A significant portion of the reviewed vulnerability indices utilise a consistent vulnerability index framework introduced by the Intergovernmental Panel on Climate Change in 2007 [105]. This framework aims to better explain population vulnerability with respect to climate, and is made up of three sections: exposure, sensitivity, and adaptive capacity. Exposure is a direct measure of household, community, or population exposure to a certain event (extreme heat, cold, air pollution, etc.). Sensitivity is defined as the susceptibility of a household, community, or population to the exposure. Finally, adaptive capacity captures the capabilities of a household, community, or population to cope with or recover from the impact of the exposure. In the creation of a vulnerability index, these sections form sub-indices which contain relevant variables and indicators. The sub-indices are used to capture different aspects of vulnerability and are ultimately used to create the final vulnerability index.

The following three tables summarise the variables included in five existing vulnerability indices which utilise the IPCC vulnerability index framework [105]. These indices are created in different locations and with different applications, including an analysis of neighbourhood-level climate change-related health hazards in Canada [75], an assessment of heat vulnerability in Sydney Australia [76], a national report published by BOM on heatwaves in Australia [57], a second heat vulnerability index developed for Sydney Australia [77], and a heat index created in Chile [78].

Variable	Source				
	[75]	[76]	[57]	[77]	[78]
Mean days per year >25°C	x				
Mean maximum temperature		x			
Mean minimum temperature		x			
Days per year >30°C		x			
Impervious land cover		x			
Population density		x		x	
Road density		x		x	
Land surface temperature			x		x
No vegetation			x		
House age			x		
High-rise dwellings			x		
Outdoor workers			x		
Days per year >38°C				x	
% Poverty rate				x	
Inequalities (Gini index)				x	

Table 9: Variables typically included in exposure sub-indices.

Exposure sub-index variables shown in Table 9 are focused mainly on temperature, with [75] and [78] including only variables relating to temperature. The included temperature-related variables are either averages or are threshold-based. Australian heatwave assessment [115] has revealed that population acclimatisation can significantly mitigate the effects of high heat. As a result, using average temperature or a temperature threshold in a national vulnerability index is likely to incorrectly frame the warmer months as consistently at high risk. For example, Tasmanian regions might never be declared vulnerable, despite spikes in temperature being more likely to overcome acclimatisation in cooler regions.

Other variables include land cover, urban variables (including house age and road density), and demographics variables (such as population density or poverty rate). It is more likely, based on the framework terminology described above, that these variables can instead be allocated to other sub-indices.

Variable	Source				
	[75]	[76]	[57]	[77]	[78]
% Children (<5 years old)	x	x	x	x	x
% Elderly (>65 years old)	x		x	x	x
% Elderly and living alone		x			
Cardiovascular diseases	x			x	
Respiratory diseases	x			x	
Renal diseases	x			x	
Mental health diseases	x			x	
General health	x				
Multiple chronic conditions	x				
% Multi-unit dwellings		x	x	x	
Low equivalised total household income			x		
Unemployment rate			x		x
Education (year 10 or less)			x		
Disability			x		x
Living alone			x	x	x
English language proficiency			x		
New arrivals/refugees			x		

Indigenous			x		
Rental housing			x	x	
Crowding			x		
No vehicle			x		
Social housing			x		
No home internet			x		
Alcohol consumption			x		
Overweight			x		
Obesity			x		
Smokers			x		
People requiring care				x	
% Women				x	

Table 10: Variables typically included in sensitivity sub-indices.

There is heavy overlap between the sensitivity sub-index variables chosen in Table 10 and the selection of variables used in the various social vulnerability indices in Table 8. A degree of care should be taken to ensure that the list of variables stays relevant to the exposure of interest.

Variable	**Source**				
	[75]	[76]	[57]	[77]	[78]
% Low education (complete year 12)	x	x		x	
% Poverty	x				
% Visible minority status	x				
% Indigenous status	x				
% Poor housing condition	x				
High population density	x				
% Impervious surfaces	x				
% Living alone	x				
Heat response plans	x				
% Speaks non-English		x			
Median home loan repayment		x			
% Home ownership		x			
Median household income		x			
% Household with internet access		x			
Current ratios of assets to liabilities of LC		x			
Per capita business rates of LC		x			
Per capita residential rates of LC		x			
Per capita community service expenses of LC		x			
Per capita environmental and health expenses of LC		x			
% Population requiring financial assistance		x			
Medical services (hospitals, GPs, emergency departments			x		x
Aged care			x		
Community services			x		
Pharmacies			x		
Psychiatry			x		
Public open space			x		
Cool places			x		
Vegetation			x		x
Water bodies			x	x	
Moved last year			x		
Moved last 5 years			x		
Volunteerism			x		

Vehicle at home	x		
Internet at home	x	x	x
Tree canopy		x	
% Parks		x	
% Local government expenses (of total expenses)			
Households with mobile phone or landline telephone			x
Water supplied to the home			x
Buildings with light materials in external walls			x
Roads			x

Table 11: Variables typically included in adaptive capacity sub-indices.

The majority of these variables relate to environmental characteristics (both urban and natural), health and government characteristics, and population demographics, with some overlap between variables in Table 11 and those in Table 8 and Table 10. It is important to note here that the adaptive capacity, and sensitivity sections of this vulnerability index framework both contain variables which can be influenced by government through policy and intervention measures.

It is clear that no two vulnerability index implementations use the same set of variables. Reasons for these differences include regional scope, data availability, application, and differences in the groupings of exposure, sensitivity, and adaptive capacity. While vulnerability index development in recent literature has a heat-health focus, there are numerous air-health research literature [135] [136] [133] with which to draw information in the development of an air quality vulnerability index. All of the indices developed as a part of AusEnHealth will utilise social vulnerability variables which are frequently included in such vulnerability assessments.

AusEnHealth Vulnerability Index Components

Information retrieved from the extensive research literature on vulnerability indices has been combined with the datasets identified through the data audit process to produce three vulnerability indices focused on heat, cold, and air quality. The following tables lists the included components grouped into exposure, sensitivity, and adaptive capacity categories (orange, blue, and green, respectively), and the themes adapted from relevant vulnerability index literature [103] [57].

Theme	Variable	Heat	Cold	Air Quality
Heat Exposure	Temperature (Heat)	x		
	Excess Heat Factor (EHF)	x		
Cold Exposure	Temperature (Cold)		x	
	Excess Cold Factor (ECF)		x	
Air Pollutant Exposure	Nitrogen Monoxide (NO)			x
	Nitrogen Dioxide (NO2)			x
	Ozone (O3)			x
	Particulate Matter (PM2.5)			x

Table 12: Exposure sub-index.

Moving population demographics to a mix of sensitivity and adaptive capacity, as well as moving urban and natural environment variables to adaptive capacity, leaves only temperature and air quality data in the exposure section. Heatwave and coldwave measures have been included in the climate exposure sub-index to incorporate additional information relating to consecutive high heat days into the overall vulnerability index. The temperature measures here are given as historical percentiles to represent out-of-the-ordinary heat on a per-region basis. This effectively includes a region-based varying threshold of heat against which to compare extreme heat in populations.

As there is no identified acclimatisation to poor air quality, only spatial percentiles are considered for the air quality index. The air quality sub-index only includes air pollutants, following the same approach and heat and cold exposure sub-indices. Notably, SO2, CO, PM1, and PM10 are absent from the air quality exposure sub-index. As previously noted, these variables were excluded after an extensive literature review and through consultation with air quality research experts in the Australian research community.

Theme	Variable	Heat	Cold	Air Quality
Socioeconomic Status	Population Density	x	x	x
	Median Income	x	x	x
	Low Income	x	x	x
	Education (no high school)	x	x	x
	Unemployment Rate	x	x	x
Household Composition	Elderly	x	x	x
	Infants	x	x	x
	Single Parents of Dependants	x	x	x
	Unpaid Childcare	x	x	x
	Persons Requiring Significant Assistance	x	x	x
	Persons with a Disability	x	x	x
	Living Alone	x	x	x
Language and Culture	English Second Language	x	x	x
	Indigenous	x	x	x
Housing Conditions	Mobile Homes	x	x	x
	Crowded Dwellings	x	x	x
	Renters	x	x	x
	Mortgage Payers	x	x	x
Health Status	Condition: Respiratory Diseases			x
	Condition: Asthma			x
	Condition: COPD			x
	Condition: Circulatory System Diseases	x	x	
	Condition: High Blood Pressure	x	x	
	Condition: High Blood Cholesterol	x	x	
	Condition: CVD	x	x	
Health Risk Factors	Condition: Overweight	x	x	x
	Condition: Obese	x	x	x
	Condition: Smoker	x	x	x
	Condition: High Alcohol Consumption	x	x	x

Table 13: Sensitivity sub-index.

Sensitivity variables are largely based on population vulnerability indices [103] [130] [104], with particular influence from the comprehensive BOM heatwave report [57]. Note that the "*health status*" and "*health risk factors*" themes are only available at LGA geographical resolution. How these missing variables are handled, as well as missing data in general, will be covered in Section 3.4.1.

Theme	Variable	Heat	Cold	Air Quality
Health Services	Hospitals	x	x	x
Cool Places	Greenspace (Parklands)	x	x	x
	Water Bodies	x	x	
	Vegetation Index	x	x	x
Social Connectedness	Dwellings with a Vehicle	x	x	x
	Access to Internet	x	x	x

Table 14: Adaptive capacity sub-index.

The adaptive capacity variable selection is quite limited in this version of the AusEnHealth vulnerability indices. Some additional datasets that could be added to this sub-index in future are road and building density data (narrowly infeasible to include in this index), as well as a range of health service data, which was temporarily unavailable at the time of vulnerability index creation. The "*social connectedness*" theme is comprised of population demographics data. These variables differ in nature from the variables in Table 13, as they inherently improve access to information and services. A more detailed description of the variables used in the creation of the vulnerability indices is given in Appendix I.

3.4.1. Methodology

The majority of vulnerability indices identified in research literature combine data from a wide temporal range to form a single index [57] [75]. When changes over time are considered in the creation of vulnerability indices, this is usually over either a short time period or only a few time points [77]. The AusEnHealth Project will present vulnerability indices at weekly, monthly, and yearly scales by incorporating daily climate and air quality data and considering changes over time to the exposure, sensitivity, and adaptive capacity sections of the created vulnerability indices. As the temporal extent of collected demographics data covers the years 2011-2019, the vulnerability indices will only be created over this time period.

The methodology for vulnerability index creation in the literature is consistent. As noted earlier in Section 3.4, index components are either equally weighted [103], or weighted using the PCA factor analysis method [130]. There is conceptual justification behind the use of equal weights: When trying to cover the broad spectrum of exposure-related health outcomes, weighting variables by comparing against any single health outcome is likely to worsen the reliability of the vulnerability index compared to a separate exposure-related health outcome. However, some of the parameters may be more important across a range of health outcomes. Also, comparing the variables to no health outcomes and assuming equal weights may lead to a vulnerability index with no reliability (depending on the underlying variables).

There are also strengths and weaknesses of the PCA method. The strength is that PCA will identify variables which do not share similar relationships with other variables, effectively reducing variation when combining the variables to form the final index. On the other hand, by not comparing the variables to a health outcome, PCA will be producing weights that optimally describes variance in a set of variables that may not be correlated to any exposure-related health outcome. I.e., Depending on the quality of underlying variables, PCA-derived weights may reduce the overall value of a vulnerability index.

Ideally, an index creation method should be applied which incorporates non-equal weights to ensure that all of the included variables add value to the overall vulnerability indices. Due to potential issues with the application of PCA, methodology has been used which connects the AusEnHealth vulnerability indices with health outcomes.

Equal Weights

The equal weights method [103] is the baseline approach applied in the AusEnHealth Project. In this approach, a sub-index is typically created by summing together spatial percentiles of its variables. Let S_{kt} be a sub-index at time t containing N_k variables, where $k = 1, 2, 3$ corresponds to exposure, sensitivity, and adaptive capacity sub-indices, respectively. With variables $n = 1, \dots, N_k$, a vulnerability sub-index, S_{ik}, for a geographical region, i, is calculated as follows.

$$S_{ikt} = \sum_{n=1}^{N_k} f_i(\underline{x}_{nt}), \tag{5}$$

where $\underline{x}_{nt}$ is a vector containing every region's observed value of variable n at time t. The function f_i computes the i^{th} region's spatial percentile of $\underline{x}_{nt}$, ignoring any missing values in $\underline{x}_{nt}$. The final vulnerability index, V_{it}, is calculated in a similar way.

$$V_{it} = \sum_{k=1}^{3} f_i(\underline{S}_{kt}), \tag{6}$$

where $\underline{S}_{kt}$ is a vector containing every region's observed value of vulnerability sub-index k. While this is a valid method for constructing vulnerability indices, some changes are necessary to facilitate large amounts of missing data.

Missing Data

The index detailed in Equation (6) breaks down if, for a given variable n, the vector $\underline{x}_n$ contains only missing values. The ranking function f_i will either fail or have no effect, and the summation component of the calculation will either fail or will cause the final index to have an inherently lower value if missing variables are ignored. These issues can be overcome if an average calculation is used in place of a summation, ignoring vectors which only contain missing values.

Hence, if there is no data for a variable over a particular period when calculating the index, the average of the spatial percentiles is calculated without the related variable. This is equivalent to the following changes to the methodology.

$$S_{ikt} = \frac{1}{N_{kt}} \sum_{n=1}^{N_{kt}} f_i(\underline{x}_{nt}), \tag{7}$$

where N_{kt} is the number of variables at time t that contain at least one non-missing observation.

$$V_{it} = \frac{1}{3} \sum_{k=1}^{3} f_i(\underline{S}_{kt}). \tag{8}$$

One immediate result from this methodology is that the "*health status*" and "*health risk factors*" themes will never be factored into the SA2, SA3, and SA4 vulnerability indices, as they are not available at those geographical resolutions. The demographics data collected for these vulnerability indices are not comprehensive, and when available, are only given at a yearly temporal resolution. Some missing demographics data have been imputed using linear interpolation (see Appendix J), however, it is not feasible to impute all values due to high potential error rates. To enable the creation of weekly and monthly vulnerability indices, weekly and monthly demographics data have been created using values available only at the yearly resolution. Filling the demographics data in this way enables allows a deeper understanding of trends in vulnerability, as more granular exposure data can now be used in the creation of the indices.

Alternate Methodology

Taking the average ranking over the summation is one simple way to alter the index creation methodology for a specific set of data. Another alteration of note is work produced by BOM in 2021, where, rather than adding the ranked sub-indices as per Equation (6), the adaptive capacity index was subtracted. This is because variables included in this index were designed to lessen vulnerability,

rather than increase. I.e., a higher adaptive capacity index is interpreted as lower population vulnerability. Taking a similar approach, AusEnHealth will assess the correlation between variables and health outcomes to identify which variables should either positively or negatively impact vulnerability. These correlations can then be applied as weights in an alternative index to also answer the question: "*how much does each variable influence vulnerability?"*

Pairwise Correlations

MORT Books 2015-2019 data has been used to create the following environmental exposure and mortality categories.

Environmental Exposure	Mortality Cause
Heat	Coronary Heart Disease
	Cerebrovascular Disease
	Heart Failure
	Cardiac Arrythmia
Air Quality	Chronic Obstructive Pulmonary
All Exposures	All Cause

Correlation using Kendall's Tau was examined between all vulnerability index variables, the overall equally weighted vulnerability indices, their sub-indices, and the above three mortality categories over the 2015-2019 period. To calculate these correlations, the average values for each of the above variables and indices at LGA geographical resolution were calculated over the time period (2011-2019). Kendall's Tau has been used over the typical Pearson correlation, as the assumptions for Pearson correlations (e.g., linearity) may not be met for all variables. A correlation plot for index components is shown in Figure 12. A correlation plot for the vulnerability indices and sub-indices is given in Figure 13.

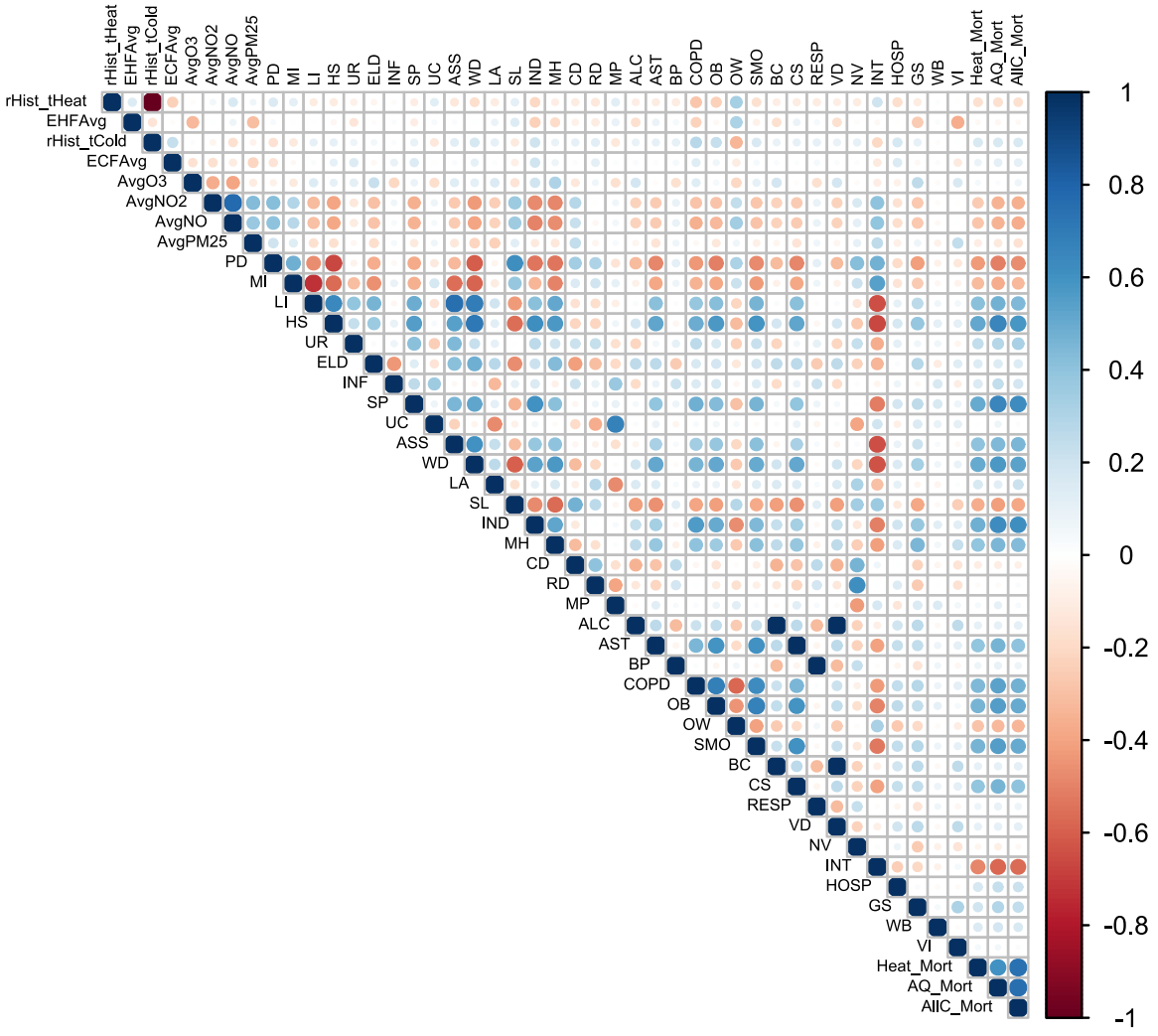

Figure 12: Pairwise correlation plot comparing vulnerability components with mortality data. Please correspond acronyms using relevant tables.

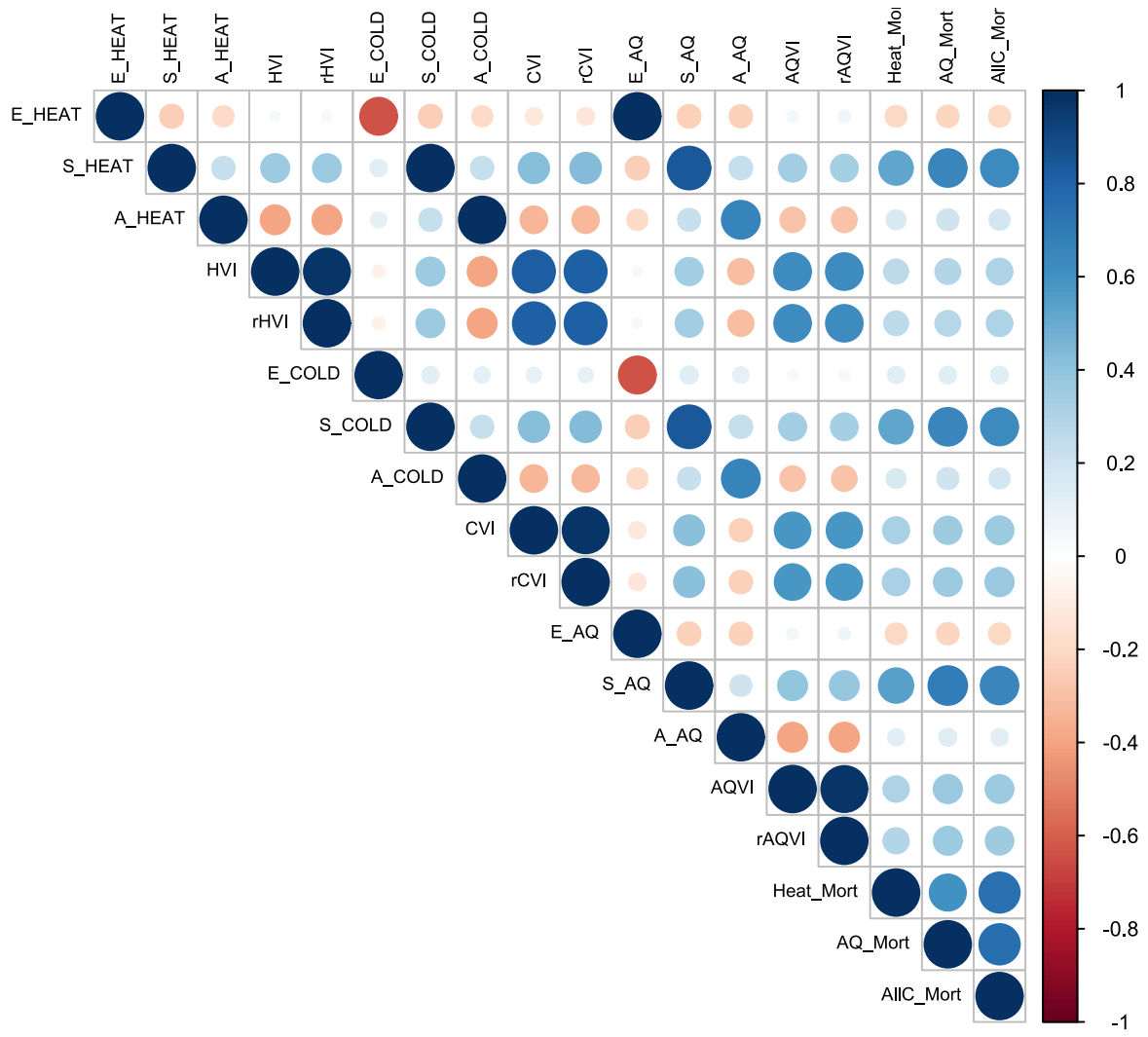

Figure 13: Pairwise correlation plot comparing equal weighted vulnerability indices and sub-indices with mortality data. E, S, and A correspond to exposure, sensitivity, and adaptive capacity sub-indices, respectively.

These correlations offer insight into the relative weights of each variable with respect to all-cause mortality. Firstly, there is not a strong correlation between any of the exposure variables and all-cause mortality. As this may be due to seasonal effects or spikes in data, the correlations weights will not be applied to exposure variables. These variables, all part of the exposure sub-index, will retain equal weighting. The other variables behave in-line with their use in similar vulnerability indices in the literature, with a single exception: "*speaks a second language at home*" (SL in Figure 12). It is expected that this variable identifies individuals for which English is their second language. While this arguably means reduced access to information and increased vulnerability, the correlation implies the opposite. This result may be due to high correlation with population density, however, exploring these relationships is left for future work. The correlation values in the last column of Figure 12 can now directly be used as weights in the following adjusted method.

First, the calculation of each weighted sub-index at time t for region i, WS_{ikt}:

$$WS_{ikt} = \frac{w_{nk}}{N_{kt}} \sum_{n=1}^{N_{kt}} f_i(x_{nt}), \tag{9}$$

where w_{nk} is the Kendall's Tau correlation between variable n and all-cause mortality in sub-index k. In this application, equal weights are being used for the exposure sub-index. That is, $w_{n1} = 1$, for $n = 1, \ldots, N_1$. The calculation of the overall vulnerability index remains the same, replacing the original sub-indices for the weighted sub-indices:

$$V_{it} = \frac{1}{3}\sum_{k=1}^{3} f_i(\underline{WS}_{kt}). \tag{10}$$

This method was used to produce a new set of vulnerability indices, as well as the following correlation plot in Figure 14.

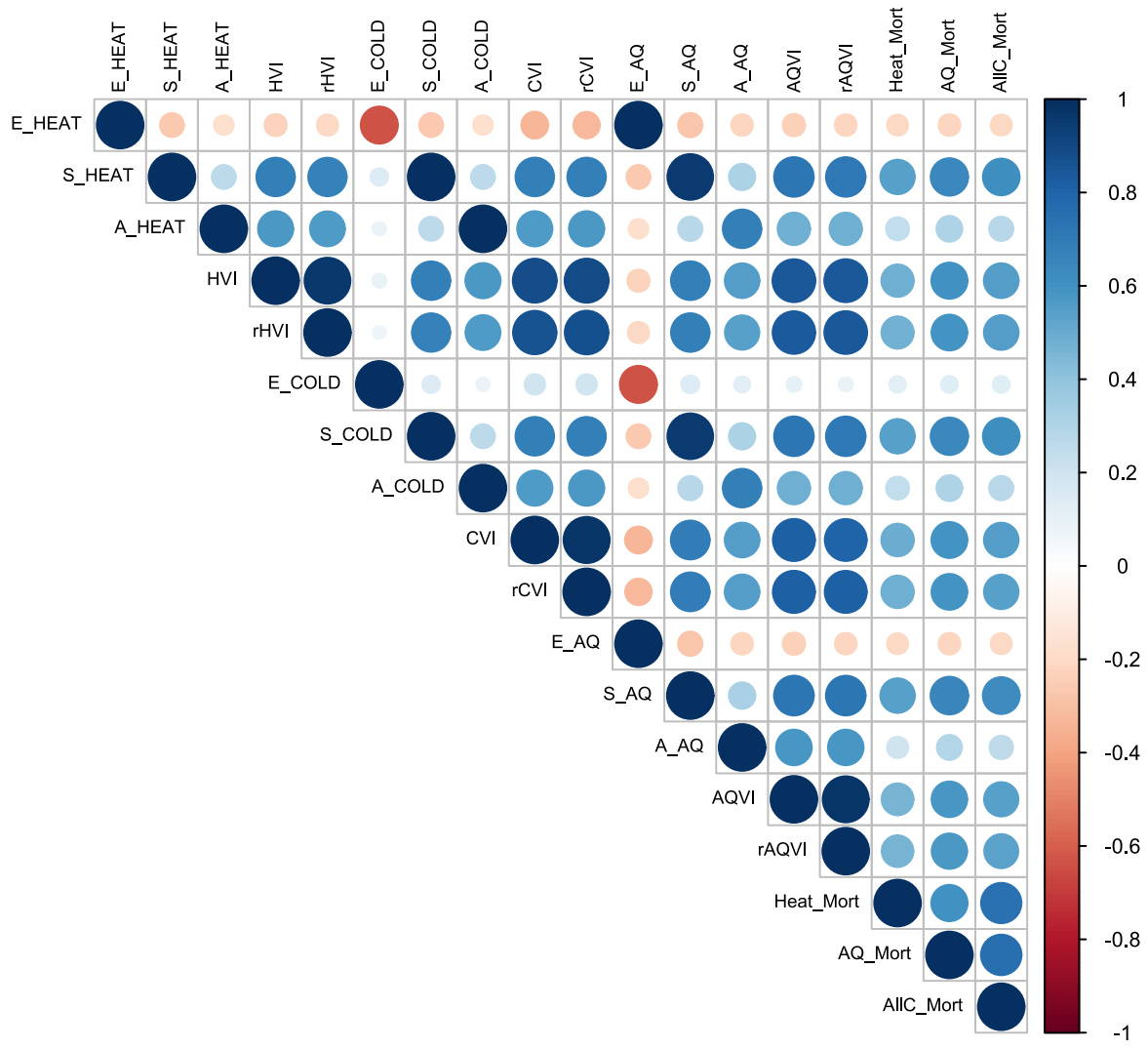

Figure 14: Pairwise correlation plot comparing weighted vulnerability indices and sub-indices with mortality data. E, S, and A correspond to exposure, sensitivity, and adaptive capacity sub-indices, respectively.

It is possible, from earlier discussion, that this approach may worsen the reliability of the vulnerability indices with respect to certain exposure-related health outcomes. However, all of the discussed methodologies have weaknesses. In future, there are many opportunities to expand data included in the indices, such as health service availability data, urban environment variables, and 2021 ABS Census data. There are also opportunities to explore the vulnerability index methodology, including a more rigorous assessment of underlying variables and related health outcomes. These potential expansions

considered, this set of vulnerability indices are the first of their kind, national vulnerability indices calculated over years, weeks, and months with a significant time period (2011-2019).

3.5. Summary Findings (Data Analysis Scope)

The data acquisition and formatting process adds significant value to the AusEnHealth pilot by taking steps to ensure that variables, indicators, and indices from existing data are displayed in the final product. Throughout the data formatting and data integration processes, several gaps and opportunities were identified which were not able to be actioned due to time restraints. Additionally, opportunities outside of the scope of the project have been identified which are worth noting for the potential future benefit of environmental health policy makers, researchers, and data custodians. The findings for the Data Analysis Scope Section are noted below and are presented alongside all other project recommendations in Appendix A.

Data Handling

One of the goals of AusEnHealth is to enable open access to the data used in the platform, so that the data can continue to provide insights past its use in the AusEnHealth pilot. The compilation of metadata has been a critical aspect of achieving this goal. Currently, access links, timescale, spatial resolution, update frequency, format, geographical extent, access restrictions, and last access date are manually collated per parameter into a metadata document. As AusEnHealth expands in future, current metadata documentation should be reviewed to ensure all relevant properties of accessed data are available to potential end users. In addition, metadata should be firmly integrated into future AusEnHealth applications, rather than housed solely in an external document.

A significant amount of time has been spent compiling data for use in the AusEnHealth pilot, including acquisition, data cleaning and formatting, and the calculation of relevant indicators and indices. There have also been significant time delays in seeking relevant nationwide health outcome data, which included the identification and compilation of relevant international classification of diseases (ICD) codes. To reduce the set-up time required in the first phase of AusEnHealth, a publicly available minimum dataset for environmental health should be created, including current data used in AusEnHealth, relevant national health outcome data, and data related to other areas in environmental health, such as vector borne disease, water quality, etc.

Analysis Methodology

As part of AusEnHealth, environmental health variables and indicators have been combined into vulnerability indices; a single value which allows for a comparison of population vulnerability across Australian statistical areas. Currently, vulnerability index methodology is limited to two main approaches; combining using equal parameter weights, or using a statistical technique called factor analysis. Significant weaknesses of these existing methodology are that neither approach is informed by health outcome data, there is currently no way to meaningfully compare between approaches, and there is no way to determine accuracy. In future, these methods should be improved to offer better insight into population vulnerability.

The first phase of AusEnHealth focuses on establishing platform architecture, a baseline data catalogue, and tools for visualising and presenting data. As a result, the AusEnHealth application offers limited capability for the analysis of associations and predictions. With a platform established, future phases of AusEnHealth should include effective data analysis and spatial modelling methods, with a focus on assessment of future trends and vulnerability assessment, including prediction of future population vulnerability.

4. Visualisation and Communication Tools

Computer-based visualisation is an essential component of the presentation of information to users, as it allows them to understand and analyse data effectively [137]. As part of the development of the AusEnHealth pilot, we have considered the importance of visualisation to the utility of our application. A pipeline from raw data to visualisations [138] consists of four main components: data modelling, data selection, visual mappings, and scene variables.

In previous sections of this report, core data processes, including data selection, data cleaning, standardisation, aggregation, and visual mappings have been discussed. In this section, we focus on how collected data is presented to end users, through a summary of numeric visualisation techniques and spatial visualisation techniques and a review of visualisation methods employed in existing applications.

4.1. Overview of Visualisation Techniques

Numeric visualisation is used in reports, websites, etc. to present relationships of varying complexity in a more digestible format for readers. The following lists contain a variety of techniques used in numeric visualisation (i.e., ignoring location) from [137] [139].

- **Graphs/Plots**:
  - Area Graph
  - Stacked bar graph
  - Bar chart
  - Scatterplot
  - Box and whisker plot
  - Bubble chart
  - Bubble graph
  - Candlestick chart
  - Density plot
  - Steamgraph
  - Error bars
  - Histogram
  - Kagi chart
  - Line graph
  - Marimekko chart
  - Multi-set bar chart
  - Open-high-low chart (OHLC)
  - Radar chart
  - Violin plot
- **Diagrams:**
  - Arc diagram
  - Flow chart
  - Network diagram
  - Sankey diagram
  - Timeline
  - Tree diagram
  - Venn diagram
  - Brainstorm
  - Chord chart
- **Tables:**
  - Heatmap (matrix)
  - Stem and leaf plot
  - Timetable
  - Gantt chart
- **Other:**
  - Donut chart
  - Dot matrix chart
  - Pie chart
  - Circle packing
  - Proportional area chart
  - Sunburst diagram
  - Word cloud

4.2. Spatial Visualisation Techniques

There are several options to consider before employing spatial visualisation. These include the distributions and frequency of data, category comparisons, relationships, and correlations, change over time, planning and development, participatory planning, risk assessment, multi-criteria decision analysis, etc. Occasionally, spatial analysis on its own can be inconclusive and requires additional analysis of numeric data (i.e., summary statistics) to reveal trends, correlations, etc. Therefore, due to its complexity, spatial data requires interactive applications to follow user-centred design principles to address cognitive aspects, e.g., presentation and exploration of numeric and spatial data for decision making. The following lists contains a variety of techniques used in spatial visualisation from various applications [138] [140] [141] [142].

- **Base map**
  - Various (provides context for referencing)
- **Thematic**
  - Choropleth (aggregated data, categorical or numerical. Comparisons, location, pattern)
  - Chorochromatic map (categorical or nominal data, colour symbols)
  - Bivariate/multivariate (two or more variables with colour and opacity)
  - Dot map (distribution, location, patterns)
  - Proportional point symbol (location, quantities, proportions)
  - Bubble map (changes over a dimension, e.g., time)
  - Linear cartogram (comparison, e.g., time, category)
- **Analysis**
  - Cluster map (density (latitude, longitude), continuous data)
  - Diagram map
  - Heatmap
  - Graph/node (networks, connectivity, relationship strength, pattern, distribution, relationships)
  - Flow (movement and flow, distribution, direction, location, relationship strength)
  - Isochrone (time)
  - 2D overlay
  - Slider (comparison, e.g., scale, time, variables, variables)
- **3D**
  - 3D overlay
  - Hexagonal binning (multiple data dimensions, e.g., quantity, range, and location)
  - Animation (elevation (z-axis) or comparison, e.g., density)
  - 3D extrusions (representation of time and/or distance from network, point or polygon, nodes (id, latitude, longitude), edges (id, from, to), areas, ways)
  - 3D environments and/or rendering (e.g., flood visualisation tool)
- **Relief**
  - Contour/Isoline
  - Slope
  - Triangulated Irregular Network (TIN)
  - Digital elevation model

4.3. Visual Variables

Techniques employed in numeric and spatial visualisations are influenced greatly by decisions made in the use of visualisation variables, such as text labels and positioning. The following lists contain a selection of prominent variables used within the above techniques and properties which can be used to further influence a user's experience with an interface [143] [138] [140] [144] [141] [142]. These lists are included to provide insight into the complexity of effective visual representation and do not capture every visual variable used in existing platforms.

Numeric data (non-spatial):

- Position (ordered, infinite values. Quantitative, ordinal, categorical, relational)
- Text labels (ordered (optional), e.g., alphanumeric, or numbered), infinite values, quantitative, ordinal, categorical, relational)
- Length (ordered, many values, quantitative, ordinal)
- Size, area (ordered, many values, quantitative, ordinal)
- Angle or orientation (ordered, few values, quantitative, ordinal)
- Pattern density, e.g., hachures (ordered, few values, quantitative, ordinal)
- Weight, boldness, e.g., line (ordered, few values, quantitative, ordinal)
- Colour saturation, brightness (ordered, few values, ordinal)
- Colour (few values (<20), categorical)
- Shape, icon (medium number of values, categorical)
- Pattern texture (medium number of values, categorical)
- Enclosure, connection (infinite values, categorical, relational)
- Line pattern (few values, relational)
- Line endings, e.g., arrows (few values, relational)
- Line weight (ordered, few values, ordinal)

Spatial data:

- Zoom
- Scale
- Tiles
- Projection
- Popup
- Overlay
- Radius
- Show/hide (zoom/scale dependent)
- Symbol rotation/orientation
- Size
- Height
- Spacing
- Colour
- Static/dynamic
- Marker/cluster
- Brushing and linking
- Interactive linking
- Grain and texture
- Symbol (e.g., pictorial, geometric)
- Visual hierarchy
- Lettering
- Spatial grouping
- Motion/direction
- Speed
- Geometric transformations (rotation, translation, scaling)

4.4. Summary of Existing Applications Review

In addition to a listing of the techniques and variables noted above, a review of existing data platforms has been carried out as part of AusEnHealth to identify attractive visualisation implementations. It is important to experience visualisation in practice to gain first-hand experience of characteristics which contribute to effective visualisation. The applications included in the review section are selected from

a variety of categories, such as population health, environmental health, population density, and pedestrian tracking. The purpose of this wide audit is to identify effective visualisation techniques which might add to the presentation of environmental health data in the AusEnHealth pilot. For brevity, only the below summary is included in this document. The full review is given in the external document "*EX. AusEnHealth Review of Existing Applications*".

Functions

- The applications reviewed include many different functions, which are dependent on the application's purpose:
    - Most applications reviewed included functions such as: interactive dashboards, brushing and linking between map views and summary statistics, which allow the user to explore data, identify relationships and patterns, easy access to additional information and share options. Other functions such as revealing additional information (e.g., through pop-up windows when clicking on an element in a map or chart) to improve understanding.
    - Some of the applications were difficult to read, particularly when the design did not allow the user to enlarge or zoom into charts, maps, or interact with the visualisation to find information.
- User requirements are needed to determine the functions to be included in the design, rather than choosing these based on their novelty. Depending on these, example functions could be simple visualisations, complex analysis tools, communication of research findings, comparisons, decision-making, making statements, sharing information, access to information and documentation, etc.
- User engagement to determine user experience and usability goals, is imperative for design to ensure the usefulness of an application for the user profile. This will determine aspects such as the application's layout and features.
    - A good layout will need to consider composition, balance, proportion, visual hierarchy (e.g., what elements are important, how much space they occupy, how and in what order does the user "*read*" the dashboard).
    - The platform should be intuitive, user friendly, with clear, easy to navigate and to find tools and other interactive elements. Too many options may lead users to confusion, while not enough may lead to decreased utility.
    - While understanding that certain applications may require complexity, especially if they are directed towards expert-level users such as the scientific community, it is helpful to have the option of a user guide (e.g., a quick tour) and a "*narrative*" to communicate with the wider audience.
- Digital platforms need to consider user friendliness and accessibility; this includes aspects such as colour-blind safe schemes for maps and charts, flexible layout and template options that allow enlarging elements (such as small charts) and minimising/removing others that may not be needed to enable the user to focus on what is needed for a specific task.

Data

- Applications reviewed mostly include quantitative data from the areas of population health, environment, climate, population density, pedestrian flow. Population health is frequently used in combination with environmental and climate data. Only a few of the analysed applications include built environment data, which is useful for understanding the complex relationship between human habitat, environment, and population health in a broader context.
- All applications include spatial data, while some of them also include numeric only and temporal data. Most frequently, spatial data is represented as choropleth and proportional point maps.

Numeric data is mostly represented through summary statistics in the form of bar charts and scatter plots.

- The most effective tools will allow brushing and linking maps, charts, tables, timelines, as well as options to filter and sort data. These features are particularly important in complex spatial-temporal datasets.
- The overall sustainability of the platform should be given careful consideration (e.g., how current is the data and how frequently will it be updated), can users upload their own data (e.g., new geographic boundaries)

Visualisation

- User requirements will guide the level of functions available and how these will draw on the data available. More complex spatial-temporal data requires more analysis tools and visualisation options than plain 2D data, such as a simple choropleth.
- In contrast to differences in function and data, the use of visual variables and their mappings to variables is more common.
    - Most visualisations include polygons, proportionate symbols, points, place markers and lines to represent spatial data. The visual variables applied to these visualisations are colours, transparency, and line weights.
    - Numeric data is mostly represented through simple bar charts and scatterplots. The visual variables applied to these visualisations are bar length and colour.
    - Spatial-temporal data is represented through timelines, sliders, or area charts.
- The most effective designs encountered in this review are based on good usability: clean and intuitive dashboards enable the user to easily understand and explore the data. A high degree of customisation and control includes features such as: zoom, selection, change base maps, switch layers on/off, ability to adjust colours, contrast, and opacity.
- Some issues identified include the use of visual variables such as colour (e.g., the use of pastel colours and transparency overlay, use of too many colours), lack of mapping conventions (e.g., missing north arrow, borders, labels, scale, etc.). Other visual variables that could be useful could include the use of patterns, motion, and speed.

Users, experience, usability, and the design process

- Functions, data, and choice of visualisation are all dependent on the purpose for the application. All the applications reviewed in this report have been designed for a specific purpose and are quite diverse. While there is no formula to create a usable application that can be universally applied to any problem, this review has revealed there are elements and combinations that could be drawn upon.
- Consideration of the application's purpose needs to address the intended user profiles and their context from which requirements can be specified for design. This may consider a range of factors such as the level of complexity (e.g., how detailed a dashboard or simple a narrative may need to be to address the purpose effectively), its functions, accessibility, how users will interact with the platform, what information is used and how is it shared, etc. Depending on the specific user profiles (e.g., scientific community, government, planners, broader community), features will vary greatly and will need to be addressed through usability and user experience.
- Understanding differences between user experience and usability is important. Usability defines how users "*achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use*" (International Organisation for Standardisation [ISO], 2019). User experience aims to address perceptions and response before, during and after interacting with the platform. Careful consideration needs to be given to user emotions, preferences, perceptions, behaviours,

and accomplishments, as these may vary depending on the user profile (ISO, 2019). A formal design process using contextual inquiry/analysis activities, such as interviews or workshops, is recommended to tackle the complexity of delivering a tool for different purposes.

4.5. Summary Findings (Visualisation and Communication Tools)

The visualisation review has identified valuable data presentation techniques to be incorporated into the AusEnHealth pilot in addition to highlighting specific design choices to improve user insights, intuition, and overall user experience. Through discussions with project stakeholders, a small number of suggestions were made which are summarised in this section as messaging and communication recommendations to improve the presentation of data and how insights are communicated to end users. These recommendations are presented alongside all other project recommendations in Appendix A.

Messaging and Communication

The digital platform review revealed that communication of data and analysis varies significantly from platform to platform. It is important for end users to understand what data they are seeing, where it came from, and how it relates to other data shown in the platform, to avoid user confusion, as well as incorrect inferences of causation. The inclusion of analysis methods in future phases of AusEnHealth must be met with clear messaging around statistics outputs and analyses, including messaging around associations between variables.

In future, one goal of AusEnHealth is to include predictive capabilities for vulnerability and health impact assessment. As these predictions result from statistical models, it is important to consider now the visualisation of uncertainty. Including uncertainty adds more detail to research end users, while potentially adding confusion to the public end user group. As clarity is an important aspect of communication, future phases of AusEnHealth should carefully consider how uncertainty can be communicated without damaging potential user insights.

Parameters used in the AusEnHealth application have been given simple definitions in a hover-over window in the main user interface, with a more detailed definition in the external metadata document. Further, users are able to access this report from the application, which provides additional technical detail. For the benefit of all end user groups, future phases of AusEnHealth should work to maintain this level of linkage between definitions and methodology, or even improve linkage by incorporating expanded digital architecture. This recommendation links strongly with messaging, as future communication of association and causation must be consistent throughout a platform which aims to assist with policy planning.

5. Platform Development Scope

Currently there is no national digital representation of environmental and health indicators at a local level, to enable policy makers, health managers and researchers to identify vulnerable populations, predict future disease burden and plan for a changing climate in a coordinated, timely manner. AusEnHealth will lay the foundation for an operational open-access Australian environmental health decision support platform by developing targeted demonstrator(s) of scenarios using real data, indicators, and indices, where possible. These demonstrators are designed to capture specific insights and are targeted towards specific end users. Developing the pilot in this way enables flexible collaboration with the data analysis team, incorporating new data into demonstrators as it becomes available. Additionally, this modular approach enables user feedback to guide the future of AusEnHealth by focusing in on any one or more demonstrators for expansion.

This section of the methods report contains an overview of the User Journey Mapping Workshops, as well as a summary of each of the contained demonstrators, including target users, scenarios, incorporated data, and planned functionality.

Technology Readiness

The AusEnHealth pilot is an experimental proof of concept, specific to the scenarios identified. The resulting pilot is not intended to be a commercial ready product and is not intended to be fully functional. The development effort was concentrated in areas deemed to provide the highest value demonstrations of a future production version of an environmental health decision support platform. The development of the Pilot targeted a Technology Readiness Level (TRL) of 3, however, development realised a TRL of 6 during development. See Figure 15 below for TRL descriptions.

Figure 15: Chart of Technology Readiness Level descriptions.

5.1. User Journey Mapping

User Journey Mapping was conducted through two separate User Journey Mapping Workshops:

- Climate and Health Outcomes Related to Heat | 1st June 2021
- Air quality and Respiratory Health Outcomes | 4th June 2021

The Primary objective of the User Journey Mapping was to identify the core insights that could be delivered by AusEnHealth, including the following:

- User (who)
  - Need to identify the specific user profiles for the AusEnHealth digital twin
- Scenarios (what)
  - Need to identify the specific scenarios that will be addressed for the respective users
- Goals and Expectations (how)
  - What the user wants to do and what they want to achieve

Figure 16 shows a process overview for the User Journey Mapping Workshops.

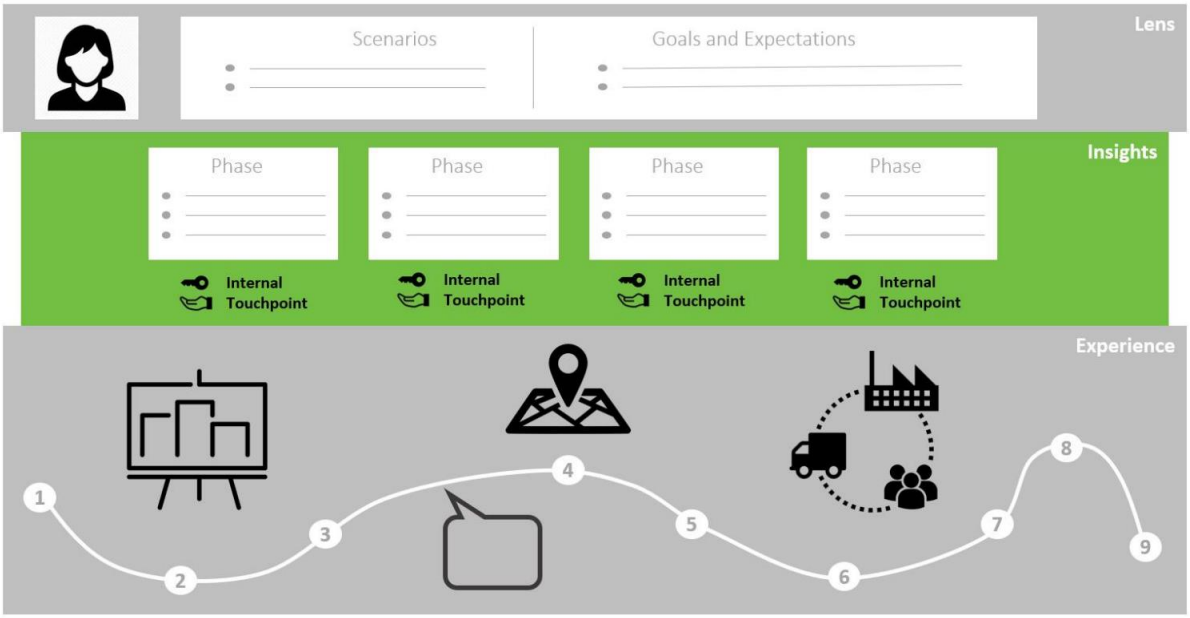

Figure 16: Overview of User Journey Mapping process.

Climate and Health Outcomes Related to Heat

Target Users

The target users for the Climate and Health Outcomes Related to Heat use case were identified by project stakeholders as:

Target User	Use
State Government	Strategic Policy and Health Service Planning
State and Local Government	Urban Planning
Researchers and Consultants	Data Driven Climate and Health Analysis

Target Scenarios

Target scenarios were identified for each target user group as the foundation for developing the AusEnHealth pilot for Climate and Health Outcomes related to Heat. The target users for the climate and Health Outcomes Related to Heat user case were identified by project stakeholders as:

- State Government
  - Strategic policy and health service planning | Vulnerability assessment
  - Strategic policy and health service planning | Service utilisation
  - Strategic policy and health service planning | Communicating climate change and health data to influence policy
- State and Local Government
  - Urban planning | Analysing urban heat forecasting
  - Urban planning | Urban vulnerability and community resilience
  - Urban planning | Identification of urban heat islands requiring adaptation measures
- Researchers and Consultants
  - Data driven climate and health analysis | Access to disparate health data including raw data
  - Data driven climate and health analysis | Flexible and scalable analysis of health data

Air Quality and Respiratory Health Outcomes

Target Users

The target users for the Air Quality and Respiratory Health Outcomes use case were identified by project stakeholders as:

Target User	Use
State Government	Strategic Advice
Public	General Awareness Raising
Community Groups	Campaign-based Awareness Raising

Target Scenarios

Target scenarios were identified for each target user group as the foundation for developing the AusEnHealth pilot for Air Quality and Respiratory Health Outcomes. The target users for the Air Quality and Respiratory Health Outcomes user case were identified by project stakeholders as:

- State Government | Strategic Advisory
    - Industry driven particle emissions analysis and communication for strategic engagement and advice
    - Transport driven particle emissions analysis and communication for strategic engagement and advice
- Public Users |General Awareness Raising
    - Broad scale Industry driven particle emission communication
    - Broad scale Transport driven particle emission communication
- Community Groups | Campaign Based Awareness Raising
    - Localised campaign Industry driven particle emission communication
    - Localised campaign Transport driven particle emission communication

5.2. Functional Requirements

The below functional requirements are applicable to the overall proposed AusEnHealth pilot solution.

Business Requirements

The following business requirements are applicable to the development of the AusEnHealth pilot:

ID	Requirement
FR01	Development of a pilot as an experimental proof of concept.
FR02	Development of an online environmental health decision support platform.
FR03	Enable users to access environmental health data.
FR04	Enable users to visualise environmental health data.
FR05	Enable users to analyse environmental health data.
FR06	Enable users to access environmental health indicators.
FR07	Enable users to visualise environmental health indicators.
FR08	Enable users to analyse environmental health indicators.
FR09	Provide tools to support adaptation planning.
FR10	Provide tools to support vulnerability assessment.
FR11	Provide tools to support decision making.
FR12	The platform will enable a range of users including the public, media, policy analysists, and decision-makers, as well as researchers and scientists.

Administrative Functions

There is no requirement for administrative function for the AusEnHealth pilot.

User Requirements

Refer to Section 5.4 for scenario specific details of user requirements.

System Requirements

The following system requirements are applicable to the development of the AusEnHealth Pilot:

ID	Requirement
FR13	The pilot will be made available as a web application to be accessed through a URL.
FR14	The pilot will require operation through a modern web browser (current version – 1, targeting Google Chrome).
FR15	The pilot will be designed for desktop use and will not be required to be mobile responsive.

5.3. Non-Functional Requirements

Usability

The following usability requirements are applicable to the development of the AusEnHealth pilot:

ID	Requirement
NFR01	The pilot will be visual and familiar
NFR02	The pilot will leverage digital mapping as a key communication mechanism
NFR03	The pilot will present consolidated data and indicators that can be further expanded.
NFR04	The pilot shall present scenarios in a targeted way to simplify communication of capability.

Availability

The following availability requirements are applicable to the development of the AusEnHealth pilot:

ID	Requirement
NFR05	The pilot is a non-critical business application and will be available as required for demonstrations purposes.
NFR06	The pilot will be available for the duration of the first phase of AusEnHealth once released and for a period of 6 months post completion.

Scalability

The following scalability requirements are applicable to the development of the AusEnHealth pilot:

ID	Requirement
NFR07	The pilot will be designed for less than 100 concurrent users.
NFR08	There is no requirement for the pilot to scale for usage.

Performance

The following performance requirements are applicable to the development of the AusEnHealth pilot:

ID	Requirement
NFR09	The pilot will need to demonstrate high performance for map visualisation to demonstrate future applications.
NFR10	The pilot will need to demonstrate high performance for geographic data analysis to demonstrate future applications.
NFR11	The pilot will need to demonstrate high performance for tabular data querying to demonstrate future applications.

Supportability

There is no requirement for ongoing support post release of the AusEnHealth pilot.

Security

There is no requirement for security for the AusEnHealth pilot.

5.4. Solution Design

The solution is designed to deliver targeted demonstrators of digital twin capability to address the use cases, target users, and scenarios that have been identified for the AusEnHealth pilot. A landing page has been delivered to provide a consolidated, branded access point for the demonstrators. The demonstrators that have been developed are listed in the table below:

Use Case	Demonstrator
Climate and Health Outcomes Related to Heat	Risk Assessment
Climate and Health Outcomes Related to Heat	Climate Change Assessment
Climate and Health Outcomes Related to Heat	Multi-Cloud Health Data Analytics
Air Quality and Respiratory Health Outcomes	Impact Analysis
Air Quality and Respiratory Health Outcomes	Public Awareness
Air Quality and Respiratory Health Outcomes	Campaign Awareness

Each demonstrator is a separate web page that will target core insights and capabilities for specific scenarios. The development of targeted demonstrators will enable clear communication of capability and vision for the project with the ability for a digital twin approach to address a range of climate and health related applications.

Landing Page

The pilot has a landing page that provides access to the targeted demonstrators (example landing page shown below in Figure 17).

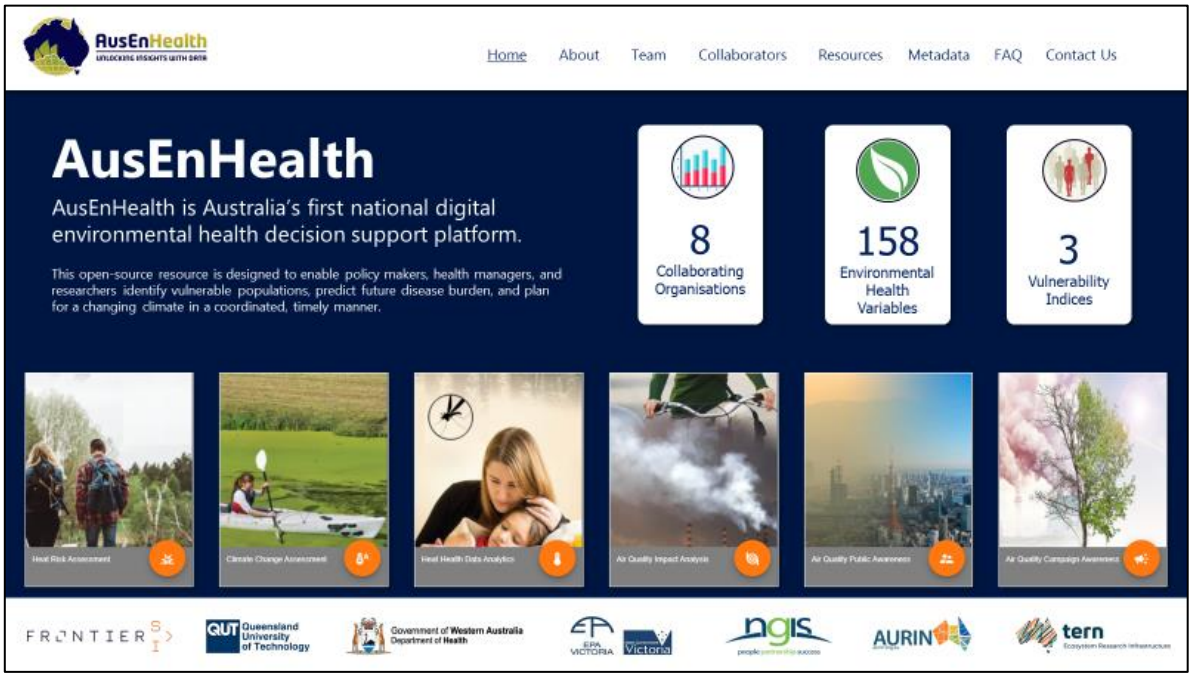

Figure 17: Example of landing page user interface.

The following user requirements are applicable to the proposed landing page:

ID	Requirement
LPUR01	The landing page will be branded in line with existing project branding (https://frontiersi.com.au/ausenhealth-digital-twin/).
LPUR02	The landing page will have a meu for accessing project information including partners and project overview.
LPUR03	The landing page will have buttons to access each targeted demonstrator.
LPUR04	Each demonstrator button will have an image (default as screenshot of the demonstrator page) and text.

5.5. Heat | Risk Assessment

This demonstrator relates to the *Climate and Health Outcomes Related to Heat* use case. This demonstrator addresses the following core insight(s):

- Which areas/communities are currently most at risk from heat?

Users

The user(s) addressed by this demonstrator:

Target User	Use
State Government	Strategic Policy and Health Service Planning

Scenario

The scenario(s) addressed by this demonstrator:

Target User	High Level Scenario
State Government	Strategic Policy and Health Service Planning | Vulnerability Assessment

Data

The variables, indicators, and indices used in this demonstrator are listed in the following table. Note any later excluded items in red italics. The final *heat risk assessment* demonstrator did not utilise the threshold-based indicators (at risk, high risk, heatwave, and severe heatwave days). These indicators were instead highlighted in the *climate change assessment* demonstrator. All other listed data was successfully integrated into this demonstrator in the AusEnHealth application on project completion.

Data Type	Data Name
Variable	Historically ranked average temperature, with heat high
Indicator	*Number of "at risk" temperature days, heat*
Indicator	*Number of "high risk" temperature days, heat*
Indicator	*Number of heatwave days*
Indicator	*Number of severe heatwave days*
Index	Heat exposure index
Index	Heat sensitivity index
Index	Heat adaptive capacity index
Index	Heat vulnerability index
Variable	Historically ranked average temperature, with cold high
Indicator	*Number of "at risk" temperature days, cold*
Indicator	*Number of "high risk" temperature days, cold*
Indicator	*Number of coldwave days*
Indicator	*Number of severe coldwave days*
Index	Cold exposure index

Index	Cold sensitivity index
Index	Cold adaptive capacity index
Index	Cold vulnerability index

User Requirements

The user requirements applicable to the proposed demonstrator are listed in the following table. Note the items in red italics are user requirements not met at the time of project completion. To maintain performance, grid data was not used in any of the AusEnHealth demonstrators. As a result, user requirement DM1UR13 was not met. All of the other user requirements were successfully implemented.

ID	User Requirement
DM1UR01	The user interface will have a map component, an information panel, and a benchmarking chart as core components.
DM1UR02	The map will be an interactive map with different basemap options (street, satellite).
DM1UR03	The information panel will have a table control to enable information to be filtered and sorted.
DM1UR04	The information panel will be collapsible to enable a full screen map.
DM1UR05	The benchmarking chart will be a bar chart with a mean trend line and will be linked to the map to display for selecting areas.
DM1UR06	There will be a location search (address, suburb, etc.).
DM1UR07	There will be a risk layer displayed on the map aggregated to statistical areas. The risk layer will display traffic light colours from low to high risk where risk is represented by the combination of vulnerability indices and climate variables and indicators.
DM1UR08	The panel will show a list of the statistical areas with values for overall risk, vulnerability indices, and climate variables and indicators.
DM1UR09	The user can select which variable, indicator, or index to use.
DM1UR10	The user can sort the panel table listing.
DM1UR11	The user can click on a statistical area on the map to display a popup with the overall risk value.
DM1UR12	The user can access further details about the variables, indicators, and indices by expanding the specific row in the panel or by clicking on a “details” link in the popup.
DM1UR13	*The user can display a grid layer representation of climate data.*
DM1UR14	The user can apply a temporal setting to the climate data.
DM1UR15	The user can click on a location to display a time series graph of climate data.

This demonstrator shares a number of features observed in existing data platforms, the most common being a digital mapping focus. The Queensland Social Atlas uses a central map panel for the user interface, linked with a tabular-based information panel on the left and a benchmarking chart at the bottom of the interface. The Victorian Women’s Health Atlas also uses a similar interface approach with a central map component, information panel on the left, and benchmarking chart with tends and averages at the bottom. A similar user interface has been designed to present data to end users in this Heat Risk Assessment demonstrator.

5.6. Heat | Climate Change Assessment

This demonstrator relates to the *Climate and Health Outcomes Related to Heat* use case. This demonstrator addresses the following core insight(s):

- Which areas/communities will be most at risk from heat in the future due to climate change?
- Which areas/communities will transform into a high-risk category for heat in the future?

Users

The user(s) addressed by this demonstrator:

Target User	Use
State Government	Strategic Policy and Health Service Planning

Scenario

The scenario(s) addressed by this demonstrator:

Target User	High Level Scenario
State Government	Strategic Policy and Health Service Planning | Communicating Climate Change and Health Data to Influence Policy

Data

The variables, indicators, and indices used in this demonstrator are listed in the following table. Note any later excluded items in red italics, and any later included items in green italics. The final *climate change assessment* demonstrator did not utilise excess heat factor or excess cold factor variables and did not incorporate any of the above indices. These data were instead highlighted in the *heat risk assessment* demonstrator. Other listed data was successfully integrated into this demonstrator in the AusEnHealth application on project completion. In addition, average temperature was added to this demonstrator after seeing the significant changes in indicator values over time.

Data Type	Data Name
Variable	*Excess heat factor*
Variable	*Excess cold factor*
Indicator	Number of “at risk” temperature days, heat (historical and forecast)
Indicator	Number of “high risk” temperature days, heat (historical and forecast)
Index	*Heat exposure index*
Index	*Heat sensitivity index*
Index	*Heat adaptive capacity index*
Index	*SEIFA IRSAD*
Variable	*Average temperature (historical and forecast)*

User Requirements

The user requirements applicable to the proposed demonstrator are listed in the following table. Note the items in red are user requirements not met at the time of project completion. With DM2UR07 implemented, it was not necessary to additionally implement change in risk in the initial design. Due to time constraints, DM2UR08 was not implemented into the final *Climate Change Assessment* demonstrator. To maintain performance, grid data was not used in any of the AusEnHealth demonstrators. As a result, user requirement DM2UR14 was not met. All of the other user requirements were successfully implemented.

ID	User Requirement
DM2UR01	The user interface will have a map component, an information panel, and a benchmarking chart as core components.
DM2UR02	The map will be an interactive map with different basemap options (street, satellite).

DM2UR03	The information panel will have a table control to enable information to be filtered and sorted.
DM2UR04	The information panel will be collapsible to enable a full screen map.
DM2UR05	The benchmarking chart will be a bar chart with a mean trendline and will be linked to the map to display for selecting areas.
DM2UR06	There will be a location search (address, suburb, etc.).
DM2UR07	There will be a risk layer displayed on the map aggregated to statistical areas. The risk layer will display traffic light colours from low to high risk where risk is represented by the combination of accessed and derived environmental health variables and indicators.
DM2UR08	*The risk layer will have symbology options for current risk, future risk, and change in risk. The default symbology will be in charge in risk due to climate change for this demonstrator.*
DM2UR09	The panel will show a list of the statistical areas with values for vulnerability index, current climate indicator, future climate indicator, and change in climate indicator.
DM2UR10	The user can select which variables or indicators to use for both vulnerability and climate, respectively.
DM2UR11	The user can sort the panel table listing of vulnerability index, current climate indicator, future climate indicator, and change in climate indicator.
DM2UR12	The user can click on a statistical area on the map to display a popup with the vulnerability index and climate variables and indicators.
DM2UR13	The user can access further details about the indicators by expanding the specific row in the panel, or by clicking on a “details” link in the popup.
DM2UR14	*The user can display a grid layer representation of climate variables or indicators.*
DM2UR15	The user can apply temporal settings to the climate variable or indicator grid layer.
DM2UR16	The user can click on a location to display a time series graph of climate variable or indicator values.

5.7. Heat | Multi-Cloud Health Data Analysis

This demonstrator relates to the *Climate and Health Outcomes Related to Heat* use case. This demonstrator addresses the following core insight(s):

- How can large scale health data analysis be performed using multiple sources?
- Future capability of a next generation health data ecosystem.

This demonstrator showcases the benefits of leveraging industry leading data analysis technology to enable data scientists to invest more time in analysis and insights as opposed to configuration and management of data repositories. Examples of some traditional data analytics model characteristics that are avoided using a serverless model are demonstrated in Figure 18.

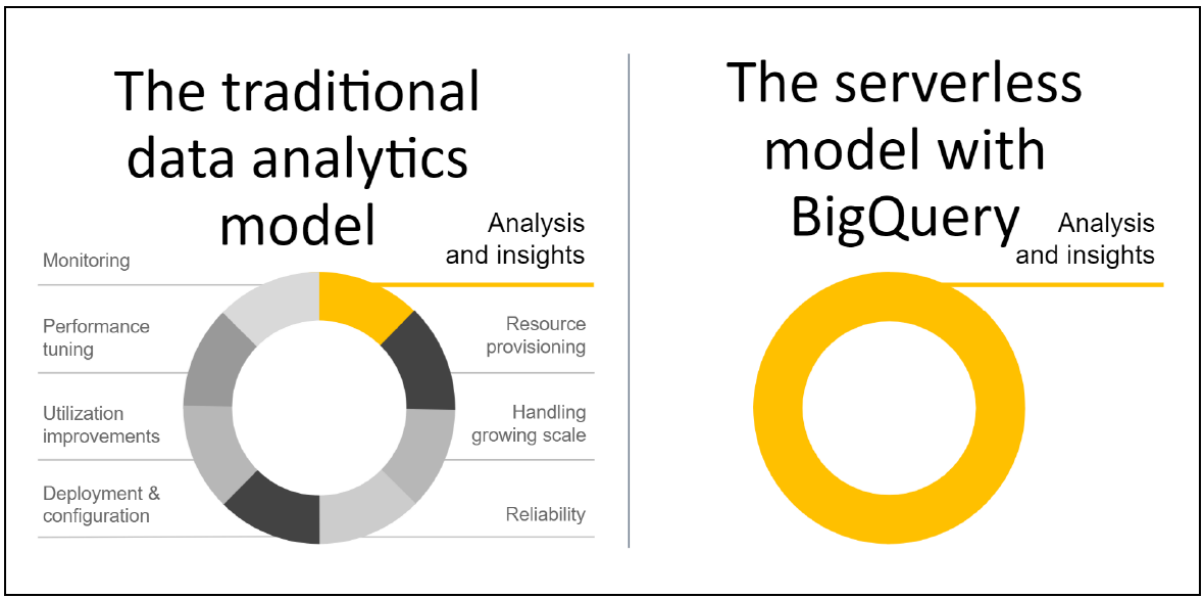

Figure 18: BigQuery data analytics model.

Users

The user(s) addressed by this demonstrator:

Target User	Use
Researchers and Consultants	Data Driven Climate and Health Analysis

Scenario

The scenario(s) addressed by this demonstrator:

Target User	High Level Scenario
Researchers and Consultants	Data Driven Climate and Health Analysis | Access to Disparate Health Data Including Raw Data
Researchers and Consultants	Data Driven Climate and Health Analysis | Flexible and Scalable Analysis of Health Data

Data

There were no specific indicators required for this demonstrator. The objective is to provide access to external raw data for analysis and modelling.

User Requirements

The user requirements applicable to the proposed demonstrator are listed in the following table. Note the items in red are user requirements not met at the time of project completion. The multi-cloud component of the demonstrator (i.e., drawing data from external data custodians) was established in the AusEnHealth application. However, the capability to visualise results was not implemented in this phase of the project. As a result, neither DM3UR05 nor DM3UR06 were met. In addition, while querying is possible, health data was not implemented or mocked up as part of this demonstrator. This means that DM3UR02 is only partially met. All of the other user requirements were successfully implemented.

ID	User Requirement
DM3UR01	The user interface will have a query component, results component, and a visualisation component.

DM3UR02	*The user will be able to construct queries from multiple cloud data repositories (raw health data) to simulate the capability to query health data sources without having to replicate the data.*
DM3UR03	The user will be able to construct and execute SQL queries and leverage multiple cloud data repositories, including spatial queries.
DM3UR04	The user will be able to view the results of SQL queries.
DM3UR05	*The user will be able to visualise results on a dynamic map for spatial queries.*
DM3UR06	*The user can customise the display of map results.*

This demonstrator showcases the capabilities of leading cloud technology to address multi-cloud ecosystems of data, enable consultants and researchers to reduce the effort required in managing data to focus on health data science. Figure Figure 19 provides an example of the query interface that will be available through this demonstrator. For the purposes of the pilot application, multiple data sources (i.e., BOM, ABS, health outcome data custodians, etc.) will be simulated to showcase capability.

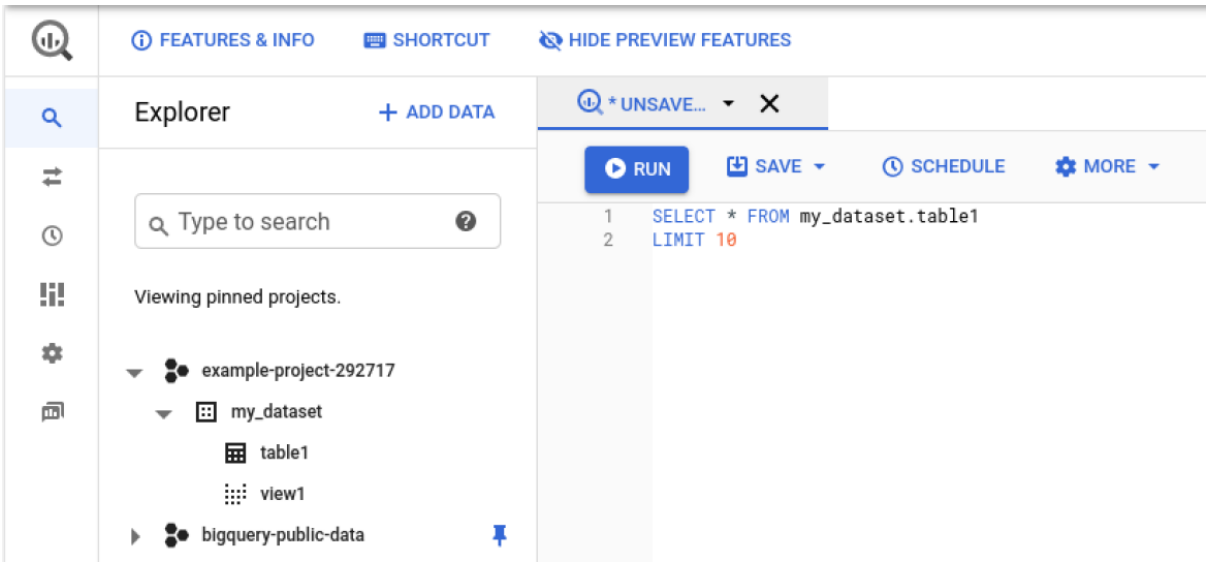

Figure 19: BigQuery user interface (http://cloud.google.com).

5.8. Air Quality | Impact Analysis

This demonstrator relates to the *Air Quality and Respiratory Health Outcomes* use case. This demonstrator addresses the following core insight(s):

- Spatial analysis of air quality data against health outcome data to identify patterns.

Users

The user(s) addressed by this demonstrator:

Target User	Use
State Government	Strategic Policy and Health Service Planning

Scenario

The scenario(s) addressed by this demonstrator:

Target User	High Level Scenario
State Government	Strategic Policy and Health Service Planning | Industry Driven Particle Emissions Analysis and Communication for Strategic Engagement and Advice

State Government	Strategic Policy and Health Service Planning | Transport Driven Particle Emissions Analysis and Communication for Strategic Engagement and Advice

Data

The variables, indicators, and indices used in this demonstrator are listed in the following table. Note any later included items in green italics.

Data Type	Data Name
Indicator	Number of "above average" ozone (O3) days
Indicator	Number of "potential risk" ozone (O3) days
Indicator	Number of "above average" nitrogen monoxide (NO) days
Indicator	Number of "potential risk" nitrogen monoxide (NO) days
Indicator	Number of "above average" nitrogen dioxide (NO2) days
Indicator	Number of "potential risk" nitrogen dioxide (NO2) days
Indicator	Number of "above average" particulate matter >2.5µm (PM2.5) days
Indicator	Number of "potential risk" particulate matter >2.5µm (PM2.5) days
Index	Air quality vulnerability index
Index	Air quality exposure index
Index	Air quality sensitivity index
Index	Air quality adaptive capacity index
Variable	*Respiratory mortality*
Variable	*Asthma mortality*
Variable	*Cardiovascular disease mortality*
Variable	*All-cause mortality*

All listed data was successfully integrated into this demonstrator in the AusEnHealth application on project completion. In addition, mortality data from the Mortality Over Regions and Time (MORT) Books were later included to offer a comparison of air quality data with health outcome data.

User Requirements

The user requirements applicable to the proposed demonstrator are listed in the following table. Note the items in red are user requirements not met at the time of project completion. The data and table components of this demonstrator have been completed, with supporting scatter plots comparing air quality parameters with relevant health outcomes. The mapping component (covering DM4UR01 and DM4UR11), as well as the more detailed plotting components (covering DM4UR13-16) were not implemented in the final demonstrator. All of the other user requirements were successfully implemented.

ID	User Requirement
DM4UR01	*The user interface will have a map component, an information panel, and a benchmarking chart as core components.*
DM4UR02	The map will be an interactive map with different basemap options (street, satellite).
DM4UR03	The information panel will have a table control to enable information to be filtered and sorted.
DM4UR04	The information panel will be collapsible to enable a full screen map.
DM4UR05	The scatter plot will show the correlation between air quality data and health outcome data.
DM4UR06	There will be a location search (address, suburb, etc.).
DM4UR07	There will be an air quality risk layer displayed on the map aggregated to statistical areas. The risk layer will display traffic light colours from low to high risk where

	risk is represented by the combination of accessed and derived environmental health variables and indicators.
DM4UR08	The panel will show a list of the statistical areas with values for vulnerability index, air quality indicators and variables, and selected health outcome data.
DM4UR09	The user can select which data to use for both air quality and health outcome data.
DM4UR10	The user can sort the panel table listing of air quality and health outcome data.
DM4UR11	*The user can click on a statistical area on the map to display a popup with the vulnerability index and priority health outcome data.*
DM4UR12	The user can access further details by expanding the specific row in the panel or by clicking on a "details" link in the popup.
DM4UR13	*The user can access a temporal graph plotting health outcome data against air quality data for a specific statistical area.*
DM4UR14	*The user can return benchmarking statistics for health outcome data and air quality data for a specific statistical area.*
DM4UR15	*The user can select multiple statistical areas to access a graph comparing air quality data over time.*
DM4UR16	*The user can select multiple statistical areas to access a graph comparing health outcome data over time.*

5.9. Air Quality | Public Awareness

This demonstrator relates to the *Air Quality and Respiratory Health Outcomes* use case. While planned, this *Public Awareness* demonstrator was not implemented into the first phase AusEnHealth application.

This demonstrator was set to focus on addressing the following core insight(s):

- Spatial distribution and benchmarking of air quality data.

Users

The user(s) addressed by this demonstrator:

Target User	Use
Public	General Awareness Raising

Scenario

The scenario(s) addressed by this demonstrator:

Target User	High Level Scenario
Public	General Awareness Raising | Broad Scale Industry Driven Particle Emission Communication
Public	General Awareness Raising | Broad Scale Transport Driven Particle Emission Communication

Data

The variables, indicators, and indices used in this demonstrator are listed in the following table. Note any later excluded items in red italics.

Data Type	Data Name
Indicator	*Number of "above average" ozone (O3) days*
Indicator	*Number of "potential risk" ozone (O3) days*
Indicator	*Number of "above average" nitrogen monoxide (NO) days*
Indicator	*Number of "potential risk" nitrogen monoxide (NO) days*

Indicator	*Number of "above average" nitrogen dioxide (NO2) days*
Indicator	*Number of "potential risk" nitrogen dioxide (NO2) days*
Indicator	*Number of "above average" particulate matter >2.5µm (PM2.5) days*
Indicator	*Number of "potential risk" particulate matter >2.5µm (PM2.5) days*
Index	*Air quality vulnerability index*
Index	*Air quality exposure index*
Index	*Air quality sensitivity index*
Index	*Air quality adaptive capacity index*

User Requirements

The user requirements applicable to the proposed demonstrator are listed in the following table. Note the items in red are user requirements not met at the time of project completion.

ID	User Requirement
DM5UR01	*The user interface will have a map component and an information panel.*
DM5UR02	*The map will be an interactive map with different basemap options (street, satellite).*
DM5UR03	*The information panel will be collapsible to enable a full screen map.*
DM5UR04	*There will be a location search (address, suburb, etc.).*
DM5UR05	*There will be an air quality risk layer displayed on the map aggregated to statistical areas. The risk layer will display traffic light colours from low to high risk where risk is represented by the combination of accessed and derived environmental health variables and indicators.*
DM5UR06	*The panel will show a list of the statistical areas with values for vulnerability index, air quality indicators and variables, and selected health outcome data.*
DM5UR07	*The user can select which data to use for both air quality and health outcome data.*
DM5UR08	*The user can sort the panel table listing of air quality and health outcome data.*
DM5UR09	*The user can click on a statistical area on the map to display a popup with the vulnerability index and priority health outcome data.*
DM5UR10	*The user can access further details by expanding the specific row in the panel or by clicking on a "details" link in the popup.*
DM5UR11	*The user can access a temporal graph plotting health outcome data against air quality data for a specific statistical area.*
DM5UR12	*The user can return benchmarking statistics for air quality data for a specific statistical area.*
DM5UR13	*The user can select multiple statistical areas to access a graph comparing air quality data over time.*

This *Public Awareness* demonstrator was planned to focus on clean communication of air quality data for public consumption, similar to the Environmental Insights Explorer from Google as shown in Figure Figure 20 below. Due to time constraints, this demonstrator was not implemented into the first phase AusEnHealth application.

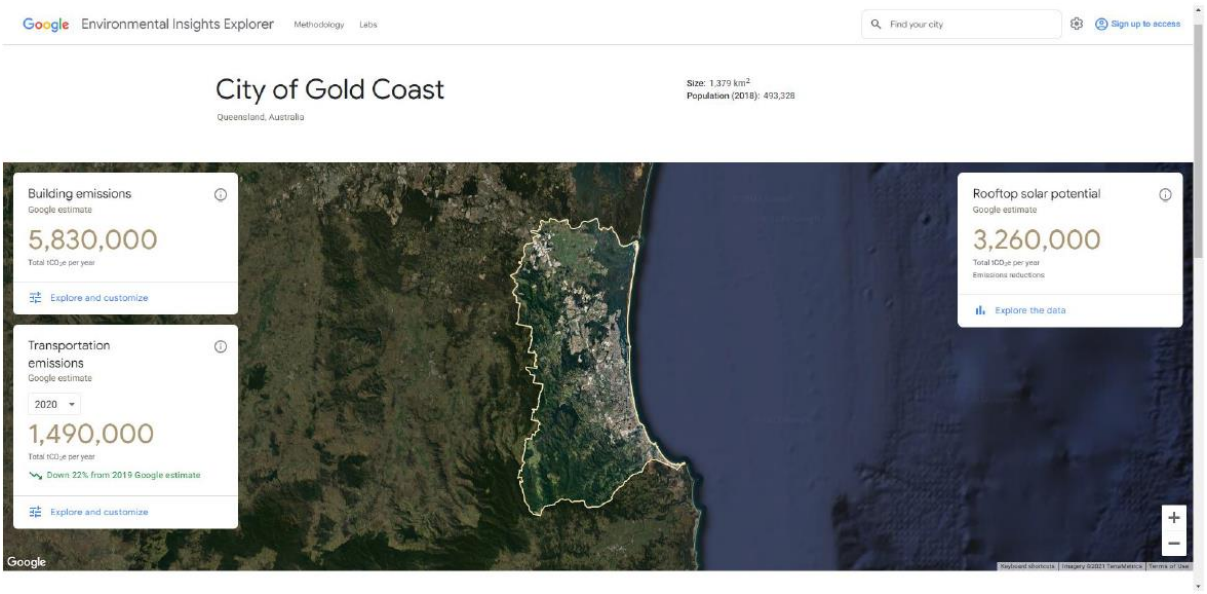

Figure 20: Google's Environmental Insights Explorer (https://insights.sustainability.google/).

5.10. Air Quality | Campaign Awareness

This demonstrator relates to the *Air Quality and Respiratory Health Outcomes* use case. While planned, this *Campaign Awareness* demonstrator was not implemented into the first phase AusEnHealth application. This demonstrator was set to focus on addressing the following core insight(s):

- Presentation of proposed development and the resulting impact on air quality.

Users

The user(s) addressed by this demonstrator:

Target User	Use
Community Groups	Campaign-based Awareness Raising

Scenario

The scenario(s) addressed by this demonstrator:

Target User	High Level Scenario
Community Groups	Campaign-based Awareness Raising | Localised Campaign Industry Driven Particle Emission Communication
Community Groups	Campaign-based Awareness Raising | Localised Campaign Transport Driven Particle Emission Communication

Data

The variables, indicators, and indices used in this demonstrator are listed in the following table. Note any later excluded items in red italics.

Data Type	Data Name
Indicator	*Number of "above average" ozone (O3) days*
Indicator	*Number of "potential risk" ozone (O3) days*
Indicator	*Number of "above average" nitrogen monoxide (NO) days*
Indicator	*Number of "potential risk" nitrogen monoxide (NO) days*
Indicator	*Number of "above average" nitrogen dioxide (NO2) days*
Indicator	*Number of "potential risk" nitrogen dioxide (NO2) days*

Indicator	*Number of "above average" particulate matter >2.5µm (PM2.5) days*
Indicator	*Number of "potential risk" particulate matter >2.5µm (PM2.5) days*
Index	*Air quality vulnerability index*
Index	*Air quality exposure index*
Index	*Air quality sensitivity index*
Index	*Air quality adaptive capacity index*

User Requirements

The user requirements applicable to the proposed demonstrator are listed in the following table. Note the items in red are user requirements not met at the time of project completion.

ID	User Requirement
DM6UR01	*The user interface will have a map component and an information panel.*
DM6UR02	*The map will be an interactive map with different basemap options (street, satellite).*
DM6UR03	*The information panel will be collapsible to enable a full screen map.*
DM6UR04	*There will be a location search (address, suburb, etc.).*
DM6UR05	*There will be an air quality risk layer displayed on the map aggregated to statistical areas. The risk layer will display traffic light colours from low to high risk where risk is represented by the combination of accessed and derived environmental health variables and indicators.*
DM6UR06	*There will be a proposed development layer (transport or industry) indicating the footprint of the physical proposed development.*
DM6UR07	*There will be high level indicators displayed on the panel showcasing air quality before development and air quality after development.*
DM6UR08	*The user will be able to adjust values (e.g., number of cars, buildings, etc.) to identify changes in air quality.*
DM6UR09	*The user can share the application of media platforms such as Twitter.*

The capability to adjust values and monitor changes in emissions would enable this *Campaign Awareness* demonstrator to shape campaigns to be tailored based on user driven information to construct and communicate impact. However, due to time constraints, this demonstrator was not implemented into the first phase AusEnHealth application. Figure Figure 21 shows an example of what localised transport related emissions could look like if sufficient data was available.

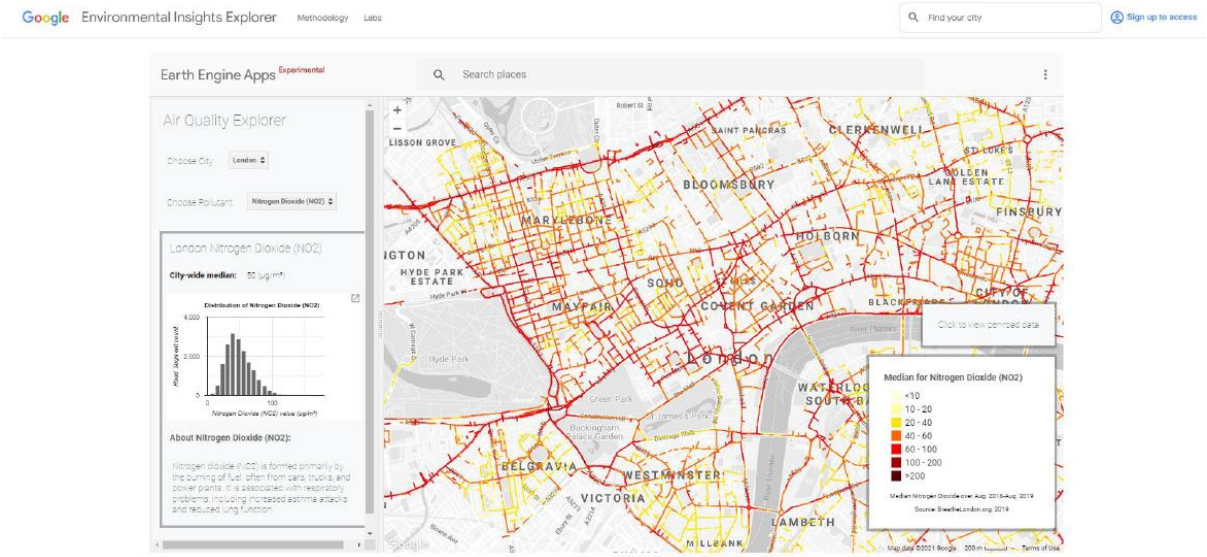

Figure 21: Google Earth Engine visualisation of air quality for transport.

5.11. Architecture

The following technology components were used to build the AusEnHealth pilot:

- **Google Cloud Platform** | Application Hosting
- **Google Earth Engine** | Raster Data and Large-Scale Analysis
- **Google Big Query** | Vector/Tabular Data and Large-Scale Analysis, Multi-Cloud Capability
- **Google Maps** | Foundation Web Mapping Service for Visualisation
- **Carto** | Off the Shelf Advanced Map BI and Large-Scale Visualisation Software

The technology has been selected for the following reasons:

- Unique Capability
  - Google Earth Engine to provide cloud-base grid analysis required for a number of use cases.
- Efficiency
  - Google Earth Engine, Cato and Big Query provide efficient deployment mechanisms.
  - NGIS has existing code that can be leveraged to expedite development.

Front-End Architecture

The proposed front-end architecture leverages modern and consistent industry standard components that will deliver a modular, secure, scalable, and reliable product (see Figure 22).

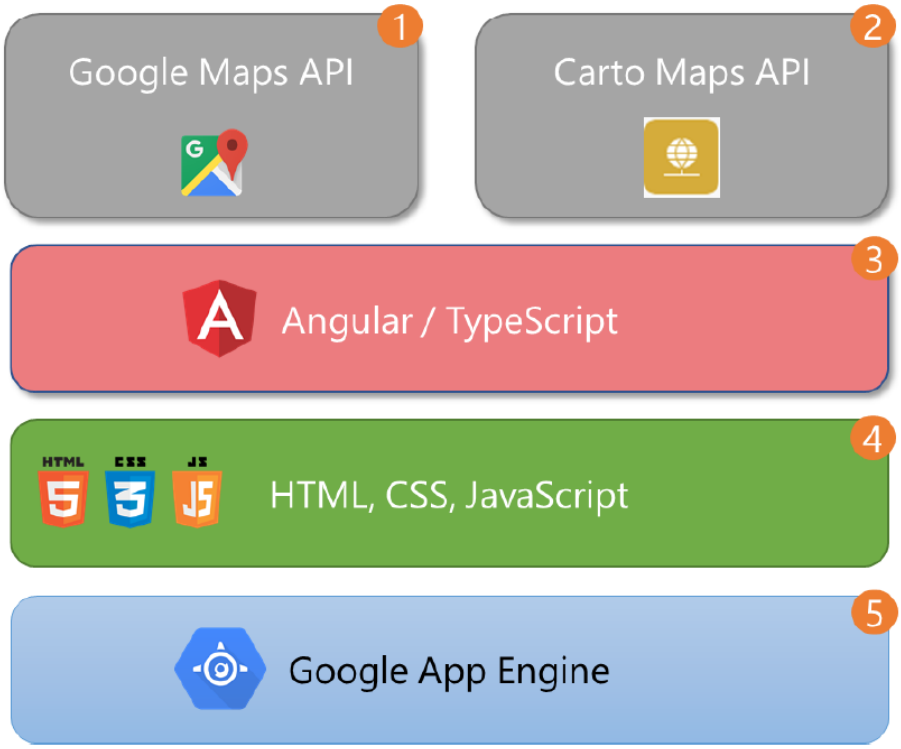

Figure 22: Front-end architecture overview.

1. The Google Maps JavaScript API will provide the foundation map visualisation layer capability including comprehensive basemap data. Google Maps API will also provide additional data services including geocoding (location search).
2. Carto Maps API is used to fetch vector data directly from Carto for rendering on the Google Maps API.
3. The application will use the latest version (current v10) of the Angular framework.

4. The Angular compiler outputs static HTML5, CSS3, and ES5 files which can be deployed to any web host.
5. The compiled Angular application will be hosted on a Google App Engine service running a Python web server that serves the static files.

Back-End Architecture

The proposed back-end architecture for the development leverages key Google products to provide high scalability and performance required to deliver on demand analytics and insights (see Figure 23).

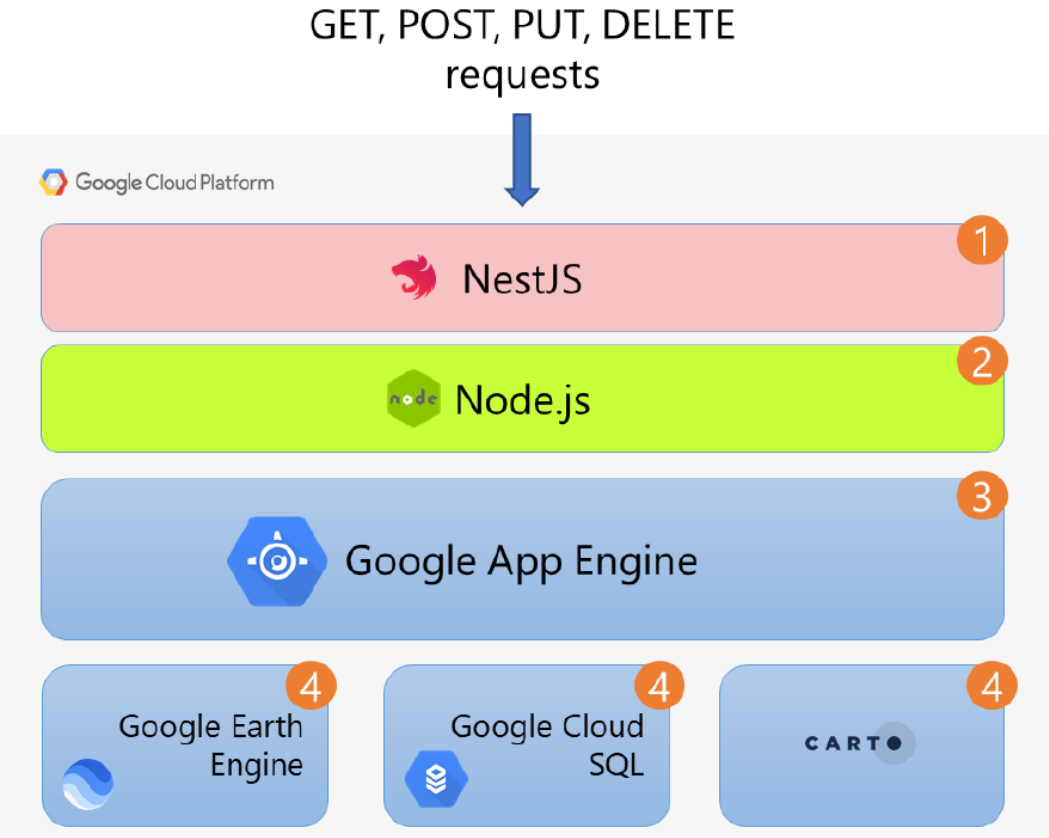

Figure 23: Back-end architecture overview.

1. NetJS is a scalable framework used to create web applications in the Node.js runtime. This layer contains the controllers and services required to access data. It also handles authentication and authorisation with minimal hassle.
2. Node.js is an open-source, cross-platform, JavaScript runtime environment (framework) that executes JavaScript code outside a web browser. It is built on Google Chrome's V8 JavaScript engine. Node.js was chosen as the back-end runtime because it supports direct access to Google Earth engine via its JavaScript API.
3. The compiled code will run on a Google App Engine instance configured to run a Node.js server in production mode.
4. The application will provide access to several data sources. These services require secret keys which are stored in the back-end configuration. The back-end is responsible for routing data request to/from the front-end.

Data Analysis Architecture

A dedicated data analysis architecture/product was required to address the Health Data Analytics demonstrator. Google Big Query Omni was used to showcase the ability to dynamically query multi-cloud data repositories without the need to copy or manage the datasets (see Figure 24).

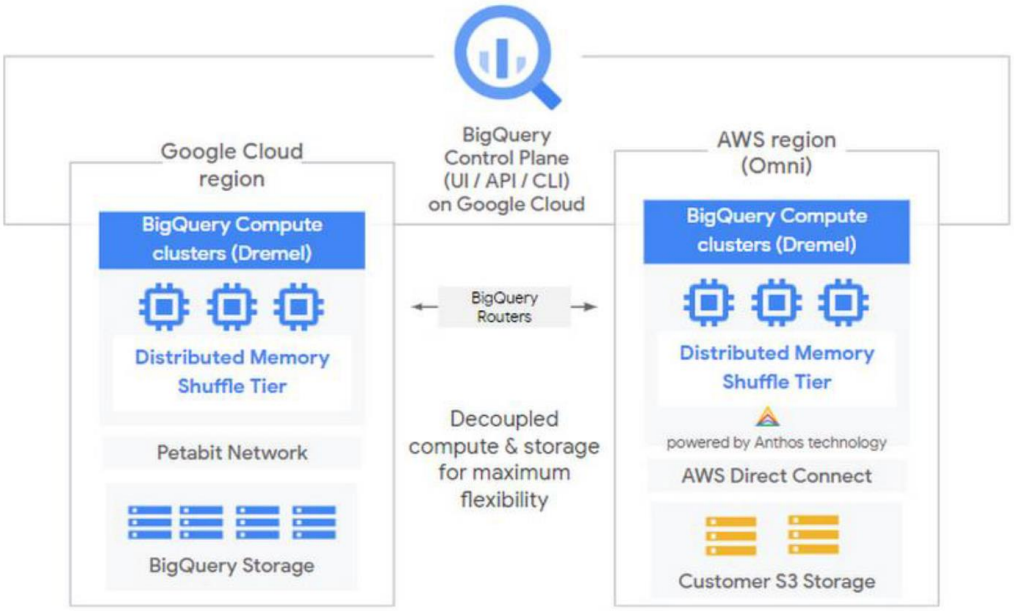

Figure 24: Big Query Omni.

5.12. Schedule

The proposed pilot development schedule aligns with the overall AusEnHealth first phase schedule. Development efforts commenced in mid-August 2021 following finalisation of the platform scope. Development sprints were conducted through to the end of October where the pilot release candidate was available for demonstration. Final release of the pilot is targeted for the end of March 2022.

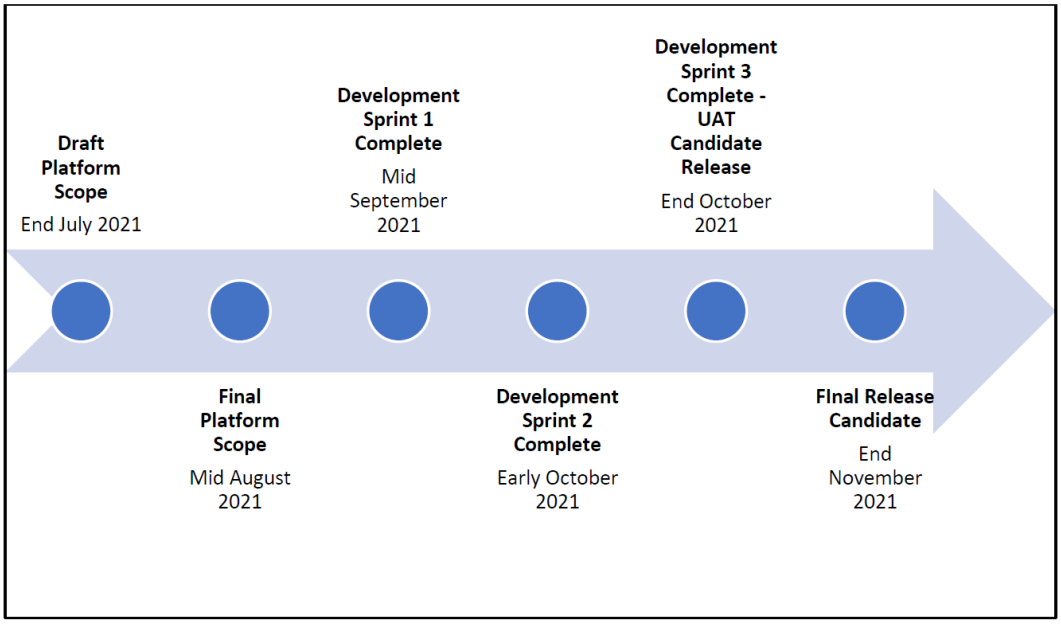

Figure 25: Schedule overview.

5.13. Summary Findings (Platform Development Scope)

The design process for the AusEnHealth pilot was iterative, taking information from this *Platform Development Scope* Section of the report, considering how this information would be communicated using information from the *Visualisation and Communication Tools* section of this report (Section 4), and then planning the platform architecture. From this process, several opportunities have been identified for implementation into future phases of AusEnHealth. Note that user testing and demonstration sessions will result in platform revisions and additional notes for future platform development. These two activities have not yet been completed at the time of writing.

Data Sharing and Integration

Significant work has been undertaken during AusEnHealth application development to integrate multi-cloud ecosystems of data into one of the demonstrators. The results of this work are promising; offering a seamless system by which to integrate pre-formatted data, ready for integration into analysis. Currently, the available data in the first phase is limited. Expanding this capability should be a priority in future to ensure that future iterations of AusEnHealth can integrate or ingest a wide range of existing public data.

One of the many goals of AusEnHealth is to promote an ongoing system of collaboration and communication around environmental health information and data. AusEnHealth has benefited greatly from multiple environmental health data custodians which house public data. As part of AusEnHealth, several environmental health parameters, indicators, and indices have been derived using this public data. In turn, this data is currently made publicly available through the pilot application. Future work in AusEnHealth should maintain this focus, where possible, to ensure open data access to other digital platforms, researchers, and the environmental health community.

One of the significant benefits of the multi-cloud system, noted above, is that integrated data is already in a valuable format. In the first phase of AusEnHealth, derived parameters have been created manually; a time-consuming process which often needed repeating for changes in geographical or temporal resolution. Long term, these processes should be automated to reduce ongoing labour costs and time required.

The first phase of AusEnHealth heavily focused on the production of the pilot application. This project identified existing national and state level environmental health data, revealed important techniques for communicating data to different end user groups, and established an understanding of differences in national and state processes, including data collection, environmental health policies, methodology, among others. While the AusEnHealth is focused on the integration and communication of data, there is also value in the collation of environmental health knowledge across Australia and internationally. While developing other use cases and otherwise expanding AusEnHealth functionality, it is recommended that a knowledge hub of environmental health policy, methodology, and research is considered for development.

6. Conclusion - Moving Towards a Cohesive Environmental Health Ecosystem

The AusEnHealth pilot application is a great first step in collating a range of environmental health data in one place, with the capacity for visualisation and analysis. The aim of such a platform is not to become the pinnacle environmental health data platform, but rather to connect the colossal existing but disconnected environmental health platforms, projects, data, researchers, and community. AusEnHealth aims to not only be a decision support platform, but also to ensure that future

environmental health research is open and accessible and to break down the ideas of data siloing and private data for various reasons by showing the value that public data can add to the environmental health community.

6.1. Environmental Health Ecosystem Recommendations

All of the activities carried out in the first phase of AusEnHealth have been carried out with consideration for the many environmental health stakeholder groups, including environmental protection agencies, health departments, local government, researchers, data collection agencies and the general public. It has become increasingly apparent as AusEnHealth progressed that stakeholder collaboration and engagement is a crucial contributor to the success of the project. It is important, therefore, to consider how AusEnHealth fits into the existing environmental health ecosystem, as well as what AusEnHealth can add to this ecosystem, both now and into the future.

AusEnHealth Integration

It is important that AusEnHealth successfully integrates into the existing environmental health ecosystem. As part of this integration, AusEnHealth must find long-term housing to ensure the resource is accessible. In future phases of AusEnHealth, this step is crucial to establish a consistent, ongoing, and reliable environmental health support platform.

The two key environmental health use cases regarding climate and air quality have demonstrated the value that AusEnHealth can add to the environmental health ecosystem. It is important to expand these use cases to ensure that AusEnHealth truly becomes an environmental health platform and is not limited to climate- and air quality- based research and policy support.

Collaborative Ecosystem

Throughout the first phase of AusEnHealth, our network of supporters and collaborators have expanded significantly. While there are different approaches, resources, and priorities across this network of researchers and policy makers, there is a shared goal: increasing our understanding of the impact of the environment on humans - and using this information to save lives. As part of moving towards an environmental health ecosystem, a community of practice should be established which focuses on communicating interests and problems, sharing best practice, and creating new knowledge.

As part of this environmental health ecosystem, AusEnHealth also proposes that strong and effective governance practices are developed to maintain a focus on the sharing of information and data for the benefit of the Australian and international environmental health community. This includes effective documentation and communication of practices and recommendations, as well as ongoing transparency as the landscape of environmental health monitoring and research continues to expand and evolve.

A common note throughout this report is that this project has relied heavily on collaboration with various environmental health researchers and organisations. It is a fitting final recommendation that collaboration in the environmental health space continues to be inclusive and cohesive, and that future phases of AusEnHealth continue to encourage a collaborative environmental health community.

Appendices

A. AusEnHealth Recommendations

The following recommendations are compiled from the findings of each main section of the report. By creating a summary of recommendations on platform development, data availability, visualisation, and data analysis methodology, this report highlights the various ways in which the AusEnHealth pilot can be made more valuable in the future.

Data Audit Recommendations

ID	Recommendation
D1	Extend the existing AIHW national health outcome data catalogue to feature publicly available targeted and high-resolution datasets at both national and jurisdiction level.
D2	Enable easier access to health data through ongoing discussion of privacy and open data, as well as the development of clearer guidelines and requirements related to health outcome data request processes.
D3	Data released from air quality monitoring stations should be provided without smoothing to enable complex modelling.
D4	Seek to increase the density of air quality monitoring stations to improve the granularity of air quality data for modelling and analysis.
D5	Ensure continued support for CAR's modelled air quality data or ensure support for an alternative in the event CAR's work is discontinued.
D6	Future efforts should be undertaken to ensure that water quality monitoring data suitable for analysis of water borne disease developments is made available both nationally and by jurisdiction
D7	Continue development of time series data relating to built environment.
D8	Work towards adherence to national and potentially global data standards.
D9	Agreement of nationally consistent environmental health indicators and metadata standards for priority use cases in Australia.

Data Analysis Recommendations

ID	Recommendation
DA1	Establish thorough metadata documentation and integration into digital platform.
DA2	Create a publicly available minimum dataset for environmental health, including health outcome data sourced using consistent disease classification codes.
DA3	Create or implement improved vulnerability index methodology.
DA4	Implement appropriate methods for analysis of associations, predictions, and spatial modelling.
DA5	Highlight necessary future focus on vulnerability assessment, health impact assessment, methods for impact of future trends (predictions), and mitigation assessment (planning).

Visualisation and Communication Recommendations

ID	Recommendation
VR1	Improve/incorporate messaging around statistical outputs and analyses, including messaging around associations between variables.
VR2	Implement methods for visualising uncertainty without damaging potential user insights.
VR3	Improve digital platform impact by clearly linking definitions, methodology, and messaging in digital platforms.

Platform Development Recommendations

ID	Recommendation
PD1	Ensure future platform can integrate or ingest existing public data.
PD2	Ensure future platform enables open data access to other digital platforms.
PD3	Implement automated data integration methods.
PD4	While developing use cases, also consider the other dimension AusEnHealth; a knowledge hub of environmental health policy, methodology, and research.

Environmental Health Ecosystem Recommendations

ID	Recommendation
EHE1	Identifying a long-term home to ensure a consistent, ongoing, and reliable environmental health support platform.
EHE2	Expansion of environmental health scope to include additional use cases (e.g., water quality, vector borne disease, etc.).
EHE3	Establish community of practice to enable better sharing of environmental health knowledge.
EHE4	Establish strong and effective governance practices.
EHE5	Promote and encourage inclusive and cohesive collaboration in environmental health research.

B. AusEnHealth Advisory Board

Team Member	Role	Organisation
Helen Spencer	Program Director, Land and Spatial Information	Queensland Department of Natural Resources, Mines and Energy
Adam Mowlam	Digital Twin Lead	Department of Environment, Land, Water and Planning, Victoria
Wayne Patterson	Direction, Spatial Operations	New South Wales Spatial Services
Ivana Ivanova	Research Fellow	Curtin University
Irina Bastrakova	Director – Spatial Data Architecture – National Location Information, Place, Space and Communities Division	Geoscience Australia
Matt Beaty	Geographer and Health Analyst	Australian Department of Health
James Spath	Geospatial Analyst	City of Swan – Local Government Representative, Western Australia
Grace Yun	Geospatial Lead	Western Australia Department of Health
Paul Merchant	Health Data Analyst	Western Australia Department of Health
Alex Xiao	Epidemiologist	Western Australia Department of Health
Peter Franklin	Air Quality Expert	Western Australia Department of Health
Nimmi Carlose	Environmental Surveillance Data Manager	Western Australia Department of Health
Angie Bone	Deputy chief Health Officer, Environment	Department of Health and human Services, Victoria
Ivan Hannigan	Data Scientist, Epidemiology	University of Sydney
Geoffrey Morgan	Associate Professor, Public Health	University of Sydney

Table 15: AusEnHealth's Advisory Group.

C. Use Case Ranking

Use Case	Ranking Options			Final Rank
	Avg. Score	Maybe + Yes	Yes Count	
Climate, Weather & Extreme Weather and Health Outcomes Related to Heat	1	1	1	**1**
Climate, Weather & Extreme Weather and Health Outcomes Related to Extreme Weather	7	2	3	6
Climate, Weather & Extreme Weather and Coronary Heart Disease (CHD) Issues / Mortality	6	2	4	7

Climate, Weather & Extreme Weather and Gastroenteritis	11	4	7	8+
Air Quality and Asthma	4	2	3	**4**
Air Quality and Chronic Obstructive Pulmonary Disease (COPD)	6	2	4	7
Air Quality and Cardiovascular Disease (CVD)	8	4	4	8+
Water Quality & Quantity and Water Borne Disease	5	1	3	**3**
Vector Distribution and Ecology and Mosquito Borne Disease	2	3	3	**5**
Vector Distribution and Ecology and Arbovirus	10	4	7	8+
Built Environment and Carbon Monoxide Poisoning	12	5	6	8+
Built Environment and Health Outcomes Related to Noise Pollution	9	4	5	8+
Built Environment and Health Outcomes Related to Heat	3	1	2	**2**

Table 16: Use case rankings with different methods applied alongside final ranks.

D. Data Catalogue Tables

Parameters from Key Data Sources

Category	Parameter	Access	Source	Time Scale	Update Freq.	Spatial Res.	Format
Climate	Temperature	Website	BOM	2000-	Daily	0.05° Raster	grid
Climate	Rainfall	Website	BOM	2000-	Daily	0.05° Raster	grid
Climate	Extreme Temperature	Request	BOM	2000-	Daily	0.05° Raster	grid
Climate	Extreme Rainfall	Request	BOM	2000-	Daily	0.05° Raster	grid
Climate	Humidity	Website	BOM	2000-	Daily	0.25° Raster	grid
Water Quality	Water Use	Website	BOM	2000-	Daily	0.25° Raster	grid
Water Quality	Groundwater Levels	Website	BOM	2000-	Daily	0.25° Raster	grid
Water Quality	Streamflow Levels	Website	BOM	2000-	Daily	Riverway Shapefile	
Water Quality	Streamflow Salinity	Website	BOM	2000-	Daily	Riverway Shapefile	
Water Quality	Groundwater Salinity	Website	BOM	2000-	Daily	0.05° Raster	grid
Water Quality	Soil Moisture	Website	BOM	2000-	Daily	0.05° Raster	grid
Water Quality	Runoff	Website	BOM	2000-	Daily	0.05° Raster	grid
Climate	Solar Exposure	Website	BOM	2000-	Daily	0.05° Raster	grid
Climate	UV Index	Website	BOM	2000-	Daily	0.05° Raster	grid
Climate	Southern Oscillation Index	Website	BOM	1870-	Monthly	1 value globally	Data frame

Category	Parameter	Access	Source	Time Scale	Update Freq.	Spatial Res.	Format
Air Quality	Nitrogen Dioxide	Website	CARDAT	2000-	Daily	Shapefile	
Air Quality	Sulphur Dioxide	Website	CARDAT	2000-	Daily	Shapefile	
Air Quality	Ozone	Website	CARDAT	2000-	Daily	Shapefile	
Air Quality	Lead	Website	CARDAT	2000-	Daily	Shapefile	
Air Quality	Carbon Monoxide	Website	CARDAT	2000-	Daily	Shapefile	
Air Quality	Particulates: PM 10	Website	CARDAT	2000-	Daily	Shapefile	
Air Quality	Particulates: PM 2.5	Website	CARDAT	2000-	Daily	Shapefile	
Air Quality	Particulates: PM 1	Website	CARDAT	2000-	Daily	Shapefile	
Built Environment	Energy Use	Website, Website	AEMO, AURIN	2019-	5-minute	Electricity Region and Station	csv
Built Environment	Energy Use	Website, Website	CER, AURIN	2001-	Monthly	Postcode	Vector Polygon
Built Environment	Energy Use	Website	ABS	2009	1 Point in time	Labour-Force Region	Vector Polygon
Built Environment	Urban Green Space	Website, Website	Geoscape, AURIN	2016-	Yearly	2m (urban), 10m (rural)	Vector Polygon, Point and Line
Built Environment	Urban Green Space	Website	AURIN, OSM	2020-	Daily	Point	Vector Point
Built Environment	Residential Density	Website, Website	Geoscape, AURIN	2020-	Yearly	2.5m (urban), 11m (rural)	Vector Polygon
Built Environment	Residential Density	Website, Website	AURIN, DataVIC	2017-	Yearly	0.5m to 25.0m	Vector Polygon
Built Environment	Residential Density (land use)	Website	AURIN	2017-	Yearly	N/A	Vector Polygon
Built Environment	Residential Density (land use)	Website	AURIN	2018-	Yearly	N/A	Vector Polygon
Built Environment	Population Distribution	Website	AURIN				
Built Environment	Public Transport Use	Website	ABS	2009	1 Point in time	Transport Zone	Vector Polygon
Built Environment	Public Transport Use	Website	Transport for NSW	2018-2019	Yearly	SA3 & LGA	
Built Environment	Solid waste generation	Website	Sustain-ability VIC	2001-	Yearly	LGA	XLSX
Built Environment	Solid waste generation	Website	QLD Gov.	2016-	Yearly	LGA	PDF
Built Environment	Solid waste generation	Website	EPA NSW	2018-2019	Yearly	LGA	XLSX
Extreme Weather	Floods	Website	GA	1986-2017	Daily	25-m grid cell	

Category	Parameter	Access	Source	Time Scale	Update Freq.	Spatial Res.	Format
Extreme Weather	Bushfires	Website	NEO-MODIS	2000-	Monthly	500-m grid cell	
Demographics	Population	Website	ABS	2016	1 Point in time	SA2	
Demographics	Age	Website	ABS	2016	1 Point in time	SA2	
Demographics	Gender	Website	ABS	2016	1 Point in time	SA2	
Demographics	Indigenous	Website	ABS	2016	1 Point in time	SA2	
Demographics	Education	Website	ABS	2016	1 Point in time	SA2	
Demographics	Income	Website	ABS	2016	1 Point in time	SA2	
Health Outcomes	Cryptosporid-iosis	Website	NNDSS	2001-	Monthly	State	
Health Outcomes	Giardia	Website	NNDSS	2001-	Monthly	State	
Health Outcomes	E. Coli (STEC)	Website	NNDSS	2001-	Monthly	State	
Health Outcomes	Influenza	Website	NNDSS	2001-	Monthly	State	
Health Outcomes	Asthma	Website	ABS, AIHW	2006, 2016	1 Point in time	SA3	
Health Outcomes	COPD	Website	ABS, AIHW	2006, 2016	1 Point in time	SA3	
Health Outcomes	Ross River Virus	Website	NNDSS	2001-	Monthly	State	
Health Outcomes	Integrated data set	Website	ABS	2006, 2011, 2016	1 Point in time	Point Data	
Health Outcomes	Dengue Fever		QUT	2005-2019	Daily	SA2	
Health Outcomes	Principal Diagnosis (ICD-10-AM)		AIHW	2008-2020		SA3	
Health Outcomes	Giardia		AIHW	2008-2020		SA3	
Health Outcomes	Asthma	Website	ABS	2001-2018	1 Point in time	State	
Health Outcomes	Emergency Department Presentations (ICD-10-AM)	Website	AIHW	2018-2019	1 Point in time	State	
Health Outcomes	Causes of Death	Website	ABS	2010-2019	1 Point in time	State	
Health Outcomes	Mortality Over Regions and Time (MORT) books	Website	AIHW	2014-2018	1 Point in time	SA3	
Health Outcomes	Burden of Disease	Website	AIHW	2011, 2015	1 Point in time	State	

Table 17: Parameters from key data sources.

Parameters from Secondary Data Sources

Category	Parameter	Access	Source	Time Scale	Update Freq.	Spatial Res.	Format
Air Quality	Nitrogen Dioxide	Website	NASA (LANCE-MODIS)	2016-	Daily	0.05° Raster	
Air Quality	Sulfur Dioxide	Website	NASA (LANCE-MODIS)	2016-	Daily	0.05° Raster	
Air Quality	Ozone	Website	NASA (LANCE-MODIS)	2016-	Daily	0.05° Raster	
Air Quality	Lead	Website	NASA (LANCE-MODIS)	2016-	Daily	0.05° Raster	
Air Quality	Carbon Monoxide	Website	NASA (LANCE-MODIS)	2016-	Daily	0.05° Raster	
Air Quality	Particulates: PM 10	Website	NASA (LANCE-MODIS)	2016-	Daily	0.05° Raster	
Air Quality	Particulates: PM 2.5	Website	NASA (LANCE-MODIS)	2016-	Daily	0.05° Raster	
Air Quality	Particulates: PM 1	Website	NASA (LANCE-MODIS)	2016-	Daily	0.05° Raster	
Built Environment	Vegetation	Website	BOM	1992-	Monthly	0.05° Raster	
Built Environment	Land Use	Website	Dept. of Ag., Water and the Env.	1992-2011	Yearly	National	
Air Quality	Nitrogen Dioxide	Website	Copernicus	2003-2020	3-hourly	0.75° Raster	
Air Quality	Sulphur Dioxide	Website	Copernicus	2003-2020	3-hourly	0.75° Raster	
Air Quality	Ozone	Website	Copernicus	2003-2020	3-hourly	0.75° Raster	
Air Quality	Lead	Website	Copernicus	2003-2020	3-hourly	0.75° Raster	
Air Quality	Carbon Monoxide	Website	Copernicus	2003-2020	3-hourly	0.75° Raster	
Air Quality	Particulates: PM 10	Website	Copernicus	2003-2020	3-hourly	0.75° Raster	
Air Quality	Particulates: PM 2.5	Website	Copernicus	2003-2020	3-hourly	0.75° Raster	
Air Quality	Particulates: PM 1	Website	Copernicus	2003-2020	3-hourly	0.75° Raster	
Air Quality		Website	CSIRO	1967-	Monthly	1 Point Nationally	
Air Quality	Greenhouse Gas	Website	OWID			1 Point Nationally	

Category	Parameter	Access	Source	Time Scale	Update Freq.	Spatial Res.	Format
Air Quality	Ozone	Website	NASA Ozone Watch	1980-	Daily	Global Raster	
Built Environment	Vegetation	Website	BOM	2000-	Daily	0.05° Raster Data	.grid
Built Environment	Land Use	Website	ABARES	2003-2018	Yearly	50m Raster	
Built Environment	Species outbreaks	Website					
Built Environment	Marine and estuarine protected areas	Website					
Built Environment	Recovery plans R 38	Website					
Built Environment	Aquatic habitat destruction	Website	Ocean Health Index	2003-2018	Yearly	Raster	
Extreme Weather	Fire Emissions	Website	Global Fire Emissions Database	2016-	Monthly	0.25° Raster	
Built Environment	Vegetation	Website	ABARES	2018	One Point in Time	100m x 100m	.grid
Built Environment	Solid waste generation	Website	GA				
Water Quality	Bathymetry Derivatives	Website	GA	2009	One Point in Time	250m x 250m	
Built Environment	Exposure	Website	GA				
Water Quality	Groundwater	Website	BOM	1880-2018			
Built Environment	Percent of land covered by forest	Website	GA	2001-2015	Monthly	250m x 250m	
Extreme Weather	Fire regimes	Website	ABARES	2011-2016	Monthly		
Built Environment	Catchment Scale Land Use of Australia	Website	ABARES	2017-	Yearly		
Built Environment	Botanical Database	Website	Australian National Herbarium				
Built Environment	Threatened Species	Website	Dept. of Ag., Water and the Env.			Point Data	
Air Quality	Greenhouse Gas	Website	Dept. Industry, Science, Energy and Resources	1990-2018	Yearly	State	

Category	Parameter	Access	Source	Time Scale	Update Freq.	Spatial Res.	Format
Built Environment	Protected Areas	Website	Dept. of Ag., Water and the Env.	2018	One Point in Time	Point Data	
Built Environment	Wetlands	Website	Dept. of Ag., Water and the Env.				
Water Quality	Algae	Website	Australian National Herbarium	Present			
Built Environment	Fisheries Data	Website	ABARES	1967-			
Water Quality	Marine Pests	Website	NIMPIS	2009-			
Built Environment	Fisheries Data	Website	FRDC	2018			
Water Quality	Pollution Events	Website	IMO	1998-			
Climate	Extreme Temperature	Website	Met Office Hadley Centre	1901-2018	Annual	1.875°×1.25° Raster	
Climate	Extreme Rainfall	Website	Met Office Hadley Centre	1901-2018	Annual	1.875°×1.25° Raster	
Extreme Weather	Fire Regimes	Website	Dept. of Ag., Water and the Env.	2020	One Point in Time	Polygons	
Air Quality	Aerosol Optical Thickness	Website	NEO-MODIS	2000-	Monthly	0.4° Raster	GeoTIFF
Air Quality	Carbon Monoxide	Website	NEO-MODIS	2000-	Monthly	0.4° Raster	GeoTIFF
Air Quality	Nitrogen Dioxide	Website	NEO-AURA	2000-	Monthly	0.4° Raster	GeoTIFF
Air Quality	Ozone	Website	NEO-AURA	2000-	Monthly	0.4° Raster	GeoTIFF
Climate	Radiation	Website	NEO-FLASHFlux	2000-	Monthly	0.4° Raster	GeoTIFF
Built Environment	Land Use	Website	NEO-AVHRR	2001-2011	Annual	0.4° Raster	GeoTIFF
Built Environment	Vegetation	Website	NEO-MODIS	2000-	Monthly	0.4° Raster	GeoTIFF
Demographics	Population	Website	SEDAC	2000-2020	5-Year Blocks	0.01° Raster	GeoTIFF
Water Quality	Chlorophyll Concentration	Website	NEO-MODIS	2002-	Monthly	4km	NetCDF
Climate	Temperature	Website	CRU	1901-2019	Monthly	0.5° Raster	NetCDF
Climate	Rainfall	Website	CRU	1901-2019	Monthly	0.5° Raster	NetCDF
Climate	Temperature	Website	TerraClim	1958-2020	Monthly	0.4° Raster	

Category	Parameter	Access	Source	Time Scale	Update Freq.	Spatial Res.	Format
Climate	Rainfall	Website	TerraClim	1958-2020	Monthly	0.4° Raster	
Climate	Humidity	Website	TerraClim	1958-2020	Monthly	0.4° Raster	
Climate	Drought	Website	TerraClim	1958-2020	Monthly	0.4° Raster	
Climate	Radiation	Website	TerraClim	1958-2020	Monthly	0.4° Raster	
Climate	Wind Speed	Website	TerraClim	1958-2020	Monthly	0.4° Raster	
Climate	Vapor Pressure	Website	TerraClim	1958-2020	Monthly	0.4° Raster	
Climate	Temperature	Website	NOAA	1979-	Daily	0.5° Raster	
Climate	Rainfall	Website	NOAA	1979-	Daily	0.5° Raster	
Climate	Solar Exposure	Website	NEO-CERES	2006-2020	Monthly	0.4° Raster	GeoTIFF

Table 18: Parameters from secondary sources.

Additional Data Collections

The websites listed in this table direct a print reader to the homepage of the data source. Electronic versions of this paper contain an extended link leading directly to the data collections page.

Source	Link to Data Collections Homepage
AIHW	https://www.aihw.gov.au/
ABS Data Packs	https://datapacks.censusdata.abs.gov.au/
WA Notifiable Diseases Register	https://www.wslhd.health.nsw.gov.au/
AURIN Data Repository	https://data.aurin.org.au/
National Map	https://www.nationalmap.gov.au/
Data.gov.au	https://www.data.gov.au/search
National Vegetation Information System	https://www.environment.gov.au/
Geoscience Australia Web Services	http://services.ga.gov.au/
Digital Earth Australia	http://www.ga.gov.au/
AIMS Data Catalogue	https://www.aims.gov.au/
Geoscience Australia Data Catalogue	https://ecat.ga.gov.au/
Atlas of Living Australia	https://collections.ala.org.au/
NCI Geonetwork	https://geonetwork.nci.org.au/
Marine Biodiversity Hub	https://www.nespmarine.edu.au/
FSDF-LINK Platform	https://link.fsdf.org.au/
SEDAC	https://sedac.ciesin.columbia.edu/
CMIP6 Environmental Projections	https://esgf-node.llnl.gov/
NOAA	https://psl.noaa.gov/

Table 19: Additional data collections.

E. Data Quality and Accessibility Heatmaps

The following tables note the key parameters for each environmental exposure domain with properties coloured red (undesirable) through to green (desirable) as a measure of quality. Also included, using the same colour scaling, are health outcome parameters. For more information, please see the AusEnHealth data catalogue.

Air Quality

Parameter / Dataset	Curator	Temporal Resolution (Extent)	Spatial Resolution (Extent)	Update Frequency	Access Restrictions
Nitrogen Dioxide	CAR	Daily (2000-)	SA2 (National)	As Accessed	Request
Sulphur Dioxide	CAR	Daily (2000-)	SA2 (National)	As Accessed	Request
Ozone	CAR	Daily (2000-)	SA2 (National)	As Accessed	Request
Lead	CAR	Daily (2000-)	SA2 (National)	As Accessed	Request
Carbon Monoxide	CAR	Daily (2000-)	SA2 (National)	As Accessed	Request
Particulates: PM 10	CAR	Daily (2000-)	SA2 (National)	As Accessed	Request
Particulates: PM 2.5	CAR	Daily (2000-)	SA2 (National)	As Accessed	Request
Particulates: PM 1	CAR	Daily (2000-)	SA2 (National)	As Accessed	Request

Climate

Parameter / Dataset	Curator	Temporal Resolution (Extent)	Spatial Resolution (Extent)	Update Frequency	Access Restrictions
Temperature	BOM, NCC	Daily (1910-)	0.05° Raster (National)	Daily	CC BY 3.0 Int
Rainfall	BOM, NCC	Daily (1900-)	0.05° Raster (National)	Daily	CC BY 3.0 Int
Extreme Temperature	BOM, NCC	Daily (2000-)	0.05° Raster (National)	Daily	Access Fee
Extreme Rainfall	BOM, NCC	Daily (2000-)	0.05° Raster (National)	Daily	Access Fee
Solar Exposure	BOM, NCC	Daily (1990-)	0.05° Raster (National)	Daily	CC BY 3.0 Int
UV Index	BOM, NCC	Daily (2000-)	0.05° Raster (National)	Daily	CC BY 3.0 Int
Humidity	BOM, NCC	Daily (1971-)	0.25° Raster (National)	Daily	CC BY 3.0 Int
Southern Oscillation Index	BOM, NCC	Daily (1870-)	Point (Global)	Monthly	CC BY 3.0 Int
Temperature	UNSW CFRC	30-minute (2016-2019)	Point (Sydney)	Bi-Monthly	CC BY 4.0 Int
Floods	GA	N/A (1986-)	25m Grid	Monthly	CC BY 4.0 Int
Bushfires	TERRA/MODIS	N/A (2000-)	500m Grid	Monthly	CC BY 4.0 Int

Water Quality

Parameter / Dataset	Curator	Temporal Resolution (Extent)	Spatial Resolution (Extent)	Update Frequency	Access Restrictions
Groundwater Salinity	BOM	Daily (2000-)	0.05° Raster (National)	Daily	CC BY 3.0 Int
Soil Moisture	BOM	Daily (2000-)	0.05° Raster (National)	Daily	CC BY 3.0 Int
Runoff	BOM	Daily (2000-)	0.05° Raster (National)	Daily	CC BY 3.0 Int
Water Use	BOM	Daily (2000-)	0.05° Raster (National)	Daily	CC BY 3.0 Int
Groundwater Levels	BOM	Daily (2000-)	0.05° Raster (National)	Daily	CC BY 3.0 Int

Built Environment

Note that the parameters and associated spatial and temporal scales listed below are based on online searches for available data. Further historic data may be accessible via discussion with data custodians, and other data may be available from less well publicised sources.

Parameter / Dataset	Curator	Temporal Resolution (Extent)	Spatial Resolution (Extent)	Update Frequency	Access Restrictions
Residential Density (Building Footprints)	VIC DELWP	Yearly (2017-)	0.5m – 25m (VIC)	Yearly	Request
Residential Density (Building Footprints)	Geoscape	N/A (2020)	2.5m, 11m (National)	Yearly	Request
Residential Density (Building Footprints)	Microsoft	N/A (2018)	Polygon (National)	N/A	ODbl
Energy Use	AEMO	5-Minute (2019-)	Electricity Region and Station (NEM)	5-Minute	Request
Energy Use	ABS	Bi-Weekly (2009)	Labour-Force Region (VIC)	N/A	Request
Energy Use	CER	Monthly (2001-)	Postcode (National)	Monthly	Request
Public Transport Use	VIC Department of Transport	Yearly (2007-2018)	LGA (Melbourne)	Yearly	Request
Public Transport Use	Translink	Monthly (2016-)	PT Route/Stop OD (East QLD)	Monthly	Request
Public Transport Use	Transport for NSW	Yearly (2018)	SA3, LGA (Sydney)	Yearly	Request
Public Transport Use	ABS	N/A (2009)	Transport Zone (Melbourne)	N/A	Request
Heat Vulnerability	UNSW CFRC	N/A (2011, 2016)	LGA (Sydney)	5-Yearly	CC BY 4.0 Int
Solid Waste Generation	Sustainability VIC	Yearly (2001-)	LGA (VIC)	Yearly	Request
Solid Waste Generation	QLD Government	Yearly (2016-)	LGA (QLD)	Yearly	Request
Solid Waste Generation	EPA NSW	Yearly (2018-2019)	LGA (NSW)	Yearly	Request
Residential Density (Land Use)	VIC DELWP	N/A (2017-)	N/A (VIC)	Yearly	Request
Residential Density (Land Use)	SA DPTI	N/A (2018-)	N/A (SA)	Yearly	Request
Land Parcels	Landgate	N/A (2021-)	N/A (WA)	Yearly	Request
Land Parcels	Geoscape	N/A (2021-)	N/A (National)	Yearly	Request
Urban Green Space	Various	Yearly (2009-2021)	LGA-STE (National)	Daily-Yearly	Various

Health Outcomes

Parameter / Dataset	Curator	Temporal Resolution (Extent)	Spatial Resolution (Extent)	Update Frequency	Access Restrictions
Mortality related to heat	WADOH	Monthly (N/A)	N/A (National)	N/A	Private Data
Hospitalisations due to heatstroke	AIHW, PHRN	Monthly (N/A)	N/A (National)	N/A	Request
Cryptosporidiosis	NNDSS	Monthly (2001-)	STE (National)	Yearly	Access Fee
Giardia	NNDSS	Monthly (2001-)	STE (National)	Yearly	Access Fee
E. Coli (STEC)	NNDSS	4 Years Aggregate (2001-)	STE (National)	Yearly	Access Fee
Influenza	NNDSS	4 Years Aggregate (2001-)	STE (National)	Yearly	Access Fee
Asthma	ABS, AIHW	4 Years Aggregate (2006, 2016)	SA3 (National)	Yearly	Open Access
COPD	ABS, AIHW	4 Years Aggregate (2006, 2016)	SA3 (National)	Yearly	Open Access
Ross River Virus	NNDSS	Daily (2001-)	STE (National)	Yearly	Access Fee
Integrated Data Set (MADIP)	ABS	Various	Various	Yearly	Access Fee
Dengue Fever	QUT	N/A (2005-2019)	SA2 (National)	As Accessed	Request
Principal Diagnosis (ICD-10-AM)	AIHW	4 Years Aggregate (2008-2020)	SA3 (National)	Yearly	Access Fee
Giardia	AIHW	Monthly (2008-2020)	SA3 (National)	Yearly	Access Fee
Asthma	ABS	4 Years Aggregate (2001-2018)	STE (National)	Yearly	Open Access
Emergency Department Presentations (ICD-10-AM)	AIHW	Monthly (2008-2019)	STE (National)	Yearly	Access Fee
Causes of Death	ABS	Monthly (2010-2019)	STE (National)	Yearly	Access Fee
Mortality Over Regions and Time (MORT) books	AIHW	Monthly (2014-2018)	SA3 (National)	Yearly	Open Access
Burden of Disease	AIHW	Monthly (2011, 2015)	STE (National)	Yearly	Open Access

F. International Classification of Diseases Code Summary

The following tables are grouped health outcomes, with ICD-9, ICD-10, and ICD-10-AM codes noted from the sources provided. ICD code modifications which are not provided are listed as "n.p.".

Respiratory Health

ICD Code	Sources
Respiratory Illness ICD-9: 460-519[1] ICD-10: J00-J99, R09.1, R09.8, P20-P28, R04.2-R06	[79] [145] [146] [147] [148] [149] [150] [151] [152] [153] [154] [59] [155]
Asthma ICD-9: 493 ICD-10: J45-J46	[156] [145] [147] [148] [150] [157] [153] [154] [158] [59] [159]
COPD ICD-9: 490-492, 494-496 ICD-10: J40-J44, J47, J67	[147] [148] [150] [153] [160] [59] [159]
Respiratory Infection ICD-9: 461-466, 480-487, 514[1] ICD-10: J00-J22 ICD-10-AM: n.p.	[156] [145] [150] [153] [158]
Pneumonia, Acute Bronchitis ICD-9: 466, 480-486[1] ICD-10: J12-J17, J18.0, J18.1, J18.8, J18.9, J20, J21	[147] [148] [158]

1 ICD Codes used in the literature seem to include general or viral illnesses which will not be exacerbated by poor air quality or extreme climate events. Examples include: 480 (viral pneumonia), 487-488 (influenza, influenza due to Avian influenza virus).

Heart Health

ICD Code	Sources
Cardiovascular Diseases ICD-9: 390-459 ICD-10: I00-I99, G45 (excl. G45.3), G46, M30, M31, R58	[79] [145] [146] [147] [148] [149] [150] [151] [153] [154] [59] [161]
Ischemic Heart Disease ICD-9: 410-413 ICD-10: I20-I25, I46, I49	[147] [148] [150] [152] [153] [159] [161]
Arrythmia ICD-9: 427 ICD-10: I46-I49	[147] [148]
Cardiac Failure ICD-9: 428 ICD-10: I50	[147] [148] [162]
Non-hypertensive CVD Diagnoses ICD-9: 410-459 ICD-10: n.p.	[145]

Brain Health

ICD Code	Sources
Cerebrovascular Disease ICD-9: 430-438 ICD-10: I60-I69, G45 (excl. G45.3), G46	[79] [147] [148] [152] [59]
All Mental Health ICD-9: n.p. ICD-10: n.p. ICD-10-AM: F01-F99	[59] [159]

Heat Health

ICD Code	Sources
Heat Related Illness ICD-9: 992 ICD-10: T67	[79] [163] [159] [164] [165] [166]
Dehydration ICD-9: n.p. ICD-10: E86	[163] [59] [159] [166]
Exposure to Excessive Natural Heat ICD-9: n.p. ICD-10: X30	[79] [163] [164] [165] [166]
Electrolyte Imbalance ICD-9: 276 ICD-10: n.p. ICD-10-AM: n.p.	[79] [159]
Exposure to Sunlight ICD-9: n.p. ICD-10: n.p. ICD-10-AM: X32	[164]

Other

ICD Code	Sources
Non-trauma Cases ICD-9: 001-799 ICD-10: A00-R99, V01-V98, except accidents, injuries, suicides, and homicides	[79] [146] [147] [148] [151] [152]
Acute Renal Failure ICD-9: 584 ICD-10: n.p. ICD-10-AM: n.p.	[79] [159]
Diabetes Complications ICD-9: n.p. (249-250) ICD-10: n.p. ICD-10-AM: n.p.	[159]
Chest Pain ICD-9: n.p. ICD-10: n.p. ICD-10-AM: R07.1, R07.3, R07.4	[59]

G. Proposed Data Integration Architecture

During AusEnHealth development, an automated data integration approach was considered. The following pipeline was proposed, beginning with a scheduled trigger (weekly, monthly, etc.) automatically collecting and preparing data, through the AWS systems and NGIS platform down to specific user queries.

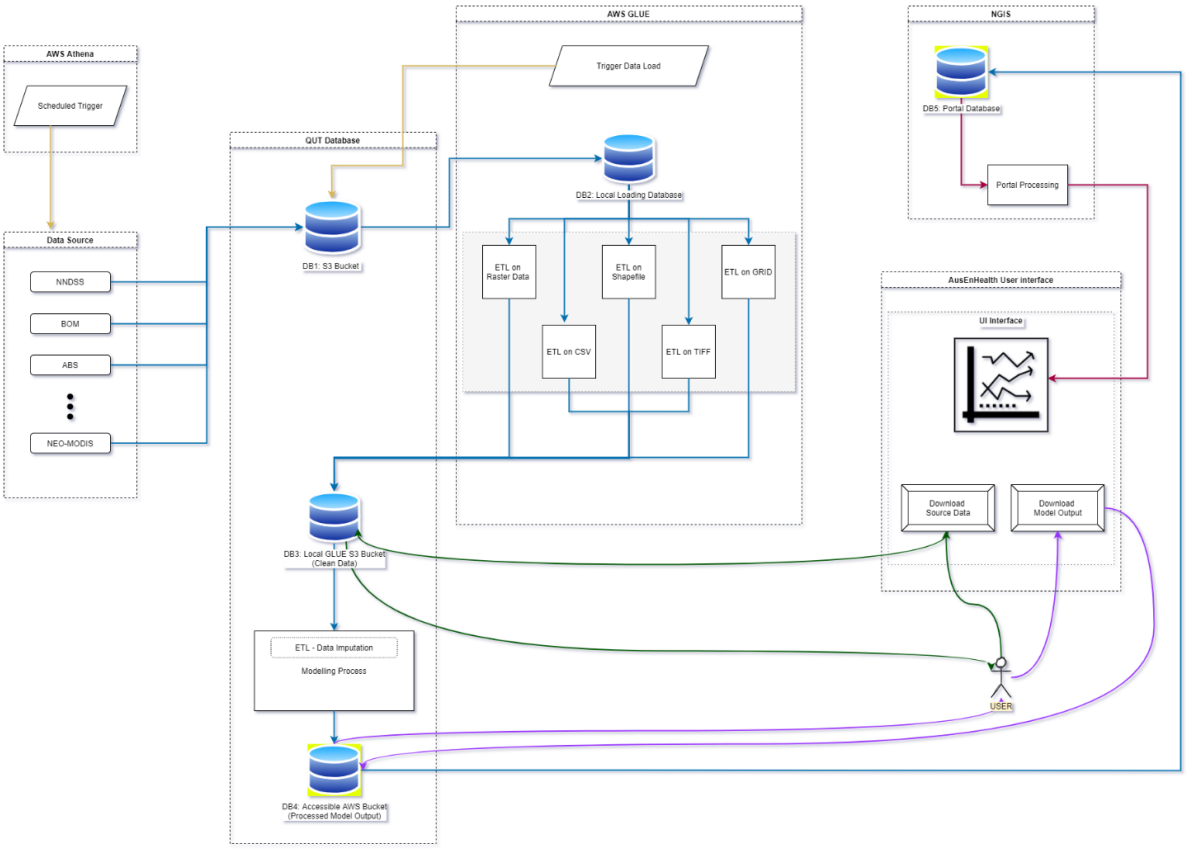

Figure 26: Proposed data pipeline, using AWS Glue.

H. Formatted Data Used in AusEnHealth Pilot

Raw Data:

f) Region Shapes

- Australian Shape Files
- State/Territory Shape Files
- Local Government Area (LGA) Shape Files
  - Contains conversions between LGA and State/Territory.
- Statistical Areas (SA) Shape Files
  - Contains conversions between SA2, SA3, SA4, Greater Capital City Statistical Areas (GCCSA), and State/Territory.

g) Demographics

- Australian Bureau of Statistics Data (relevant columns only)
  - Population Dataset (SA2, LGA)
  - Income Dataset (SA2, LGA)
  - Education & Employment Dataset (SA2, LGA)
  - Family & Community Dataset (SA2, LGA)
  - Health & Disability Dataset (SA2, LGA)
  - Economy & Industry Dataset (SA2, LGA)
  - Motor Vehicles by Dwelling Dataset (SA2, LGA)
- Public Health Information Development Unit (PHIDU) Data (relevant columns only)
  - Social Health Atlas Dataset (LGA)

h) Climate

- Maximum Temperature (Grid)
- Minimum Temperature (Grid)
- Daily Average Temperature (SA2, LGA)

i) Air Quality

- Nitrogen Monoxide, NO (Grid, SA2, LGA)
- Nitrogen Dioxide, NO_2 (Grid, SA2, LGA)
- Sulphur Dioxide, SO_2 (Grid, SA2, LGA)
- Carbon Monoxide, CO (Grid, SA2, LGA)
- Ozone, O_3 (Grid, SA2, LGA)
- Particulate Matter, <10µm (Grid, SA2, LGA)
- Particulate Matter, <2.5µm (Grid, SA2, LGA)
- Particulate Matter, <1µm (Grid, SA2, LGA)

j) Built Environment

- Hospital Counts (SA2, LGA)
- Green Space, Parklands (SA2, LGA)
- Water Bodies (SA2, LGA)
- Vegetation Index (Grid, SA2, LGA)

k) Health Outcomes

- MORT Books (SA3)
- AIHW Mortality Data (SA3)
- AIHW Morbidity Data (SA3)
- WADOH Mortality Data (LGA)
- WADOH Morbidity Data (LGA)

I. Vulnerability Index Variable Definitions

The following list of variables are used in the creation of the various AusEnHealth vulnerability indices. Expanded definitions are listed here for additional context.

Theme	Variable
Heat Exposure	Sum of daily historical heat percentiles
	Average non-zero EHF
Cold Exposure	Sum of daily historical cold percentiles
	Average non-zero ECF
Air Pollutant Exposure	Average Nitrogen Monoxide (µg/m^3)
	Average Nitrogen Dioxide (µg/m^3)
	Average Ozone (µg/m^3)
	Average Particulate Matter <2.5µm (µg/m^3)
Socioeconomic Status	Population Density
	Median Income ($)
	Low Income (%)
	Highschool Education (%)
	Unemployment rate (%)
Household Composition	Elderly Population (%)
	Infant Population (%)
	Single Parents of Dependants (%)
	Adults Providing Unpaid Childcare (%)
	Adults Requiring Assistance for Core Activities (%)
	Adults with a Disability (%)
	Population Living Alone (%)
Language and Culture	Speaks a Second Language at Home (%)
	Indigenous Population (%)
Housing Conditions	Mobile Homes (%)
	Crowded Dwellings (%)
	Rented Dwellings (%)
	Mortgage Payers (%)
Health Status	Condition: Respiratory Diseases (ASR per 100)
	Condition: With Asthma (ASR per 100)
	Condition: With Chronic Obstructive Pulmonary Disease (ASR per 100)
	Condition: Circulatory System Diseases (ASR per 100)
	Condition: With High Blood Pressure (ASR per 100)
	Condition: High Blood Cholesterol (ASR per 100)
	Condition: Heart, Stroke, and Vascular Disease (ASR per 100)
Health Risk Factors	Condition: Overweight but not Obese (ASR per 100)
	Condition: Obese (ASR per 100)
	Condition: Current Smokers (ASR per 100)
	Condition: More than Two Alcoholic Beverages per Day (ASR per 100)
Health Services	Hospitals (per 1000)
Cool Places	Greenspace (Parkland Mesh block %)
	Water Bodies (%)
	Vegetation Index
Social Connectedness	Dwellings with a Vehicle (%)
	Access to Internet (%)

Table 20: Extended vulnerability index variable descriptions.

J. Demographics Imputation Assessment

The following table represents the available demographics data (at SA2, SA3, SA4, and LGA geographical resolutions) and the imputed data for use in vulnerability indices.

Variable		Data Availability (by year)								
		2011	2012	2013	2014	2015	2016	2017	2018	2019
PD	Population Density									
MI	Median Income									
LI	Low Income									
HS	Highschool Education									
UR	Unemployment Rate									
ELD	Elderly									
INF	Infant									
SP	Single Parents									
UC	Unpaid Childcare									
ASS	Requiring Assistance									
WD	With Disability									
LA	Living Alone									
SL	Speaks a Second Language									
IND	Indigenous									
MH	Mobile Home									
CD	Crowded Dwellings									
RD	Rented Dwellings									
MP	Mortgage Payers									
NV	No Vehicle									
INT	Internet									
ALC	Excess Alcohol Consumption									
AST	With Asthma									
BP	High Blood Pressure									
COPD	With COPD									
OB	Obese									
OW	Overweight but not Obese									
SMO	Smokers									
BC	High Blood Cholesterol									
CS	Circulatory System Diseases									
RESP	Respiratory System Diseases									
VD	Vascular Diseases									

Figure 27: Available and imputed ABS demographics data. A dark green colour for a given variable and year represents observed data in that year. A light green colour indicates data that has been imputed using region-based linear interpolation. No colour represents no observed or imputed data in the final product.

Bibliography

[1] Environmental Health Indicators New Zealand, “What is environmental health?,” [Online]. Available: https://www.ehinz.ac.nz/indicators/overview/what-is-environmental-health/.

[2] A. Pruss-Ustun, J. Wolf, C. Corvalan, R. Bos and M. Neira, “Preventing disease through healthy environments: a global assessment of the burden of disease from environmental risks,” World Health Organization, 2016.

[3] Healthy People, “Environmental health,” [Online]. Available: https://www.healthypeople.gov/2020/topicsobjectives/topic/environmental-health.

[4] National Environmental Health Association, “Definitions of environmental health,” [Online]. Available: https://www.neha.org/about-neha/definitions-environmental-health.

[5] Australian Government Department of Health, “Overview of environmental health,” [Online]. Available: https://www1.health.gov.au/internet/main/publishing.nsf/Content/health-pubhlth-strateg-envhlth-index.htm.

[6] enHealth, “National Environmental Health Strategy Implementation Plan,” Department of Health and Aged Care. Commonwealth of Australia, 2000.

[7] A. D’Addona, N. Mullan and J. Dodds, “Investigation, scoping and planning of a national environmental health tracking system,” West Australian Department of Health for enHealth Committee. Environmental Health Directorate, 2010.

[8] Australian Institute of Health and Welfare, “National health reform: performance and accountability framework,” 2011. [Online]. Available: https://www.aihw.gov.au/getmedia/ea9b2361-38de-43f3-9426-8705fcc8f1da/performance-and-accountability-framework.pdf.aspx.

[9] Centers for Disease Control and Prevention, “Indicators and data,” [Online]. Available: https://ephtracking.cdc.gov/showIndicatorPages.

[10] Department of Communications and the Arts, “National map,” [Online]. Available: https://nationalmap.gov.au/.

[11] Public Health Information Development Unit, “Social health atlases,” [Online]. Available: https://phidu.torrens.edu.au/social-health-atlases.

[12] Geoscience Australia, “Digital Earth Australia hotspots,” [Online]. Available: https://hotspots.dea.ga.gov.au/.

[13] Western Australia Department of Health, “Atlas of environmental health,” [Online]. Available: http://doh.gaiaresources.com.au/.

[14] Small Area Health Statistics Unit, “The environment and health atlas for England and Wales,” [Online]. Available: http://www.envhealthatlas.co.uk/homepage/.

[15] National Institute of Public Health and the Environment, “Environmental health atlas,” [Online]. Available: https://www.atlasleefomgeving.nl/en.

[16] S. B. Thacker, D. F. Stroup, R. G. Parrish and H. A. Anderson, “Surveillance in environmental public health: issues, systems, and sources,” *American Journal of Public Health,* vol. 86, no. 5, pp. 633-638, 10996.

[17] C. Corvalan, M. Nurminen and H. Pastides, “Linkage methods for environment and health analysis: technical guidelines: a report of the health and environment analysis for decision-making (HEADLAMP) project,” World Health Organisation, 1997.

[18] European Environment Agency, “EEA core set of indicators: guide. EEA technical report,” 2005. [Online]. Available: https://www.eea.europa.eu/publications/technical_report_2005_1.

[19] Organisation for Economic Co-operation and Development,, “OECD core set of indicators for environmental performance reviews: a synthesis report by the group on the state of the environment,” *Environment Monographs,* vol. 83, pp. 1-39, 1993.

[20] B. Edokpolo, N. Allaz-Barnett, C. Irwin, J. Issa, P. Curtis, B. Green, I. Hanigan and M. Dennekamp, “Developing a conceptual framework for environmental health tracking in Victoria, Australia,” *International journal of environmental research and public health,* vol. 16, no. 10, p. 1748, 2019.

[21] I. Hanigan, C. Geromboux, J. Horsley, S. Phelan, E. Jegasothy, K. Heathcote and G. Morgan, “Environmental health indicators for selected environmental risk/health outcome pairs in New South Wales,” *Unpublished.*

[22] Ministry of Health, “Environmental health indicators for New Zealand 2008,” 2008.

[23] S. Kingsland, “The Victorian environmental health indicator project: a discussion paper on the development of environmental health indicators for Victoria,” Department of Human Services, 2006.

[24] I. Bastrakova, “Reducing lives lost from heatwave events,” *Unpublished.*

[25] A. Haines and K. Ebi, “The imperative for climate action to protect health,” *New England Journal of Medicine,* vol. 380, no. 3, p. 263–273, 2019.

[26] Health Effects Institute, “State of global air 2020,” [Online]. Available: https://www.stateofglobalair.org/.

[27] J. Spickett, H. Brown and D. Katscherian, “Health impacts of climate change: adaptation strategies for Western Australia,” Western Australia Department of Health, 2008.

[28] Australian Institute of Health and Welfare, “METeOR Home,” Australian Government, 2015. [Online]. Available: https://meteor.aihw.gov.au/content/index.phtml/itemId/181162. [Accessed March 2022].

[29] S. Solomon, D. Qin, M. Manning, Z. Chen, M. Marquis, K. Averyt, M. Tignor and H. Miller, Summary for Policymakers. In Climate Change 2007: The PHysical Science Basis. Contribution

of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change, Cambridge, United Kingdom and New York, NY, USA: Cambridge University Press, 2007.

[30] J. Hanse, M. Sato, R. Ruedy, K. Lo, L. David and M. Medina-Elizade, "Global temperature change," *Proceedings of the National Academy of Sciences,* vol. 103, no. 39, pp. 14288-14293, 2006.

[31] P. Bi and K. Parton, "Effect of climate change on Asutralian rural and remote regions: What do we know and what do we need to know?," *Australian Journal of Rural Health,* vol. 16, pp. 2-4, 2008.

[32] D. Patel, L. Jian, J. Xiao, J. Jansz, G. Yun, T. Lin and A. Robertson, "Joint effects of heatwaves and air quality on ambulance services for vulnerable population in Perth, Western Australia," *Evironmental Pollution,* vol. 252, pp. 532-542, 2019.

[33] B. Wondmagegn, J. Xiang, K. Dear, S. Williams, A. Hansen, D. Pisaniello, M. Nitschke, J. Nairn, B. Scalley, A. Xiao, L. Jian, M. Tong, H. Bambirkc, J. Karnon and P. Bi, "Increasing impacts of temperature on hospital admissions, length of stay, and related healthcare costs in the context of blimate change in Adelaide, South Australia," *Science of the total Environment,* vol. 773, p. 145656, 2021.

[34] National Institute of Environmental Health Sciences (NIEHS), "A human health perspective on climate change: a report outlining the research needs on the human health effects of climate change," Environmental Health Perspectives, NC, 2010.

[35] J. Berko, D. Ingram, S. Saha and J. Parker, "Deaths attributed to heat, cold, and other weather events in the United States, 2006-2010, national health statistics report," US Department of Health and Human Services, 2014.

[36] A. Gasparrini, Y. Guo, M. Hashizume, E. Lavigne, A. Zanobetti, J. Schwartz, A. Tobias, S. Tong, J. Rocklov, B. Forsberg and M. Leone, "Mortality risk attributable to high and low ambient temperature: a multicountry observational study," *The Lancet,* vol. 386, no. 9991, pp. 369-375, 2015.

[37] S. Vardoulakis, K. Dear, S. Hajat, C. Heaviside, B. Eggen and A. McMichael, "Comparative assessment of the effects of climate change on heat-and cold-related mortality in the United Kingdom and Australia," *Environmental Health Perspectives,* vol. 122, no. 12, pp. 1285-1292, 2014.

[38] P. Kinney, J. Schwartz, M. P, P. Elisaveta, A. Le Terte, S. Medina and R. Vautard, "Winter season mortality: will climate warming bring benefits?," *Environmental Research Letters,* vol. 10, no. 6, p. 064016, 2015.

[39] Bureau of Meteorology, Record-breaking La Niña events, Melbourne, Victoria: Bureau of Meteorology, 2012.

[40] Bureau of Meteorology, "Long-range weather and climate," 2021. [Online]. Available: http://www.bom.gov.au/climate/. [Accessed 08 2021].

[41] Bureau of Meteorology, “Thermal conformat observations,” 2010. [Online]. Available: http://www.bom.gov.au/info/thermal_stress/#wc. [Accessed 08 2021].

[42] R. Steadman, “Norms of apparent temperature in Australia,” *Australian Meteorological Magazine,* vol. 43, pp. 1-16, 1994.

[43] R. Cao, Y. Wang, J. Huang, J. He, P. Ponsawansong, J. Jin, Z. Xu, T. Yang, X. Pan and T. Prapamontol, “The mortality effect of apparent temperature: a multi-city study in Asia,” *International Journal of Environmental Research and Public Health,* vol. 18, no. 9, p. 4675, 2021.

[44] S. Almeida, E. Casimiro and J. Calheiros, “Effects of apparent temperature on daily mortality in Lisbon and Oporto, Portugal,” *Environmental Health,* vol. 9, no. 1, pp. 1-7, 2010.

[45] Copernicus, “CAMS global reanalysis (EAC4),” 2021. [Online]. Available: https://ads.atmosphere.copernicus.eu/cdsapp#!/dataset/cams-global-reanalysis-eac4?tab=overview. [Accessed 08 2021].

[46] C. Haenfling, “Environmental health indicators programme: indicators of drought, climate change, and health,” Massey University of New Zealand, Wellington, 2019.

[47] IQAir, “Air quality in Australia,” IQAir, 2021. [Online]. Available: https://www.iqair.com/au/australia. [Accessed 09 2021].

[48] Australian Institute of Health and Welfar, “Australian burden of disease study: impact of illness and death in Australia,” AIHW, Canberra, 2011.

[49] Australian Government, “National Clearn Air Agreement,” Commonwealth of Australia, 2015.

[50] L. Frank and P. Engelke, “Multiple impacts of the built environment on public health: walkable places and the exposure to air pollution,” *International Regional Science Review,* vol. 28, no. 2, pp. 193-216, 2005.

[51] AURIN, “Australian Urban Health Indicators,” 2021. [Online]. Available: https://aurin.org.au/about-aurin/network/ausurb-hi/. [Accessed 08 2021].

[52] Victoria State Government, “Landmark investment to create state-wide digital twin,” 2021. [Online]. Available: https://www.land.vic.gov.au/maps-and-spatial/maps-and-spatial-news/landmark-investment-to-create-state-wide-digital-twin. [Accessed 08 2021].

[53] Australian Bureau of Statistics, “What the Census is,” 2021. [Online]. Available: https://www.abs.gov.au/census/learn/about. [Accessed 08 2021].

[54] Geoscience Australia, “DEA Waterbodies,” 2021. [Online]. Available: https://www.ga.gov.au/dea/products/dea-waterbodies. [Accessed 08 2021].

[55] NASA, “MODIS Vegetation Index Products (NDVI and EVI),” 2021. [Online]. Available: https://modis.gsfc.nasa.gov/data/dataprod/mod13.php. [Accessed 08 2021].

[56] TERN, “Seasonal ground cover - landsat, JRSRP algorithm, Australia coverage,” 2021. [Online]. Available: https://portal.tern.org.au/seasonal-ground-cover-australia-coverage/22022. [Accessed 08 2021].

[57] M. Beaty and B. Varghese, “Reducing illness and lives lost from heatwaves,” Bureau of Meteorology, Canberra, 2021.

[58] Australian Bureau of Statistics, “Socio-economic indexes for areas (SEIFA),” Australian Bureau of Statistics, Canberra, 2016.

[59] Australian Institute of Health and Welfare 2020, “Australian Bushfires 2019-20: Exploring the Short-term Health Impacts,” AIHW, Canberra, 2020.

[60] Public Health Information Development Unit, “Notes on the data: chronic diseases and conditions,” 2021. [Online]. Available: https://phidu.torrens.edu.au/notes-on-the-data/health-status-disability-deaths/est-diabetes. [Accessed 08 2021].

[61] N. Seltenrich, “Between extremes: health effects of heat and cold,” *Environmental Health Perspectives,* vol. 123, no. 11, pp. 276-280, 2015.

[62] I. Manisalidis, E. Stavropoulou, A. Stavropoulos and E. Bezirtzoglou, “Environmental health impacts of air pollution: a review,” *Frontiers in Public Health,* vol. 8, p. 14, 2020.

[63] IPCC, “: Summary for Policymakers. In: Climate Change 2021: The Physical Science Basis,” IPCC, 2021.

[64] P. Kinney, “Climate change, air quality, and human health,” *American Journal of Preventive Medicine,* vol. 35, no. 5, pp. 459-467, 2008.

[65] World Health Organisation, “ICD purposes and uses,” WHO, 2020. [Online]. Available: https://www.who.int/standards/classifications/classification-of-diseases. [Accessed 09 2021].

[66] Australian Consortium for Classification Development, “ICD-10-AM/ACHI/ACS,” ACCD, 2017. [Online]. Available: http://www.accd.net.au/icd-10-am-achi-acs/. [Accessed 09 2021].

[67] Australian Institute of Health and Welfare, “Mortality over regions and time (MORT) books,” AIHW, 2021. [Online]. Available: https://www.aihw.gov.au/reports/life-expectancy-death/mort-books/contents/mort-books. [Accessed 09 2021].

[68] Australian Institute of Health and Welfare, “Australian bushfires 2019-20: exploring the short-term health impacts,” AIHW, Canberra, 2020.

[69] A. Fotheringham and D. Wong, “The modifiable areal unit problem in multivariate statistical analysis,” *Environment and Planning A,* vol. 23, pp. 1025-1044, 1991.

[70] B. Gurjar, A. Jain, A. Sharma, A. Agarwal, {. Gupta, A. Nagpure and J. Lelieveld, “Human health risks in megacities due to air pollution,” *Atmospheric Environment,* vol. 44, no. 36, pp. 4606-4613, 2010.

[71] C. Keijzer, D. Agis, A. Ambros and G. Arevalo, “The association of air pollution and greenness with mortality and life expectancy in spain: A small-area study,” *Environment International,* vol. 99, pp. 170-176, 2017.

[72] D. Hondula, R. Davis, M. Leisten, M. Saha, L. Veazey and C. Wegner, “Fine-scale spatial variability of heat-related mortality in Philadelphia County, USA, from 1983-2008: a case-series analysis,” *Environmental Health,* vol. 11, no. 1, pp. 1-11, 2012.

[73] C. Rinner, D. Patychuk, K. Bassil, S. Nasr, S. Gower and M. Campbell, “The role of maps in neighborhood-level heat vulnerability assessment for the city of Toronto,” *Catography and Geographic Information Science,* vol. 37, no. 1, pp. 31-44, 2010.

[74] C. Reid, M. O'Neill, C. Gronlund, S. Brines, D. Brown, A. Diez-Roux and J. Schwartz, “Mapping community determinants of heat vulnerability,” *Environmental Health Perspectives,* vol. 117, no. 11, pp. 1730-1736, 2009.

[75] J. Yu, K. Castellani, K. Forysinski, P. Gustafson, J. Lu, E. Peterson, M. Tran, A. Yao, J. Zhao and M. Brauer, “Geospatial indicators of exposure, sensitivity, and adaptive capacity to assess neighbourhood variation in vulnerability to climate change-related health hazards,” *Environmental Health,* vol. 20, no. 1, pp. 1-20, 2021.

[76] A. El-Zein and F. Tonmoy, “Assessment of vulnerability to climate change using a multi-criteria outranking approach with application to heat stress in Sydney,” *Ecological Indicators,* vol. 48, pp. 207-217, 2015.

[77] University of New South Wales (UNSW), “Heat Vulnerability Index for Sydney,” City Futures Research Centre, [Online]. Available: https://cityfutures.be.unsw.edu.au/cityviz/heat-vulnerability-index-sydney/. [Accessed 09 2021].

[78] L. Inostroza, M. Palme and F. de la Barrera, “A heat vulnerability index: spatial patterns of exposure, sensitivity and adaptive capacity for Santiago de Chile,” *PLOS One,* vol. 11, no. 9, p. e0162464, 2016.

[79] C. Reid, J. Mann, R. Alfasso, P. English, G. King, R. Lincoln and H. Margolis, “Evaluation of a heat vulnerability index on abnormally hot days: an environmental public health tracking study,” *Environmental health perspectives,* vol. 120, no. 5, pp. 715-720, 2012.

[80] Vermont Department of Health, “Health vulnerability in Vermont,” Vermont Department of Health, 2016.

[81] B. Balasubramani and I. Cruz, “Spatial Data Integration,” in *Encyclopedia of Big Data Technologies*, Springer International Publishing, 2018, pp. 1--8.

[82] P. Vassiliadis, “A Survey of Extract-Transform-Load Technology,” in *Integrations of Data Warehousing, Data Mining and Database Technologies*, 2011.

[83] T. a. G. G. a. L. L. a. O. G. a. P. N. W. Furche, “Data Wrangling for Big Data: Challenges and Opportunities,” in *EDBT*, EDBT, 2016, pp. 473--478.

[84] F. a. P. H. Endel, “Data Wrangling: Making data useful again,” *IFAC-PapersOnLine,* vol. 48, no. 1, pp. 111--112, 2015.

[85] J. a. B. Y. a. S. J. Song, “A Triggering and scheduling approach for ETL in a real-time data warehouse,” in *2010 10th IEEE International Conference on Computer and Information Technology*, IEEE, 2010, pp. 91--98.

[86] S. a. B. R. Goel, “Data replication strategies in wide-area distributed systems,” in *Enterprise service computing: from concept to deployment*, IGI Global, 2007, pp. 211--241.

[87] C. L. a. L. Y. a. C. R. H. Abad, “DARE: Adaptive Data Replication for Efficient Cluster Scheduling,” in *2011 IEEE International Conference on Cluster Computing*, IEEE, 2011, pp. 159-168.

[88] S. Guo, Z. Yuan, A. Sun and Q. Yue, “A new ETL approach based on data virtualization,” *Journal of Computer Science and Technology,* vol. 30, no. 2, pp. 311--323, 2015.

[89] A. Bologa and R. Bologa, “A Perspective on the Benefits of Data Virtualization Technology,” *Informatica Economica,* vol. 15, no. 4, 2011.

[90] Segment, “What is data orchestration & why it’s essential for analysis,” Segment, [Online]. Available: https://segment.com/resources/data-strategy/what-is-data-orchestration/. [Accessed 09 2021].

[91] K. Sudhakar, “Amazon Web Services (AWS) GLUE,” September 2018. [Online]. Available: https://www.ijmra.us/project%20doc/2018/IJMIE_SEPTEMBER2018/IJMRA-14276.pdf. [Accessed 22 September 2021].

[92] B. Harenslak and J. Rutger de Ruiter, Data Pipelines with Apache Airflow, Simon and Schuster, 2021.

[93] ANDS, “Australia National Data Service,” [Online]. Available: https://www.ands.org.au/working-with-data/metadata/storing-metadata.

[94] L. van den Brink, P. Barnaghi, J. Tandy, G. Atemezing, R. Atkinson, B. Cochrane, Y. Fathy, R. Garc{\'\i}a Castro, A. Haller and A. Harth, “Best practices for publishing, retrieving, and using spatial data on the web,” *Semantic Web,* vol. 10, no. 1, pp. 95-114, 2019.

[95] M. D. Wilkinson, “The FAIR Guiding Principles for scientific data management and stewardship,” p. 160018, 15 3 2016.

[96] ANZLIC Committee on Surveying & Mapping, “ICSM ISO19115-1 metadata good practices guide,” ICSM, 2019.

[97] DataCite, “DataCite Metadata Schema,” [Online]. Available: https://schema.datacite.org/meta/kernel-4.4/.

[98] Information/Geomatics, ISO/TC 211 Geographic, “ISO/CD 19115-3,” ISO, 2020.

[99] W3C, “Data Catalog Vocabulary (DCAT) - Version 2,” [Online]. Available: https://www.w3.org/TR/vocab-dcat-2/.

[100] OSGeo Project, “GeoNetwork,” [Online]. Available: https://geonetwork-opensource.org/. [Accessed 09 2021].

[101] CKAN, “The world's leading open source data management system,” CKAN, [Online]. Available: https://ckan.org/. [Accessed 09 2021].

[102] World Health Organisation, “Health in sustainable development planning: the role of indicators,” World Health Organisation, Geneva, 2002.

[103] B. Flanagan, E. Gregory, E. Hallisey, J. Heitgerd and B. Lewis, “A social vulnerability index for disaster management,” *Journal of Homeland Security and Emergency Management,* vol. 8, no. 1, 2011.

[104] Australian Bureau of Statistics, “Technical paper: Census of population and housing: socio-economic indexes for areas (SEIFA),” Australian Bureau of Statistics, Canberra, 2001.

[105] M. Parry, M. Parry, O. Canziano, J. Palutikof, P. Van der Linden and C. Hanson, “Climate change 2007: impacts, adaptation and vulnerability,” Cambridge University Press, Cambridge, 2007.

[106] L. Vescovi, M. Rebetez and F. Rong, “Assessing public health risk due to extremely high temperature events: climate and social parameters,” *Climate Research,* vol. 30, no. 1, pp. 71-78, 2005.

[107] A. Pezza, P. Van Rensch and W. Cai, “Severe heat waves in Southern Australia: synoptic climatology and large scale connections,” *Climate Dynamics,* vol. 38, no. 1-2, pp. 209-224, 2012.

[108] J. Nairn and R. Fawcett, “The excess heat factor: a metric for heatwave intensity and its use in classifying heatwave severity,” *International Journal of Environmental Research and Public Health,* vol. 12, pp. 227-253, 2015.

[109] D. Gaffen and R. Ross, “Increased summertime heat stress in the US,” *Nature,* vol. 396, no. 6711, pp. 529-530, 1988.

[110] R. Davis, D. Hundula and A. Patel, “Temperature observation time and type influence estimates of heat-related mortality in seven US cities,” *Environmental Health Perspectives,* vol. 124, no. 6, pp. 795-804, 2016.

[111] S. Sharma and P. Mujumdar, “Increasing frequency and spatial extent of concurrent meteorological droughts and heatwaves in India,” *Scientific Reports,* vol. 7, no. 1, pp. 1-9, 2017.

[112] K. Laaidi, A. Zeghnoun, B. Dousset, P. Bretin, S. Vandentorren, E. Giraudet and P. Beaudeau, “The impact of heat islands of mortality in Paris during the August 2003 heat wave,” *Environmental Health Perspectives,* vol. 120, no. 2, pp. 254-259, 2012.

[113] A. Cardil, D. Molina and L. Kobziar, “Extreme temperature days and their potential impact on southern Europe,” *Natural Hazards and Earth Systems Sciences,* vol. 14, no. 11, pp. 3005-3014, 2014.

[114] T. Chen, S. Sarmat, A. Grundstein, A. Winquist and H. Chang, “Time-series analysis of heat waves and emergency department visits in Atlanta, 1993 to 2012,” *Environmental Health Perspectives,* vol. 125, no. 5, p. 057009, 2017.

[115] J. Nairn and R. Fawcett, “Defining heatwaves: heatwave defined as a heat impact event servicing all community and business,” The Centre for Australian Weather and Climate Research, 2013.

[116] R. Trancoso, J. Syktus, N. Toombs, D. Ahrens, K. Koon-Ho Wong and R. Dalla Pozza, “Heatwaves intensification in Australia: A consistent trajectory across past, present and future,” *Science of the Total Environment,* vol. 742, no. 140521, 2020.

[117] B. Scalley, T. Spicer, L. Jian, J. Xiao, J. Nairn, A. Robertson and T. Weeramanthri, “Responding to heatave intensity: excess heat factor is a superior predictor of health service utilisation and a trigger for heatwave plans,” *Australia and New Zealand Journal of Public Health,* vol. 39, no. 6, pp. 582-587, 2015.

[118] B. Scalley, T. Spicer, L. Jian, J. Xiao and T. Weeramanthri, “Heatwave calculations: a comparative analysis of heatwave calculations,” Western Australia Department of Health, Perth, 2014.

[119] J. Xiao, T. Spicer, L. Jian, G. Yun, C. Shao, J. Nairn, R. Fawcett, A. Robertson and T. Weeramanthri, “Variation in population vulnerability to heat wave in Western Australia,” *Frontiers in Public Health,* vol. 5, no. 64, pp. 1-10, 2017.

[120] J. Nairn and R. Fawcett, “The Excess Heat Factor: A Metric for Heatwave Intensity and Its Use in Classifying Heatwave Severity,” *International Journal of Environmental Research and Public Health,* vol. 12, no. 1, pp. 227-253, 2015.

[121] Australian Governmnent Federal Register of Legislation, “National Environment Protection (Ambient Air Quality) Measure,” 25 February 2016. [Online]. Available: https://www.legislation.gov.au/Details/F2016C00215. [Accessed 09 2021].

[122] World Health Organisation, “WHO air quality guidelines for particulate matter, ozone, nitrogen dioxide and sulfur dioxide,” World Health Organisation Press, Geneva, 2005.

[123] G. Zosky, S. Vander Hoorn, M. Abramson, S. Dwyer, D. Green, J. Heyworth, B. Jalaludin, J. McCrindle-Fuchs, R. Tham and G. Marks, “Principles for setting air quality guidelines to protect human health in Australia,” *The Medical Journal of Australia,* vol. 214, no. 6, pp. 254-256, 2021.

[124] National Resilience Taskforce, “Profiling Australia's vulnerability: the interconnected causes and cascading effects of systemic disaster risk,” Australian Government, Department of Home Affairs, Canberra, 2018.

[125] Intergovernmental Panel on Climate Change, “A special report of working groups I and II of the intergovernmental panel on climate change (IPCC),” Cambridge University Press, Cambridge and New York, 2012.

[126] G. Burton, S. Rufat and E. Tate, “Social vulnerability: conceptual foundations and geospatial modelling,” in *Vulnerability and Resilience to Natural Hazards*, Cambridge, Cambridge University Press, 2018, pp. 53-81.

[127] B. Wisner, P. Blaikie, T. Cannon and I. Davis, At risk: natural hazards, people's vulnerability, and disasters, Routledge: Taylor & Francis Group London, 1994.

[128] A. Dwyer, C. Zoppou, O. Nielsen, S. Day and S. Roberts, “Quantifying social vulnerability: a methodology for identifying those at risk to natural hazards,” Geoscience Australia, Canberra, 2004.

[129] NAPSG Foundation, “Guidance on risk, resilience, and vulnerability indices,” National Alliance for Public Safety GIS Foundation, [Online]. Available: https://arcg.is/1Cze510. [Accessed 08 2021].

[130] S. Cutter, B. Boruff and W. Shirley, “Social vulnerability to environmental hazards,” *Social Science Quarterly,* vol. 84, no. 2, pp. 242-261, 2003.

[131] J. Bao, X. Li and C. Yu, “The construction and validation of the heat vulnerability index, a review,” *Internation Journal of Environmental Research and Public Health,* vol. 12, no. 7, pp. 7220-7234, 2015.

[132] L. Li, G. Lin, H. Liu, Y. Guo, C. Ou and P. Chen, “Can the air pollution index be used to communicate the health risks of air pollution?,” *Environmental Pollution,* pp. 153-160, 2005.

[133] S. Dasgupta, S. Lall and D. Wheeler, “Spatiotemporal analysis of traffic congestion, air pollution, and exposure vulnerability in Tanzania,” *Science of the Total Environment,* vol. 778, p. 147114, 2021.

[134] S. Marvel, J. House, M. Wheeler, K. Song, Y. Zhou, F. Wright, W. Chiu, I. Rusyn, A. Motsinger-Reif and D. Reif, “The COVID-19 pandemic vulnerability index (PVI) dashboard: monitoring county-level vulnerability using visualisation, statistical modelling, and machine learning,” *Environmental Health Perspectives,* vol. 129, no. 1, p. 017701, 2021.

[135] M. Ghose, R. Paul and S. Banerjee, “Assessment of the impacts of vehicular emissions on urban air quality and its management in Indian contect: the case of Kolkata (Calcutta),” *Environmental Science & Policy,* vol. 7, no. 4, pp. 345-351, 2004.

[136] P. Romero-Lankao, H. Qin and M. Borbor-Cordova, “Exploration of health risks related to air pollution and temperature in three Latin American cities,” *Social Science & Medicine,* vol. 83, pp. 110-118, 2013.

[137] T. Munzner, Visualization analysis and design (AK Peters visualzation series) (1st ed.), CRC Press, 2014.

[138] T. Slocum, R. McMaster, F. Kessler and H. Howard, Thematic cartography and geovisualisation, 3rd edition, Prentice Hall, 2008.

[139] S. Ribecca, “The data visualisation catalogue,” 2021. [Online]. Available: https://datavizcatalogue.com/. [Accessed 09 2021].

[140] M. Kraak and F. Ormeling, Cartography: visualization of geospatial data, Harlow, England: Prentice Hall, 2003.

[141] ESRI, “Style and create data visualizations,” ArcGIS Developers, 2921. [Online]. Available: https://developers.arcgis.com/documentation/mapping-apis-and-services/visualization/. [Accessed 09 2021].

[142] deck.gl, “Layers,” 2021. [Online]. Available: https://deck.gl/examples/. [Accessed 09 2021].

[143] N. Iliinsky, “Properties and best uses of visual encodings,” 2012. [Online]. Available: http://complexdiagrams.com/wp-content/2012/01/VisualPropertiesTable.pdf. [Accessed 09 2021].

[144] V. Agafonkin, “Leaflet - an open source JavaScript library for interactive maps,” 2021. [Online]. Available: https://leafletjs.com/. [Accessed 09 2021].

[145] S. Henderson, M. Brauer, Y. MacNab and S. Kennedy, “Three measures of forest fire smoke exposure and their associations with respiratory and cardiovascular health outcomes in a population-based cohort,” *Environmental health perspectives,* vol. 119, no. 9, pp. 1266-1271, 2011.

[146] D. Royé, R. Codesido, A. Tobías and M. Taracido, “Heat wave intensity and daily mortality in four of the largest cities of Spain,” *Environmental research,* vol. 182, p. 109027, 2020.

[147] F. Johnston, S. Purdie, B. Jalaludin, K. Martin, S. Henderson and G. Morgan, “Air pollution events from forest fires and emergency department attendances in Sydney, Australia 1996--2007: a case-crossover analysis,” *Environmental health,* vol. 13, no. 1, pp. 1-9, 2014.

[148] K. Martin, I. Hanigan, G. Morgan, S. Henderson and F. Johnston, “Air pollution from bushfires and their association with hospital admissions in Sydney, Newcastle and Wollongong, Australia 1994--2007,” *Australian and New Zealand journal of public health,* vol. 37, no. 3, pp. 238-243, 2013.

[149] H. Crabbe, “Risk of respiratory and cardiovascular hospitalisation with exposure to bushfire particulates: new evidence from Darwin, Australia,” *Environmental geochemistry and health,* vol. 34, no. 6, pp. 697-709, 2012.

[150] I. Hanigan, F. Johnston and G. Morgan, “Vegetation fire smoke, indigenous status and cardio-respiratory hospital admissions in Darwin, Australia, 1996--2005: a time-series study,” *Environmental Health,* vol. 7, no. 1, pp. 1-12, 2008.

[151] F. Johnston, I. Hanigan, S. Henderson, G. Morgan and D. Bowman, “Extreme air pollution events from bushfires and dust storms and their association with mortality in Sydney, Australia 1994--2007,” *Environmental research,* vol. 111, no. 6, pp. 811-816, 2011.

[152] P. Nafstad, L. Håheim, T. Wisløff, F. Gram, B. Oftedal, I. Holme, I. Hjermann and P. Leren, “Urban air pollution and mortality in a cohort of Norwegian men,” *Environmental health perspectives,* vol. 112, no. 5, pp. 610-615, 2004.

[153] F. Johnston, R. Bailie, L. Pilotto and I. C. Hanigan, “Ambient biomass smoke and cardio-respiratory hospital admissions in Darwin, Australia,” *BMC public health,* vol. 7, no. 1, pp. 1-8, 2007.

[154] A. Merrifield, S. Schindeler, B. Jalaludin and W. Smith, “Health effects of the September 2009 dust storm in Sydney, Australia: did emergency department visits and hospital admissions increase?,” *Environmental Health,* vol. 12, no. 1, pp. 1-7, 2013.

[155] R. Tham, B. Erbas, M. Akram, M. Dennekamp and M. Abramson, “The impact of smoke on respiratory hospital outcomes during the 2002--2003 bushfire season, Victoria, Australia,” *Respirology,* vol. 4, no. 1, pp. 69-75, 2009.

[156] S. Grineski, T. Collins, J. Chakraborty and Y. McDonald, “Environmental health injustice: Exposure to air toxics and children's respiratory hospital admissions in El Paso, Texas,” *The Professional Geographer,* vol. 65, no. 1, pp. 31-46, 2013.

[157] N. Arriagada, J. Horsley, A. Palmer, G. Morgan, R. Tham and F. Johnston, “Association between fire smoke fine particulate matter and asthma-related outcomes: systematic review and meta-analysis,” *Environmental research,* vol. 179, p. 108777, 2019.

[158] Commonwealth of Australia, “Overcoming Indigenous Disadvantage,” 2014.

[159] Physical Environment Research Network, “Reducing Illness and Lives Lost from Heatwaves:,” Australian Government Data Integration Partnership for Australia, Canberra, 2021.

[160] A. Haikerwal, M. Akram, M. Sim, M. Meyer, M. Abramson and M. Dennekamp, “Fine particulate matter (PM 2.5) exposure during a prolonged wildfire period and emergency department visits for asthma,” *Respirology,* vol. 21, no. 1, pp. 88-94, 2016.

[161] A. Haikerwal, M. Akram, A. Del Monaco, K. Smith, M. Sim, M. Meyer, A. Tonkin, M. Abramson and M. Dennekamp, “Impact of fine particulate matter (PM 2.5) exposure during wildfires on cardiovascular health outcomes,” *Journal of the American Heart Association,* vol. 4, no. 7, p. e001653, 2015.

[162] C. Wong, A. Brooks, D. Leong, K. Roberts-Thomson and P. Sanders, “The increasing burden of atrial fibrillation compared with heart failure and myocardial infarction: a 15-year study of all hospitalizations in Australia,” *Archives of Internal Medicine,* vol. 172, no. 9, pp. 739-741, 2012.

[163] A. Schaffer, D. Muscatello, R. Broome, S. Corbett and W. Smith, “Emergency department visits, ambulance calls, and mortality associated with an exceptional heat wave in Sydney, Australia, 2011: a time-series analysis,” *Environmental Health,* vol. 11, no. 1, pp. 1-8, 2012.

[164] B. Scalley, T. Spicer, L. Jian, J. Xiao, J. Nairn, A. Robertson and T. Weeramanthri, “Responding to heatwave intensity: excess heat factor is a superior predictor of health service utilisation and a trigger for heatwave plans,” *Australian and New Zealand journal of public health,* vol. 39, no. 6, pp. 582-587, 2015.

[165] J. Madrigano, S. McCormick and P. Kinney, “The two ways of assessing heat-related mortality and vulnerability,” *American journal of public health,* vol. 105, no. 11, p. 2212, 2015.

[166] Y. Zhang, M. Nitschke and P. Bi, “Risk factors for direct heat-related hospitalization during the 2009 Adelaide heatwave: a case crossover study,” *Science of the total environment,* vol. 442, pp. 1-5, 2013.