MASTERS STANDARD FOR IR QUALITY MONITORINGGGGGGGGGGGGGGGG

SOCY

8001

First Edition

2021-05-14

 MASTERS STANDARD



**3AI DATA STANDARDS BODY**

ORGANISME DE NORMALISATION DES DONNÉES 3AI

ОРГАН ПО СТАНДАРТИЗАЦИИ ДАННЫХ 3AI

**Data sharing standard and interchange formats – Naming and labelling standards for air quality sensor data**

*Normes de partage de données et formats d'échange - Normes de dénomination et d'étiquetage pour les données des capteurs de qualité de l'air*

Reference number:

SOCY 8001 : 2021 (3A)

**ISO 8001 : 2021 (3A)**

**Foreword**

The **3Ai Data Standards Body (3ADSB)** is fictional Australian based standards body located within the School of Cybernetics at the 3A Institute. The 3ADSB committee consists of students from the 3Ai Masters Cohort of 2021. The work of the committee in standards development is conducted in consultation with the Build Team from the faculty of the School of Cybernetics.

Draft Masters Standards proposed by the committee are circulated to the School of Cybernetics for approval before their acceptance as Masters Standards.

The 3A Institute, by delegated authority from the College of Computer Science and Engineering (CECS) at the Australian National University, acts as a co-regulator of this standard which is part of a series of standards in the Open Air Quality Data Standard Series (OAQDSS).

Masters Standard SOCY 8001 was prepared by 3ADSB, *Defining a Data Standard for a Sensor Network.* This is version 1.2.1.

Users should note that all Masters Standards undergo revision from time to time and that any reference made herein to any other Masters Standard implies its latest edition, unless otherwise stated.

**© 3AI Data Standards Body, 2021**

**MASTERS STANDARD SOCY 8001 : 2021 (3A)**

**Naming and labelling standard for air quality sensor data to logging server**

**0 Purpose**

To create a naming and labelling standard for the naming and representation of information for data collected by PM2.5 air quality sensors, to be uploaded and retrieved from a Data Logging Server within the School of Cybernetics, Australian National University.

This standard is part of the Open Air Quality Data Standard Series (OAQDSS) and applies only to the naming convention for data naming and labelling for interactions with the server. Further data formats are not covered here.

While this standard is intended for use within the School of Cybernetics, it is envisaged that the standard may be adopted more broadly beyond members of the 3A Institute as a way of facilitating the open exchange of air quality data exchange and contributing to building cyber-physical systems that leverage these data sets.

**1 Authority**

This standard is maintained by the 3Ai Data Standards Body (3ADSB), with the Data Standards Chair as the decision maker. The 3ADSB is part of the 3A Institute, School of Cybernetics and comprised of the Masters Cohort of 2021.

The 3A Institute, by delegated authority of the College of Computer Science and Engineering (CECS) at the Australian National University, acts as a co-regulator of this Open Air Quality Data Standard (OAQDS) along with the 3ADSB.

**2 Standards Principles**

The following principles, classified as Outcome Principles and Operational Principles, are the basis for the development of the standards for the Open Air Quality Data Standard Series (OAQDSS).

**2.1 Outcome Principles:**

* OAQDSS contributes to the New Branch of Engineering (NBE);
* OAQDSS is a prototype that aims to extend in beyond the school of Cybernetics;
* OAQDSS is easy to consume for purposes such as visualisation and application development; and,
* OAQDSS uses open standards where possible.

**2.2 Operational Principles:**

* OAQDSS is implementation agnostic;
* OAQDSS is usable/simple: the full implementation will seek to be as simple as possible but no simpler to decrease the time/cost of implementation and ensure that the standard is widely used as possible;
* OAQDSS is consistent: where possible common data structures and patterns should be defined and reused;
* OAQDSS is performant: the implementation will consider performance requirements when designing and making changes to the full implementation; and,
* OAQDSS is version controlled and backwards compatible to enable future development while not breaking future application.

**3 Data Labelling and Unique ID**

These naming conventions relate to the labelling of data sent to the logging server within the School of Cybernetics.

**3.1 Unique ID, Name and IP Address**

The unique identifier (UID) must be the individual student or staff ID of the user in conjunction with the first name and student IP address.

Noting that an Application Programming Interface (API) is being created may be accessible by users outside of the 2021 cohort, the approach for the naming convention is based on usability as a metric, as well as the potential for multiple unique users.

There is a trade-off with using the student and staff ID for the UID, that being security and the potential for identification. However, as our primary metric is usability, a decision has been made to use staff and student IDs, as they already fulfill that function. Users external to the ANU will be provided with an automatically generated ID.

The student and staff ID must include a lowercase ‘u’ to align with the University’s existing naming convention.

Basic Format: {

 “uid” : “Unique ANU Student ID”,

 “student\_name” : “Your First Name”,

 “student\_ip” : “192.168.1.XX”,

“date” : formatted\_date

}

For “date”, please add the following syntax:

from datetime import date

today = date.today()

formatted\_date = today.strftime(“%d/%m/%y”)

*Example: {*

 “uid”: “u4847416”,

 “student\_name” : “Chloe”,

 “student\_ip” : “192.168.1.19”,

“date” : formatted\_date

}

**3.1 Server Request String**

Description here XXX

*Basic Format:*

“uid,student\_name,student\_ip,date,particles”

*Example:*

 **“u4847416, Chloe,192.168.1.19, 13/05/21, 510”**

**3.3 Metadata**

Metadata to be linked with each recording of sensor data is date collected, time collected and latitude/longitude and, where available, room number collected (in that order)

**5.1 Date**

To be logged in the format of [day][month][year]

For example: 040521

**5.2 Time:** is to be logged in 24-hour format [hour][minutes][seconds] for example 12:20:59

**5.3 Lat/long:** is to be recorded as per ISO 6709. Where the available equipment precludes obtaining an accurate lat/long inside a building it is acceptable to use the lat/long of the main entry of the building in which sensors are located plus. For example the lat/long of the front entry to the ANU Cybernetic Institute would be logged as [to be added]

Room numbers are an optional but recommended field. Where a room number is not available or not applicable “000” is to be recorded. The naming convention for the relevant building is to be used as the room number. For example the room number for the Studio is S136 and would be logged as “S136”

Metadata is to be separated by commas for example: 040521, 12:20:59, [lat long for 3Ai front door here], S136 [do we need this bit??]

*…..[incomplete]*

**3.1 Naming Convention**

Clear language that uses limited technical terms. Where technical terms are used, clear definitions and explanations are required.

**INITIAL DRAFT BELOW**

Version 1.0.0 (Correct as at 1300 04 May 2021)

Task definition ([from F5 Homework](file:///C%3A/Users/katet/Downloads/CECS8001_F5_HW_Network_Protocol.pdf)):

*‘This task asks that you define and implement a standard for your sensor network. As a cohort, define the data sharing standard sufficiently to ensure consistency of the data shared by the different sensors/users’.*

Who’s doing what for our standard:

* **Purpose/Principles:** Adrian
* **Operational Considerations:** Chloe and Myrna
* **Sensor and metadata:** Jake and Jules
* **Equipment used + systems map:** Erika and James
	+ **Data and time/frequency:** Erika and James
* **Labels (naming conventions and Unique ID**: Sarah and Matt (Chloe, Myrna, Erika)
* **Data visualisation:** Kate

**Purpose**:

To create a standard for the inter-change of air quality information within the School of Cybernetics. The standard covers the collection and classification of data relating to air quality within the School of Cybernetics, as well as enumerating all information required to understand the provenance or lineage of that information.

These standards are maintained by the **Prototype 3Ai Data Standards Body (3ADSB)**, with the Data Standards Chair as the decision maker. The 3ADSB is part of the School of Cybernetics and the Masters Cohort of 2021. The work of standards development is conducted in consultation with the faculty of the School of Cybernetics who acts as a co-regulator of the **Open Air Quality Data Standard (OAQDS)**, along with the 3ADSB.

While this standard is intended for use within the School of Cybernetics, it is envisaged that potentially the standard may be adopted more broadly as a way of facilitating the open exchange of air quality data and contributing to building CPS that leverage these data sets.

The standards are required to be published by the 3ADSB as well as any version changes on a timely basis.

**Standards**:

This document represents version 1.0.0 of the standard. Please see the versioning section for more information regarding versions. This standard has been inspired in part by:

<https://consumerdatastandardsaustralia.github.io/standards/#introduction>

<https://www.standards.org.au/>

**Versioning**:

The standards documentation will be versioned using three version parts <major>.<minor>.<bug fix>. This version will be used to describe updates in the change log. The standards will be backwards compatible to at least one major version.

**Principles:**

The following principles, classified as Outcome Principles and Operational Principles, are the basis for the development of the standards for the Open Air Quality Data Standard (OAQDS).

**Outcome Principles:**

* OAQDS contributes to the New Branch of Engineering (NBE)
* QAQDS is a prototype that potentially be extended beyond the school of Cybernetics.
* OAQDS is easy to consume for purposes such as visualisation and application development.
* OAQDS uses open standards where possible

**Operational Principles:**

* OAQDS is implementation agnostic
* OAQDS is usable/simple: the full implementation will seek to be as simple as possible but no simpler to decrease the time/cost of implementation and ensure that the standard is widely used as possible.
* OAQDS is consistent: where possible common data structures and patterns should be defined and reused.
* OAQDS is performant: the implementation will consider performance requirements when designing and making changes to the full implementation
* OAQDS is version controlled and backwards compatible to enable future development while not breaking application
* OAQDS is extensible as this is a prototype only and will potentially be extended into a production usage in future

**OPERATIONAL CONSIDERATIONS (Chloe & Myrna)**

**Purpose of System:** To collect data from the PM2.5 Sensor and upload to and retrieve from the Data Logging Server.

**Time period of data collection:** 24 hours

**Frequency of data collection:** Every 1 minute

**Access frequency of server:** 10 minutes (to reduce access events to the server at any one time)

**Security Concerns:**

Data privacy: use of unique identifier can lead to personal identification through the ANU system.

Cyber-attack and hacking the sandbox network (wifi: roboroo i'm a robot robot)

Possible data manipulation if accessed externally.

**Sensor data generated and metadata**

[Jake and Julian – sensor/ metadata]

Sensor to be used in this system is: Adafruit PM 2.5 sensor PMS5003.

**Data sheet:**

<https://cdn-shop.adafruit.com/product-files/3686/plantower-pms5003-manual_v2-3.pdf>

Sensor will collect data at a rate of 9600 bits per second.

Data is streamed out at a rate of once per second.

Data includes:

* A measure of the concentration of particulates in the air.
* A measure of the mass of particulates in the air.

The measure of particulate matter per 0.1L air is categorized into 0.3um, 0.5um, 1.0um, 2.5um, 5.0um and 10um size bins

Concentration is measured in particulate matter (measured in microns) in a volume of .1L

**Metadata**

Metadata to be linked with each recording of sensor data is **date** collected, **time** collected and **latitude/longitude** and, where available, **room number** collected (in that order)

Date is to be logged in the format of [day][month][year] for example: 04052021

Time is to be logged in 24 hour format [hour][minutes][seconds] for example 12:20:59

Lat/long to be recorded as per ISO 6709. Where the available equipment precludes obtaining an accurate lat/long inside a building it is acceptable to use the lat/long of the main entry of the building in which sensors are located plus. For example the lat/long of the front entry to the ANU Cybernetic Institute would be logged as [to be added]

Room numbers are an optional but recommended field. Where a room number is not available or not applicable “000” is to be recorded. The naming convention for the relevant building is to be used as the room number. For example the room number for the Studio is S136 and would be logged as “S136”

Metadata is to be separated by commas for example: 04052021, 12:20:59, [lat long for 3Ai front door here], S136 [do we need this bit??]

[who is doing the bit about how often the data gets sent to the server and how often sensor turns on ?]

Change Log:

**LABELS – NAMING CONVENTIONS AND UNIQUE ID (SARAH & MATTHEW)**

**Naming Convention:**

Clear language that uses limited technical terms. Where technical terms must be used, clear definitions and explanations are required.

**Unique ID:**

Noting that an Application Programming Interface (API) is being created that will presumably be accessible by users outside of the 2021 cohort, the approach for the naming convention is based on usability as a metric, as well as the potential for multiple unique users. The unique identifier (UID), will therefore use individual student and staff IDs. Users that are external to the ANU will be provided with an automatically generated ID.

There is an identified trade-off with using the student and staff ID for the UID, that being security and the potential for identification. However, as our primary metric is usability, the judgement has been made to use staff and student IDs, as they already fulfill that function. **[Need to add further explanation here].**

{“uid”: “Unique ANU Student ID”}

Example: {“uid”: “u4847416”}

The student and staff ID must include a lowercase ‘u’ to align with the University’s existing convention.

**Data Visualisation Method**

The preferred data visualisation method for reporting purposes is Tableau.

ERIKA & JAMES

What data are we collecting?

* TO UART (FROM PM2.5 Air Quality Sensor)
	+ Format: 9600 baud data stream
	+ Frequency: Once per second
	+ Information:
		- PM1.0, PM2.5, PM10.0 concentration in standard and environmental units
		- Particulate matter per 0.1L air, categorized into 0.3um, 0.5um, 1.0um, 2.5um, 5.0um and 10um size bins
* TO RPi (FROM UART)
	+ Format: JSON
	+ Information:
		- PM1.0, PM2.5, PM10.0 concentration in standard and environmental units
		- Particulate matter per 0.1L air, categorized into 0.3um, 0.5um, 1.0um, 2.5um, 5.0um and 10um size bins
		- Metadata – where able
* TO Server (FROM RPi)
	+ Format: TCP (Sockets)
	+ Information:
		- Metadata – where able
* TO RPi (FROM Server)
	+ Format: .csv
	+ Information:
		- All data stored in server
* To Data Logging Server
	+ Equipment: Raspberry Pi
	+ Format:
	+ Frequency:
* Read by User
	+ Format: csv. file
	+ Interface: http through web browser....
	+ Access frequency: open access by all
	+ Information: data from Air Quality Sensor

SYSTEM ARCHITECTURE

**Universal asynchronous receiver-transmitter (UART):** a computer hardware device for asynchronous serial communication in which the data format and transmission speeds can be configured.

* Equipment:

**Air Quality Sensor:** a sensor that monitors the presence of particles 2.5 microns or smaller in diameter. This sensor uses laser scattering to radiate suspending particles in the air, then collects scattering light to obtain the curve of scattering light change with time.

* **Equipment**: Adafruit PM2.5 Sensor and Adapter Kit
* Ref: <https://learn.adafruit.com/pm25-air-quality-sensor>

**Raspberry Pi (RPI):** a small single-board computer

* Equipment:

**Data Logging Server:** an electronic device that contains a log file (or several files) with a list of activities performed. This should be a python based server running on a Raspberry Pi

* Equipment: Raspberry Pi 4(? – check model)

**Router:**

DIAGRAM TITLE



**\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\***

**Method analysis:**

**Decisions made during the drafting of the Standard:**

* We leveraged a template that will define the standard to a level of granularity to ensure interoperability of the systems that use the OAQDS
	+ discounted but were informed by an ISO/Australian template
* We discounted adding in the type of sensor into the meta-data as this is only a prototype and could be added later.
* We increased the amount of sensor readings from 1 per hour to 1 per 10 minutes
* We reduced the data analysis period from 72 hours to 24 hours.
* We decided not to use first names as a UID as university IDs are already a UID and due to future cohorts coming through which may have the same first names (ie there are already being two Matthews at the School.