Open Source Data Management System (OSDMS) for Air Quality

Expression of Interest
from the **WAQI** project

(In response to cleanairfund.org/../Clean-Air-Fund-Open-DMS-EOI-Final_19.06.24.pdf)

July 11th 2024, version 1.2

1/ Strategic Approach

How will you approach the project? Please outline your broad approach, including key activities, milestones, and expected interim outcomes

The underlying approach for building the Clean Air Fund Open Source Data Management System (CAF OSDMS) is to leverage the existing data platform solution from aqicn.org and waqi.info, which has been operating since 2015.

The activities will consist after having identified and nominated the COMPASS team leads, of establishing horizontal challenges, led by twin-pairs leads from the COMPASS team, and creating a vibrant eco-system around the OSDMS based on the concept of "horizontal" plugins.

We expect that, based on reusing the existing work done for the aqicn.org data platform, it should be possible to complete the initial outcome 8 months after the project starts, and functional MVP with initial plug-ins 12 months after the project starts (although we do expect the first MVP demos to be available earlier, and progressively, following the intermediate workshops outcomes). During the second phase, we will strengthen the community by funding the development of additional plugins, to make the OSDMS an even more vibrant eco-system.

2/ Community Building

How will you ensure active participation and strong membership of the COMPASS, including diverse actors across different regions?

We will identify the members directly from the users of the aqicn.org data platform, which currently comprises users from the different domains: Countries with low/mid income, aka LMIC, (eg Laos), the Private sector (eg Black Rock), Accademia (many, worldwide), Cities or Counties (many worldwide), NGO (many worldwide) and individual users.

There is already a very strong participation on the aqicn.org data platform, with several millions of API calls made every day and hundreds of custom user requests every month, coming from all sectors.

However, to ensure an even stronger membership of the core COMPASS team, we will call for electing leads in each sector - the leads will however not be asked to represent their "vertical

business domain", but rather horizontal and cross-business-domain problematics. The leads will be twinned-paired with another lead from a different sector, to ensure that the problem is seen from a cross-business domain point of view, thus benefiting the overall community.

The initial ideas for the horizontal problem to be addressed are the following:

* Real-time data: providing a uniform model, especially for different sources having different formats, units, etc.
* Historical data: Whether sensor data, remote sensing data, hindcasting, etc…
* Data visualisation: This is about providing means and recipes for connecting to different viz tools, rather than developing the viz tools.
* Augmentation: processing and transforming data from different sources, from spatial and temporal perspectives, to produce synthetic data.
* Augmentation: the ability for users to inject their own data sources and data sets, either privately or publically, with all the associated problems (authorisation, validation, etc…)
* AirGPT (NLP): Making the data queries human-friendly, and thus addressing custom user requests in a more streamlined way.

The intent is to have at least 6 horizontal problems so that 2*6 leads can be nominated. Each twinned pair will be given the budget for organizing the necessary workshops. Each pair may propose new cross-functional problems to be addressed.

3/ Capacity and Expertise

What are your relevant qualifications and experiences for delivering this project successfully, especially in managing collaborative initiatives, open-source initiatives and achieving scale? Please provide examples of previous projects that demonstrate your capability to deliver a scalable and sustainable system.

All the work done for aqicn.org and waqi.info are examples that hopefully prove that we could successfully lead this OSDMS project.

To further emphasise the viability of the OSDMS, the data platform used by aqicn.org has been developed in a way to minimize operating costs, both from the cloud provider and user operation and maximize the usefulness. For reference, the aqicn.org data platform is cloud-hosted for a budget of less than 1K USD a month, yet able to deliver millions of API calls every day, able to collect several millions of sensor data points every hour and process gigabytes of gridded datasets (eg forecast) every day.

We do believe the data platform is the right base for an Open Source DMS since it has been optimized, from Day one, for lean and low-cost operations, which will allow many users to deploy their instances at an efficient cost-benefit ratio.

Another key to the OSMDS is to extend the DMS horizontally, via means of cross-functional plug-ins, rather than horizontally by the business domain (eg storage, data processing): This allows not only a better collaborative community, but also learner extensions, by making it possible for small teams to change/customize the entire system.

4/ Equity and Inclusivity

How your organisation will ensure a participatory and equitable approach in building and managing the project.

To ensure participation in developing the OSDMS, we will reserve a budget for developing plugins by different teams. We estimate that it should be possible to allocate 25% of the initial budget for those plugins, aiming at funding the development of between 10 to 100 plugins, for an estimated budget of 1K~10K USD per plugin. The allocated funds will be used by the plug-in development team themselves, under the supervision of the twinned leads.

To ensure inclusivity and equitability, we will ensure that the plug-in teams/project proposals come from different countries in the world, and from all continents. We will do our best to ensure a balanced number of men and women on the project. We will also try to establish virtual project teams to allow actors from different backgrounds to work together (eg a data scientist, a policy maker and a software engineer).

Last, we will also try to add a mechanism for the community to propose plug-ins, and vote for the plug-ins that would benefit the community. For example, a setup similar to a hybrid between crowdfunding and Reddit voting to elevate valuable content from the community.

5/ Budget and Timeline

What are the Indicative costs and timeline of delivering the activities (split by phase).

Out of the 400.000 USD maximum budget, the estimated planned usage is the following:

Phase 1:
- 25% for twinned-pair workshops.
- 25% for the DMS blueprint and initial MVP engineering.

Phase 2:
- 25% for community plugin development.
- 25% kept as a buffer, for alignment workshops and DMS engineering support.

We estimate that the workshop's outcome could be ready 8 months after the project starts and that the initial MVP (based on the workshop's outcome) could be ready 12 months after the project starts. The initial ~10 plugins could be ready 16 months after the project starts. The full initial plugins should be completed no later than 24 months after the project starts.

We will furthermore seek for additional budget, especially for phase 2, with institutional and private actors agreeing to fund the plug-in development effort. The objective is to strengthen the community, by creating more plugins and a more vibrant eco-system which will fulfil the needs of not only the plugin developers but also the institution and private actors.

Establishing a global, open-source Data Management System for Air Quality

Request for Expressions of Interest (EOI)
June/July 2024

Project Summary

The Clean Air Fund invites expressions of interest for a new project to develop an open-source air quality Data Management System (DMS). This will be realised in 2 phases: first through establishing a Community of Practice for Air Quality Systems (COMPASS) which will secure the partnerships, broad user buy-in and co-funding necessary for long-term success, and second through developing and testing the prototype Open-Source Data Management System. We are seeking a partner organization/consultant or consortia of partners with a strong background in open source software development, air quality, and global community engagement, to lead this work.

Background to the project

A Data Management System (DMS) is software that collects, stores, quality assures, visualizes, and distributes air quality data from monitoring networks and instruments. A DMS is an essential part of any monitoring network to efficiently manage data, aiding in the analysis and implementation of air quality measures.

Despite its importance, the current approach to data management is plagued by a lack of coordination and consistency, leading to duplication of effort, increased costs, systems that are not built-for purpose, inconsistent quality control and constrained data sharing. Growing data volumes and diverse data formats further exacerbates these challenges, especially with the increasing uptake of low-cost monitors, satellite data, and sophisticated modelling. Consequently, governments, institutions, academia and communities face significant hurdles in effectively managing air quality data, which reduces the impact of air quality monitoring efforts.

Developing an open-source DMS – one that is built both *for* and *in collaboration with* a community of users – could resolve these issues. A single, unified and open-source platform that can be iterated and built on over time can streamline data management, enhance transparency, allow adaptability to local needs, bring consistency and ultimately improve the value of air quality data for all stakeholders.

The Clean Air Fund, recognizing the critical need for a standardized, scalable solution, is seeking to support the development of an open-source DMS to benefit the global community. To this end, we conducted a scoping study in 2023 to understand current practices, future needs, and the best way forward. This exercise identified two main recommendations to overcome current data management challenges worldwide: to 1) create a global Community of Practice to ensure the right stakeholders are inputting to the design and long-term viability of the DMS; and 2) develop the open-source DMS prototype based on the Community of Practice recommendations.

This project seeks to take forward these recommendations. Proposers should review the final report of the scoping study, (see the bottom of this web page), before submission.

Background to Clean Air Fund

Launched in 2019, the Clean Air Fund is a philanthropic initiative with a mission to tackle air pollution around the world. Our aim is to help build and support a powerful global movement for clean air. We achieve this by bringing together funders, researchers, policy makers and campaigners working on a wide range of issues to find and scale solutions that will provide clean air for all.

More information about the Clean Air Fund can be found on our website: www.cleanairfund.org

Project outcomes and prospective activities

This Expression of Interest (EOI) invites organizations, experienced consultants, and/or networks of partners to register their interest and provide a high level summary of their approach, relevant experience and indicative budget. Further guidance is provided below. This will be followed by a request for a more detailed proposal from the top 1 or 2 bidders.

We welcome and encourage collaborations / consortia of organizations to apply that together bring the necessary expertise (for example, open-source software development, air quality expertise, stakeholder engagement and management).

The below activities are indicative: we welcome creativity in the high-level approach proposed in order to achieve the outcomes.

Outcome 1: **A Community of Practice for Air Quality Systems (COMPASS) brings together the necessary stakeholders to drive the development of an impactful, fit-for-purpose, widely used and sustainable Open-source DMS prototype.**

Indicative activities:

- *Develop a term of reference to set out the remit, approach and management of the COMPASS.*
- *Recruit founding members from diverse sectors (government, academia, NGOs, etc.). Identify how these members will be engaged and appropriately compensated for their involvement, to ensure there is a high level of engagement.*
- *Organize workshops/webinars to advise and refine on the open-source DMS prototype, including system requirements, where and how it will be housed, a long-term (at least 5 year) funding plan, governance and future development of the DMS to ensure its long-term viability.*
- *Finalize the open-source DMS blueprint with COMPASS members, building on recommendations from the scoping phase.*
- *Identify potential pilot locations, projects and/or users (2-5) and work with COMPASS members (including CAF) to secure commitments for co-funding to progress the development of open-source DMS.*

Outcome 2: **An open-source DMS prototype is fully functional, serves diverse end-user need and has widespread uptake.**

- *Based on the COMPASS recommendations, develop a fully open-source DMS prototype (minimum viable product) that is most useful to the widest relevant audience.*

- *Pilot the prototype with the projects/users identified in phase 1 and feedback the learnings to COMPASS and other relevant stakeholders.*
- *Develop documentation for developers to support the broader community to add to/update/build modules to the DMS and adapt it to their needs, as well as user friendly training materials.*
- *Identify the long-term role of the COMPASS and, if needed, expand membership to additional stakeholders who can support for further development, scale up and funding.*
- *Develop a long-term sustainability plan and support its delivery before the project end.*

Project duration and budget

This project will be delivered in two distinct phases, to ensure there is sufficient progress and buy-in from Outcome 1 before moving on to Outcome 2. For that reason, we invite proposers to outline an indicative approach and budget separately for each outcome.

Clean Air Fund has budgeted for a maximum contribution of $400,000 USD for both phases. We expect this to go a significant way to setting up the long-term viability of the DMS and to encourage additional fundraising to achieve the objectives of outcome 2.

We request that bidders provide an indicative split of these funds between Outcome 1 and Outcome 2 in their proposal. We are flexible to the approach and will fund the proposal that most compellingly demonstrates use of these funds to maximise our intended outcomes.

We are also flexible on timeline but would expect the project to move on to deployment (Outcome 2) within 8-12 months.

Ongoing across all activities is frequent (approximately monthly) engagement with Clean Air Fund.

In parallel to this project, Clean Air Fund will likely launch project(s) to develop data platforms / management systems in support of individual cities as part of the Breathe Cities initiative. Running these projects in parallel provides the opportunity to learn from and build on each other, and further highlights the need for a standardised open source system to avoid future duplication of effort. Proposers are welcome to bid for both pieces of work (when published).

What is the Clean Air Fund looking for?

The following criteria will be used to assess the proposals submitted.

1. **Experience & commitment to successful open-source community-led projects**: demonstrable achievements in supporting (and ideally leading) open-source community projects, as well as the ability to work across diverse stakeholders, work globally, and develop trusting relationships.
2. **Commitment to openness**: demonstrable commitment to developing a system that is fully open, collaboratively built and strictly not for profit, with a willingness to lead 'behind the scenes'.
3. **Technical suitability**: a track record and expertise (or the ability to bring in the relevant expertise) in air quality data management or a related topic

4. **Creativity**: in proposing an approach that meets the brief and ensures the success of the work
5. **Value for money**: ensuring the long-term viability that can succeed without ongoing (long-term) Clean Air Fund support

What to include in your expression of interest

Clean Air Fund is structuring this project by first inviting expressions of interest with the intention of understanding the range of different proposers and approaches that could deliver this work.

Please consider the following points in your EOI response:

1) **Strategic Approach**: How will you approach the project? Please outline your broad approach, including key activities, milestones, and the expected interim outcomes.
2) **Community Building**: How will you ensure active participation and strong membership of the COMPASS, including with diverse actors across different regions?
3) **Capacity and Expertise**: What are your relevant qualifications and experience for delivering this project successfully, especially in managing collaborative initiatives, open-source initiatives, and achieving scale? Please provide examples of previous projects that demonstrate your capability to deliver a scalable and sustainable system.
4) **Equity and Inclusivity**: How will your organization ensure a participatory and equitable approach in building and managing the project?
5) **Budget and timeline:** What are the indicative costs and timeline of delivering the activities (split by phase)?

How to submit your expression of interest and process

To submit your expression of interest, please send your response to the five questions above and any additional information deemed relevant in document format to data@cleanairfund.org by the deadline set out below. This should be limited to 3 pages: we are looking for a high level concept at this stage. For the expression(s) of interest that are the best fit for this funding, we will invite the submission of a more detailed proposal.

Timeline

Activity	Deadline
Publish Expression of Interest	19th June
Deadline for questions	3rd July
Deadline for Expression of Interest	11th July
Review of Expressions of Interest	12th July– 25th July
Final Clean Air Fund decision to invite full proposal	26th July