

Phase II Competition - Finalist Application Form

Created: 10/24/2016

Last updated: 11/21/2016

In this form, the six finalist teams for the Open Science Prize are asked to describe the work they have done to develop their prototypes, and make the case for why they should be considered for the phase II Prize.

The information you submit on this form will be considered alongside the prototype you have developed in deciding which team will receive the Phase II Prize.

Please note that, unless otherwise stated, the information you submit on this form will be made available publicly via the Open Science Prize website (under a [CC BY 4.0 license](#)), so that it can be assessed as part of the public voting process. All fields, except the final box for additional information are mandatory.

Your application must be completed by 11:59pm Pacific Standard Time on 21 November 2016 . You may edit this form as many times as you like before the deadline.

If you have any questions about the Prize or the review process or if you would like to provide any further information that you would not wish to be made public, please contact David Carr (d.carr@wellcome.ac.uk) or Elizabeth Kittrie (elizabeth.kittrie@nih.gov). Any technical questions regarding this form or the web platform should be directed to (openscience@wellcome.ac.uk).

Page 1

Executive Summary

Please provide a brief Executive Summary of no more than 150 words for the public voting page on the Open Science Prize website. This should be suitable for an informed lay audience, and should briefly describe your prototype and why it should be considered for the phase II Prize. [150 words]

Air pollution is one of the biggest health threats of our time, responsible for one out of eight deaths each year. In the most polluted places, basic air quality data are critical for science to advance at policy-relevant, actionable levels. Our grassroots community of scientists, journalists, software developers and open data lovers has created the world's only open-source platform that aggregates, freely shares, and visualizes global air quality data from disparate sources (openaq.org). To date, we have opened up 28 million data points from 33 countries. However, a dataset is only powerful if people are using it, so we foster a virtual community and hold in-person

workshops to get people using these data. To meet community demands, we now seek to add satellite and citizen-scientists' low-cost sensor data, increase our real-time dataset to 70 countries and scale our workshops to be held in more communities facing air inequality.

Weblink for prototype

Please provide the public URL for your prototype tool or service (this will be viewed by the public for purposes of public voting):

<https://openaq.org>

Your Prototype

Purpose and need

Please provide a brief summary description of the purpose of the prototype you have developed and the key challenges or needs it is seeking to address [200 words]

From Bangkok to Los Angeles, meaningful access to air quality data can arm scientific and non-technical communities to combat air pollution. Many governments, researchers and citizen scientists publicly share air quality data - to the tune of 5-8 million data points per day - but often in disparate and sometimes temporary forms. Many groups that would like to use these data for a variety of purposes (e.g. public health research, policy analysis, or public engagement) have been forced to build ad hoc systems to collect a portion of these data. Others are never able to use these data to advance their work because the barrier to entry is too high.

The purpose of OpenAQ is to fill this gap between air quality data shared by public sources and the many sectors that would like to use these data. We do this by capturing these data before they disappear and put them in a universal format for anyone to access in a highly available manner. By creating this foundational data layer in a scientifically robust and transparent way, OpenAQ makes access to these data dramatically easier - or even possible - for an ecosystem of users.

Please summarise the work you have taken forward to develop your prototype since you were awarded the Phase I Prize in April 2016.

i. Progress

The key milestones achieved and the extent to which the goals and challenges you set out to address in your original application were delivered [400 words]

Milestones:

(1) Build exploration tools that make it easier for a scientist (or citizen) to explore and visualize the data.

Prior to the Phase I Prize, we had built a system that was capable of fetching air quality data from various sources, injecting them into a database, and serving them out via an Application Programming Interface. With Phase I funding, we built a new user-friendly platform around this system that allows users to download real-time and historical data in multiple ways, including a point-and-click system that allows users to specify pollutants, regions, and time-intervals to download.

The tool: <https://openaq.org/#/locations>

We have also created a tool that enables real-time and recent comparisons between stations in different cities across the world for different pollutants.

The tool:

<https://openaq.org/#/compare/US%20Diplomatic%20Post%3A%20Dhaka/San%20Borja/FRANCONIA?parameter=pm25>

Additionally, we have improved our map: <https://openaq.org/#/map>

This new platform launched on September 1. From that point until November 19th, the number of users who have accessed our online platform has tripled compared to the same period last year.

The platform and all tools are open-source: github.com/openaq

(2) Expand the number of countries from which we aggregate real-time data, from 13 to 30.

We have added 20 more countries, for a total of 33 countries since the start of Phase I. Our dataset has grown by a factor of 7 and currently contains 28 million data points, adding roughly 130,000 measurements per day.

(3) Prototype a tool to allow researchers and government agencies to upload historical air quality data to enable wider sharing of data collected by the scientific community.

We first worked with colleagues who run the SPARTAN Network (spartan-network.weebly.com/) to add in their research-grade data from locations around the world manually. Based on this dataset and an Indian dataset provided by researchers who wish to more broadly share data, we have now created a tool through which one can upload data into our system.

This tool will be launching in the next few weeks at: <https://upload.openaq.org>

(4) Hold a workshop to connect with a local science, technology and journalism community and test the new platform capabilities.

We held a workshop in Delhi, India (Nov 25 & 26) for 40 participants from Delhi and around the country. The workshop generated more interest than we could meet, with more than 70 applications from scientists, software developers, policy organizations, international non-governmental organizations, national-level journalists, and students.

A write up on the workshop materials and results: <https://medium.com/@openaq/delhi-openaq-workshop-info-materials-and-results-2bd74b88bee6#.33opbuoae>

ii. Team Contributions

The contributions of the team members to the development of the prototype [200 words]

Christa Hasenkopf (co-founder OpenAQ/USA) is the overall lead. She strategizes on long-term project sustainability, serves as a community manager, organizes external engagements, and provides scientific input.

Michael Brauer (University of British Columbia/Canada) has provided scientific advice, long-term strategic sustainability advice, and network connections for prototyping the tool for a user to upload research and government grade data.

Joe Flasher (co-founder OpenAQ/USA) is the overall lead on technical aspects of the platform, from the fetching mechanism to database design to building the API. He also serves as a community manager and provides expertise on best practices for open data, open-source projects.

Olaf Veerman is a project lead Development Seed (USA & Portugal) who has overseen the technical development of the new web platform and user interface.

Michael Hannigan (University of Colorado/USA) has provided input on the next potential Phase II on considerations for adding in low-cost sensors to our system.

Sarath Guttikunda (Urban Emissions/India) is a new partner who is a world leader in air quality emissions and source appointment. He is providing ongoing scientific expertise in the expansion of our system and is also a partner in our Delhi Workshop.

iii. Significant Achievements

Any significant achievements or key success metrics you wish to highlight - this might include, for example, numbers of users, key endorsements or engagements with users, new partnerships, external funding, and so forth [200 words]

Platform Usage:

- Our API typically receives 500,000 monthly requests for data and received more than 700,000 in the past 30 days.
- More than 700 research organizations and universities have accessed our online platform.
- The platform has been accessed by users in 1526 cities in 122 countries.

Community Highlights:

- The OpenAQ Community was invited to write a commentary on the importance of open air quality data for the Winter 2016 issue of The Clean Air Journal, a South African-based journal. Twelve scientists in the OpenAQ Community from Canada, Egypt, Ghana, India, Kenya, Mongolia, New Zealand, Rwanda, Spain and USA collaborated on the commentary. Volunteers will translate it into 8 languages. The translations will be posted on: medium.com/@openaq
- Our Slack channel has grown from 40 people to more than 100.

- Hawa Badlo powers their AQI bot with OpenAQ-aggregated data. Hawa Badlo is an Indian grassroots movement fighting air pollution that has reached 6.6 million people digitally.

Funding:

- We continue to receive in-kind support from Amazon Web Services for our platform infrastructure.

- We have received support from Echoing Green through a Fellowship that has enabled Hasenkopf to work full-time on building OpenAQ as an independent organization since July 2016.

Learning Points

Please briefly highlight any key learning points you took from the work that you undertook to develop your prototype [200 words max]

Key learning points our community has gained during Phase I:

(1) Before building our user-friendly interface, we conducted research amongst potential users (representing public health, software development, epidemiology, atmospheric chemistry, and journalism) to understand what our community most wanted, in terms of data exploration and visualization tools. Because we thought the results would be of interest to other open data, open-source communities, we shared what we found:

<https://medium.com/@openaq/a-report-what-does-the-open-air-quality-community-need-in-order-to-be-awesome-73586bf6f45#.w2xzr8ou>

(2) As we have been interacting virtually and in person with various communities around the world, we have been amassing a wish list of what analytical tools community members would like built off of the platform and what enhanced capabilities they would like for the platform itself. The current Community Wish List: <https://medium.com/@openaq/whats-on-the-openaq-community-wish-list-846ef2a78dc0#.4bf117avm>

(3) People are both professionally and personally invested in this platform existing and these data being shared. Recently, we attended an air quality conference, where we gave a presentation on the OpenAQ Platform and Community. Afterwards, an air pollution exposure assessment professor came up to the community member presenting, and asked if she could give them a hug. She said our platform was the first time she would be able to access air quality data in her country.

Case for Phase II Prize

Please make the case for why your prototype should be considered for the Phase II Open Science Prize against the following key criteria [100 words each]:

i. Impact

The current and potential future impact of the tool or service in terms of advancing research and generating health and societal benefit

Embedded in this dataset are innumerable studies, PhD theses and stories. For instance, the data in aggregate provide new opportunities for regional and global air pollution exposure assessments to estimate public health and economic impacts. The dataset also provides previously inaccessible avenues for data critical to epidemiological studies. Besides the dataset itself, fostering the community allows an ecosystem of open-source tools, open data advocacy platforms (e.g. our upcoming OpenAQ Community commentary in the Clean Air Journal), and public engagement techniques to emerge.

ii. Innovation

The degree of innovation associated with the tool or service

We are currently providing the only way in the world to openly and freely access many of these air quality measurements, as well as the only platform through which to access them in a universal format. These features, as well as our entirely open-source platform (github.com/openaq), has enabled a community to form around the platform and an ecosystem of tools, apps, and research to emerge.

iii. Utility

The level of demand and utility associated with the proposed service or tool

See “Significant Achievements” section for current Phase I platform usage.

For Phase II, we receive 1-3 requests each week from scientists and private sector organizations, requesting to put their low-cost sensor data onto our platform. We have also received inquiries from those working with satellite data to provide access to these datasets on our site, as well. Community members have recently expressed interest for workshops in Bangalore and Sarajevo.

For these reasons, it is a clear extension of our Phase I work to expand our platform to incorporate low-cost sensors and satellite data, as well as scale our workshops.

iv. Feasibility & Technical Merit

The feasibility and technical merit of the prototype

Adding the capability to share and visualize low-cost sensor and satellite data will not be trivial, but it is entirely feasible. Low-cost sensor data are typically shared at a much higher time resolution, with lower data quality, and require more metadata for most use-cases. Satellite data will need to be shared in a fundamentally different way, given that the data points will not be discrete sources but rather grids with spatial resolution. Adding in these disparate sources will further the platform's main goal of providing standard access across data inputs.

Development & sustainability plan

Please briefly describe your vision, and any tangible steps you have taken, to develop your prototype into a sustainable tool or service that advances the goals of open science [400 words]

We are deeply and passionately committed to sustaining the existence and growth of the OpenAQ platform and the open-source ecosystem of users and tools developing around it. Our short-term vision is to become the world's go-to air quality database for ground measurements (government-level, research-grade, citizen science low-cost sensors) and satellite measurements. Our long-term vision is to act as a convener of open air quality efforts in the scientific and policy communities.

Technical Sustainability and Scaling

Our current system has nearly evolved out of the prototype state and is now a production-level system (i.e. it already has a heavily cached data-layer capable of meeting global traffic spikes, a multi-region database with automatic failover, and an auto-scaling application layer). Our system is already capable of ingesting the global set of government-level real-time data available, and well-positioned to expand to include the even more data-dense sources from low-cost sensors and satellites.

Community Sustainability and Scaling

We have prototyped workshops that engage multiple sectors working on air quality and health, and we are seeking to expand them in Phase II. We seek to do this by conducting more in-person workshops, and also by creating resources and toolkits for other individuals and entities to use

and adapt in their own communities, independent of our physical presence.

Organizational Sustainability

We are seeking sustainability for our platform through several avenues. Two of the ways are through building partnerships with entities producing low-cost sensor measurements and producing satellite estimates of ground level pollution. Because we already host other government and research-grade ground monitoring data, we serve as an attractive platform for these entities to share their data openly to the public.

We are also seeking to partner with organizations with a vested interest in raising public awareness of air pollution within a local community. Our workshops typically receive national-level media attention and attract entities and individuals who act as key players in engaging the local public on air pollution. Such organizations that may share this interest and be willing to fund and sponsor our work could include: international environmental organizations (non-profits, NGOs) and companies that benefit from raising this awareness (e.g. indoor air filter and mask companies). We are also seeking partnerships with U.S. universities to invest in the long-term sustainability of our platform and include their students in our engagement with local science, policy and media sectors through workshops.

Final comments

Please use the box below to provide any further information you would like to add, that has not been addressed in the questions above [200 words]

We are currently perfectly positioned to expand the capabilities of the OpenAQ platform to incorporate low-cost sensor and satellite data, as well as scale our workshops. We have received funding from Amazon Web Services that will continue to support our platform infrastructure and help us explore new ways to make the data available. Organizational lead Christa Hasenkopf has received funding that allows her to work full time on OpenAQ through July 2018. Phase I funding was instrumental in getting the platform off the ground and has led to a marked increase in usage across the world. Phase II funding would be critical for our team and community to build the new capabilities mentioned above and increase our impact in the field of air quality and health, from science to policy.