esipfed.org/collaborate

UNBOUND FOR AIR QUALITY

Prepared by Earth Science Information Partners (ESIP)
May 2023
NASA Grant 80NSSC21K0365

Recommendation Report: Understanding Needs to Broaden Outside Use of NASA Data for Air Quality (UNBOUND-AQ)

UNBOUND FOR AIR QUALITY

ESIP Team:
Susan Shingledecker
Kelsey Breseman (contractor)
Charley Haley (facilitator)
Allison Mills
Patty Allen
Megan Carter
Annie Burgess

NASA Team:
Katherine Saad
Jenny Bratburd
Jennifer McGinnis
Elizabeth Joyner
Leah Schwizer

UNBOUND-AQ Workshop Participants:
Achieving Community Tasks Successfully - Bridgette Murray, Rachel White-Roy
Brown University - Heejeong Kim, Vivien Chen
Concerned Citizens of Cook County - Shenyue Jia, Kibri Hutchison
Earthjustice - Robyn Winz
Environmental Data and Governance Initiative - Gretchen Gehrke
Environmental Defense Fund - Courtney Grimes, Libby Mohr
FracTracker Alliance - Kyle Ferrar, Matt Kelso
Health Effects Institute - Allison Patton
Hyphae Design Laboratory - Brent Bucknum, Daniel Fleischer
IQAir North America - Amirhosein Mousavi
NY State Department of Environmental Conservation - Jeongran Yun, Ruby (Yuhong) Tian
TCEQ - Chola Regmi
Thriving Earth Exchange, American Geophysical Union - Alisha Saley, Brittany Keyes
Virginia Tech - Wendy Stout

Workshops:
Workshop One - Data Discovery: Friday, March 24, 2023, 1-3 p.m. ET
Workshop Two - Data Exploration: Wednesday, March 29, 2023, 1-3 p.m. ET
Workshop Three - Data Use: Friday, April 7, 2023, 1-3 p.m. ET

TABLE OF CONTENTS

OVERVIEW

What is UNBOUND?

NASA's Understanding Needs to Broaden Outside Use of NASA Data (UNBOUND) Program works to engage and understand the needs of potential users of NASA's wealth of data and information. In this workshop series, UNBOUND-Air Quality (AQ), data practitioners applied to participate in a series of three workshops. The workshops' primary goal was to help NASA understand participant needs to identify barriers to using NASA data in their work related to air quality. Participant organizations were compensated for their participation.

UNBOUND-AQ Workshops

Workshop One - Data Discovery:
Friday, March 24, 2023, 1-3 p.m. ET

Workshop Two - Data Exploration:
Wednesday, March 29, 2023, 1-3 p.m.

Workshop Three - Data Use:
Friday, April 7, 2023, 1-3 p.m.

Understanding Needs to Broaden Outside Use of NASA Data (UNBOUND)

UNBOUND-AQ Objectives

In the spring of 2023, 24 air quality data practitioners from 14 organizations participated in a series of three workshops designed to discover specific user needs to expand the use and usability of NASA data.

Workshop Focuses:

1. **Data discovery:** Can practitioners and users in the air quality field identify NASA data they might use?
2. **Data exploration:** Are air quality practitioners and users able to download, open, manipulate, and understand NASA data?
3. **Data use:** Are participants able to apply NASA data to their air quality work?

During each workshop, participants were broken into small groups and tasked with goals relevant to NASA air quality data and their work. Non-participant observers and participants themselves took notes on the experience, paying particular attention to challenges, frustrations, and ideas for improvement. Non-participant observers were also tasked with answering whether each group accomplished specific objectives for each workshop.

Raw information is included in the Appendices for potential further analysis.

TOP RECOMMENDATIONS FOR NASA EARTH SCIENCE

1. Reduce technical skill requirements for access and use of NASA data.

Make NASA data available, accessible, and understandable to people who do not have strong coding or other technical manipulation abilities. **Top suggestions** would include the use of standard formats such as CSV and GIS-friendly file types; more straightforward downloading through a standardized subsetting tool; easy "download" buttons that do not require the use of command line tools like wget.

2. Reduce hardware and internet requirements for access to and use of NASA earth data.

Reducing physical or hardware constraints increases both *who* is able to access NASA data and *where* NASA data can be accessed. Multiple participants were unable to work with the NASA data due to both hardware constraints on the computers they had access to and the bandwidth available to them. Some participants mentioned only having access to cheaper computer hardware. Others were located in areas of the country where high bandwidth was not available. Still others mentioned that it would be useful to be able to access NASA data from a field lab. **Top suggestions** would be to make subsetting tools more effective and accessible across NASA datasets so that people can download only the data they will use. Several participants independently imagined an ideal search tool that would let you select region and timeframe, then let you filter what's available (with information about file type, size, and organization, as well as data parameters, provenance, and resolution) and only download the subset of the data that fits the specified parameters. (Google Flights search was suggested as a potential model.) Many expected this result from the Worldview interface, and they were surprised to find themselves downloading data outside of the parameters they had specified in the interface.

3. Use standard and familiar formats.

File formats such as CSV and GIS-friendly file types such as .shp (see APPENDIX). If using less familiar formats (e.g. NetCDF, .he5), keep different datasets as similar as possible and create references or examples for how to open and use the data.

TOP RECOMMENDATIONS

4. Make metadata more available, especially before download.

Participants expressed strong need for more information *about* the files, for both discoverability and more targeted downloading, as well as to understand data provenance. SEDAC, for example, was interesting to many participants, but participants noticed that the data in the files had many sources, and expressed frustration that the data cleaning steps were not clearer and more reproducible. Other participants described a desire to be able to have a "details" view on files before download that would list information such as file type, size, number of files in a set, column names, dates of data collection, spatial and temporal resolution, link to datasheet, and any other information that might help them decide whether the file was appropriate for their work before needing to download, open, or manipulate it.

5. Make data available in standard libraries.

Data discoverability was a major challenge for all participants at some point in their process, often even when NASA did have a dataset appropriate to their needs. NASA already publishes large lists of data, guides such as Pathfinders, and other approaches. But not only did participants find these confusing and difficult to navigate (and in many cases discovered broken links), but also had trouble discovering that there was NASA-produced data in the research area at all. NASA sources were rarely top-of-list in web searches for data about various chemical concentrations or other common air quality queries. **Top suggestions:** Data practitioners often use go-to websites that act as clearinghouses for datasets that are appropriate to their skills and toolsets (e.g. ESRI Living Atlas (especially desired), Kaggle, etc.). NASA should consider listing existing datasets, especially common pre-subsetted datasets; if these are unknown, consider installing metrics on the subsetting tools, using Google Keyword Planner to discover common search queries in the air quality space (see APPENDIX). These sites often already have strong search engine optimization and a broad user base that will help to broaden use of data outside of NASA.

Hear it in their own words.

Throughout the report, we highlight direct quotes and feedback from participants — in their own words — shown in maroon text.

BACKGROUND

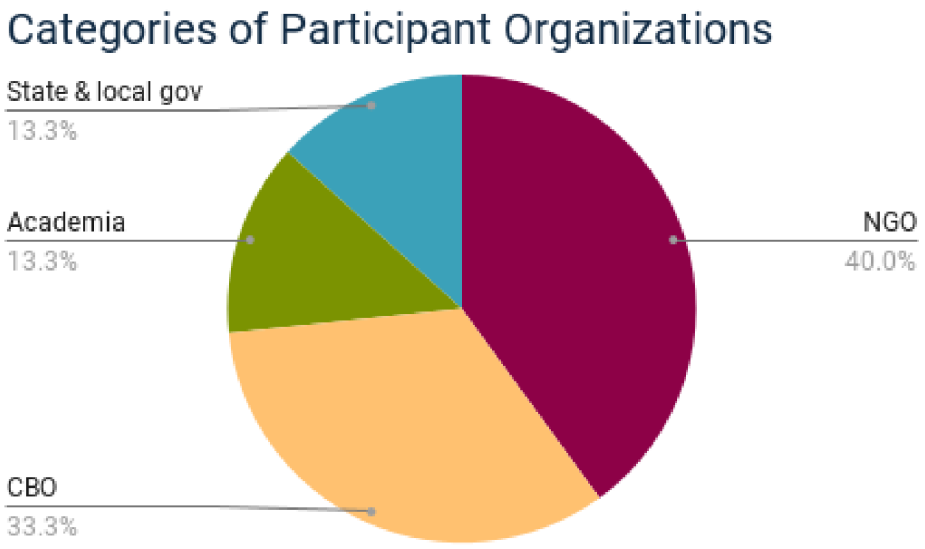
Participant Selection

In order to attract a diverse set of stakeholders, the application to participate in UNBOUND-AQ was advertised through several networks, including the Earth Science Information Partners (ESIP), the Environmental Data and Governance Initiative (EDGI), Coming Clean, and more. This brought in 76 applications from across environmental justice, state and local government, academia, for-profit corporations, nonprofits, and more.

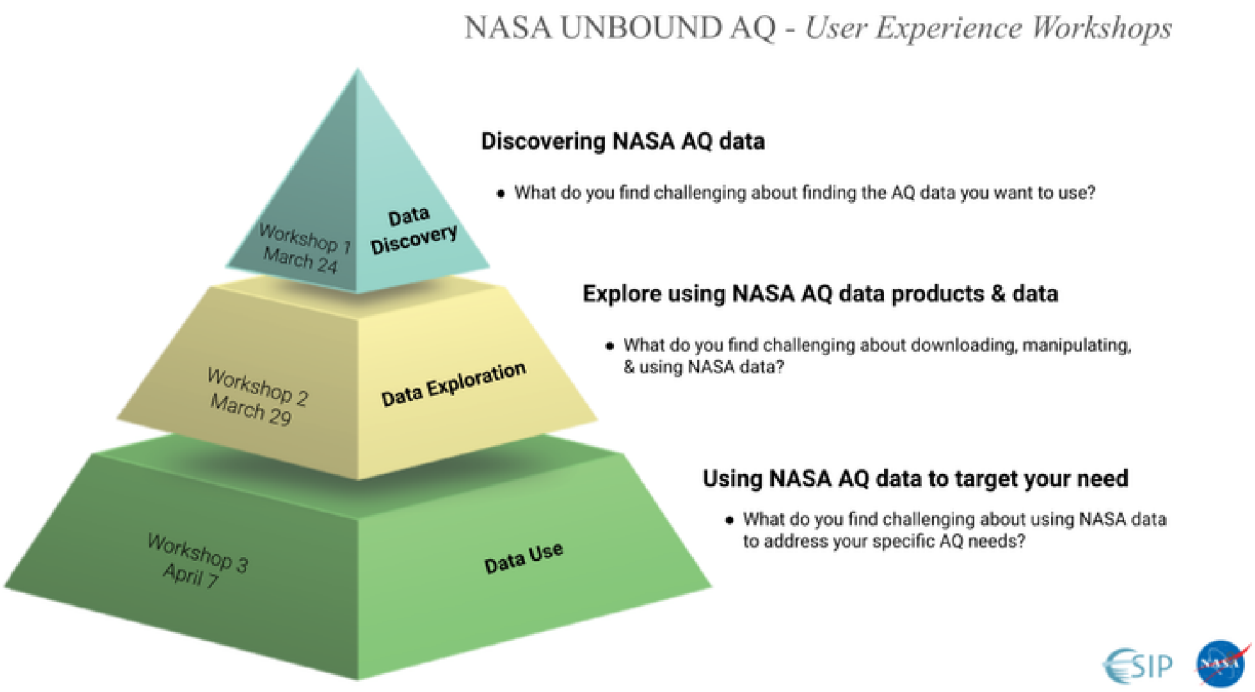
We downselected to 22 organizations based on the following screening criteria:

- Due to the nature of the project's funding, UNBOUND participants from outside the U.S. could not be compensated and were disincluded.
- The goal of the UNBOUND program is to expand the use of NASA data. For this reason, we selected participants who had not previously used NASA data.
- We know from UNBOUND-Environmental Justice workshop series that NASA's data is challenging for users who do not have high data manipulation/technical abilities. For this reason, we selected participants reporting strong technical and data skills (though we would like to note that this significantly reduces the potential audience of NASA data).
- Finally, we selected participants who were able to articulate a clear and current area of research to which they might apply NASA air quality data.

This yielded a reasonably diverse set of participants in terms of organization type, mission, and location, which spanned ten states in the U.S. Because of the different scales of focus, we separated larger non-governmental organizations (NGOs) from smaller, more localized community-based organizations (CBOs).



WORKSHOP SERIES



The workshop series followed three major phases of data analysis: data discovery (identifying and accessing an appropriate dataset), data exploration (getting to understand the data and its limitations), and data use (data cleaning and applying the data to its intended use).

Each of the three workshops followed a similar format. Participants were given an objective, then divided into breakout rooms to work on the objective together as a team, typically in specific roles. Groups debriefed in their breakout rooms, then reported back to the main room. In each case, there were non-participant observers in each breakout room taking notes as participants shared screens and talked aloud about what they were doing and why. Participants were also invited and encouraged to take notes, with at least one participant taking on an observer and note taking role, especially noting challenges they encountered.

Over the course of the workshops, as they grew more comfortable with the format and with NASA tools, participants were invited into increasingly self-defined work. This allowed us to see the specific needs, projects, and toolsets of the individual participants.

WORKSHOP ONE: DATA DISCOVERY

"With NASA data, you have to do the footwork. With other datasets, the search engine usually does the footwork with you."

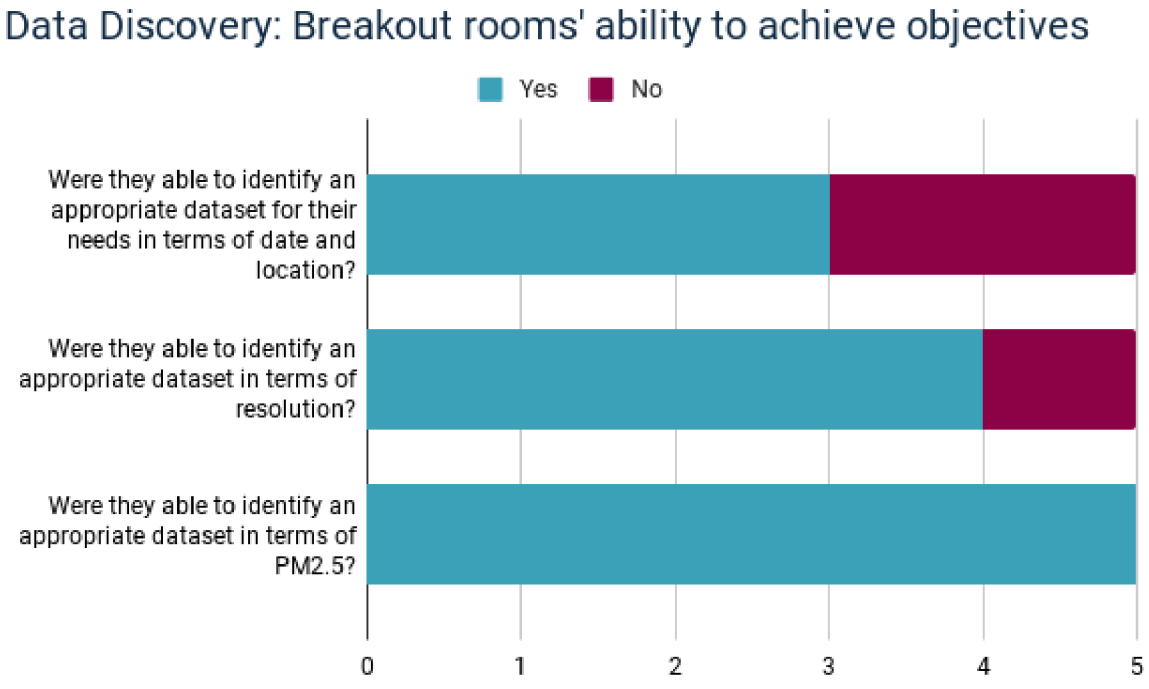
In the first workshop, participants were given a specific objective (shown below) and invited to approach it without resources or prompting (e.g. many started with Google; we did not direct them to specific sites. If the group was still struggling after several minutes, non-participant observers were allowed to suggest Worldview).

Objective

Working together in your roles, find and download: NASA PM 2.5 data in a 1km grid over Texas, with a daily resolution, for all of 2011.

Observer assessment (per breakout room)

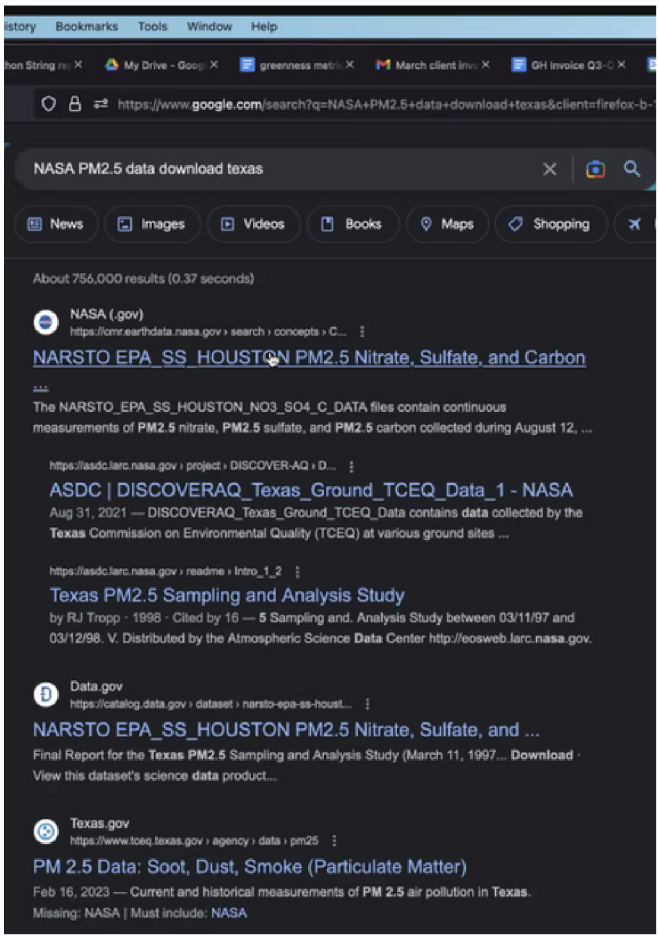
- Were they able to identify an appropriate dataset in terms of PM2.5? 5 yes
- Were they able to identify an appropriate dataset in terms of resolution? 4 yes; 1 no
- Were they able to identify an appropriate dataset for their needs in terms of date and location? 3 yes; 2 no



WORKSHOP ONE: DATA DISCOVERY

Major themes:

- Participants found it difficult to discover NASA data on the internet when not explicitly looking for it. Participants discovered that adding the term “Earth data” increased likelihood of seeing NASA data in search results. Even when specifying “NASA”, it was unclear which NASA links would be the best for finding download links. Participants recommended the use of stronger search engine optimization, tagging, and cross-linking from other sites such as EPA and CDC. “I might not even know that NASA has the data I need. Try to own that journey from the beginning.”
- Filtering was often confused with subsetting. Many participant groups applied temporal and spatial filters to the data in Earth Data Search, and then were surprised to discover that the data did not match the specified parameters and was still suggested for download.



- The time spans of NASA missions was not obvious to participants. They did not know that NASA does missions over specific periods rather than continuous monitoring, so it was frustrating for them to discover that the right data exists, but not in the last X years. “Why does the data only go to 2016?”
- Not all practitioners were familiar with the idea of columnar data, so limitations and processing steps could be clearer.
- Many held the expectation that there should be one place to go to find all NASA data.

“My first response: This is not available. But clearly it is, so let’s go!”

WORKSHOP ONE: DATA DISCOVERY

Notes from Participants

- Unsure where to start to find NASA data. NASA homepage? Many started googling. Some found NASA Data CMR, some found Earth Data Search, some found Worldview.
- Acronyms & terminology confusing, e.g. “AOD” familiar to some but not all practitioners.
- User Agreement for data download turned away participants. They assumed that this meant they were unauthorized to access the data.
- Expected more familiar filetypes, e.g. CSV files
- Felt that more background knowledge was necessary in order to know which files would be relevant to download.
- “Would be good to have code with the data to open the data and give some description of the code.”
- Some users started their data search over several times after following perceived dead ends to try to download data. For example, one group started at earthdata.nasa.gov, navigated to Pathfinders, clicked “Find Data” from earthdata.nasa.gov/topics/atmosphere/air-quality, then restarted at NASA Worldview, which linked them back to their original earthdata.nasa.gov, where they eventually found and used Earthdata Search.
- Several groups had someone with a bit of NASA data familiarity present in the room, and noted that this was pivotal in their ability to efficiently find appropriate datasets. Others in these rooms often noted that they “would never have found it”.
- “Every one of us would have started in a different place” depending on background and how you think about the problem: ESRI vs CSV vs AOD vs PM
- “What’s a granule?”

"NASA has offered tons of useful data products for air quality and beyond. I would very much like to see the publication and tutorials on the best practices to use these data, starting from ground zero, and without much assumptions on users' background. In many times, I often get stuck while I find I need to "generate a script" for batch process, or find there is one piece of information missing that made it difficult to proceed, and this piece of information was missing because it is "taken for granted" by more experienced users. We are so close to make NASA data more useful if we can just make a further step to close this gap."

HOMEWORK: DOWNLOADING DATA

"How much of this will I have to download to know if I'm getting what I want?"

"Three hours later, we'll either have the data we need, or we won't."

After the data discovery exercise in Workshop One, each participant was assigned a specific dataset and asked to download it before the next session, partly due to the large size of data files. Participants received links to datasets where there was a clear "Download" button on the page. The datasets selected for participant download were:

- https://asdc.larc.nasa.gov/project/MAIA-Sim/MAIA_L4_GFPM_VSIM001
- https://sedac.ciesin.columbia.edu/data/set/aqdh-pm2-5-o3-no2-concentrations-zipcode-contiguous-us-2000-2016/data-download
- https://disc.gsfc.nasa.gov/datasets/M2I3NVAER_5.12.4/summary?keywords=M2I3NVAER_5.12.4
- https://asdc.larc.nasa.gov/project/SCOAPE/SCOAPE_Ground_Data_1
- https://disc.gsfc.nasa.gov/datasets/OMAEROe_003/summary?keywords=OMAEROe_v003

Major themes:

- Many of the file download instructions made big assumptions on people being able to use scripting tools like Python and command line.
  - They don't necessarily have or know these (e.g. some participants are more GIS-oriented).
  - People are willing to try them but need more instructions (e.g. "type this in your command line").
  - A simple click-to-download button would make a big difference.
- Download times are incredibly long due to large file sizes.
- Because of the need for wget in some instances and for fast internet in others, many participants were unable to complete the homework (downloading data).
- Login to download was off-putting for participants. "I assumed I wasn't authorized."
- Participants want to know more about the files before downloading: what file type, how big, what area, what resolution
- Clear & intuitive subsetting would make a big difference. In order to reduce the download times, storage requirements, compute, and data cleaning, barriers to NASA data could be reduced with preprocessing in the cloud.

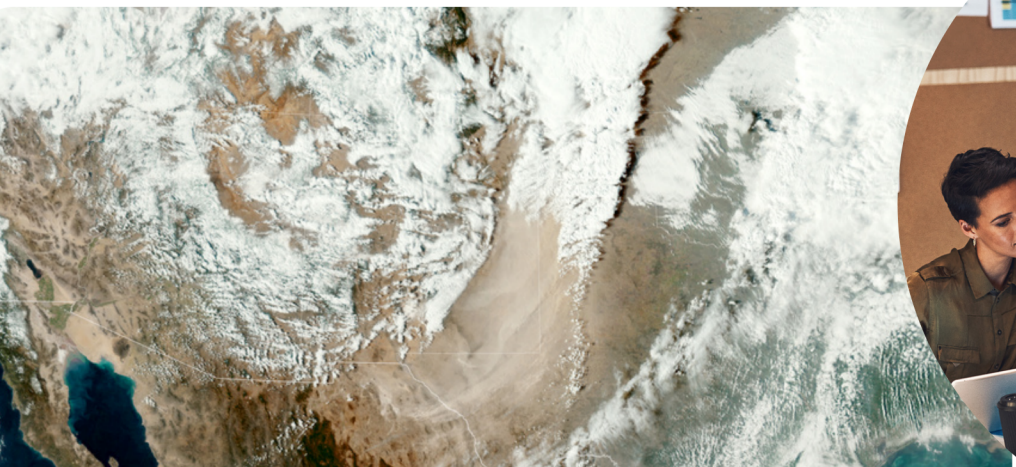
HOMEWORK: DOWNLOADING DATA

Participants were invited to comment on the download process. The following are direct quotes from participants:

“If I click each link, I can download it. Do I need to download all of the links manually? And then the file name is .he5 and if I try to open in my Mac, I can’t open it. Please let me know how I can download the files.” — participant attempting to download: https://disc.gsfc.nasa.gov/datasets/OMAEROe_003/summary?keywords=OMAEROe_v003

“I am having issues with my data download. The first option that the website recommends takes me to a webpage that says "no longer found". The next option takes me to a site and a download link tutorial for Python (which I have never used). I have tried to see if there is a way for me to download the files in R since there is not an auto click option for the full file directly from the site. Also, I tried using all of the search terms from the map function but didn't get any hits on the dataset. Can you advise me on the best way to move forward?” — participant attempting to download: https://asdc.larc.nasa.gov/project/MAIA-Sim/MAIA_L4_GFPM_VSIM001

“I found a wget tool challenging to install with multiple zipping and unzipping procedures. After downloading one of the lists (3.5 GB). I had to open a .nc4 file. I asked chatGPT [sic] to provide a method, and it provided multiple options that were free or through python and MATLAB. I selected Panoplay which was free and easy to install, but it needed a Java runway environment. So, I switched to NCO netCDF Operators. Since this one also was a line-command tool, I changed my mind to install Java on my PC. That was a pain, too! And was not able to install it on a corporate laptop. So, I changed my strategy and told chatGPT to create a script in R to convert .nc4 to CSV and helped me troubleshoot the combining section. It took more than I expected to run the code given the size of the file.” — participant attempting to download: https://disc.gsfc.nasa.gov/datasets/M2I3NVAER_5.12.4/summary?keywords=M2I3NVAER_5.12.4



WORKSHOP TWO: DATA EXPLORATION

"It's a challenge to open these files. Especially if an average American opened this file and they don't have a coding background. It's challenging for us and we have experience with code!"

In the second workshop, participants were grouped so that all participants in the room had been asked to download the same data. They were invited to suggest software and tools that they were comfortable with and had on their computers to open and explore the downloaded data.

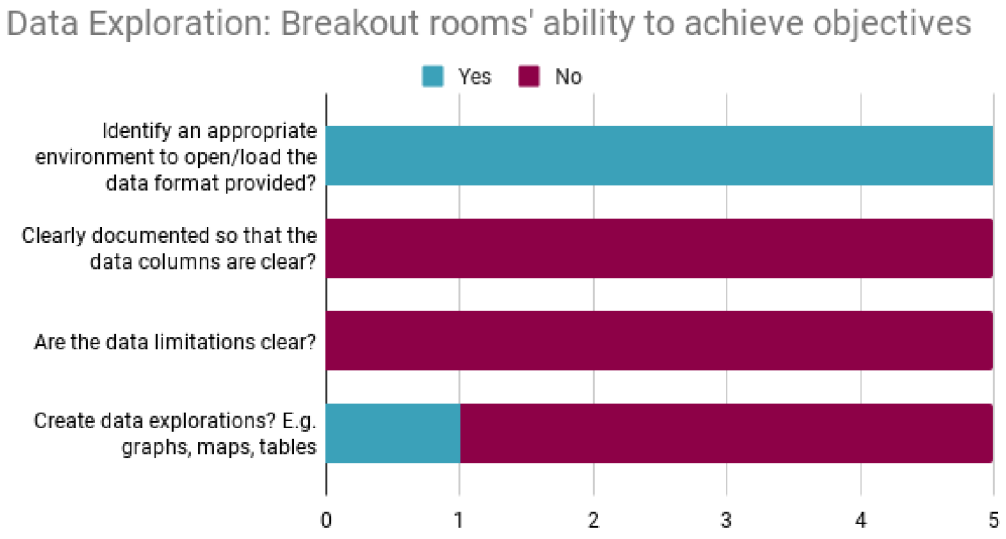
Objective

Work together to see what you can do with the specific data you downloaded.

- *What is the data format? Can you open it? What do you use to open it?*
- *If there are columns of data, what are they? Can you pull out any important characteristics?*
- *Are there visualizations you can create? If so, please share screenshots!*
- *What is this dataset useful for and what are the limitations?*

Observer assessment (per breakout room)

- Is the group able to identify an appropriate environment to open/load the data format provided? 5 yes
- Is the data comprehensible so that the data columns are clear? 5 no
- Are the data limitations clear? 5 no
- Is the group able to generate some appropriate data explorations? E.g. graphs, maps, tables 1 yes; 4 no



WORKSHOP TWO: DATA EXPLORATION

Major themes:

- Several participants' computers crashed trying to process large files.
- Many file formats available are unfamiliar or inconvenient. It would be helpful to have the data pre-converted into several available formats, especially .shp and .csv formats (see APPENDIX). In some cases, participants used scripts to convert the files into familiar formats, which was time consuming.
- If other file types cannot be made available, examples of usage would be helpful for the less common file formats.
- In particular, a how-to guide for a smooth transition from netCDF format with georeference information to geospatial software and coding tools would be helpful—participants were not able to correctly project the data onto georeferencing coordinates.
- Particularly for GIS users, it is important to include metadata as an attribute so that files can be merged easily. Participants specifically mentioned a desire for column headers Year, FIPS, and reference to shapefiles.
- It would be very helpful to have shapefiles accompany location-based data.

"Most of the group felt like if they had to for work, they could see the pathway to getting the data, loading it, understanding its source/limits, transforming it to a more easily manipulated format, and then visualizing it. The barriers were that, due to the data's format/structure (and lack of readily available metadata), that it would be time consuming and frustrating to process or visualize, and it would be a very inefficient process overall or almost impossible if not using a tool like R or Python."

WORKSHOP THREE: DATA USE

"I don't really want to learn a new subsetting tool just to download the right data."

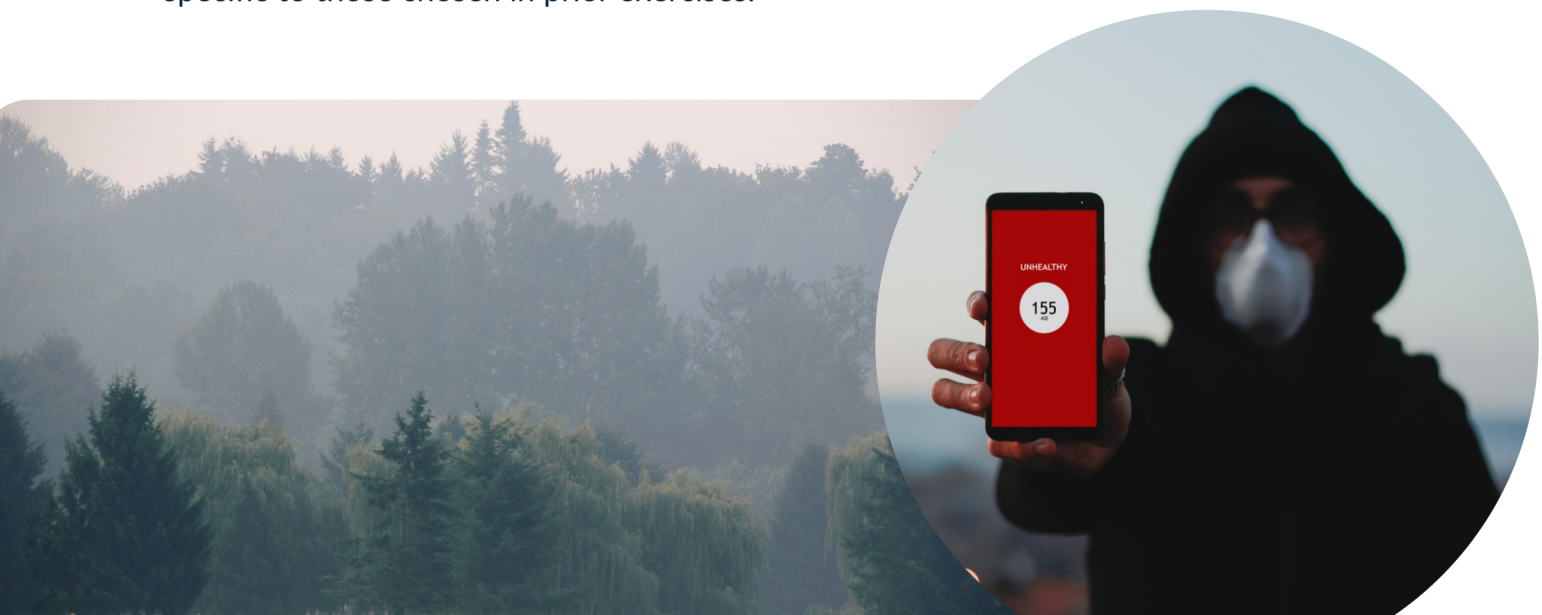
"If it wasn't readily available in the formats I already used, to try to dig down in the other formats took a lot of time."

Workshop Three was very different in format from the prior two workshops. Between Workshops Two and Three, participants were given a list of datasets that might be applicable to their specific work (see APPENDIX), and invited to use the intervening two weeks to apply those datasets to their own projects. In the workshop itself, they were given space to continue working on these projects with NASA subject matter experts on hand to help them progress.

Objective
Individually, apply NASA datasets to your active areas of work.

Observer assessment
N/A, individual work

Major themes
In the first two workshops, coordinators set the objectives. This workshop was designed to test whether participants would encounter novel challenges when pursuing their own work with NASA data. This workshop confirmed and reinforced existing themes from the prior two workshops. Overall, participants reported the same and similar barriers as in prior workshops, which demonstrates that these barriers are common across many datasets, not specific to those chosen in prior exercises.



WORKSHOP THREE: DATA USE

Notes from participants:

- "I would be really partial to a QGIS plugin. As soon as you're exiting the ESRI universe, QGIS plugins become such a valuable resource."
- Code snippets (especially in multiple languages) showing example use of each dataset would be extremely useful. Some participants were able to use ChatGPT to assist (though often the code had to be corrected in order to run).
- Several participants found broken links and 401 errors navigating GES DISC and Earth Data Search
- "I'm just a GIS person and I'm not a coder. Downloading the data from the NASA website was so hard for me. … if it wasn't in a GEOTIFF, or a geodatabase, or a shape file, for me it was so much harder… I had to dig down into layers to find what I understood, because I don't understand all the other formats. … If it wasn't readily available in the formats I already used, to try to dig down in the other formats took a lot of time. You have to try four or five different pathways." "Some datasets …I couldn't even find a format that I recognized or understood."
- "I'm excited to dig into the Discover-AQ data, and the descriptions on the Atmospheric Science Data Center (e.g. this page) are helpful. However, when I click the link to 'Get Dataset' I am taken to an earthdata.nasa.gov page where I have 197 options with titles like 'DISCOVERAQ-LACO-PINEHP-P12_P38_20130116_R1.ict' and no descriptions, and I have no idea what this means or which I should download. I also see an option to "add granule" and I have no idea what that means."

WORKSHOP THREE: DATA USE

Specific data and tools participants desired but were unable to locate or use:

- Real time or near-real time data
- CSDAP - https://www.earthdata.nasa.gov/esds/csda commercial smallsat data required authorization
- Currently, only NASA-affiliated researchers can use NASA's access to commercial satellite data, it would be helpful to extend this to community groups that might not otherwise have the resources to access
- Airplane flyover data about methane
- PM at different resolutions
- Cloud-based processing tools

Participants found the following particularly frustrating:

- Lag in data availability, sometimes by several years
- Low spatial and temporal resolution
  - Methane data was particularly mentioned as frustrating for low spatial resolution
- Difficulty finding sufficient information about the data's units, sources, and limitations
- TROPOMI changed its spatial resolution, so data from different years cannot be compared
- File descriptions and other navigational elements in Earth Data Search were unintuitive
- File names are unintuitive/difficult to follow/understand. For example, products with different resolutions have almost the same names.
- Coming from different backgrounds and wanting to research the same data - code vs GIS. Our approaches would be totally different.

Datasets and tools noted as useful by participants:

- Google Earth Engine
- TROPOMI/OMI, particularly for O3
- MERRA2
- MAIA
- SEDAC
- ArcGIS
- Air Quality Data Pathfinder
- HAQAST draft flowchart for identifying datasets
- TEMPO
- Table of suggested datasets (see APPENDIX)

Participants found the following particularly helpful:

- ASDC website provides more information prior to download than Earth Data Search
- GDAL for transforming file between different types
- 1x1km scale in SEDAC
- Anything that could be directly accessed from R, Python, QGIS, ArcGIS

PARTICIPANT REFLECTIONS

"NASA has an enormous amount of amazingly useful data that communities concerned about air quality can use in myriad ways, but being aware of, locating, obtaining, and analyzing this data is still challenging for most people, even for those with advanced technical backgrounds in remote sensing. Workshops like this can both help users understand the scope of NASA's data offering, and help NASA address the complex challenges of providing such a large and diverse collection of data to a large and diverse set of users."

At the end of the final workshop, participants were given time to complete a survey that asked for both qualitative and quantitative reflection across the three workshops.

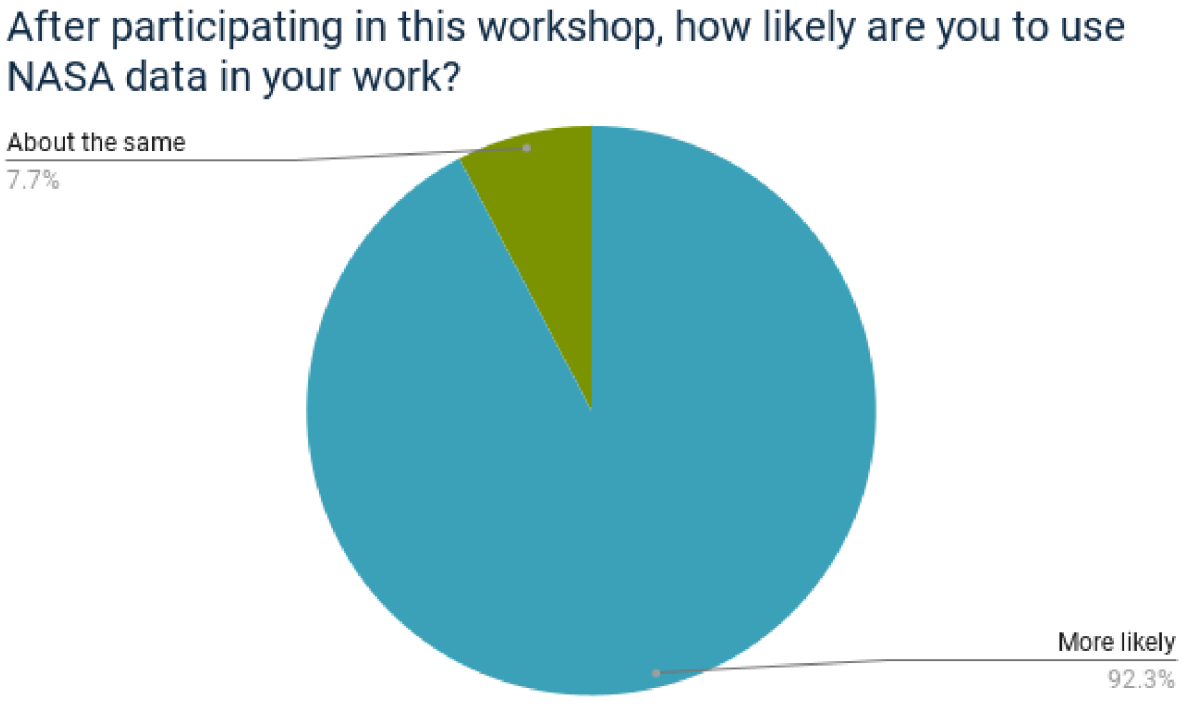
Overall, participants responded positively to learning about the available NASA data and were keen to use them further in their work.
The themes emphasized above were reinforced: strong interest in the data available, desire to have more discoverability and guides/guidance on data use (both from a data provenance perspective and from a technical perspective), and frustration around accessibility issues — especially download sizes and technical skill requirements.

"Part of what was so helpful about this workshop was learning the different kinds of data that exist and how to access them. I honestly didn't know the kinds of air quality data available, and now that I do know, I will try to use them (though I've run into some notable accessibility problems over the course of this workshop as I use a computer that is a few years old and on the cheaper end of the spectrum)."

PARTICIPANT REFLECTIONS

In each workshop, participants expressed appreciation for doing this work in a group setting and sharing back, because they felt vulnerable about how difficult it was for them to accomplish the objectives. Many said that they felt relieved to hear that “it wasn’t just me” struggling to identify, select, download, and use NASA data. Participants felt challenged throughout the workshop series, and many expressed the idea that this data was “not for them” but rather for someone with specific other knowledge. For example, someone with more coding knowledge, someone more familiar with satellite data, or someone with an existing login to NASA’s data services. All participants were technically competent and expressed interest in using the datasets they found in their teams, but many noted that they would not have continued to attempt to access this data due to the barriers presented. Working through these challenges together helped them to feel less alienated by the difficulties they faced in attempting to access the data.

“I really appreciated getting to participate in this workshop. I'm definitely in the category of people who would love to use some of the kinds of data NASA has, but without coding expertise, I've been remiss at even trying. Now I know there are some things I can do with my background, and I can advocate for things to be more accessible to people like me.”



APPENDICES

Appendix: Research Questions of Air Quality Practitioners

Raw responses from participants (from survey):
Please state your current research question(s) for satellite data. What question(s) do you hope to answer with satellite data? What specific gaps are you hoping satellite data can fill?

We would like to compare NASA PM 2.5 satellite with the clarity low cost network in Harris County.

We hope to obtain a time series of satellite data for a selected area of interest and identify the past and current hot spots of air pollution. Our area of interest is relatively small and we are not sure if the current resolution of NASA satellite products on air pollution is good enough. This analysis will be incorporated in our final mapping tool to inform residents and other relevant stakeholders.

I look at hourly concentrations of criteria pollutants (primarily ozone) at a national scale. There are large swaths of land where ozone monitors don't exist, as monitors are concentrated mostly in population centers. Ozone affects trees and other plant life as well as people, so satellite data could help us get a handle on what concentrations of ozone are throughout the year/ozone season in areas that don't have monitors. This data would help us advocate to EPA for stronger secondary (aka affecting plant life) ozone standards.
One thing we study is the disjuncture between environmental protection as a function of addressing emissions as opposed to addressing exposures. Hypothetically one should lead to the other, but the gap between the two is wide. We are interested in what can be done to identify and measure multiple pollutants at scale simultaneously, and at different altitudes, and to understand how the atmospheric chemistry of airborne pollutants is affected by climate change and how that might influence measurements. We hope satellite data can support our understanding of these issues.

I study air toxics that affect communities living near industrial facilities. Ground monitoring of VOCs is often sparse or non-existent around these facilities, making it difficult to identify hot spots, assess community exposure and health impacts, and estimate emissions. We are interested in the potential for satellite data (specifically HCHO columns) to help us identify major industrial sources, detect major emissions events, understand industrial emission trends over time and space, constrain VOC emissions, and understand health impacts.
I study the contribution of oil and gas extraction operations to local and regional air quality degradation.

APPENDICES

(con't) Raw responses from participants (from survey):
Please state your current research question(s) for satellite data. What question(s) do you hope to answer with satellite data? What specific gaps are you hoping satellite data can fill?

We fund research on ground-level air quality (PM2.5, NO2, ozone, and many other pollutants) and health at local and global scales to inform policy. Satellite data provides greater spatial coverage than ground monitors.
We investigate human health outcomes mediated by air quality in an urban context specifically with intention to propose urban design interventions. We use satellite data to build our models of existing conditions. We are researching and developing ways to infer leaf area from orthographic imagery in visible light as well as other spectra.
I study how the built environment can be changed to improve human health. This includes traffic sourced air pollution impacts on near-road communities. We hope satellite data can provide insights into how built environment parameters affect both air pollution and the human impact of it, as well as how built environment parameters change over time.
For the Breathe Providence project, we would be interested in comparing satellite air pollution measurements (PM2.5, CO, NOx, O3, CO2) to our hyperlocal measurements to fill in spatial gaps in monitoring and provide feedback on our network-based sensor calibrations. This could also help us determine if there are spatial gaps in the city where it would be beneficial for us to locate an additional continuous stationary monitor.
I study spatial and temporal variation of PM2.5, PM10 and gaseous pollutant in the global scale with hourly resolution. We hope that satellite data could help us refine our hourly estimates/forecasts.
We study surface NO2 and HCHO that affect surface O3. We hope satellite data can provide spatial distribution of NO2 and HCHO, and interannual variability of NO2 and HCHO. 1) There are lots of cloud in daily data. How to fill these GAP in the data? 2) Satellite data derived from different version of the algorithms should be co-calibrated before they are used to study the interannual variability. How could I know whether TROPOMI NO2 and HCHO data from 2017 to current are co-calibrated (i.e., TROPOMI data spatial resolution was changed in 2019, the algorithm was updated too as we remembered)? 3) Is there a simple way to derive surface NO2 and HCHO data based on TROPOMI column NO2 and HCHO data?

We study surface PM2.5. We hope to find the satellite data that could relate to surface PM2.5 at spatial resolution of 4km or 1km. Is there such kind of satellite data?

APPENDICES

(con't) Raw responses from participants (from survey):
Please state your current research question(s) for satellite data. What question(s) do you hope to answer with satellite data? What specific gaps are you hoping satellite data can fill?

When the fire plume is transported to an area, which satellite data could be used to quantify the present of the smoke plume over this area?
I want to map satellite measurement of criteria pollutants (eg Ozone, NO2, SO2, HCHO ect) from selected area with nearby ground-based monitoring data. I know they are not equivalent in magnitude/representation but can have some sort of correlation. Understanding this relation may insight about air pollution in other areas that do not have ground-based monitors. This skill and knowledge could be very useful in future NASA data from TEMPO too.

We are studying particulate matter concentrations in Beloit, WI in neighborhoods whose residents may have increased health vulnerability to power plant emissions. The wind direction between the plant and Beloit--which sits just to the south--in combination with particulate matter emission trajectories, are presently unknown in our communities. We hope that satellite data may be able to provide a better idea of the potential impacts of meteorological factors on the dispersal and residence times of particulate matter emissions. We hope that these data sources become more accessible to local health care providers and individual community members so that we can empower all people to be able to have the tools to advocate for their health.
I study impacts of pollutants on air quality that affects Hampton Roads Virginia . We hope satellite observations and ground-based air airborne measurement data can provide build a more complete picture of urban environments and how people locally experience these environments insights.

APPENDICES

Appendix: Datasets Suggested as Relevant to Participants

Workshop coordinators sent the list of workshop participants' research questions (APPENDIX) to various contacts within NASA and requested that they suggest potentially relevant NASA datasets. Workshop coordinators then compiled a list of suggested datasets for each participating organization, which they were given access to between workshops two and three so that they could explore on their own.

The following is a list of suggested datasets along with the number of participant organizations it was suggested to, which served as a way of suggesting the most potentially relevant datasets for air quality data practitioners.

Dataset	How many participant organizations was this potentially relevant to?
SEDAC: PM 2.5, O3, NO2; 2000-2016; 1km and ZIP	8
MAIA L4: PM 2.5; 2018; 1km	6
MISR: PM 2.5; 2000-2021; 275m, 1.1km	7
DSCOVR EPIC: AOD, aerosol properties; 10km CAL_LID_L3_Tropospheric_Apro_CloudFree-Standard-V4-21 CAL_LID_L3_Tropospheric_Apro_AllSky-Standard-V4-21 CAL_LID_L3_Tropospheric_Apro_CloudySkyOpaque-Standard-V4-21 CAL_LID_L3_Tropospheric_Apro_CloudySkyTransparent-Standard-V4-21	3

APPENDICES

Dataset	How many participant organizations was this potentially relevant to?
MERRA2: Hourly PM 2.5 Monthly	4
AERONET: Ground-based AOD	7
M2I3NVAER: 3-hourly instantaneous PM 10	2
More AOD: OMI/Aura OMAEROe_v003, Level 3 daily global gridded, 0.25x0.25 degree (2004.10-present) OMAERO_v003, Level 2 swath, 13x24 km, 2004.10-present TROPOMI/Sentinel-5P S5P_L2__AER_AI_v1, Level 2 swath, 7x3.5 km, 2018.06 - present GSFC LAADS DAAC: Deep Blue Aerosol L2 6-minute Swath Data at 6 km; Daily AOD data from MODIS, VIIRS SNPP, VIIRS N20 at 1*1 degree grid; Monthly AOD at 1*1 degree; Dark Target Aerosol AOD: 6-minute Swath at 6 km; GEO-LEO Dark Target Aerosol AOD Data Swath at 6 km MODIS Atmosphere Profiles: 5-minute L2 Swath	8

APPENDICES

Dataset	How many participant organizations was this potentially relevant to?
ACTIVATE: trace gas	4
SCOAPE: NO2, O3, CH4, CO2, CO, VOC, AOD, black carbon, meteorological variables; offshore oil exploration	4
Fire Influence on Air Quality	1
Comparing satellite to ground measure ments of O3, NO2, CH20	6
AJAX: CA, NV, and coastal Pacific measurements of O2, HCHO, CO2, CH4, 3D wind	4
NH3 (Ammonia): AIRS: AIRSAC3MNH3 TROPESS: TRPSDL2NH3AIRSFS, TRPSDL2NH3CRSAUS, TRPSDL2NH3CRSFS, TRPSDL2NH3CRSWCFHI , TRPSDL2NH3CRSWCF, TRPSYL2NH3CRS1FS	3

APPENDICES

Dataset	How many participant organizations was this potentially relevant to?
HCHO (Formaldehyde): OMI/Aura OMHCHOG_v003 OMHCHO_v003 TROPOMI/Sentinel-5P S5P_L2__HCHO___1 S5P_L2__HCHO_HiR_1 S5P_L2__HCHO_HiR_2	6
CO (Carbon monoxide) AIRS/Aqua AIRS3STD_v7.0 AIRS2RET_v7.0 MLS/Aura ML2CO_v004 TROPOMI/Sentinel-5 S5P_L2__CO___v1 S5P_L2__CO_____HiR 1 S5P_L2__CO_____HiR 2 MERRA-2 M2T1NXCHM_5.12.4 M2TMNXCHM_5.12.4 MOPITT- CO	5

APPENDICES

Appendix: Datasets Suggested as Relevant to Participants

Dataset	How many participant organizations was this potentially relevant to?
NO2 (Nitrogen dioxide): OMI OMNO2d_v003 OMNO2_v003 TROPOMI S5P_L2__NO2___1 S5P_L2__NO2____HiR 1 S5P_L2__NO2____HiR 2	7
O3 (Ozone): OMI/Aura: OMDOAO3_003, OMDOAO3Z_003, OMDOAO3G_003, OMDOAO3e_003, OMO3PR_003, OMTO3_003, OMTO3G_003, OMTO3e_003, OMTO3d_003 OMPS /Suomi-NPP OMPS_NPP_NMTO3_L2_2 OMPS_NPP_NMTO3_L3_Daily_2 OMPS_NPP_NPBUVO3_L2_2 OMPS_NPP_LP_L2_O3_DAILY_2 TROPOMI/Sentinel-5P S5P_L2__O3_TOT_1, S5P_L2__O3_TOT_HiR_1, LS5P_L2__O3_TOT_HiR_2 Lake Michigan Ozone Study, Long Island Sound Ozone Chesapeake Ozone TRACER: O3 over East Houston	5

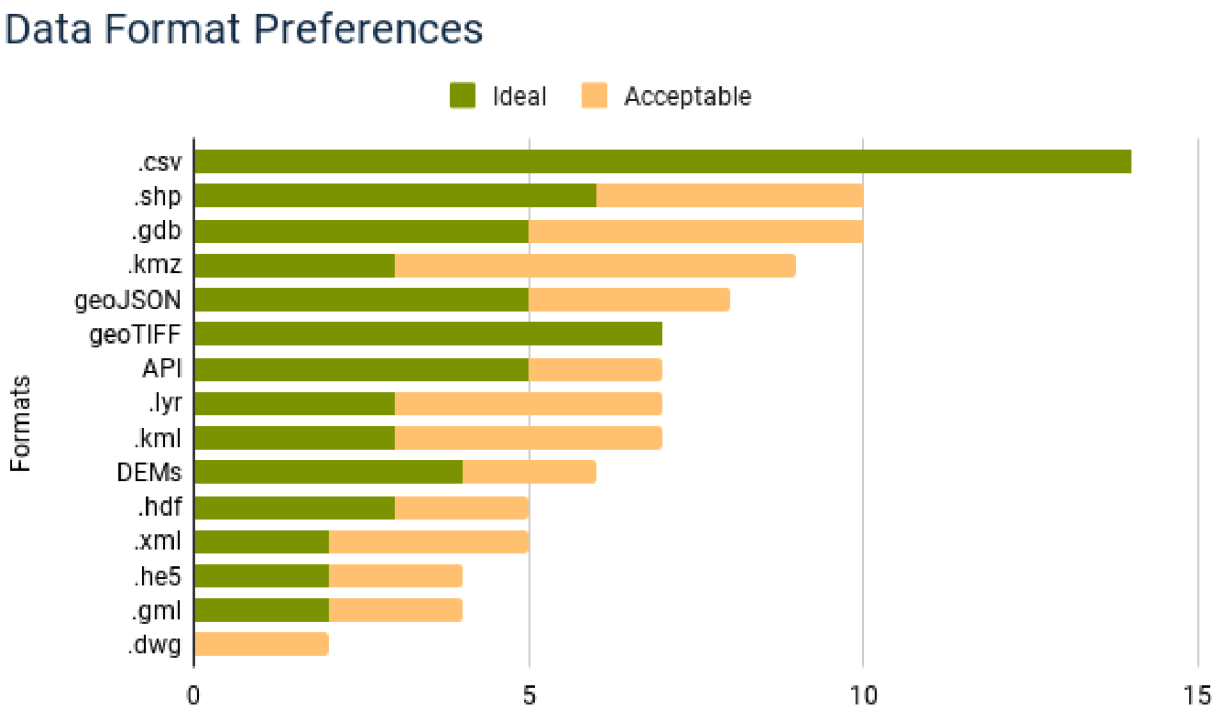
APPENDICES

Appendix: Familiarity of Tools and Formats to Air Quality Practitioners

Survey Data

Before the workshops, participants were asked to complete a survey ranking different tools and data formats by their familiarity and ability to use them in their work.

(note: “Ideal” is here considered to be mutually exclusive with “Acceptable.")

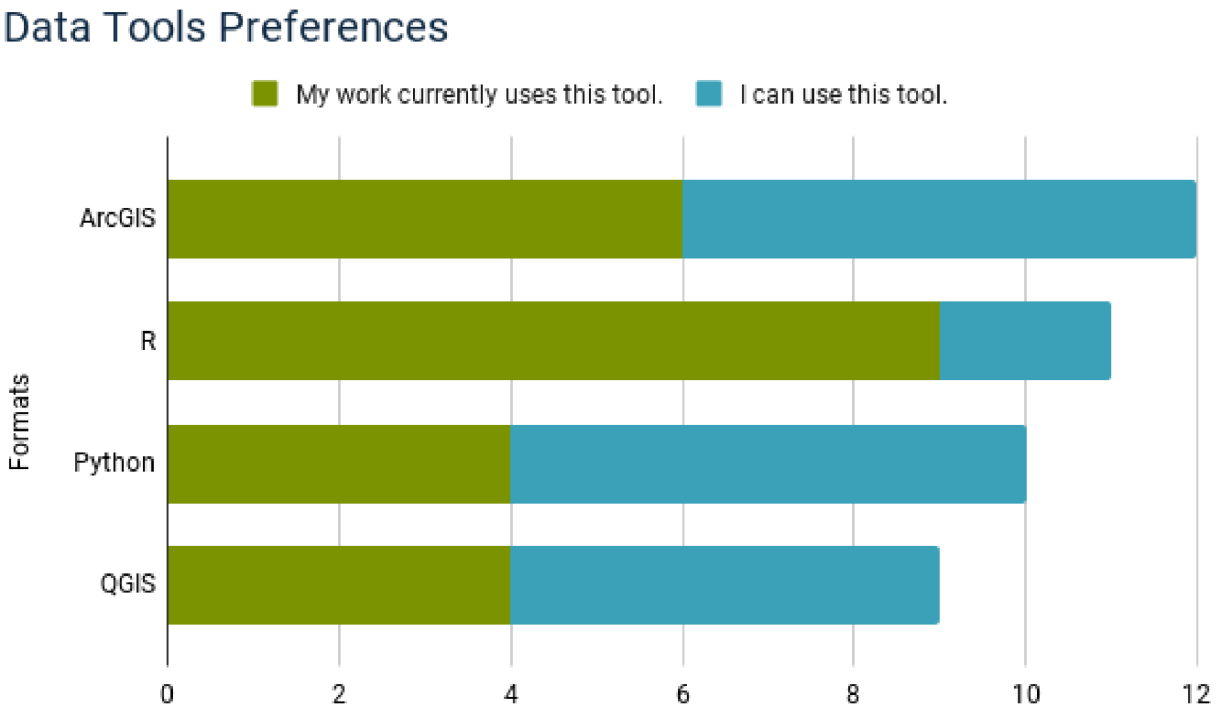


Other formats suggested by participants were:

- NetCDF
- .xls
- Geopackage
- WMS
- Web-ready geoTIFF

File formats that users particularly struggled with in workshops included: .RDS, .he5, NetCDF (NetCDF was familiar to some users and challenging for others.)

APPENDICES



(note: “I can use this tool” is here considered to be included in “My work currently uses this tool”. For example, if a participant marked both, that entry was counted as “My work currently uses this tool” but not also counted as “I can use this tool”.)

Other tools participants suggested in the survey:

- Excel
- Google Earth Engine
- Google BigQuery
- SAS
- Shell scripts
- F90
- NCL
- Orfeo Toolbox
- Whitebox
- Grass
- GDAL

APPENDICES

Participant-Suggested Software

In the Workshop 2 breakouts, participants were invited to volunteer software they were familiar with and had on their computers for potential use in opening and analyzing datasets.

Software	Breakout Rooms Where Suggested
R	5
ArcGIS	4
Python	3
WGET	2
Excel	2
Panoply	2
QGIS	2
Paraview	1
ESRI	1
MATLAB	1
Stata	1
F90	1
NCL	1
ArcPro	1

APPENDICES

Appendix: Participants' Top Recommendations for NASA

Raw participant responses, unordered, from the participant reflection survey:

What is the top priority from your perspective that NASA could do to make you more likely to use NASA data in your work?

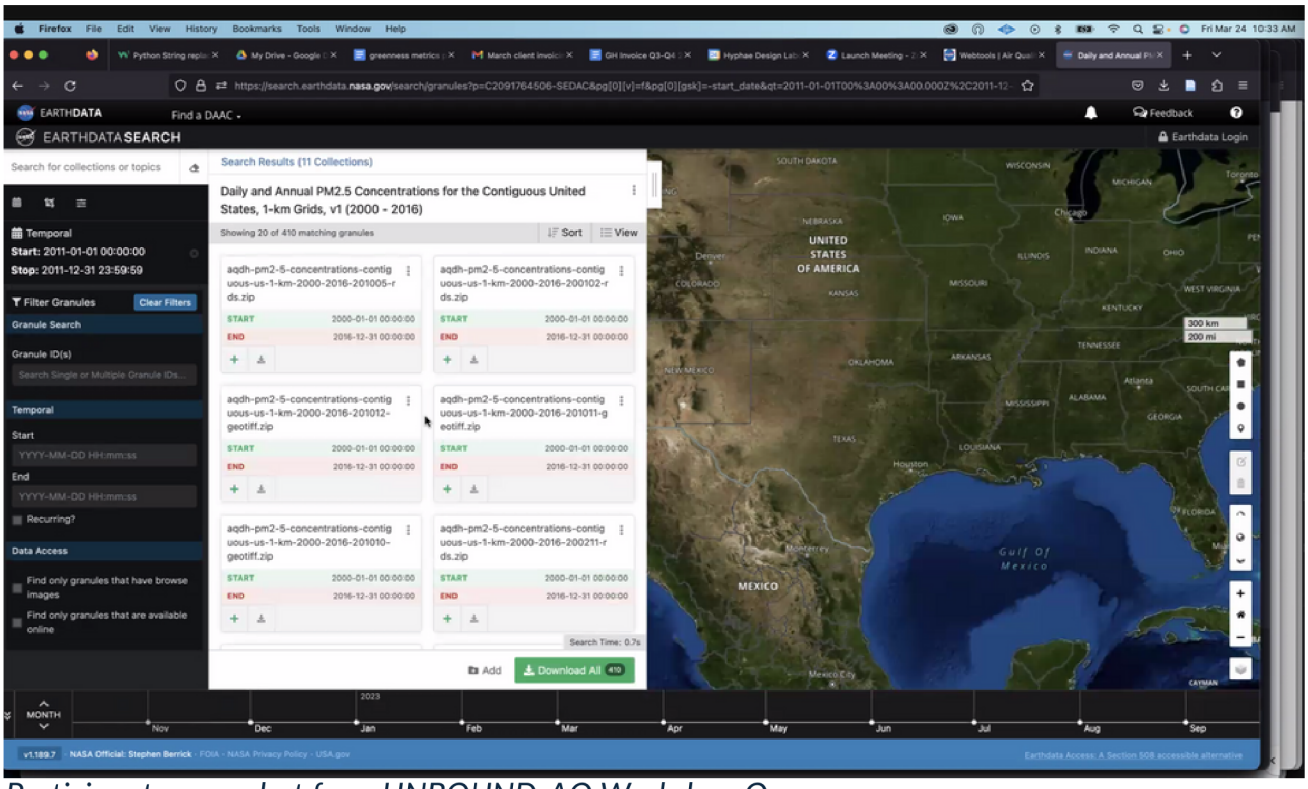
- Cloud data capabilities, air quality field -familiar data formats (.shp, .csv)
- Data Accessibility and Data Formats and Standards
- More clear guidelines on best practices to use NASA products
- Include more data in ESRI Living Atlas
- Detailed metadata and ability to use it efficiently without a high-powered computer.
- Getting higher spatial resolution - either by launching more sophisticated sattellites or buying data from Maxar/Planet/Airbus and making it available to environmental justice communities and economically disadvantaged non-profits
- Subsetting data through portal on website that allows direct downloads of data files
- Easy data format, ability to sub setting data in term of spatial and temporal way.
- making it more organized and accessible to a non-technical user
- More packages and tutorials for accessing, processing, visualizing, and analyzing NASA data in R, Python, and/or Google Earth Engine.
- finer spatial resolution (1km or less) and hourly observations of PM2.5, O3, NO2 and VOCs with QA information
- The flow chart was great, but an overview of what data is accessible, as a type of database guidance doc would be super helpful. Not knowing what is available from NASA is a major hurdle. Also a reference guidance on file types and how to get started on types commonly usaed by nasa and how to get them into r, python, or looking like a csv would be helpful.
- There are two primary things that would make me much more likely to use NASA data in my work. The first is making it standard to be able to obtain sub-sets of data such that I can actually download them onto my computer; customizable sub-sets would be ideal, but even pre-set sub-sets would help. The second is consistently making CSV a file format option. Most statistical software can import data from CSVs, and especially for people like myself who are physical scientists and policy advocates but not data scientists, creating files in the most common lay-person-adjacent formats like CSVs would be very helpful.

APPENDICES

Appendix: Feedback on Specific Tools

Earth Data Search

- "NASA needs a page of instructions."
- Helpful flow from air quality to particulates, intuitive. In general, participants were able to find some relevant data quickly.
- Difficult to choose between results in filtered list: "There are a lot of options here." Satellite & mission names are not helpful at this point, more descriptive info e.g. column names, resolution, file types would be more effective for filtering and downselecting to appropriate data. Especially, it would be very helpful to include filters on file type.
- Expected to be able to see data previews on map before needing to download. Felt compelled to "Download All" because it's unclear what's in the specific files.
- Dataset & data file names not intuitive to understand.
- Confusing to be kicked out to another site to download the data
- "The search is really not intuitive. When you enter search terms, it's really hard to tell if it's applying."



Participant screenshot from UNBOUND-AQ Workshop One.

APPENDICES

GES DISC subsetting tool

- Subset by variable option not immediately apparent. Not clear if the subset capability is for the variable, the region or the time
- Would be helpful for lower resolution data to have the option to select a city or region, rather than having to enter the coordinates (the GES DISC subsetter didn't provide a map to draw a bounding box, so in order to subset for e.g. a certain city, participants had to look up the coordinate points and enter them manually). Confusion about what shapes were possible e.g. bounding box vs. polygon
- Very little help/instruction is provided, more would make a big difference

Worldview

- "If the Worldview didn't show up in the Google search, maybe it should be cross-linked from other websites more."
- Took some time to understand how to use the filters. "Did that filter apply?" - didn't know whether to look for an "apply" button or just expect it to apply on selection. Confusion related to idea that Worldview subsets the data to the selected parameters.
    - Participant notes: "Huge pain point was the filter system on Worldview. We landed on "Daily Annual PM2.5 Concentrations for the Contiguous U.S., 1-km Grids, v1 (2000-2016), and tried to filter it using the Temporal filter (for granules) on the left (Jan. 2011 to Dec. 2011). "Did that filter apply?" Still the same number of results showed (appeared to be 410), though some text at the top to the left of the filter said "Search Results (67 Collections)", but clicking that link took us back to the different collections of data."
- Drop pin vs draw polygon confusing (expected dropping pin to select e.g. the state it was in, not just the very specific area.
- Spent some time drawing polygon, but it would be helpful to have pre-selectable shapes e.g. countries, states, ZIP codes, census blocks rather than needing to hand-draw these.
- Confused when results list gives data irrelevant to specified parameters. Mental model is that Worldview does subsetting based on specified time & location.
- Difficult to choose between results in filtered list: "There are a lot of options here." Satellite & mission names are not helpful at this point, more descriptive info e.g. column names, resolution, file types helps filter and downselect to appropriate data.
- Option overwhelm – no visual clarity over where we are supposed to go, cool icons but not with explanations. Lots of assumptions of prior knowledge.
- "Interfaces that are helpful let me upload a shapefile, like a boundary of Texas, then find data within that shape."
- Many checkboxes "screams jargon - my confidence plummeted, now I have to understand what each of these boxes mean."

APPENDICES

Appendix: Feedback on Specific Datasets

MAIA

- Projection issues—is this correctly georeferenced? Opened in QGIS and Panoply, tried reprojecting coordinates, did not succeed.
- NetCDF data is not very convenient for geospatial tools
- It would be helpful to be able to download this in a more point-and-click method
- AWS access is helpful, it would be even better to be able to access via different cloud vendors

SEDAC

Accessing and downloading files

- File sizes are too large to download; many participants were not able to download over their internet. Other users succeeded but were dismayed by download times. “I don’t think my internet is that bad, but it took four and a half hours to download. It should not take that long.” It would be helpful to be able to subset based on the research, e.g. a long time period over a small area, or a low time resolution over a small area.
- Description says data is available by month and year, but participants only able to access data by month (not year)
- More information (especially a visual preview) of data is desired prior to download
- Arriving at this dataset from Earth Data Search was confusing because the site looks different from other NASA sites. “I wasn’t prepared for that handoff”. Is this NASA? The website says it’s Columbia.

Opening and understanding the data

- RDS filetype is unfamiliar to many users (including R users)
- CSV files should not have “rds” in the file name
- ZIP code column format was challenging for multiple groups, parsed as a number and not a string.
- Participants had difficulty with the units, source, provenance, and prior cleaning of the data. E.g. “Units difficult to discern for Ozone: is it in parts per billion? What is the averaging timespan?” The averaging timespan is important to some air quality practitioners because of regulatory standards. “I would have to read several academic papers to understand how the data were arrived at.”
- The number of files was challenging to participants, e.g. you would need to open and join 20 files to see a 20-year trend. Possibly more helpful to have this data available in different slices, or to have example scripts for joining files.

APPENDICES

SEDAC (con't)
Using the data

- It would be extremely helpful to have shapefiles accompany this data, corresponding to the ZIP code column
- It would be helpful to have this data available by other geographies as well, e.g. state
- It would be helpful to have Year as an attribute of the data, not just a part of the file name.
- FIPS codes would be useful to include as a column for ArcGIS users

Appendix: Other Datasets Air Quality Practitioners Seek to Integrate with NASA Data

Raw responses from participants (from survey):

Many of you have indicated wanting to combine NASA tools with existing datasets. Please link or send any specific datasets you would like to combine with NASA tools. Please send or attach as much detail as possible: scripts you are currently using, data formats, challenges you hope NASA data can help you solve.

- We will overlay NASA data with analytical results from AERMOD and other widely used air pollution data.
- Hourly data for ozone (44201) https://aqs.epa.gov/aqsweb/airdata/download_files.html#Raw
- I would be interested in combining NASA tools with satellite data from TROPOMI, satellite data from TEMPO (once operational), HAPs ground monitoring data from EPA's ambient monitoring archive (https://www.epa.gov/amtic/amtic-ambient-monitoring-archive-haps), ground monitoring data from mobile monitoring campaigns, and data from NASA's Pandora spectrometers. One of the challenges we face is that L3 satellite data is not always available at our desired spatial and temporal resolution (e.g., fine spatial resolution, annual averages). I'd love to learn what algorithms and tools are most appropriate for creating custom L3 products from L2 products.
- I'm interested in combining NASA data with estimations of emissions inventories from area sources like oil fields. Also interested in screening for sources of uncontrolled emissions from oil and gas extraction operations. Excited to see how CARB uses the NASA data as well and have been participating in the methane task force data source prioritization process.
- The datasets our funded investigators use usually are housed in secure centers and cannot be sent to other users. To link data, the NASA data would have to be transferred to the health researchers with latitudes and longitudes for linking.
- intersection with LiDAR datasets from USGS

APPENDICES

Raw responses from participants (from survey - con't)

Many of you have indicated wanting to combine NASA tools with existing datasets. Please link or send any specific datasets you would like to combine with NASA tools. Please send or attach as much detail as possible: scripts you are currently using, data formats, challenges you hope NASA data can help you solve.

- Aclima has a lot of nice data: https://www.aclima.io/aclima-pro , which I would like to combine with built environment time-series satellite data.
- Breathe Providence data has a time resolution of one minute; we are collecting measurements of PM2.5 (ug/m3); NO, NO2, O3, and CO (voltage), and CO2 (vaisala carbocap). Corrections are still underway and we do not currently have data publicly available for me to link here. We currently perform corrections in R and/or Python.
- We use JSON formats from APIs all around the world. (Government and private party).
- At https://drive.google.com/drive/folders/1ElzeF_HuJR3caRKEeB9VFz6CiX6muGA1, to find data named "HR2DAY_LST_ACONC_EQUATES_v532_12US1_2018.nc". Could you interpolate NASA satellite TROPOMI 2018 NO2 and HCHO data to this EPA's CMAQ output's grid?
- How to download TROPOMI current data (for example, today is March 17, 2023, could I find and download TROPOMI yesterday's data, and combine several swaths to one image quickly)?
- I am just started to learn these satellite products and do not have any specific data. My curiosity is can there be a way to add AQS data in NASA Geovanni Tool ?? such that we can plot the eg: NO2 data from satellite (that we can do) plus get AQS data that best match the selection for satellite data. Finally download only selected data and AQS data as .csv file for any further analysis on interest.
- Population Health Metrics for respiratory illnesses
- I primarily use GIS data through ESRI Living Atlas and geodatabase data. I am hoping NASA can help me understand how to transform other formats into data that I can use in my GIS.

ACKNOWLEDGEMENTS

Thank you to all the participants for their contributions and insights. Thank you as well to the workshop organizers and ESIP staff and the NASA subject matter experts who helped support workshop development and answer participant questions.

“Those who contemplate the beauty of the Earth find reserves of strength that will endure as long as life lasts. There is something infinitely healing in the repeated refrains of nature — the assurance that dawn comes after night, and spring after winter.”

- Rachel Carson,
Silent Spring

UNBOUND - AQ

The 2023 workshops were a NASA collaboration facilitated by Earth Science Information Partners (ESP)

EARTH SCIENCE INFORMATION PARTNERS