Atmospheric data Community Toolkit (ACT) Roadmap

Atmospheric data Community Toolkit

The Atmospheric data Community Toolkit (ACT) is an Atmospheric Radiation Measurement User Facility (ARM) supported python library for working with atmospheric time series x N-dimension datasets. ACT has functions for every aspect of the scientific process. Current modules include:

- Discovery – Access APIs for accessing data
- IO – Read/Write data
- Plotting – Easy visualization of data
- QC – Working with existing QC and applying new tests
- Corrections – Apply known corrections necessary for using data
- Retrievals – In depth calculations that aren't suited for their own repo
- Utils – General calculations/utilities not suited for the other modules

ACT aims to reduce the redundant coding performed by individual researchers and focus that effort into developing capabilities that can be utilized by the broader community. This document is a guiding roadmap for development of ACT.

ACT grew out of a need to easily work with atmospheric data across sources and groups. Knowledge gained from the development and maintenance of the Python ARM Radar Toolkit (Py-ART) (Helmus and Collis 2016) was leveraged to ensure a solid framework was developed on which to advance the toolkit. ACT is currently being supported and maintained by ARM.

Version 1.0.0 of ACT was released in June 2020 after some extensive updates to the code structure and documentation for consistency as well as added testing to bring the unit test coverage up to 90%. As the structure of ACT has solidified over the past year, the roadmap will adjust and start to focus on larger scopes of work.

Value of ACT to the Community

The avenue of fostering collaboration and sharing of code will be the largest benefit of ACT to the community. ACT will allow individual groups to break out of their individual development efforts and connect with development from other groups. It is then that the science and infrastructure communities can start to remove redundant efforts and effectively advance, building off of one another as opposed to recreating all the building blocks from the ground up for their effort. ACT will also help to facilitate transparency in codes and decisions made in processing.

ACT has been adopted as part of the ARM DQ Office processing and the ARM Data Center quick look generation processes. This has consolidated the base code for plot generation between two groups which has been a long-term goal. This has streamlined the process for making the DQ Office codes and processes more transparent but also allows for users to make use of these codes and seamlessly contribute back to ARM processing.

ACT Roadmap Survey

With support from ARM's communication team, a survey was released to gather feedback on community priorities. The survey was open for several months and was highlighted during multiple conferences but there were only 14 total responses with only 30% (4) of the respondents having used ACT previously. With a limited sample size, the input is valuable but will be taken into account in setting priorities in a reasonable manner.

In response to roadblocks they faced in using ACT, a reoccurring theme was that there are some problems with ACT on windows machines and while there were requests for code in other languages (IDL, Matlab). The latter is outside the scope of ACT and would require a larger effort.

For each of ACT's modules, respondents could provide responses on additional features they would like to see.

- Discovery
    - For the discovery module, being able to retrieve data from the NOAA FTP servers (3) and air quality data from EPA's AIRNOW portal (3) were of interest as well as model data from Earth System Grid Federation and others (2), satellite data (1) from NASA EarthData, and state governments.
- File Reading and Writing (I/O)
    - On the ability to read different formats of data, ICARTT (5) was of interest as well as GRIB and expanding the ability to read different NOAA formats.
- Corrections

    - Only 2 users responded with requests on functions for corrections. One is for ship corrections, which are already included for wind speed and direction. Corrections for profiling instruments like ARM's shipcor VAPs could be explored. The other request was for a function to calibrate the radars using cloudsat or disdrometers or a way to calibrate the lidars
- Quality Control (QC)
    - There were no requests for the QC modules but just a general comment that the user wants to explore and see how it applies QC variables to the associated variables.
- Retrievals
    - The only request for added retrievals were to calculate liquid water path from brightness temperatures and radiative heating rates.

Six respondents provided additional feedback which varied. One user noted that it was hard to install python and corresponding firewalls behind corporate firewalls and that Matlab was easier to use since it was an executable. Another respondent recommended that ARM dedicate effort to pulling data into a CHORDS server instead of working on a python package. One last note was that a user had troubles running all the examples of which there was also a previous report of compatibility with windows which may be the same issue. The full survey results are available at the end of the report.

Additional feedback from the Ameriflux 2021 meeting included expanding ACT for working with aircraft or drone data, adding scripts to pull more NOAA data, add interactive plotting options, bridge Python with R, and develop a toolbox in the documentation of other related software of note. Additionally, functions for the MFRSR quality control (FFT) and PBL calculations were of interest to multiple users.

Proposed Governance Structure

As with the first roadmap, there is a need to ensure that this effort is responsive to the needs of stakeholders that rely on ACT. The proposed governance structure needs to be flexible and have the ability to expand as the capabilities of the toolkit expand. The roles have not changed since the first major roadmap. The roles required are:

Science Lead: Provides high level leadership for the toolkit, organizes outreach and education, and coordinates contributor and stakeholder input to form a long-term vision for the project. The science lead will also coordinate reviews of the science behind a pull request to ensure accuracy to the literature. The science lead will make a judgement on if a pull request requires more review or can be accepted as is.

Lead Developer: Responsible for overall architecture of the project. Final arbiter in what pull requests to accept. Develops the required style guidelines and coordinates the associate developers. Coordinates contributions from associated developers to a Contributors Guide (and

contributes as well). Responds to users on the GitHub issue tracker with the assistance of the associate developers.

Associate Developer: Responsible, as time allows, for doing an initial check of pull requests for suitability and adherence to the Contributors Guide. Associate developers should come from a diverse background to ensure there is no single point of failure in providing support to ACT. The number and areas of expertise of the associate developers will need to increase and adapt as ACT grows.

Measuring Impact

The impact of ACT can be measured in multiple ways

1) **Growing the number of users and installs.** Success for ACT would mean growing the traffic and user base over time from the period of the first roadmap. Efforts have been made in 2021 to present ACT to the broader atmospheric science communities with posters at the 2021 Ameriflux and ARM/ASR PI meetings and a presentation at SciPy. As shown in Figure 1, while the visitors to the ACT repo were very high during SciPy, there were a lot more clones resultant from the Ameriflux meeting. The traffic recorded on a random basis in Table 1, shows a general increase in unique visitors with a consistent number of views and cloners. As seen in Table 1, there is an ebb and flow to these statistics which make it somewhat import to incorporate regular tracking as part of this process. At time of writing (Dec 20, 2021), the total number of downloads through Conda-Forge were over 21,000.

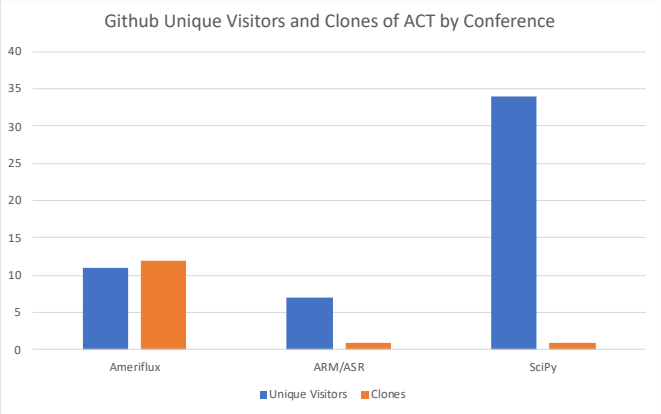


Figure 1. Number of unique visitors and clones by 2021 conferences

Table 1. Record of Github traffic for weeks around roadmap development.

Date Range	Views	Unique Visitors	Unique Cloners
7/16-7/29 2019	128	11	12
8/15-8/28 2020	389	23	40
5/29-6/11 2021	489	41	30
12/7-12/20 2021	141	34	11

2) **Number and success of dependent projects.** Success can also be measured by an increase to the number of dependents that are using ACT. ACT has seen an increase in number of dependents, including the use of ACT in other DOE programs.

Table 2. Record of dependent repositories by roadmap version.

Roadmap	Number of Dependent Repositories
V1.3	9
V2	14

3) **Papers and presentations using ACT.** Publications are treated by many user facilities as a metric of scientific impact. In an effort to be fully inclusive, all contributors are included in the DOI. ACT is using an auto-generating DOI through Zenodo that provides an updated DOI with all contributors as authors upon each new version release. Currently, there are 10 listed authors of ACT. There is additionally, one publication citing ACT with at least one more in preprint.

Results from First Roadmap

The first major ACT roadmap was updated on a 6–12-month basis due to the major development efforts that were ongoing, but the general goals and results were as follows.

Increase unit test coverage

At the time of the first roadmap, the code coverage was 74% with a goal to bring the unit coverage up to 90% or better in 12-months. That goal was accomplished, and the unit test coverage is currently at 91%.

Figure 2. Unit test coverage percentage over time.

Discovery

It was proposed to continue to expand the discovery module to include more data web services. A more cautious approach has been employed here as there are a number of other python libraries that can do this as well and it is important not to duplicate effort.

I/O

A general goal of expanding the number of data readers was accomplished. The I/O module saw a number of updates to include functions for reading a variety of National Oceanic and Atmospheric Administration (NOAA) datasets as well as improvements for writing data to netcdf files that follow CF-convention.

QC

While the number of additional QC tests added to ACT has been minimal compared to expectations, there have been a large number of improvements to the QC functions. There was also an additional effort to more broadly communicate how one can use ACT for interrogating data quality through an ARM blog post. https://arm.gov/news/blog/post/69939

Retrievals

The goal was to start to develop retrieval functions in concert with the science community but there has not been as much engagement as would have been ideal so some efforts to code up retrievals based on ARM Value-Added Products (VAPs) has been started such as the calculation of the boundary layer height from radiosondes.

Education and Outreach

While the ACT developers have not developed many tutorials as part of the ACT documentation, there has been a lot of education and outreach through conferences, working with others in the programs, and providing sample codes to summer interns at multiple labs.

Overall, not all the goals were fully realized but the roadmap was still very successful in driving development efforts. It should be noted that the ACT budget is for the general maintenance, pull request response, and education and outreach. Development efforts have generally been performed with any remaining effort which is part of the reason for not fully realizing the roadmap.

Goals and Priorities Summary Table

The development team will prioritize the library maintenance, acceptance of pull requests, and perform strategic development that adds the features outlined the following subsections. As with the previous roadmap, effort available for development will depend on general library maintenance and upkeep. The high priority items will require development effort from the ACT team but as always, will prioritize contributions from external users. They will take precedence over the medium and low priority features. Medium and low priority features will be worked on as time allows and will be dependent on priority and effort required.

Feature	Priority	Effort
Aircraft or UAS Related Functionality With ARM's continued UAS related efforts and a new aircraft on the horizon, along with requests from the Ameriflux meeting, the features for visualizing, applying QC, and other functionality will be useful.	High	High
Retrievals With the objective to appeal to the scientific user base, effort will be put forth to add additional retrieval functions with a focus on ARM VAPs.	High	High
Windows Compatibility Due to none of the development team having windows machines for testing, there has been a gap in troubleshooting and diagnosing issues from some of the users. Effort will be added to ensure windows compatibility.	High	Medium
ARM Data Surveyor As part of the efforts to incorporate ACT into the ADC quick look processing, a command line tool was developed to replace what was previously being done. If will be important to continue maintenance and support of that tool.	High	Medium
Performance Improvements Dask is already being used in some functions to improve speed, but there will be further investigation and development to improve the speed of ACT loading,	High	Medium

processing, and visualization. Additionally, efforts to optimize the ACT package will be implemented.		
Statistics Tracking A process to better track statistics of the repository will be developed in conjunction with Py-ART to provide more regular statistics.	Medium	Low
I/O Improvements A number of survey respondents indicated that the ability to read ICARTT and GRIB would be beneficial. There are python libraries already in existence so the effort is expected to be low but could increase if there are integration problems.	Medium	Low
Tutorial and Example Development There was a large effort to overhaul ACT's documentation which saw major improvements to the auto-documentation process. To further improve user interactions, additional tutorials and examples will be produced.	Medium	Low
Discovery Improvements Carrying on from the first roadmap and responding to requests from the survey to add functions to download data from the NOAA FTP sites and EPA's AirNow portal, the discovery functions will be expanded where appropriate.	Low	High
Visualizations While ACT already has a lot of plotting capabilities, there are areas in which it could be expanded, including interactive plotting and 3-D plotting.	Low	Medium

If unforeseen priorities arise in the next three years, the roadmap will be updated to reflect the additional priorities in the interest of transparency.

References

Giansiracusa, M., (2019). Python scripts for ARMLIVE Web service. Github: https://code.ornl.gov/ofg/armlive_getfiles

Helmus, J.J. & Collis, S.M., (2016). The Python ARM Radar Toolkit (Py-ART), a Library for Working with Weather Radar Data in the Python Programming Language. Journal of Open Research Software. 4(1), p.e25. DOI: http://doi.org/10.5334/jors.119

2021 ACT Survey Report

Prepared by Stacy Larsen

October 2021

2021 ACT Survey

Open June 21, 2021 – October 15, 2021

Survey Responses: 14

Survey Results Summary

Atmospheric data Community Toolkit (ACT) is a python toolkit for exploring and analyzing time-series atmospheric science datasets. A goal of ACT is to provide a platform to enable the broader sharing of code between ARM, other scientific organizations, and the scientific community. Feedback was sought to help set development priorities.

While most of the survey participants responded that they have not yet used ACT, many also reported that there are no barriers to using it.

Notable Results

- 70% have not used ACT before
- 64% do not believe there are any barriers to prevent them from using it
- 56% are interested in the ICARTT file format.

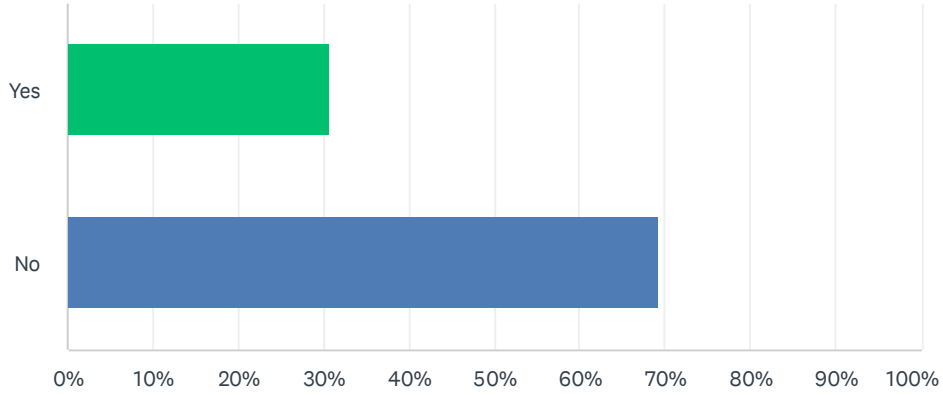
Q1 Name (optional):

Answered: 6 Skipped: 8

#	RESPONSES	DATE
1	Bryan Raney	10/7/2021 12:20 PM
2	David Delene	10/6/2021 12:06 PM
3	Josh Howie	10/6/2021 11:46 AM
4	Maxwell Levin	10/6/2021 11:37 AM
5	Michael Klatt	7/16/2021 9:51 AM
6	John Stevens	6/21/2021 10:57 AM

Q2 Have you used ACT before?

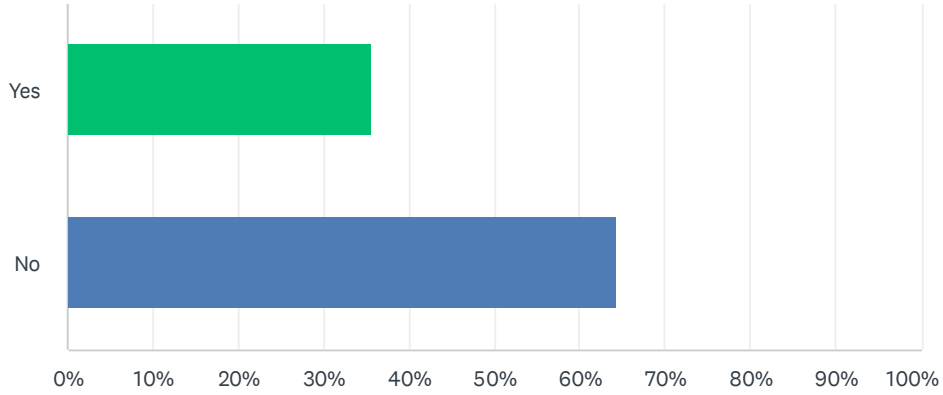
Answered: 13 Skipped: 1



ANSWER CHOICES	RESPONSES	
Yes	30.77%	4
No	69.23%	9
TOTAL		13

Q3 Are there barriers to you using ACT?

Answered: 14 Skipped: 0

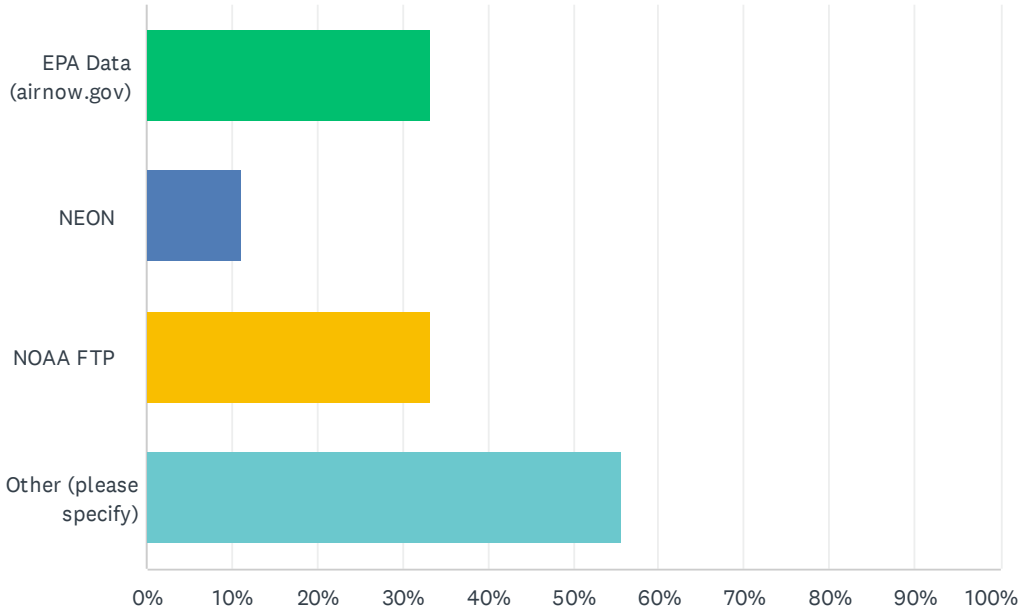


ANSWER CHOICES	RESPONSES	
Yes	35.71%	5
No	64.29%	9
TOTAL		14

#	IF YES, PLEASE EXPLAIN.	DATE
1	Python	10/7/2021 8:36 PM
2	On my windows computer the code does not recognize or contain all modules needed to run the examples.	10/6/2021 11:45 AM
3	Some of the code isn't compatible with windows machines.	10/6/2021 11:22 AM
4	"Not Invented Here" and lots of legacy code at my workplace. This attitude is changing, thankfully.	7/16/2021 9:51 AM
5	It only works in Python, it will be good to make similar codes in IDL and Matlab. Also expand the code base, it does very simple tasks as of now.	6/21/2021 10:57 AM

Q4 ACT’s discovery module houses scripts for downloading data from web services like ARM’s data live web service. Are there specific organizations that you would like to have this functionality for?

Answered: 9 Skipped: 5

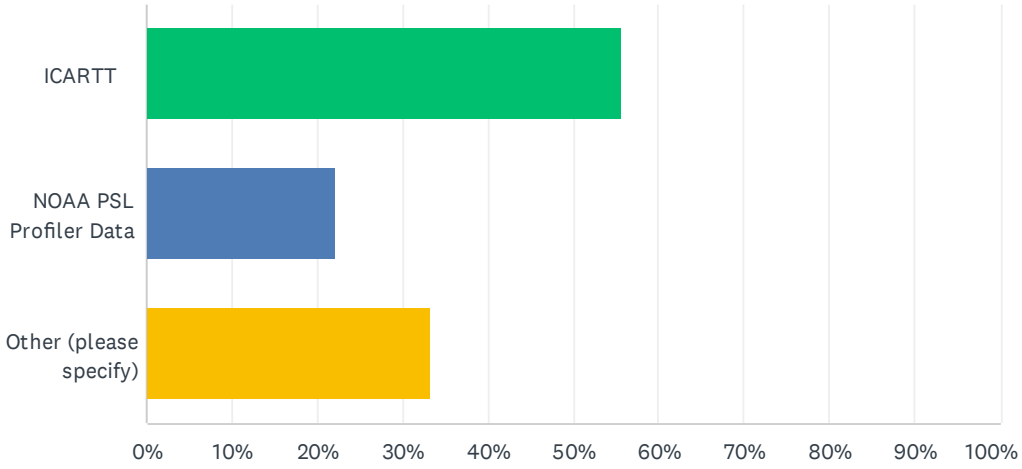


ANSWER CHOICES	RESPONSES	
EPA Data (airnow.gov)	33.33%	3
NEON	11.11%	1
NOAA FTP	33.33%	3
Other (please specify)	55.56%	5
Total Respondents: 9		

#	OTHER (PLEASE SPECIFY)	DATE
1	Esgf	10/7/2021 8:36 PM
2	ERA5, CMIP	10/7/2021 12:20 PM
3	state goverment	10/6/2021 11:45 AM
4	N/A	7/21/2021 11:16 AM
5	NASA EarthData	6/21/2021 10:57 AM

Q5 ACT's IO module contains scripts for reading ARM netcdf files, ascii or csv files, NOAA GML data, and uses an external library to read Micropulse Lidar binary files. Are there other formats or specific datasets that you would be interested in?

Answered: 9 Skipped: 5



ANSWER CHOICES	RESPONSES	
ICARTT	55.56%	5
NOAA PSL Profiler Data	22.22%	2
Other (please specify)	33.33%	3
Total Respondents: 9		

#	OTHER (PLEASE SPECIFY)	DATE
1	The power is in the data format that can be used; however, the format, like ICARTT, change also.	10/6/2021 12:06 PM
2	GRIB	7/16/2021 9:51 AM
3	grib	6/21/2021 10:57 AM

Q6 ACT’s correction module contains scripts for correcting data (example: micropulse lidar afterpulse, range, overlap corrections). Are there any specific corrections that you would be interested in?

Answered: 2 Skipped: 12

#	RESPONSES	DATE
1	Ship corrections	7/21/2021 11:16 AM
2	It will be great if a module is included that can be used to calibrate the ARM radars using CloudSat or disdrometer, calibrate the lidars.	6/21/2021 10:57 AM

Q7 ACT’s QC module contains scripts for applying basic tests, such as limit or delta changes, to more advanced tests, such as the Fast-Fourier transform (FFT) method, for detecting shading in shadowband radiometers. Are there specific QC tests that you would find useful?

Answered: 1 Skipped: 13

#	RESPONSES	DATE
1	Not having used ACT, I'd like to see how it applies existing QC variables to the associated data variable, or if it does this. So no request, I just need to explore the software.	10/6/2021 11:46 AM

Q8 ACT’s retrieval module contains scripts for calculations/retrievals (example: wind profiles from Doppler lidar PPI scans). Are there retrievals or other calculations that you would be interested in?

Answered: 1 Skipped: 13

#	RESPONSES	DATE
1	It will be great if liquid water path can be calculated using the recorded brightness temperatures. Also the radiative heating rates.	6/21/2021 10:57 AM

Q9 Do you have any other comments or suggestions?

Answered: 6 Skipped: 8

#	RESPONSES	DATE
1	Pulling a lot of exotic modules and packages is very hard behind strict corporate firewalls. Using python and installing all of this by hand is a huge barrier and bottleneck because it is not already an executable, ready to use, like Matlab.	10/7/2021 8:36 PM
2	I don't see much usefulness with a python plotting application. Once you have a ingest code for the data file format you are working with, python provides necessary tools with just a few lines of code. Setting up the python environment is a pain, a more accessible and unversity interface like a Website is better in my opinion. Hence, I believe that being able to pull data into CHORDS and Grafana (display) is better. How about working on code to pull ARM data into a CHORDS sever (see https://earthcubeprojects-chords.github.io/chords-docs/).	10/6/2021 12:06 PM
3	The examples did not work for me due to the code not finding all modules.	10/6/2021 11:45 AM
4	github.com/tsdat/tsdat relies on ACT for its quality filter capabilities and various utility functions, so we would like to see ACT's continued growth and success :)	10/6/2021 11:37 AM
5	I just learned of this toolkit today and need to work with it awhile before I have any opinions or suggestions.	8/16/2021 4:09 PM
6	Great effort overall.	6/21/2021 10:57 AM