20 21 THE ANACONDA ANNUAL DIVIDEND PROGRAM REPORT

ANACONDA
DIVIDEND PROGRAM

Table of Contents

ANACONDA
DIVIDEND PROGRAM

A LETTER FROM PETER WANG

Why We Developed the Anaconda Dividend Program

Peter Wang
Co-Founder and CEO, Anaconda

I started using scientific Python in the early 2000s as a software consultant and was blown away by the power of these open-source tools to solve real-world problems. When I co-founded Anaconda nearly ten years ago, my co-founder Travis Oliphant and I wanted to promote the use of Python for a broader set of data analysis needs, but doing so was very difficult. At the time, Wes McKinney started pushing pandas for a DataFrame library. The Jupyter notebook was new, and we saw people wanting to do more significant analysis using Python. But the biggest problem was that our numerical software tools were too difficult to install properly. That's why we decided to create our packaging system, to ensure that everyone, from absolute beginners to advanced experts, could easily access the entire robust ecosystem around open-source data science.

Anaconda has always been about open-source innovation. Over the years, we have invested nearly $30 million into incubating and maintaining a wide variety of open-source projects, including pandas, Dask, Numba, Bokeh, HoloViews, Panel, Intake, and–most recently–the Pyston project. Most of this open-source funding has been through employee time allocation, with occasional direct donations and event sponsorship. However, we felt that we could do much more. To scale our contributions as we grow the company, we proudly launched the Anaconda Dividend Program to directly give a portion of our revenue dollars back to the projects and the community that make all of this possible.

A LETTER FROM PETER WANG

What is the Anaconda Dividend Program?

The Anaconda Dividend Program is a formal initiative where we donate a portion of Anaconda's revenue to open-source organizations and projects that support data science innovation and the community that's given so much back to us and our users.

As an extension, we also launched a fee exemption program for non-profit research institutions to gain access to our commercial tools. We believe it is important to support projects that address societal needs or global challenges and ultimately make the world better.

ANACONDA.
DIVIDEND PROGRAM

4

Reflecting on the First Year

We launched the program in October 2020 in partnership with NumFOCUS, a U.S.-based nonprofit organization that provides essential administrative services and operational support for 44 open-source scientific computing projects. I am proud to have played an instrumental role in the founding of NumFOCUS in 2012 and continue to volunteer with the organization by serving on their Advisory Council.

I believe it's a part of our responsibility and our users' responsibility to support and maintain the open-source community so that we can continue to reap the benefits of data science innovation. Therefore, the premium version of Anaconda has been our mechanism to give back. Our enterprise customers and individual users are now helping to support open-source development in our space.

Our goal was to donate a minimum of $25,000 committed for the whole year, and we're on track to exceed this goal significantly. These funds will help provide support and resources to projects like Matplotlib, pandas, SciPy, Bokeh, TARDIS, mlpack, Dask, and more.

How I See the Program Evolving

I hope to increase our contributions in years to come. We will coordinate with NumFOCUS and others to develop a governing board for the Anaconda Dividend Program funds, expand our beneficiaries, and support additional projects in the Python ecosystem. As Anaconda approaches its ten-year anniversary, I am proud to see how much this community has grown and will thrive in the next decade.

Sincerely,

Peter Wang, Co-Founder and CEO, Anaconda

ANACONDA.
DIVIDEND PROGRAM

DIVIDEND CONTRIBUTIONS

About NumFOCUS

NumFOCUS is a 501(c)(3) non-profit organization whose mission is to promote open research, data, and scientific computing practices by sponsoring open-source projects and organizing community-driven educational programs. NumFOCUS offers a stable and professional home for open-source projects that power scientific and business processes. They aim to ensure the longevity of projects in the scientific data stack by providing funding and resources. NumFOCUS supports commonly known data science tools like pandas and Jupyter and helps organize PyData, a community engagement program for users and developers of open-source scientific computing tools.

Anaconda is a Silver Sponsor of NumFOCUS, helping to fund projects needing additional resources and supporting the diversification of the open-source community. What is unique about the Anaconda and NumFOCUS relationship is that Anaconda is only one of two corporate sponsors that offer recurring quarterly support to the organization. We're honored to give back to the community that uses our tools and support its vibrant ecosystem of users, maintainers, and contributors.

"The Dividend Program positions Anaconda among the industry leaders in support of open-source data science. We are extremely grateful for their partnership in ensuring the continued growth and innovation of our projects. Anaconda's support is a direct investment in NumFOCUS's efforts to provide an operational foundation from which these essential project communities can succeed."

Leah Silen, NumFOCUS Executive Director

Donations

We launched the Anaconda Dividend Program in partnership with NumFOCUS, kicking off the program with a sponsorship of $10,000. Starting in October 2020 and through 2021, we committed to funding the dividend with a portion of our revenue per quarter, with a minimum of $25,000 committed for the year.

Quarterly contribution breakdown

QUARTER AND YEAR	DOLLARS DONATED
Q4 2020	$1,285
Q1 2021	$6,500
Q2 2021	$10,700
Q3 2021	$16,250

Since launching the program, Anaconda has donated $44,735 to NumFOCUS to help support open-source project development and is honored to be a Silver Sponsor of their program initiatives.

We have always prioritized empowering individuals with the right tools and access to open-source innovation. With the Anaconda Dividend Program, enterprises now have the opportunity to give back to the community through their investment.

ANACONDA®
DIVIDEND PROGRAM

TARDIS

BACKGROUND

Known for its collection of tools that help analyze observations from exploding stars, this project is widely adopted in the astrophysical community and is used primarily in higher education research and teaching. The Numba project, an open-source JIT compiler that translates a subset of Python and NumPy code into fast machine code, uses TARDIS as a benchmark.

TARDIS' USE CASE WITH NUMBA

Supernovae are the energetic death throes of stars that can be seen across vast cosmic distances. These explosions are the hearths that transform primordial hydrogen to the elements crucial to our existence such as iron and silicon. Despite decades of studies, many mysteries about these events are unsolved. TARDIS is a code that can simulate this event and produce synthetic observations that, when compared to real observations, can be used to unravel some of these mysteries.

"TARDIS is a completely community-developed and supported code – a first in this field. We have collaborated with the Numba team to replace the C-portion of the TARDIS code with Numba. The Numba code is essentially as fast as the C-version but is much easier to maintain. The flexibility of Numba offers us to try new and even more accurate ways to simulate supernovae using state-of-the-art simulations and trying out new hardware choices such as GPUs. In turn, the Numba team can use our open-science code to try out new ways to accelerate complex codes such as TARDIS."

Wolfgang Kerzendorf, Assistant Professor of Astrophysics & Computational Mathematics, Science, and Engineering at Michigan State, Project Lead, TARDIS

You can learn more about the project, donate directly, or contribute code.

Learn more

ANACONDA.
DIVIDEND PROGRAM

8

NumFOCUS Projects the Dividend Program Helps Support

BOKEH

BACKGROUND

If you're looking to quickly and easily connect powerful PyData tools and pipelines to interactive plots, dashboards, and data applications, Bokeh is for you. It's a visualization library that allows users to turn complex data sets into versatile, interactive graphics for modern web browsers. It is used in a variety of industries including government use cases, business and industry applications, higher education research, and teaching.

One of the highlights of Bokeh's recent 2.4 release is support for scientific math notation. This long-requested feature allows users to integrate mathematical symbols and formulas directly into Bokeh visualizations. Mathematical notations in LaTeX and MathML can now be used in several elements of visualization, specifically on axis labels, axis tick labels, Div widgets, and Paragraph widgets. Other highlights of the 2.4 release include documentation improvements and improvements to Bokeh's SVG backend.

"We are very glad to finally have math notation capabilities directly in Bokeh, a key improvement for our users in research and science. Going forward, we plan to expand those capabilities and also add several improvements for in-browser rendering and integrating Bokeh with common web frameworks. We also continue to focus on improving our documentation and making it easier for new contributors and contributors from more diverse backgrounds to join the team."

Timo Metzger, Bokeh Core Team Member

BOKEH

BOKEH USE CASE

Bokeh is used across government, academia, and professional spheres including finance, medicine, oil and gas, data science, and more.

On a more granular level, Bokeh's new math notation features allow users to provide more detailed information and explanations directly in Bokeh. These two plots, for example, demonstrate the use of mathematical formulas and symbols on axis labels as well as in explanatory paragraphs below the plot. The first plot reflects Planck's law, and the second demonstrates an ideal Gaussian distribution.
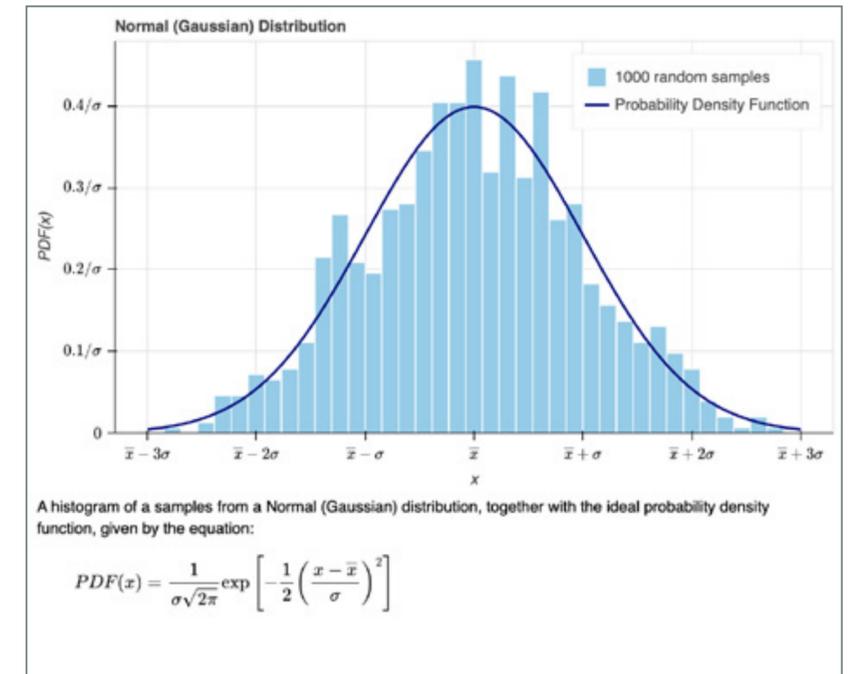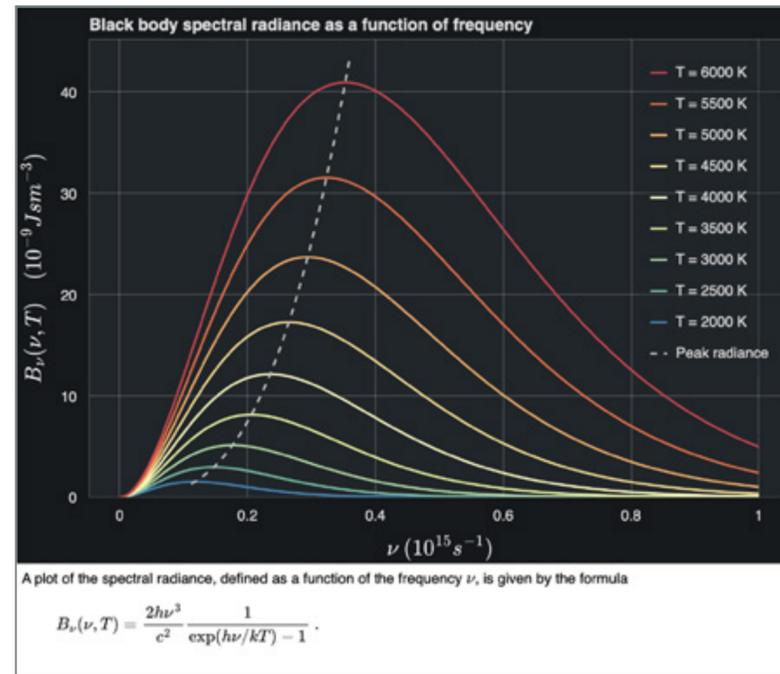

Black body spectral radiance as a function of frequency. A plot of the spectral radiance, defined as a function of the frequency ν, is given by the formula

$$B_\nu(\nu, T) = \frac{2h\nu^3}{c^2} \frac{1}{\exp(h\nu/kT) - 1}.$$

Normal (Gaussian) Distribution. A histogram of a samples from a Normal (Gaussian) distribution, together with the ideal probability density function, given by the equation:

$$PDF(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{x - \bar{x}}{\sigma}\right)^2\right]$$

You can learn more about the project, donate directly, or contribute code.

Learn more

ANACONDA.
DIVIDEND PROGRAM

MLPACK

BACKGROUND

mlpack is a fast and flexible machine learning library and is suitable for both data science prototyping and deployment. It is used primarily for business and industry applications, as well as higher education research and training. It's written in C++ and aims to provide fast, extensible implementations of cutting-edge machine learning (ML) algorithms. mlpack provides said algorithms as command-line programs, Python bindings, and C++ classes that can be integrated into larger-scale ML solutions.

MLPACK USE CASE

mlpack's academic background has led to mlpack being used in many scientific publications both inside the machine learning community and in adjacent fields.

mlpack's neural network toolkit has been used by NASA and the University of Maryland for the purposes of communications with spacecrafts and satellites. Reinforcement learning techniques from mlpack were used as an integral part of a used-in-production space communications system.

"In the past year, we've honed in on the importance of making our efficient machine learning algorithms lightweight and deployable, and we've been hard at work refactoring and retooling mlpack and its dependencies to support this. We recognize that deploying machine learning models seamlessly to all kinds of different environments is a tremendous need that is also somewhat underserved in today's open-source software ecosystem. As such, the mlpack developer community is very excited at making our vision a reality—and then demonstrating that reality to our users. Anaconda's support is integral to these efforts."

Ryan Curtin, mlpack Developer

You can learn more about the project, donate directly, or contribute code.

Learn more

THE YEAR AT A GLANCE

Anaconda's Commitment to Open Source

In addition to the dividend contributions, Anaconda provides free storage, networking, and infrastructure support to large community packaging channels like conda-forge and bioconda. We launched conda signature verification earlier this year, based on The Update Framework (TUF). We also provided training to other packaging projects on how best to implement and adopt similar trust and signing features. We believe that the overall challenges of packaging open source can only be solved by collaboration with the broader community. To this end, our team organized and sponsored Packaging Con, to bring together individuals working on software packaging across languages and ecosystems, to facilitate learning and cross-pollination.

Beyond software packaging and distribution, Anaconda has given back to the open-source community through dedicated time, resources, and team members. In August of 2021, Anaconda welcomed the Pyston team to our ranks. The Pyston developers, Kevin Modzelewski and Marius Wachtler, joined Anaconda's large team of other fully-funded OSS developers who contribute to Dask, Numba, Intake, HoloViews, Bokeh, and many other projects. This year, we also launched a new project called "Kerchunk" to enable cloud-scale in-situ access to archival scientific data formats that are not already cloud friendly. While this may sound niche, this is of extreme importance for facilitating open science and will be an integral part of many real-world scientific workflows.

ANACONDA.
DIVIDEND PROGRAM

The Year at a Glance

Learn more about Anaconda's impact on open source via the key stats below.

25+ million
active Conda users in the last 12 months

The Anaconda team spent
17,000 hours
building and maintaining open-source packages

2,467
practitioners contributed to the Anaconda Dividend Program

In 2021...

5.2+ billion
package downloads

17 petabytes
of package data downloaded

500+ million
package downloads each month!
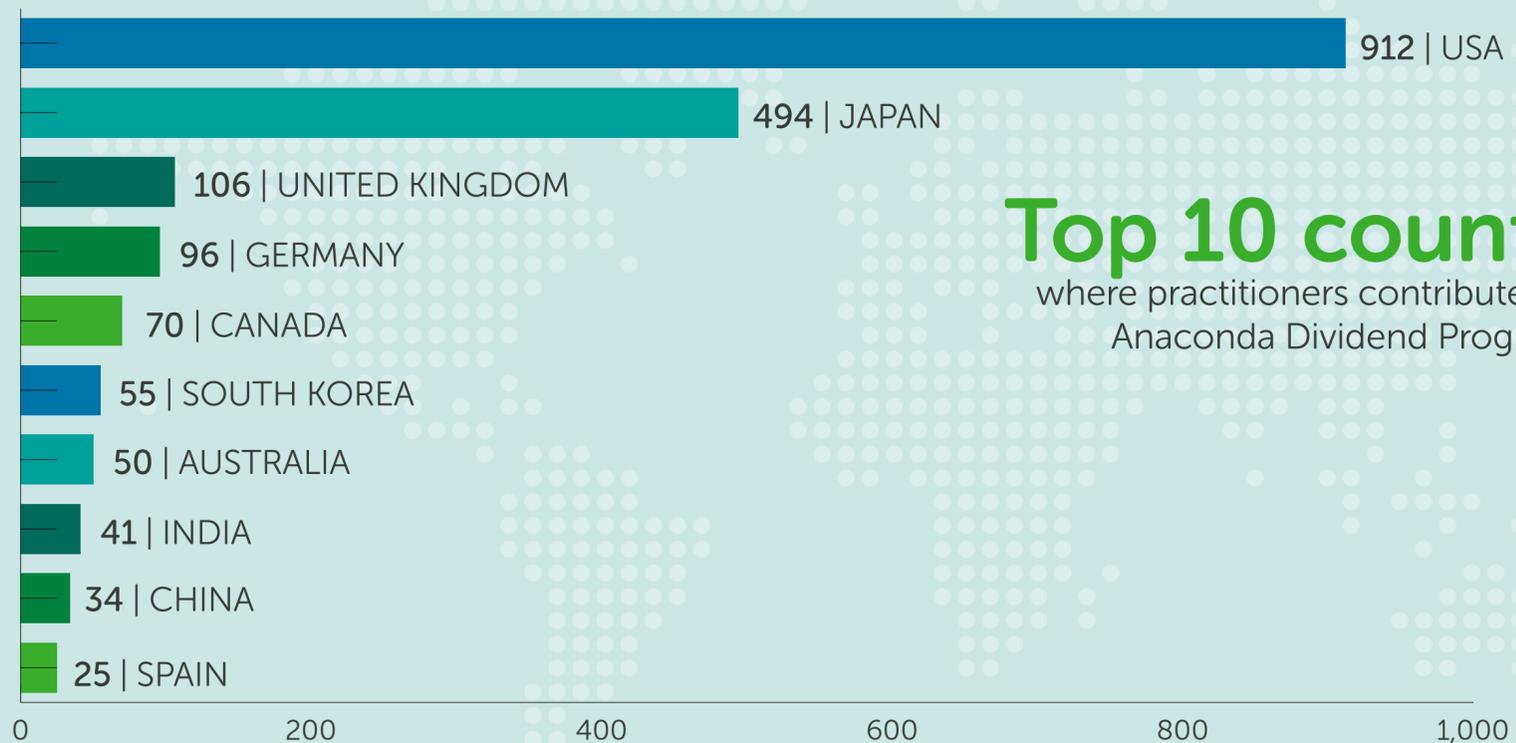
Top 10 countries
where practitioners contributed to the Anaconda Dividend Program

Country	Value
912	USA
494	JAPAN
106	UNITED KINGDOM
96	GERMANY
70	CANADA
55	SOUTH KOREA
50	AUSTRALIA
41	INDIA
34	CHINA
25	SPAIN

0   200   400   600   800   1,000

ANACONDA.
DIVIDEND PROGRAM

The Year at a Glance

Learn more about Anaconda's impact on open source via the key stats below.

305 enterprise customers
contributed to the Anaconda Dividend Program
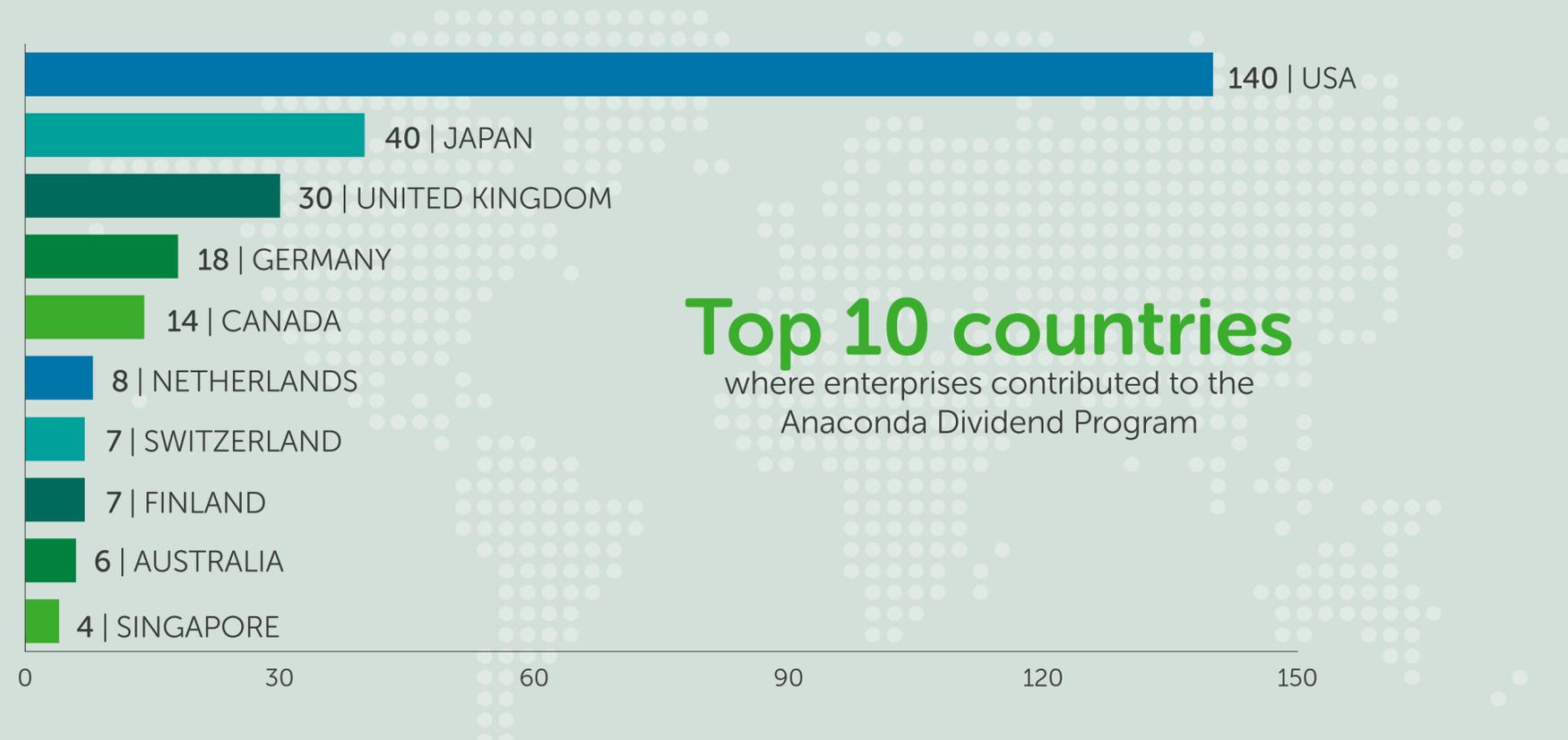
Top 10 countries
where enterprises contributed to the Anaconda Dividend Program

Country	Value
140 | USA	
40 | JAPAN	
30 | UNITED KINGDOM	
18 | GERMANY	
14 | CANADA	
8 | NETHERLANDS	
7 | SWITZERLAND	
7 | FINLAND	
6 | AUSTRALIA	
4 | SINGAPORE	

0   30   60   90   120   150

41%
of the U.S. enterprise contributions were driven by Fortune 500 companies

ANACONDA
DIVIDEND PROGRAM

ROADMAP FOR 2022 AND BEYOND

Evolving the Dividend Program

ADVISORY COMMITTEE AND GOVERNANCE

As the program continues to scale and grow, Anaconda will develop an Advisory Committee which will advise on community matters and oversee the administration of the program, which will include seats from both Anaconda and NumFOCUS.

ALLOCATION OF FUNDING

For the initial rollout of the program, all funds generated have benefited NumFOCUS. As the program continues to evolve, Anaconda and the Advisory Committee may decide in the future to allocate funds to other programs or organizations.

Open Source Commitment

EVENT SPONSORSHIP

Anaconda was proud to sponsor multiple open-source events hosted by NumFOCUS in 2021, including the PyData Global Conference and Packaging Con. We look forward to continuing to be involved in these events that help shape the future of open source.

PROJECT SUPPORT

As always, we plan to allocate time toward critical projects that are instrumental for data science and Python users. At Anaconda, we have team members who support various projects like Numba, Dask (this team works on a multitude of projects related to distributed computing, cloud computing, and data storage: Dask, Intake, fsspec, fastparquet, s3fs, gcfs, adlfs, etc.), Metagraph, and Pyston. They work with open-source software users, contributors and adjacent projects, external maintainers, makers, partners, and sponsors to help support the long-term research and development of open-source projects and tools.

ANACONDA.
DIVIDEND PROGRAM

ABOUT ANACONDA

With more than 25 million users, Anaconda is the world's most popular data science platform and the foundation of modern machine learning. We pioneered the use of Python for data science, champion its vibrant community, and continue to steward open-source projects that make tomorrow's innovations possible. Our enterprise-grade solutions enable corporate, research, and academic institutions around the world to harness the power of open-source for competitive advantage, groundbreaking research, and a better world.

Visit https://www.anaconda.com to learn more.

ANACONDA®
DIVIDEND PROGRAM