GORDON & BETTY MOORE FOUNDATION
DATA-DRIVEN DISCOVERY GRANT DEVELOPMENT FORM

Organization	*The Astropy Project*
Project Title	*Sustaining and Growing the Astropy Project*
Amount	*$901,000*

Proposal

1) Grant Purpose:

a) Brief Purpose Statement: *In 2-3 sentences describe the objectives of the proposal and expected outcomes and deliverables.*

The Astropy Project provides an open and community-driven framework for core computational tools to enable data-driven research and discovery in astronomy. This project has laid the foundation for a groundbreaking astronomy software ecosystem, but lacks dedicated funding to sustain it in the long-term. This proposal aims to turn the Astropy Project into a sustainable effort by: 1) using targeted funding to ensure that critical parts of the Project continue to be maintained by retaining core team members who may otherwise leave, 2) building a self-sustaining long-term *user to contributor to maintainer* pipeline, and 3) formalizing our governance and management structure which will make the Project more compelling to federal funding agencies.

b) Description of Project: *Use this section to describe why you are seeking this grant and how it will achieve the stated goals*.

Background

The field of astronomy—broadly defined as encompassing astrophysics and cosmology, along with planetary, solar, and space sciences—has undergone a revolution over the last few decades. Advances in computational power and instrumentation have ushered in an era where both the data volume and the computational complexity required to analyze the data has increased dramatically. For example, the (ground-based) Large Synoptic Survey Telescope and the (space-based) Wide Field Infrared Survey Telescope will generate datasets hundreds or even thousands of times larger than those of present-day surveys, creating both challenges and opportunities for applying existing and new techniques to such large datasets. Hence, the discoveries in the 21st century and beyond will critically depend on the tools, techniques, and algorithms available to analyze the wealth of data available.

GORDON & BETTY MOORE FOUNDATION
DATA-DRIVEN DISCOVERY GRANT DEVELOPMENT FORM

Over the last decade, the Python programming language has become the most used language throughout astronomy[1]. In addition to being an easy-to-learn and powerful high-level programming language, it is free and open source. This allows scientists, developers, students, and amateur astronomers around the world to use, write, and share analysis tools without being excluded by the high cost of proprietary software, and has led to the development of a strong ecosystem of scientific Python packages (including NumPy, SciPy, and Matplotlib).

The Astropy Project was started in 2011 by researchers and software developers with an ambitious twofold goal: 1) develop a single core Python package to provide the foundational tools for astronomy research; and 2) build an ecosystem of usable, interoperable, and collaborative astronomy Python packages. The first goal is of significant practical importance in that it provides the astronomy community with a shared "base" of standard functionality that is understandable to and useful for everyday working scientists. Meanwhile, the second goal represents a transformative element of the Astropy Project, recognizing the inevitable proliferation of scientist-developed or institution-specific packages. To maximize the community benefit from these packages, the Project has built and maintains a robust development infrastructure. These software tools and supporting documentation simplify and unify the documentation, testing, and distribution of individual packages. Packages that use this infrastructure and meet minimum standards for interoperability, uniqueness, documentation, and testing are formally recognized by the community as "Astropy Affiliated Packages."

The Astropy Project has been successful in meeting both of its initial goals. The **`astropy`** core package is now mature in many regards and provides required core tools such as coordinate and time transformation, data table and N-d image manipulation, and I/O support for the myriad of file formats used in astronomy. Built on top of the **`astropy`** core package, the community has created a high quality ecosystem of over 40 Affiliated Packages[2] that have become the basis of a significant portion of astronomical research. Over the last five years, the Astropy ecosystem and the tools built on top of it have been used for research presented in thousands of refereed publications.

One of the unique aspects of the Project is that it is not developed by one particular institution, nor solely by software developers—it is a true community effort in which contributions are made by participants with a mix of backgrounds. The developer community includes researchers who are volunteering their time to work on the Project, students who work on it as part of internships, and software developers at different institutions who have been assigned to work on some aspects of the Project for part of their time. It is even seen in some circles as a "neutral ground" where academic institutions that might otherwise see themselves as competitors can work together in areas of common interest.

Despite these successes, the Project does not have any funding that it can use to prioritize development—instead, development is done following the priorities of the contributors and the institutions funding some of their developers to work on Astropy. This has resulted in some areas of the Project becoming at-risk in terms of sustainability. In fact, some key components of the **`astropy`**

[1] see Momcheva & Tollerud 2015 - https://arxiv.org/abs/1507.03989
[2] http://www.astropy.org/affiliated/

core package lack active maintainers (due to the original authors being unable to continue contributing), and other parts of the ecosystem rely on single individuals who may not be able to continue to contribute on a voluntary basis in the future.

Funding from the Moore Foundation would provide the Astropy Project with the resources needed at this critical point in time to transition to a self-sustaining project where its success no longer hinges on a specific set of key volunteers. In particular, we are requesting funding so that we can support critical members of the Project, not only to maintain existing infrastructure and functionality while also developing crucial missing functionality (*Output 1.1*), but also to help grow and increase the diversity of the Project by actively recruiting and mentoring individuals who can become contributors and eventually core maintainers of the Project (*Output 1.2*). In addition, we want to formalize the governance and management structure of the Project, which is a requirement if we are to obtain federal funding to provide long-term resources for the Project (*Output 1.3*). This will give Astropy the stable base it needs to provide the astronomy community with a fertile space to build the advanced analysis tools necessary for the coming decade of data-driven science, which will be essential for enabling scientific discoveries.

Ensuring a well-maintained Astropy Project ecosystem (*Output 1.1*)

The `astropy` core package is a large, maturing code base.. As for any package of its size, there is a distinct need for continued maintenance to address existing and new bugs (sometimes urgently), respond to user queries (both usage questions and potential bugs), review contributions and pull requests, write and clarify documentation, and develop incremental improvements. Even though these tasks are critical to maintaining the code at a level that can keep up with modern astronomy and its often very large datasets, day-to-day maintenance is not viewed as glamorous or "fun." Hence, volunteers often prefer to work on developing new functionality rather than maintaining existing code. In addition, even if volunteer effort is available, some bugs or code reviews require a highly experienced developer. As a result, reports of bugs/issues and reviews of pull requests can sometimes remain unaddressed for long periods of time (years in some cases). Over the last few years, the number of open issues in the core `astropy` repository has increased to over 800 and a continued increase is not sustainable. Similarly, there are over 60 open pull requests for the core package, and we are currently severely limited in terms of expert reviewers who have time to volunteer to review these.

To remedy this, we are requesting funding so that we can pay developers as consultants to work on very specific tasks (bug fixes, code reviews, and/or targeted performance enhancements) for which we are unable to find volunteer effort. This will allow us to ensure that the code in the `astropy` core package and other packages in the ecosystem remains well maintained and that we stay on top of submitted contributions. These consulting developers would be identified from within the existing Astropy community. We believe that this approach will be the most efficient use of funds, as no one person possesses the expertise to address issues across all of Astropy. In addition, hiring a single developer is much less sustainable since it creates a single point of failure if they leave the Project in the future.

This same consultant funding pool may also be applied to targeted efforts in specialized astronomy domains currently lacking support in the Python ecosystem. The Project's "Affiliated Package" model allows the consultant framework to work as well for the development of initial work as for prototype packages. This application of funds would be focused in areas where it is likely that longer-term support would be possible via either other grant funding or academic institutions—for example, development of a package that is known to be applicable for a particular observatory's instrument, if that observatory has already expressed interest in developing Python software but lacks funding for that area. In this way, the consulting framework would foster sustainability through the growth of funding for these efforts.

One of the most glaring deficits in the Astropy ecosystem is the underdevelopment of functionality for the analysis, reduction, and Python representation of astronomical spectra. Several years and multiple community-input workshops have produced an overall plan for what that functionality should look like, and some narrowly-focused development has occurred associated with specific missions. But critical elements of the plan still have not been implemented due to lack of funding and volunteer time. Users currently have to resort to using older, unmaintained tools in other languages, which is not sustainable moving forward. To address this shortcoming, we are requesting dedicated funding to support consulting developers who can work on building a robust and generally-applicable set of affiliated packages for the representation, reduction, and analysis of spectra.

In addition to the user-facing packages we provide, a critical part of the Astropy Project ecosystem is the infrastructure that we maintain for testing, documentation, and package management. While this infrastructure is functionally adequate, it needs a constant level of maintenance, and we need to continuously investigate issues and adapt our infrastructure to changes in the services or third-party packages that we use. For instance, we regularly need to investigate issues in the testing infrastructure that arise from new releases of other packages in the Python ecosystem or issues due to bugs in those packages. Another important part of our infrastructure that needs regular maintenance is a Python package template[3] that makes it convenient for scientists to get started with sharing their code and tries to encourage best practices, along with a set of helpers which facilitate installing packages, running tests, and building documentation—the package template has been used to set up over 350 repositories on GitHub and astropy-helpers is used in almost 400 repositories.

Maintaining astropy-helpers to resolve issues and stay compatible with the rest of the scientific Python ecosystem requires an investment of time that is currently not funded. In addition to maintenance, we regularly add new features and capabilities to different parts of the infrastructure in order to support the different packages in the Astropy ecosystem. For example, we have recently needed to update the astropy-helpers[4] package in order to make it possible for packages to compile C/Cython extensions using OpenMP for multi-threading, and we anticipate having to make similar improvements in the future. Finally, we believe that some effort should be spent in further generalizing as much of our infrastructure as possible so that it can be used beyond the Astropy Project, which would help improve sustainability as we could then share the maintenance burden with other projects. Maintaining and

[3] https://github.com/astropy/package-template
[4] https://github.com/astropy/astropy-helpers

extending all of this infrastructure requires a high level of expert effort which is currently not funded. We are therefore requesting funding to support five existing core contributors—who have previously been carrying out this work on a voluntary basis and are at risk of not being able to continue—for 25% of their time each.

Community infrastructure to ensure long-term contributor sustainability (*Output 1.2*)

As discussed above, one of the bottlenecks in the Project is the small number of core maintainers who are able to perform such tasks as fixing critical bugs, reviewing pull requests in detail, and maintaining the infrastructure. While our current group of core maintainers is small, there is in fact a large pool of individuals who are enthusiastic about contributing to the Project. The core package has had contributions from nearly 300 individuals to date, but many of these contributions are one-off additions that are relatively small. In order to improve the sustainability of the Project, we need to convert some of these one-time contributors into core maintainers, at the level where they can drive the development of a package, provide expert reviews, and help train more maintainers.

While we have made a good effort to run *user to first-time contributor* workshops, where we teach scientists how to use `git` and make their first pull request, we have not had the resources to then follow up in a manner that would bring promising contributors to the next level, where they could start to hold some responsibility in the maintenance of the package. We believe that a one-on-one mentoring program is needed to facilitate the transition from contributor to maintainer. The idea of this mentoring program is to identify contributors who we think would benefit from being paired up with a funded core maintainer, and ensuring that they have regular calls or meetings to go through any issues and discuss the maintenance of specific functionality together. A key point is for the mentoring process to be formally established as a compensated duty instead of a voluntary task. Along with this, we would allocate funds for travel so the mentor and mentee could interact personally (sprinting on code, learning about process) for intervals of up to a week. As with many scientific open source projects, the diversity amongst the core maintainers and team members does not reflect the diversity of the wider research community, and so we will prioritize improving diversity and inclusivity as one of the core goals of the proposed mentoring program.

A sustainable model for managing Astropy into the future (*Output 1.3*)

Since its inception, the development of the Astropy Project has been open and community-driven. There is no formal process for joining the Project, and anyone can participate in discussions or suggest changes to the code of any of the packages, regardless of geographical location or area of work. Contributions are peer-reviewed to maintain strict standards for package quality and stability. The Project is steered by a four-person coordination committee and we have defined a set of formalized roles in the Project that are currently filled by approximately 30 individuals[5]. These roles name, for example, a community engagement coordinator, release coordinators, and lead and deputy maintainers for various parts of the `astropy` core package. The aim of these roles is threefold: 1) it ensures that we know who has volunteered to be in charge of different aspects of the Project; 2) it allows us to see gaps

[5] http://www.astropy.org/team.html

in our management coverage; and 3) it allows contributors (junior scientists in particular) to get recognition for their contributions.

While the creation of these roles was a step in the right direction, a number of aspects of the Project management could and should be formalized, particularly in relation to the Project governance and the handling of funding. This is especially important since having a more formalized management structure has been cited as one of the aspects of the Project that was necessary but missing when we have applied for federal funding from the NSF in the past. We are therefore interested in paying a management consultant who has experience with scientific software projects and community efforts to advise us on and help us navigate the changes required to the Astropy Project that would enable us to receive and seamlessly distribute funding to our developers in a transparent way.

In addition to the management consultant, we would like to pay an individual to help on a part-time basis with project administration and management, specifically for tasks that fall outside of the scope of NumFOCUS responsibilities. For instance, this person would help us keep track of contractors work output, organize meetings, and help with Project-specific administrative details related to submitting grant proposals. Without additional funding this would continue to be done by a mix of the coordination committee and other Project members on a volunteer basis, which is becoming increasingly difficult to sustain.

2) Expertise in the Field:

Please describe your (or your organization's) expertise in this specific field and its relevance to the project.

Over the last few years, the Astropy Project has grown to become the largest open source collaboration in professional astronomy. With nearly 300 code-level contributors to the core package alone, Astropy has a broad and ever-growing developer base. Many of these contributors so far have been researchers and students who have applied their domain-specific knowledge to a particular relevant area of Astropy. While the core development team itself is a mix of such researchers and people who do software development as their primary day job, one of the main goals of the Project is to combine these areas of expertise together.

The proof of the applicability of this approach and expertise is given by its uptake in the astronomy community. At the large-scale, both the `astropy` core package and other parts of the ecosystem are used for a number of observatories or large missions. Examples include the Large Synoptic Survey Telescope (LSST), whose Science Pipelines include the `astropy` core package as a dependency, and NASA's flagship observatory, the James Webb Space Telescope (JWST), whose calibration pipelines and data analysis tools strictly depend on the `astropy` core package and some of the Astropy Affiliated Packages. Many other telescopes and missions that rely on Python are also using the Astropy ecosystem. At a more individual level, the first paper published about the Astropy Project[6] has been cited over 1,900 times (and is currently the third-most cited astronomy paper of 2013), which shows

[6] The Astropy Collaboration, A&A, 558 (2013) A33. DOI: 10.1051/0004-6361/201322068

that the Project is also widely used by a large body of active astronomy researchers. It is also the second-most popular software package on GitHub with the tag "science," demonstrating the broad interest in the Project as an open-development software project.

Astropy has also had an impact beyond astronomy—for example, it has been used by a team of astronomers and ecologists to help monitor rare endangered species[7], and is used by the Norwegian Mapping Authority as part of a set of tools for geodetic data analysis[8]. In addition, it has inspired similar projects in other fields, such as the PyMedPhys Project in medical physics[9].

3) Budget Justification:

a) Explanation: *Explain your budget request in further detail with justification.*

Funding for core/infrastructure developers

We are requesting $125,000 per year to contract with five core developers each at a ~25% level. These developers will be existing team members who have been performing critical work in terms of infrastructure and maintenance of the Project on a voluntary basis and are at risk of not being able to continue without funding. They will be responsible for continuing to maintain all of the infrastructure the Project uses, as well as provide timely reviews on pull requests whenever possible (liaising with other expert developers if needed). They will also be responsible for responding to issues and triaging them to in order to keep the number of issues at a manageable level. The cost for this is based on the typical salary for experienced software developers in the U.S. (~$100,000 for a full time job).

Funding for consulting developers

While the core/infrastructure developers mentioned above will be responsible for the smooth functioning of the Project overall, they will not necessarily always have the expertise to fix specific bugs, implement critical features, and carry out detailed reviews, depending on the pull requests. We are therefore requesting funding to be able to pay expert developers as consultants on an hourly basis. This funding could also be used to provide seed funding for developers to build prototypes of specialized packages that could then be used to apply for support from funding agencies. Unlike the core/infrastructure developers, for which there are clear candidates that could start to be paid from the time the funding becomes available, we want to ramp up the funding for consulting developers over the course of three years to give time for team members to adjust to the funding availability. We are therefore requesting funding for 200, 300, and 400 hours of development from year one to three. Assuming a consulting rate of $150 per hour, this means $30,000, $45,000 and $60,000 in years one, two, and three respectively.

---

[7] https://www.ljmu.ac.uk/about-us/news/articles/2018/4/4/astro-ecology
[8] https://kartverket.github.io/where/
[9] *"The PyMedPhys Project aims to develop a common core package for Medical Physics in Python and foster an ecosystem of interoperable Medical Physics packages. It is inspired by our physics peers in astronomy and their Astropy Project"* - https://pymedphys.com/en/stable/

GORDON & BETTY MOORE FOUNDATION
DATA-DRIVEN DISCOVERY GRANT DEVELOPMENT FORM

Funding for developers to work on spectroscopic functionality

We are requesting funding of $22,500 per year for consulting developers to work specifically on developing spectroscopic functionality, which has been identified as being the most important missing functionality from the Astropy Project, as well as being time-sensitive for the astronomy community due to impending retirement of an older generation of tools. Assuming a rate of $150 per hour, this will provide 450 hours of development in total, which will be sufficient to build the foundations for Astropy Affiliated Packages for the representation, reduction, and analysis of spectra.

Funding for contributor workshops and workshop/mentoring coordinator

The establishment of a mentoring program will be critical to training new core maintainers for packages in the Astropy Project. We therefore propose to fund a coordinator who will design and run the mentoring program. Assuming that this is done on a consulting basis, at a rate of $150 per hour, and working on average two hours a week, we request $15,000 per year.

In addition, since face-to-face interactions are often highly valuable when training new maintainers, we are requesting $10,000 per year to cover travel for mentors and mentees to meet—either individually or at joint events where several mentors and mentees can work in a common environment for a few days. Since we are interested in improving the geographical diversity of the Project, we are assuming an average cost of $2,000 for travel and accommodation, which would allow us to support travel for five mentors and/or mentees each year.

Finally, we are requesting $6,000 per year in addition for existing core maintainers to travel internationally (either to existing conferences or to specific institutions) to continue running workshops to train and recruit new first-time contributors. Assuming $2,000 per trip, this will allow us to run three workshops every year, and with ~10 participants at each workshop, we will train 90 new contributors over the course of the funding period—from which we could then mentor the most promising contributors to become core maintainers as mentioned above.

Management consulting

We request $6,000 per year to cover the costs of a management consultant (assuming a rate of $150 per hour and 40 hours per year) who will help us formalize the management structure of the Project. We have already identified a person who is recently retired from managing a large astronomy project who would be well suited for this task and has come highly recommended.

Project administration/management

We request $19,500 per year (assuming a rate of $150 per hour and 130 hours per year) to support an individual who will handle project administration management tasks that fall outside of the scope of NumFOCUS responsibilities.

GORDON & BETTY MOORE FOUNDATION
DATA-DRIVEN DISCOVERY GRANT DEVELOPMENT FORM

Astropy coordination meeting

We request $24,000 per year to support an annual five day, face-to-face Project Coordination Meeting (two days), code sprint (two days), and management training (one day) for 30 attendees. Not all attendees will require travel support from this award and this request includes full travel support for 12 participants, assuming an average cost of $2,000 per participant (since some participants will be traveling internationally).

b) Outside support: *State other sources of support you already have, have applied to, or are expecting to receive.*

The only direct funding that the Astropy Project has received to date is a contract with the American Astronomical Society (AAS) to support specifically the development of time series functionality in the Astropy Project. The Project currently also receives indirect support from the Space Telescope Science Institute to develop and maintain educational resources such as tutorials and documentation.

Budget

Include a budget using the provided Excel spreadsheet template. Other budget formats will not be accepted. Please note that overhead costs must not exceed 12.5% of the budget.

Organization Details & Grant Contacts

Organization Details

Address:	NumFOCUS, P.O. Box 90596, Austin, TX 78709
Website:	https://numfocus.org/
Phone:	512-831-2870
Fax:	N/A

Signatory

An Executive or Board Member who would sign the legal contract between GBMF and the grantee. For universities, this may include an officer from the Office of Sponsored Projects.

Name:	Leah Silen
Title:	Executive Director
Email:	leah@numfocus.org
Phone:	512-831-2870
Fax:	N/A

GORDON & BETTY MOORE FOUNDATION
DATA-DRIVEN DISCOVERY GRANT DEVELOPMENT FORM

Finance Lead

Person who is knowledgeable of the institution/department's budget and finance policies.

Name: Lynn Brubaker
Title: Project Finance Manager
Email: lynn@numfocus.org
Phone: 512-337-9573
Fax: N/A

Project Lead

Person who will be in charge of the programmatic content of the project

Name: Erik Tollerud
Title: Astropy Coordination Committee Member
Email: etollerud@stsci.edu
Phone: 410-338-6761
Fax: N/A

Other

Any other person(s) who should be copied on important matters relating to the legal, financial, or programmatic aspects of the grant. This may include other scientists, other administrators, and/or the person who submits financial reports (if this is not the finance lead).

Name: Tom Aldcroft, Kelle Cruz, Thomas Robitaille
Title: Astropy Project Coordination Committee
Email: coordinators@astropy.org
Phone:
Fax: