PLUTOPY: AN OPEN-SOURCE COMMUNITY FOR LEARNING REPRODUCIBLE PLANETARY SCIENCE.

C. J. Tai Udovicic[1] and A. L. Boivin[2], [1]Department of Physics and Astronomy, Northern Arizona University, Flagstaff, AZ (cjt347@nau.edu) [2]Solar System Exploration Group, Department of Earth Sciences, University of Toronto, Toronto, ON, Canada.

Introduction: Planetary science is a data intensive field, and the tools used to process and extract valuable insights from data are constantly evolving. The rapid increase in the volume and complexity of available planetary data has given rise to a proportionate increase in the volume and complexity of the computational workflows used to process that data. Combined with a lack of code review during publication, this often leads to scientific results which are difficult or impossible to reproduce [1].

The Plutopy community is a new initiative to educate planetary scientists about the tools available to facilitate high quality, reproducible planetary science. It imparts best practices of writing repeatable planetary data analysis through a combination of self-directed learning, worked examples, hands-on collaboration, and community generated feedback. This learning community is built as an open-source collaboration and encourages members to contribute their domain knowledge back to community, which will make it robust to changes in industry tools and standards.

Plutopy: The community exists as a Git repository, a collection of files which is shared via the popular Git source control management tool. Plutopy is hosted on GitHub, a thriving Git repository hosting platform which also hosts several tools used by the planetary science community, including ISIS3, GDAL, astropy, scikit-learn, etc. [2-5]. The primary programming language used for data analysis is Python, which has gained significant traction in planetary science for reasons described in [6-7]. The data analysis and worked examples are presented as Jupyter (formerly IPython) notebooks, which combine narrative text, code, and figures into a legible file which can be shared and executed line-by-line (see Fig 1) [8-9]. Jupyter notebooks have quickly gained popularity as tools for sharing reproducible scientific research, as they allow for full transparency of the code that produces figures from raw data [9-10].

Scientific value: The Plutopy repository is structured as an open-source Python package used analyze New Horizons mission data [11]. While this repository could result in software that is of scientific value to planetary science, this is not the primary intent. The code repository format was chosen to expose the development process of software tools that planetary scientists use on a day-to-day basis (e.g., [2-5] among others), and to serve as an example repository for any

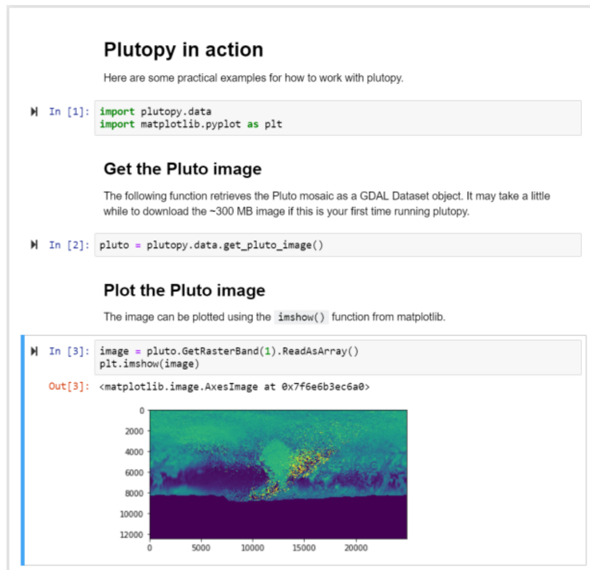

Fig 1: A Jupyter notebook explaining how to display an image using the plutopy Python package.

future collaborations, supplementary material for publications, or other shared analysis code that Plutopy members may write in the future. Plutopy is released under the MIT open license, so that all code and examples are free to be re-used and distributed.

Learning outcomes: A typical participation cycle of a new Plutopy member occurs in 5 steps (see Fig 2).

1) Git basics: Learn about the Git source control management tool and how to contribute to an open-source repository. Example task: copy the Plutopy repository using Git, add name to the contributors list, then submit the updated file to the main repository on GitHub.
2) Scientific coding with Python: Practice using Python and common planetary data analysis packages to analyze Pluto image data. Example task: extract an interesting region from an image of Pluto and plot it using Python, then contribute the new code to the repository.
3) Reproducible analysis with Jupyter: Use a Jupyter notebook to present the results of analysis done with Plutopy. Example task: Detail in a Jupyter notebook the steps needed to apply a noise reduction filter to an image and then plot the noisy and smoothed images side-by-side. Contribute the notebook to the repository to grow the collection of worked examples.

4) Open-source collaboration with GitHub: Become a Plutopy collaborator and practice Git collaboration skills by mentoring more junior members. Example task: review contribution requests of junior members or draw on domain experience to add new functionality that junior members can use in steps 2 and 3.
5) Producing reproducible research: Incorporate best practices learned in each of the previous steps to active planetary research. Example task: Use the Plutopy worked examples as a reference for writing reproducible analysis for active research, or create a code repository like Plutopy to share work with collaborators or to supplement a publication.

Results: Plutopy is currently live and in its early stages at https://github.com/cjtu/plutopy/. The first plutopy members joined as part of a scientific coding course run at Northern Arizona University in the fall of 2018. All 6 members were able to copy the repository and submit their first contributions to Plutopy, thus completing step 1. Most members then began work on the image analysis task, but as of writing of this abstract, no member has submitted a contribution for step 2. On-going work to improve the Plutopy landing page and to support members in making their first scientific coding contributions is under way.

Conclusion: The Plutopy repository leverages modern software and data analysis tools to teach planetary scientists how to conduct high quality reproducible research. Through practical experience collaborating on a planetary data analysis project, Plutopy members will learn how to build their own reproducible analysis that they can share with the community or use to strengthen their publications. The format of the learning is entirely self-paced, and as the repository grows, more members will contribute modules and examples to improve the learning outcomes of more junior members. In time, Plutopy has the potential to be a powerful resource and interactive learning community for any planetary scientists interested in improving the reproducibility of their work.

Future Work: As an active open-source community, Plutopy will be built upon by the members who participate in it. One challenge that this project could face is engaging senior members in providing feedback and improvements to the repository. In future work, we will develop a project roadmap that will outline goals for the repository so that all members have a common task to work toward while learning the skills they need to improve their own work.

The authors are also actively seeking senior planetary scientists interested in technology education in order to improve the learning outcomes of new members. The on-going feedback from Plutopy members will inform how the repository evolves to better suit the needs of the planetary science community.

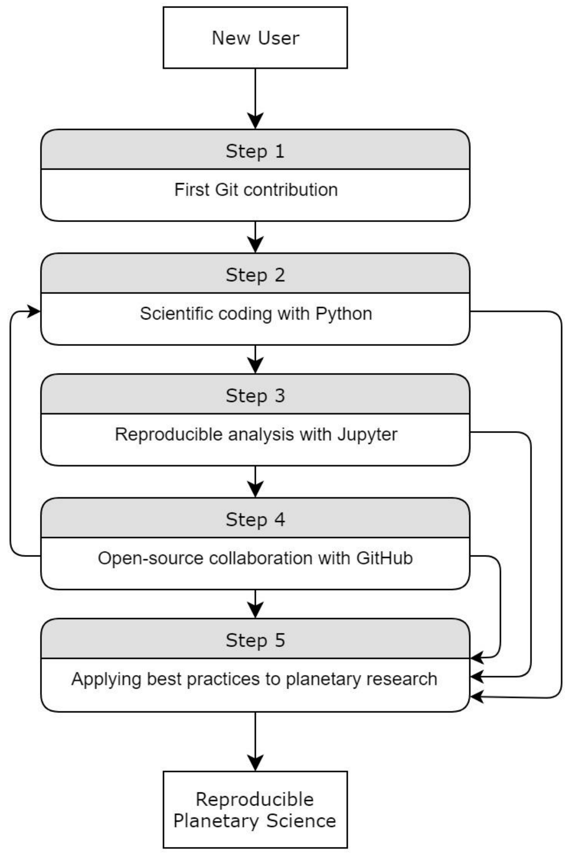

Fig 2: Participation cycle of a Plutopy member.

References: [1] Peng R. (2011) Science, 1226-1227. [2] Anderson J. A., et al. (2004) LPSC XXXV, Abstract #2039. [3] GDAL/OGR contributors (2018) OSGeo Foundation. [4] Robitaille T., et al. (2013) A&A, A33. [5] Pedregosa F. et al. (2011) JMLR, 2825-2830. [6] Laura J. et al. (2013) LPSC XLIV, Abstract #2226. [7] Laura J. et al. (2016) LPSC XLVI, Abstract #2208. [8] Perez F. and Granger B. (2007) Comput. Sci. Eng., 21-29. [9] Kluyver T., et al. (2016) ELPUB, 87-90. [10] Piccolo S. R. and Frampton M. B. (2016) GigaScience, 30. [11] Moore J. M., et al. (2016) Science, 1284-1293.