

Community Concerns on the Adoption of Open Science in Astrophysics

APAC Discussion

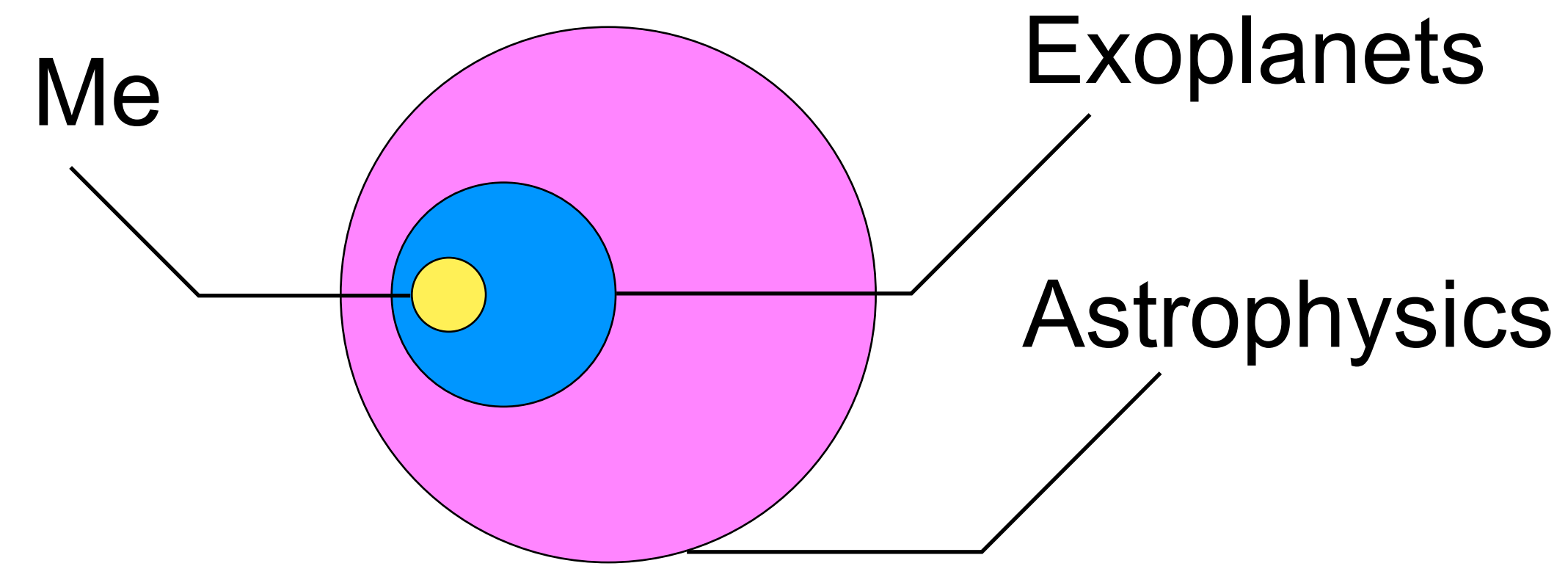


Natasha E. Batalha (NASA Ames Research Center) March 21, 2023

My OS experience: TOPS Champion, former Project Scientist for TOPS OpenCore, led OpenCore Model on Open Results, creator of many OS tools/codes, Planetary Data System User Community

OpenCore: <https://nasa.github.io/Transform-to-Open-Science/take-os101/>

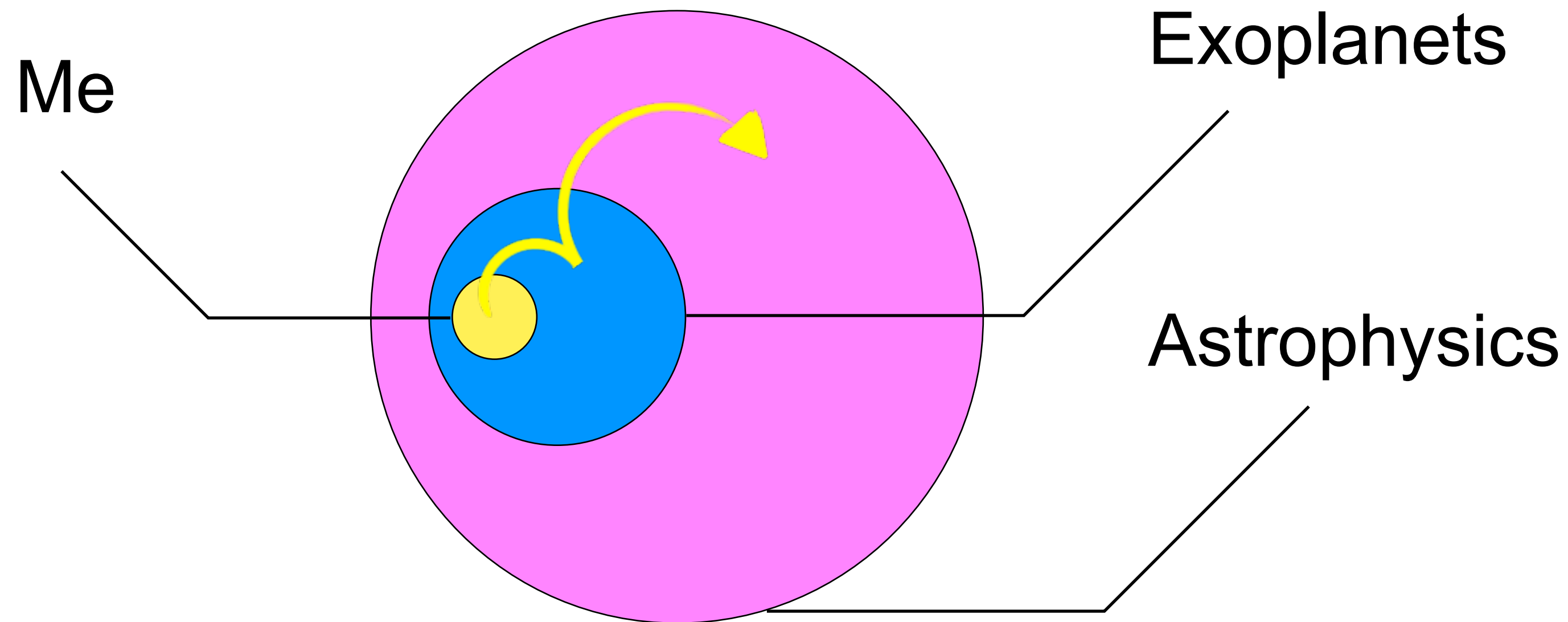
Where have I gathered these community concerns?



	Area of expertise	Career level	Where do you receive funding	Where you receive data	Where you receive tools
M	Exoplanet atmospheres	PhD: 2017 Perm. Job: 2019	JWST, HST, ROSES (XRP, PDART, ICAR, TCAN)	Create: models Consume: JWST, HST, HITRAN	Create: open software Consume: open software
E	Exoplanets	Skewed ECR	Space, ground, ROSES, NSF	Open and closed dbs (skewed open)	Open and closed tools (skewed open)
A	All astrophysics	All levels		Open and closed dbs	Open and closed tools

***Other factors that influence community thoughts:** where you publish, residing country, BIPOC identification, institution classification, exposure to OS principles, and more

Where have I gathered these community concerns?



The following talk *is not* a statistical aggregation of thoughts and concerns by the astrophysics community.

This talk *is* a collection of thoughts I've heard from various OS related outreach (TOPS OpenCore development, providing OpenCore trainings, OS forums, community input solicitation via forms, TOPS Champion) all subject to my own interpretation

Four Major Concerns Facing Adoption of OS

***with all aforementioned caveats**

1. Concern that there are no realistic avenues to support maintenance/infrastructure of open source software, open database, libraries
2. Concern that the development of OS requirements have not been met with the development of new incentives for adoption
3. Concern that there are no standards regarding the definition and requirements of “an open reproducible paper”
4. Concern with open data, elimination of exclusive access periods and being scooping

Concern:

No realistic avenues to support maintenance/infrastructure of open source software, open database, libraries

Option 1: F14. High Priority Open-Source Science & F7. Support for Open-Source Tools, Frameworks, and Libraries

- Very high bar to entry for Open Science knowledge
- Codes have to have demonstrated usage. E.g. major strengths and weakness from a rejected proposal for a code used within exoplanets:

MS: "The proposal clearly described how <the code> directly addresses the goals of NASA SMD and how the code is being used in their community"; MW: "<the code> does not have significant usage in the NASA community"

These funding elements are for super-duper open-science developer

Concern:

No realistic avenues to support maintenance/infrastructure of open source software, open database, libraries

Option 2: F8. Supplemental Open Source Software Awards

- The program award amount (~\$50k for 1 year) is not enough for an average community member to make meaningful headway on OS program goals (e.g. turning existing legacy software into modern code, publicly releasing new software package, etc)
- A non-OS expert might have a vision for how they want to implement OS to their framework, but might not know how to execute this.

Common questions include: 1) who do I hire? 2) if I hire a grad student or an intern, how do I advise them? 3) are grad students even the right people to be doing this work? 4) am I expected to continue maintaining this code after 1 year?

Supplemental awards may not meaningfully help people adopt OS

Proposed “Solutions”:

No realistic avenues to support maintenance/infrastructure of open source software, open database, libraries

- Create an opportunities like “NASA GPU Hackathon” where over the course of 3 days open science experts are paired with mentees that want to improve their open science infrastructure. Should be geared toward novices in OS. Should engage both students and advisors together. There is **huge** benefit to having expert advice over 3 day period.
- Instead of getting ~\$50k through a supplemental award, create an option to buy out “OS expert” for certain FTE levels. In this way community can rely on cost-shared support and have direct access to the right kind of expertise. Would allow for more rapid progress in a 1 year time frame.
- Better expertise would enable development of resources that had longer shelf-life. NASA has to acknowledge that most codes will not have long term maintenance.
- Educate astrophysics community (partly being done through TOPS OpenCore) on what the “minimum viable product” is for OS compliance

Concern:

Development of OS requirements have not been met with the development of new incentives for adoption

TOPS OpenCore is trying to educate community surrounding the benefit to OS adoption but it does not directly create incentives.

Consider these two groups writing papers:

- Group 1: rapidly develops half-baked script with small number of authors and writes paper with results (effort required: 1 lead author, 3 contributing authors)
- Group 2: coordinates effort to develop OS software with checks for reproducibility, writes paper with better results (effort required: 1 lead author, often dozens of contributors)

Current incentive structures (e.g. postdoc applications) reward “Group 1” model of science and penalize contributors to “Group 2” model

Concern:

Development of OS requirements have not been met with the development of new incentives for adoption

TOPS OpenCore is trying to educate community surrounding the benefit to OS adoption but it does not directly create incentives.

Consider these two groups writing proposals:

- Group 1: Develop new method and release niche OS code. Write proposal to do science leveraging code.
- Group 2: Writes proposal leveraging new method by “Group 1” to do science.

Current proposal structures (e.g. DAPR) do not prioritize Group 1 for funding over Group 2. Group 2 has more to write proposal while Group 1 maintains code base.

Proposed “Solutions”:

Development of OS requirements have not been met with the development of new incentives for adoption

- Add avenue for students/postdocs to articulate their commitment and contributions to OS in NASA Postdoctoral Program & Hubble Program (create model for other postdoctoral programs)
- Add avenue for proposers to articulate their commitment and contributions to OS in ROSES (other than OSDMP which is not used in scientific eval)
- Professional development societies (e.g. AAS, DPS) should create career awards for contributions to OS
- For top-down cultural shift we need to continue pushing community to take training via TOPS OpenCore, especially senior leaders (initial trainees are skewed ECR). Will help with teaching folks how to credit contributions, when to share to ensure credit can be received, etc.

Concern:

No standards and requirements regarding the definition of “an open reproducible paper”

TOPS OpenCore provides definition for reproducibility: *Ability to recreate the same results as the original study (including tables, figures, and quantitative findings), using the same input data, computational methods, and conditions of analysis. (FORRT www.forrt.org)*

- This definition is hugely different depending on your specific view of “ability to recreate”. Consider:

Group “*click and run all*”: complies with reproducibility standard by creating Google Colab Jupyter notebook that reproduces original study from start to finish

Group “*dump and run*”: complies reproducibility by posting code on Zenodo that can “theoretically” be run but would require huge work effort to spin up, rerun, and reproduce figures

Both groups have technically complied with SPD-41, the definition for reproducibility, could easily get papers published.

Proposed “Solutions”:

No standards and requirements regarding the definition of “an open reproducible paper”

- Through TOPST/ScienceCore or other avenues, coordinate the development of reproducibility standards that are community specific by expertise or observational or computational (this has been done in other communities e.g. [“Geoscience Paper of the Future”](#)).
- These efforts should be coordinated with major publishers so that reproducibility standards can be enforced via peer review and data editors
- These reproducibility standards could also offer tiers of compliance. Realistically we do not all have to create papers that utilize stunning user-friendly libraries. However, for people that go “above and beyond” it should be explicit they are (e.g.) “Tier 1 compliant”.
- Creating clear hierarchies would allow non-experts and students the ability to evaluate reproducibility of papers and provide additional incentive structure for OS adoption

Concern:

Open data, elimination of exclusive access periods and being scooped

With fully open data and the elimination of an exclusive access period, a team other than the original proposing/awarded team could produce a publication with no involvement, acknowledgment, or credit given to that original team.

Because of the cultural emphasis we still place on *novelty*, this could be particularly detrimental to:

- 1) early career researchers
- 2) members of marginalized groups within the research community
- 3) members of non-R1 institutions
- 4) quality of scientific papers

Incentive structures are not aligned to combat detrimental effects potentially associated with elimination of EAPs

Concern:

Open data, elimination of exclusive access periods and being scooping

Fear of being scooped also applies to creators of models and code. Without any “period of exclusivity” for new modeling/code advancements, people who spend time writing code do not have enough time to benefit from open code development efforts.

Consider PhD student who spent 4 years developing novel code and now plans on using the code during postdoctoral time to establish themselves in the field.

Incentive structures are not aligned for model providers to benefit more from releasing open code, opposed to keeping code to themselves

Proposed “Solutions”:

Open data, elimination of exclusive access periods and being scooping

- Consider an avenue that allows for fully open data with no EAP, but still leaves “first to publish” rights with the original proposing team within an certain period (e.g. 1 year). Would require new “open-but-restructured-use” data license that would force publishers to check for compliance with license. Overall, community benefits from open data access. Could create similar model with software.
- Leverage OpenCore or other training materials to educate and create trust in communities (this is the documented way in which other communities have assuaged the fear of scooping (Laine 2017)). *Though, concerns that many of our communities are far too large to “self-police”.*

Four Major Concerns Facing Adoption of OS

***with all aforementioned caveats**

1. Concern that there are no realistic avenues to support maintenance/infrastructure of open source software, open database, libraries
2. Concern that the development of OS requirements have not been met with the development of new incentives for adoption
3. Concern that there are no standards regarding the definition and requirements of “an open reproducible paper”
4. Concern with open data, elimination of exclusive access periods and being scooping

And these are all fully addressable!