

Collaboration Challenges and Opportunities in Developing Scientific Open-Source Software Ecosystems: A Case Study on Astropy



Jiayi Sun¹



Dr. Aarya Patil²



Youhai Li¹



Dr. Jin Guo³



Dr. Shurui Zhou¹

¹University of Toronto

²Max Planck Institute for Astronomy

³McGill University



UNIVERSITY OF
TORONTO



MAX PLANCK INSTITUTE
FOR ASTRONOMY



McGill
UNIVERSITY



Scientific Open Source Software (Sci-OSS)

Scientific Software: “source **code files**, **algorithms**, **scripts**, computational **workflows**, and **executables** created during the research process or for a research purpose”.



Astronomy



Drug discovery



Seismology



Genomics



Meteorology

Scientific OSS: Scientific software developed openly and collaboratively, with source code freely available for use, modification, and contribution.

Event Horizon Telescope (EHT) Project

M87 - the first image of a black hole

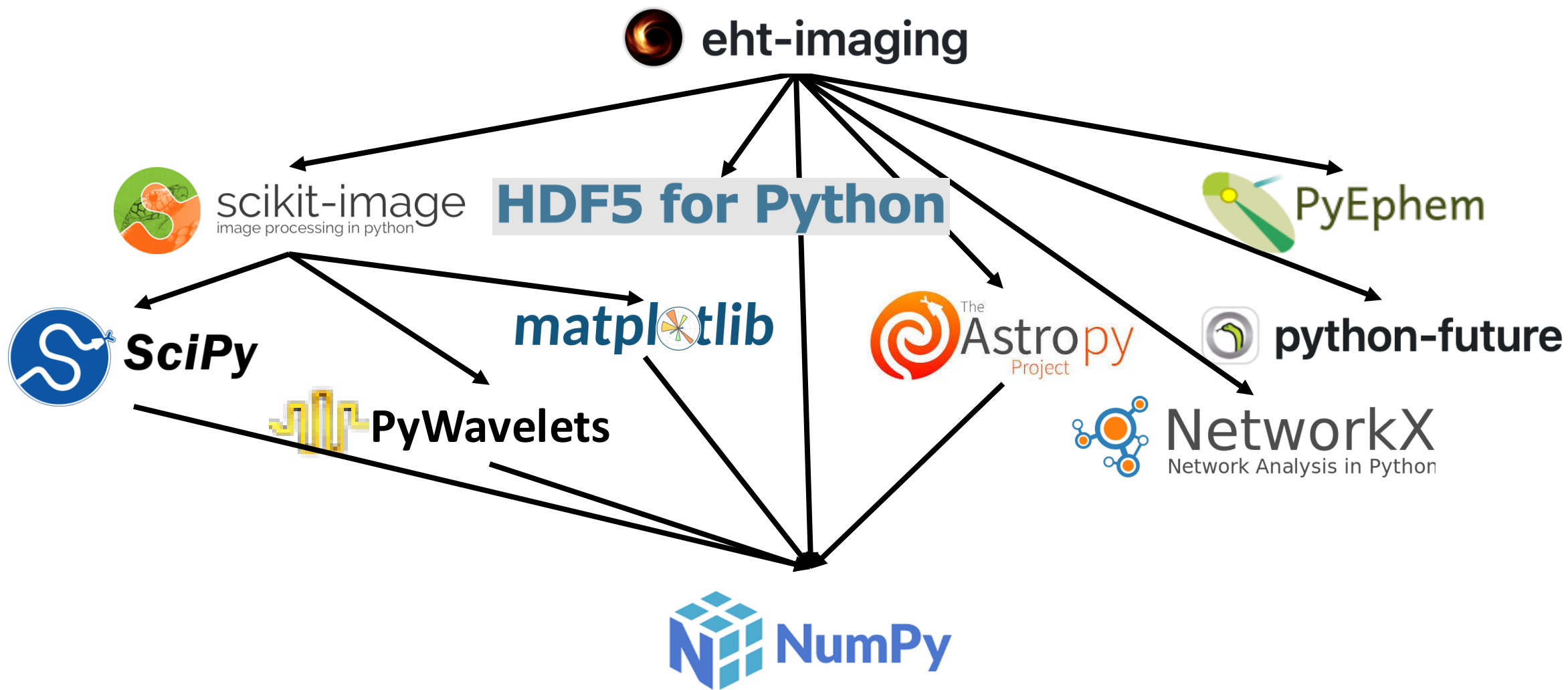


Event Horizon Telescope (EHT) Project

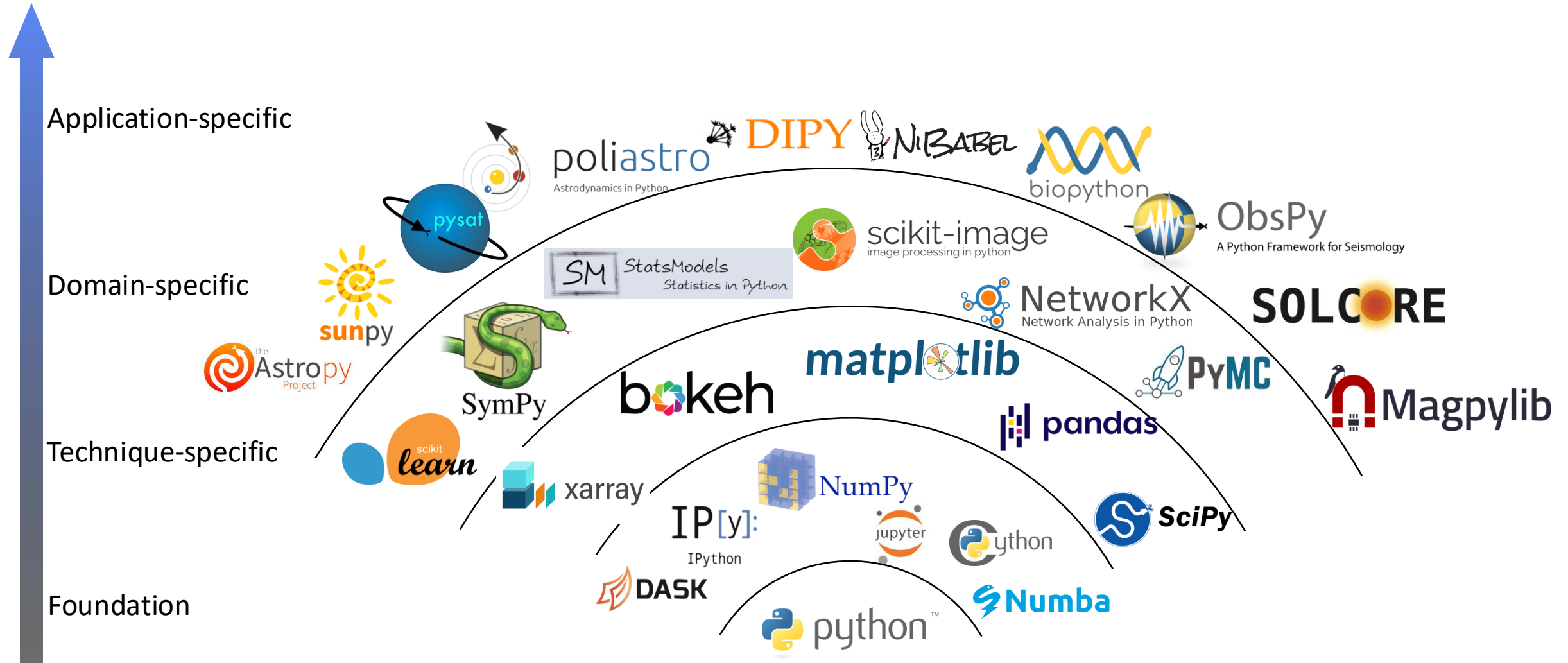
M87 - the first image of a black hole



Sci-OSS in EHT Project



Landscape of Scientific OSS Ecosystem

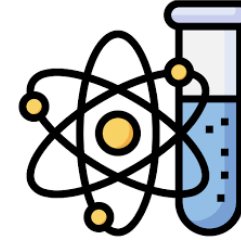
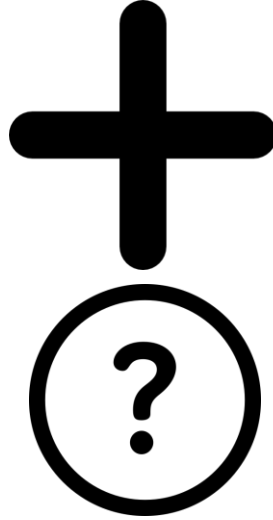


Adopted from Jake VanderPlas. 2017 | Harris., et al.. 2020

Challenges on OSS and Scientific Software



OSS
development



Scientific
software



Distributed & voluntary



Burnout



Toxicity



Complex dependency

Lack of SE training



Communication barriers



Incentives misalignment



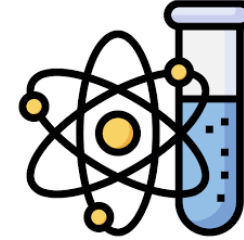
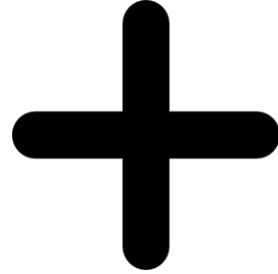
Unevaluated “extra work”



What We Don't Yet Know



OSS
development



Scientific
software

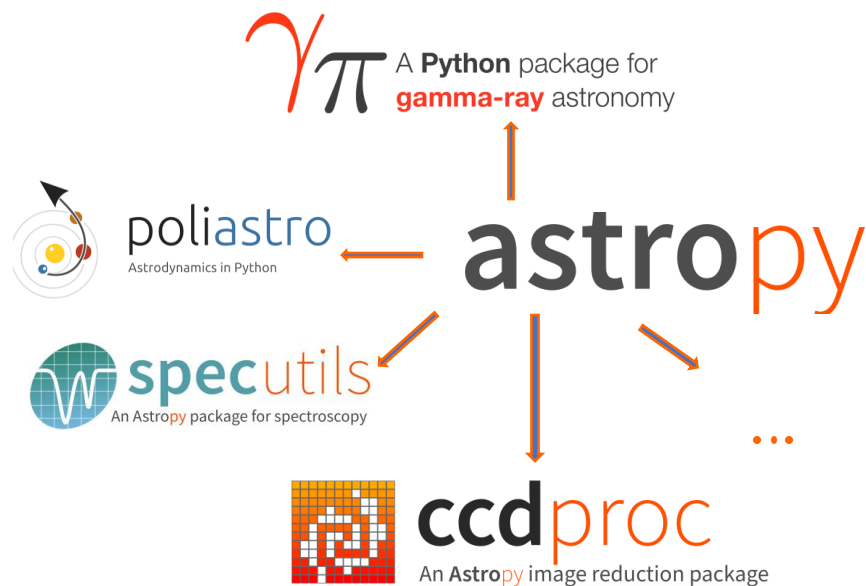
- Prior work mostly focuses on **small, co-located teams** working on single projects
- General OSS collaboration studies often miss **interdisciplinary challenges** unique to science
- **Ecosystem level collaboration** limited to dependency related issues

We Conducted a Case Study

to explore the collaborations in scientific OSS ecosystems in depth



Study Subject -- The **Astropy** Project



- Python software ecosystem for astronomy

astropy



Fork

1.9k



Star

4.6k

- One core package:
 - ≥ 10 years of age
 - > 400 contributors
- 50 other interoperable packages



Maturity



Community-oriented



Popularity



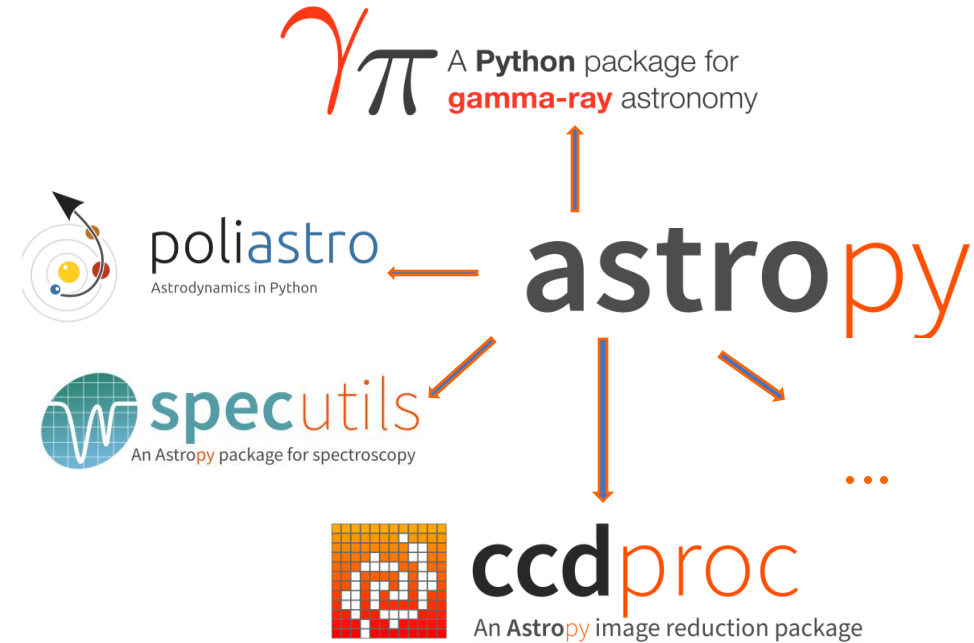
Development artifacts

Research Questions



RQ1: (within-project collaboration)

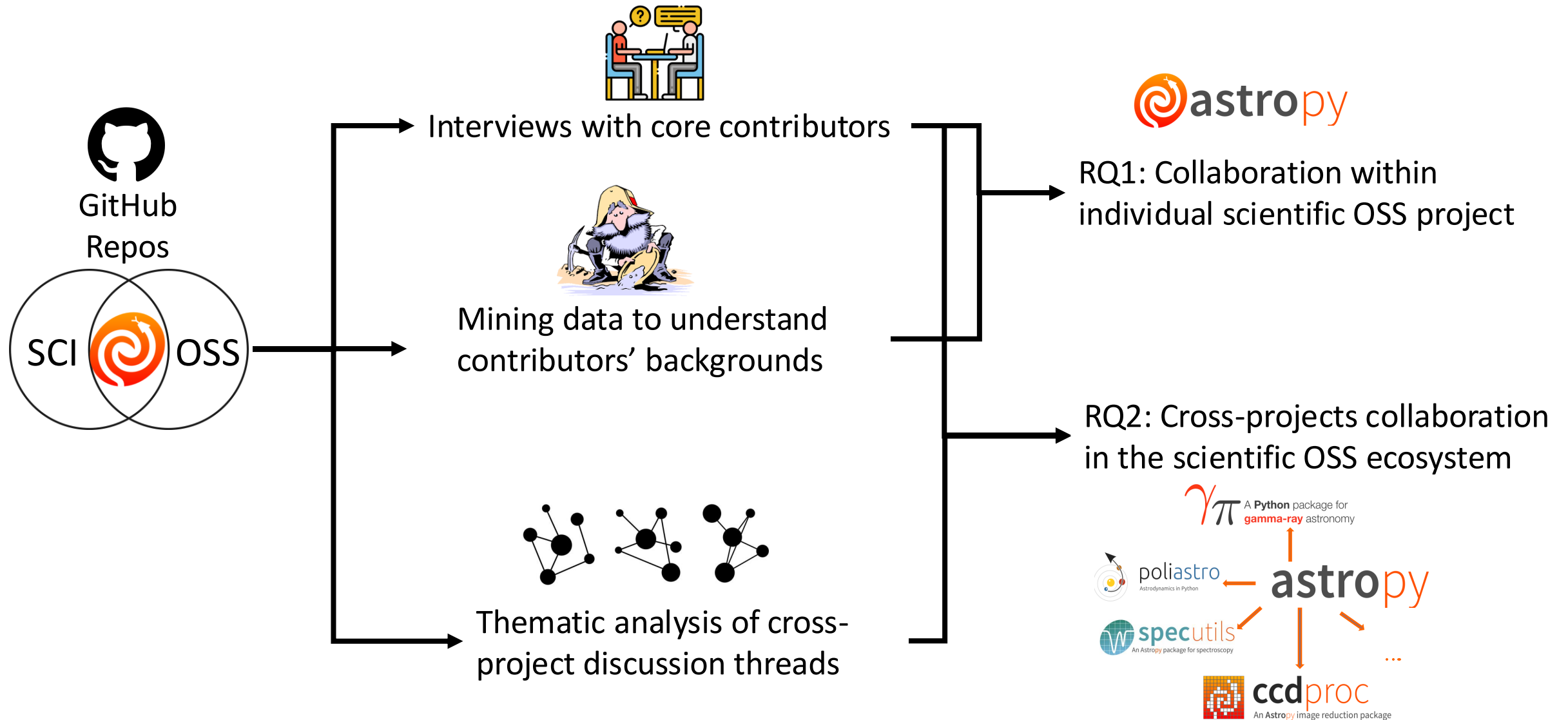
What are the major challenges when interdisciplinary experts collaborate to develop and maintain scientific OSS in the Astropy?



RQ2: (cross-project collaboration)

What are the intentions and the corresponding challenges in cross-project collaboration within the Astropy ecosystem?

Method Overview

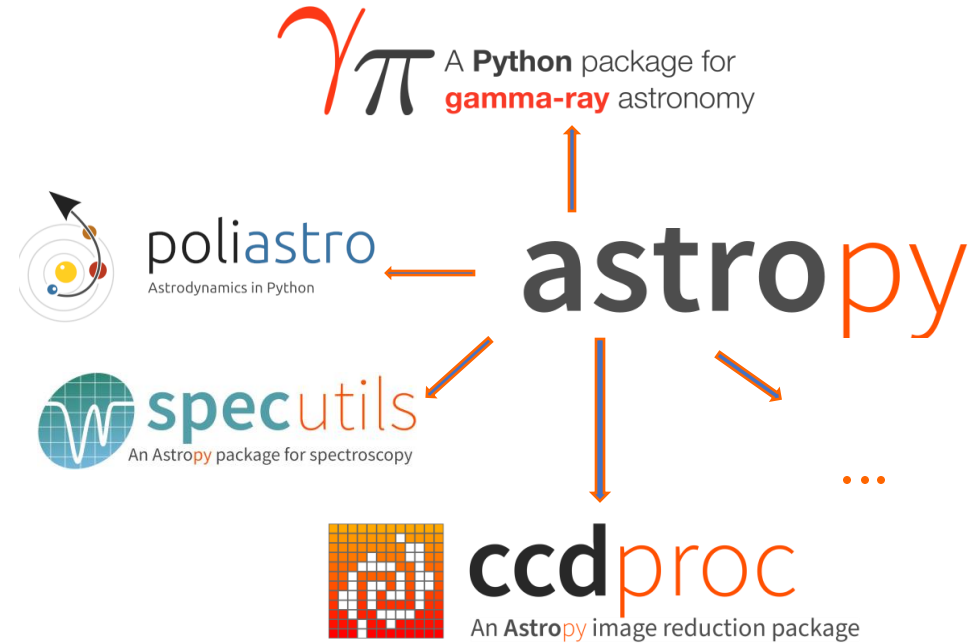


Research Questions



RQ1: (within-project collaboration)

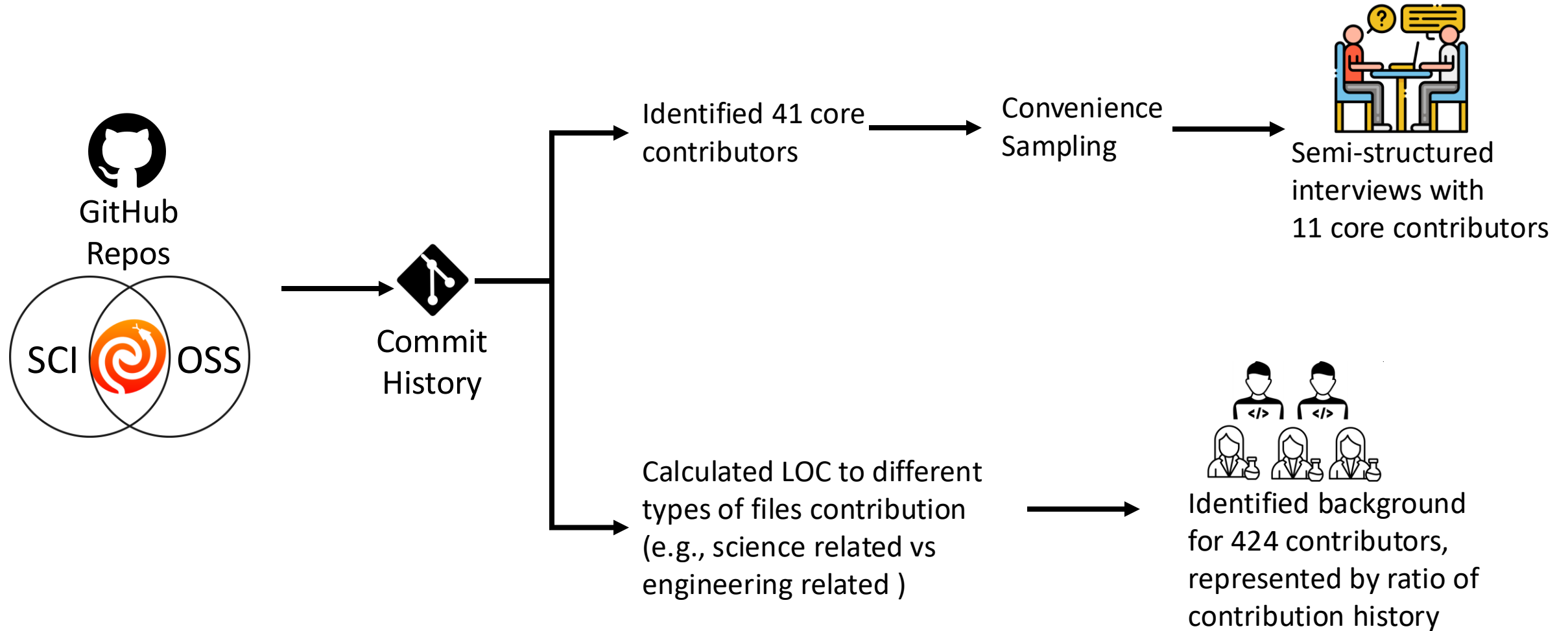
What are the major challenges when interdisciplinary experts collaborate to develop and maintain scientific OSS in the Astropy?



RQ2: (cross-project collaboration)

What are the intentions and the corresponding challenges in cross-project collaboration within the Astropy ecosystem?

RQ1 Method: Within-project Collaboration

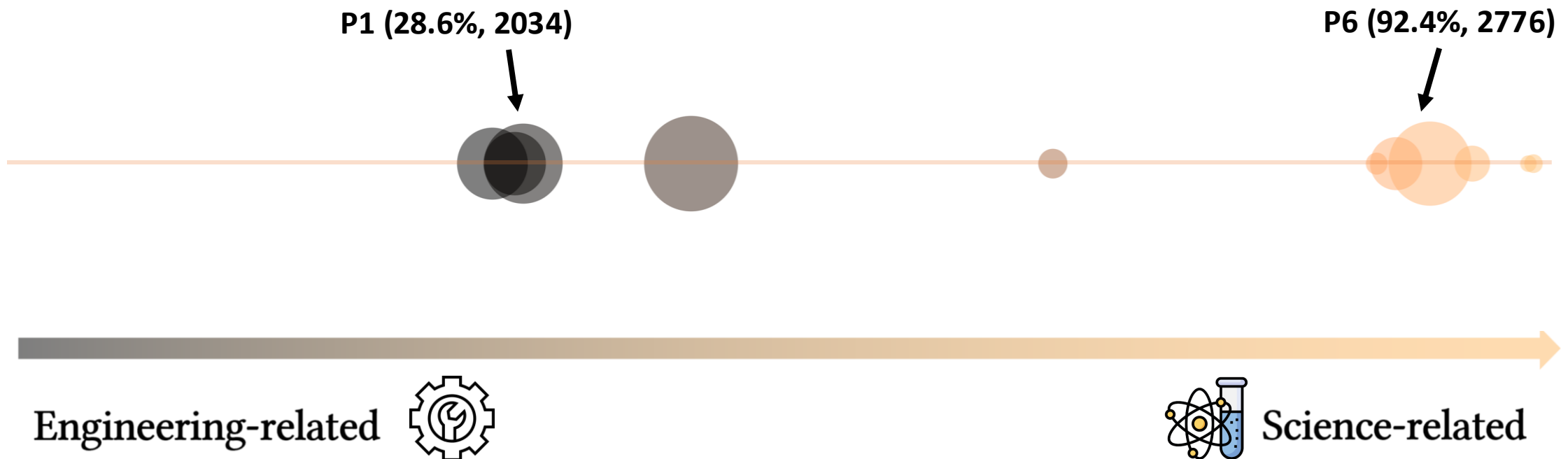


RQ1 Interdisciplinary Team Composition

11 core contributors interviewed:

- Scientists with self-taught software skills to professional software developers with scientific background.

Dot size: #commits
X-axis: Ratio of LOC contribution to code files



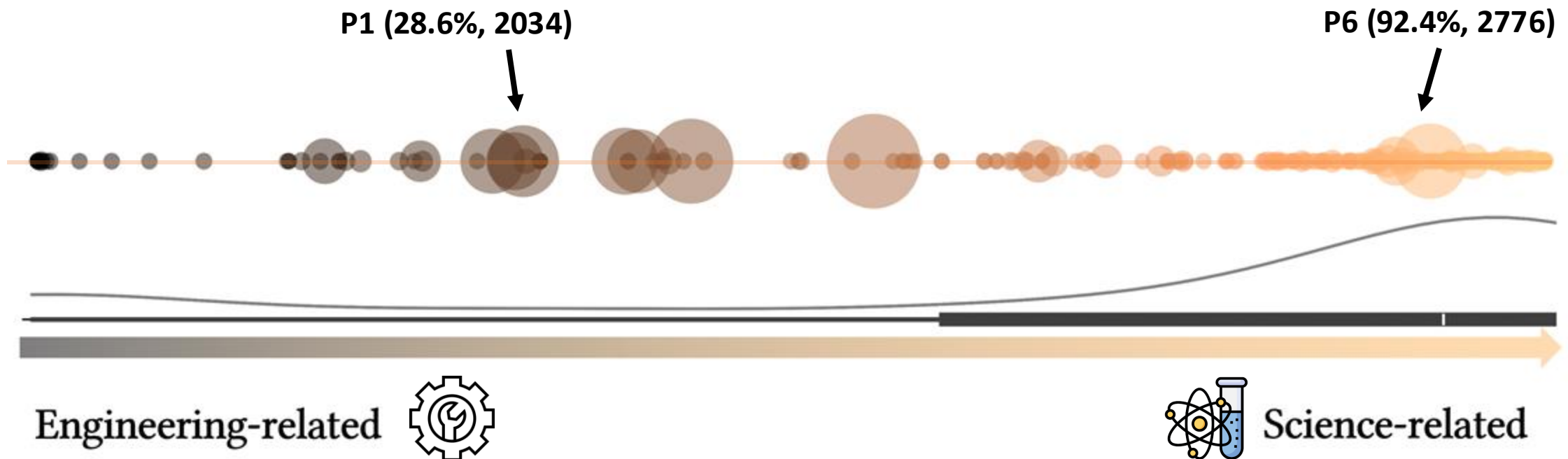
RQ1 Interdisciplinary Team Composition

11 core contributors interviewed:

- Scientists with self-taught software skills to professional software developers with scientific background.
- Contributors' backgrounds span across a spectrum of expertise

Dot size: #commits

X-axis: Ratio of LOC contribution to code files



RQ1 Reduced Communication Barriers

Deep domain knowledge requirement,
even for scientists.

*“Astronomy is a very big field, even though it seems small... if a **pull request** say “we want to update the **world coordinate system (WCS)**”, you have to **read at least three papers to even understand what that one was about.**”*

Interviewee with a
Msc deg. in astronomy



Use of the FITS World Coordinate System by STEREO/SECCHI

W.T. Thompson · K. Wei

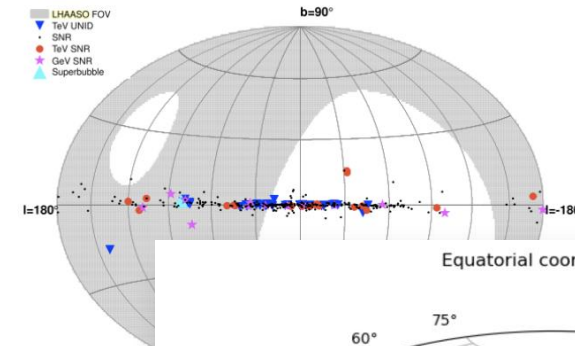
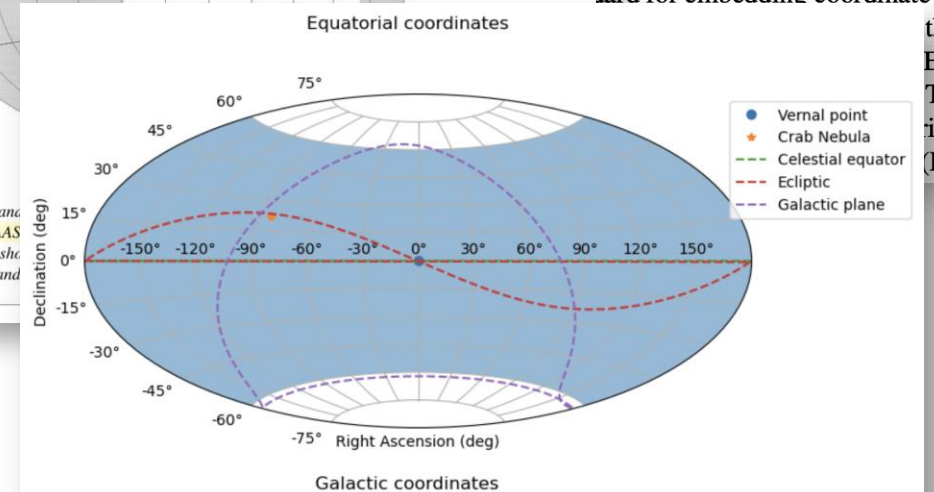


Figure 10: Locations of SNRs and field-of-view of LHAASO and magenta stars show TeV γ -ray sources, and bands.



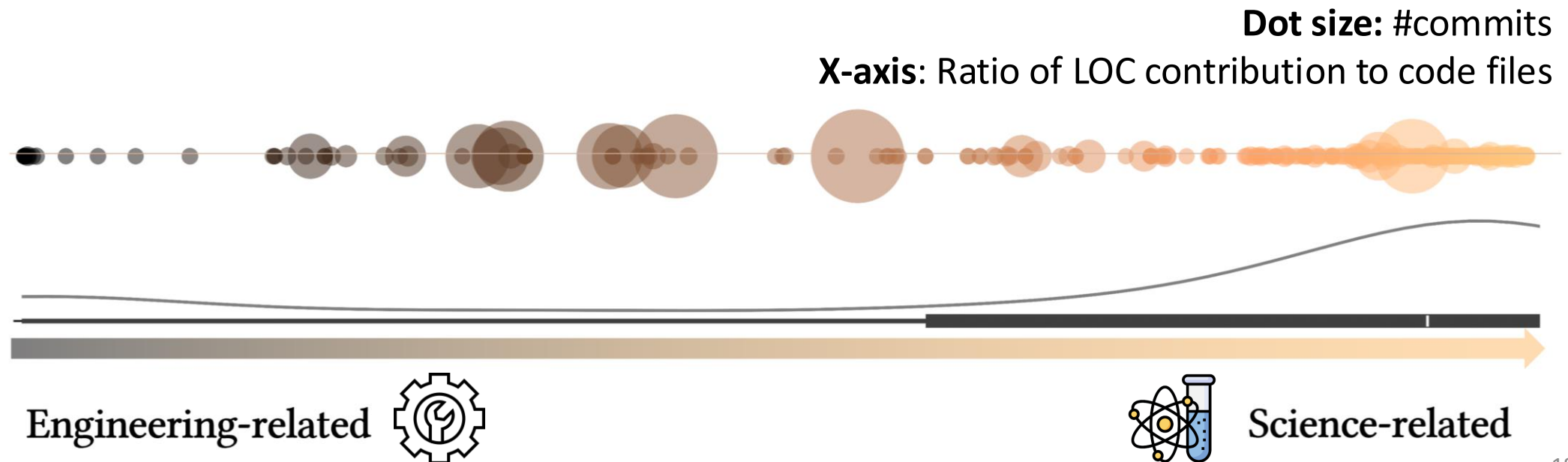
line: 15 December 2009

ard for embedding coordinate in-
thin
EC-
The
ical
(HI)

RQ1 Reduced Communication Barriers

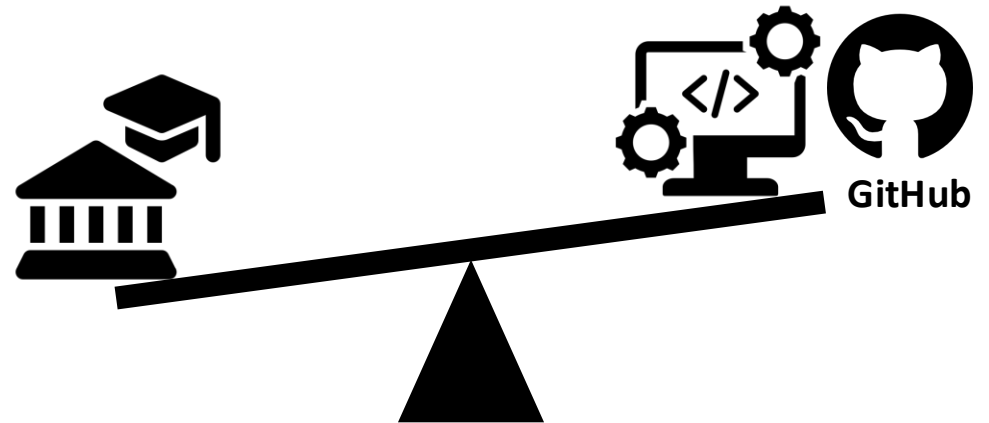


*“there is **always someone you can talk to** who understands the domain knowledge of what you are working on but also has a lot of knowledge in the software structure.”*



RQ1 Tensions between Different Mindsets

- **Task prioritization:** code quality vs flexibility of scientific collaboration
- **Perception of seniority:** academic culture can influence decision making in OSS development

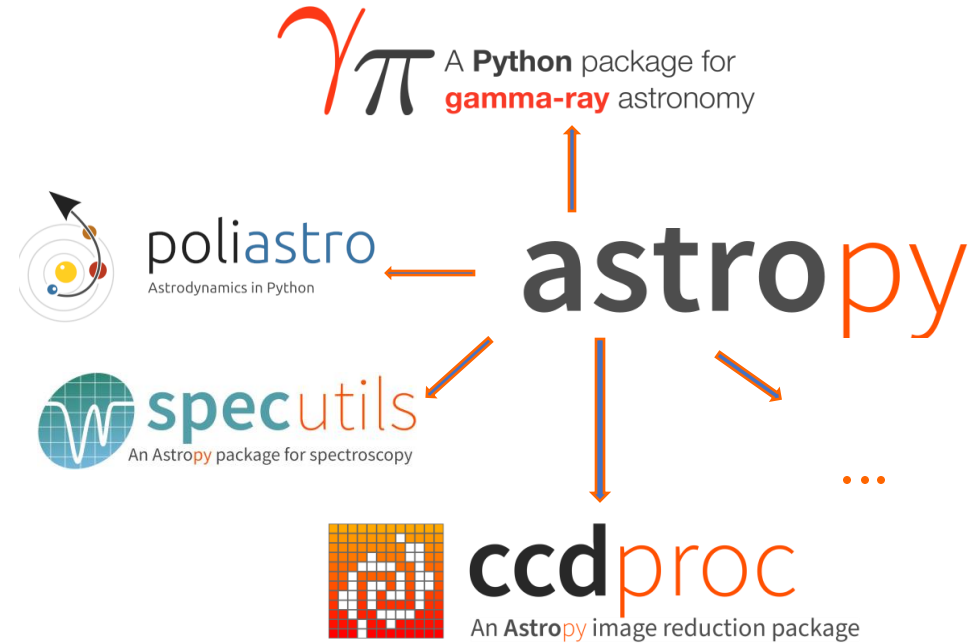


Research Questions



RQ1: (within-project collaboration)

What are the major challenges when interdisciplinary experts collaborate to develop and maintain scientific OSS in the Astropy?



RQ2: (cross-project collaboration)

What are the intentions and the corresponding challenges in cross-project collaboration within the Astropy ecosystem?

RQ2 Method: Cross-project Collaboration

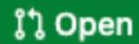


RQ2 Cross Reference Links

NASA-Planetary-Science/sbpy



Add Dastcom5 Service #115



shreyasbapat wants to merge 16 commits into **NASA-Planetary-Science:main** from



Conversation 64



Commits 16



Checks 0



Files changed 8



shreyasbapat commented on Feb 27, 2019 • edited ▾

As discussed in chats of astroquery: <https://astropy.slack.com/archives/C8U8VGQFM/p>

And from a PR in astroquery: [astropy/astroquery#1339](#)



I am adding this module here.



Sbpy #115



**Astroquery
#1339**

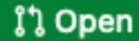


RQ2 Cross-Project Reference Graph (CRG)

NASA-Planetary-Science/sbpy



Add Dastcom5 Service #115



Open

shreyasbapat wants to merge 16 commits into **NASA-Planetary-Science:main** from

Conversation 64

Commits 16

Checks 0

Files changed 8



shreyasbapat commented on Feb 27, 2019 • edited

As discussed in chats of astroquery: <https://astropy.slack.com/archives/C8U8VGQFM/p>

And from a PR in astroquery: [astropy/astroquery#1339](#)



I am adding this module here.

... other events ...



This was referenced on Feb 27, 2019

Add Dastcom5 Module [astropy/astroquery#1339](#)



Contribute NEOs code to astroquery [poliastro/poliastro#364](#)

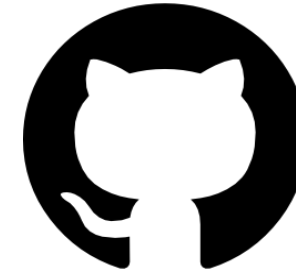


... other events ...



astrojuanlu mentioned this pull request on Mar 29, 2019

[WIP] Remove Neos from poliastro forever [poliastro/poliastro#605](#)



GitHub

Sbpy #115

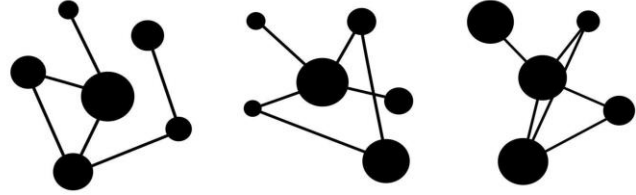
Poliastro
#364

Poliastro
#605

Astroquery
#1339



RQ2 Cross-Project Reference Graph (CRG)



- **node**: issues/PRs
- **edge**: directional cross reference link (source issue/PR -> target issue/PR)
- **node color**: repo (project)



GitHub

Sbpy #115

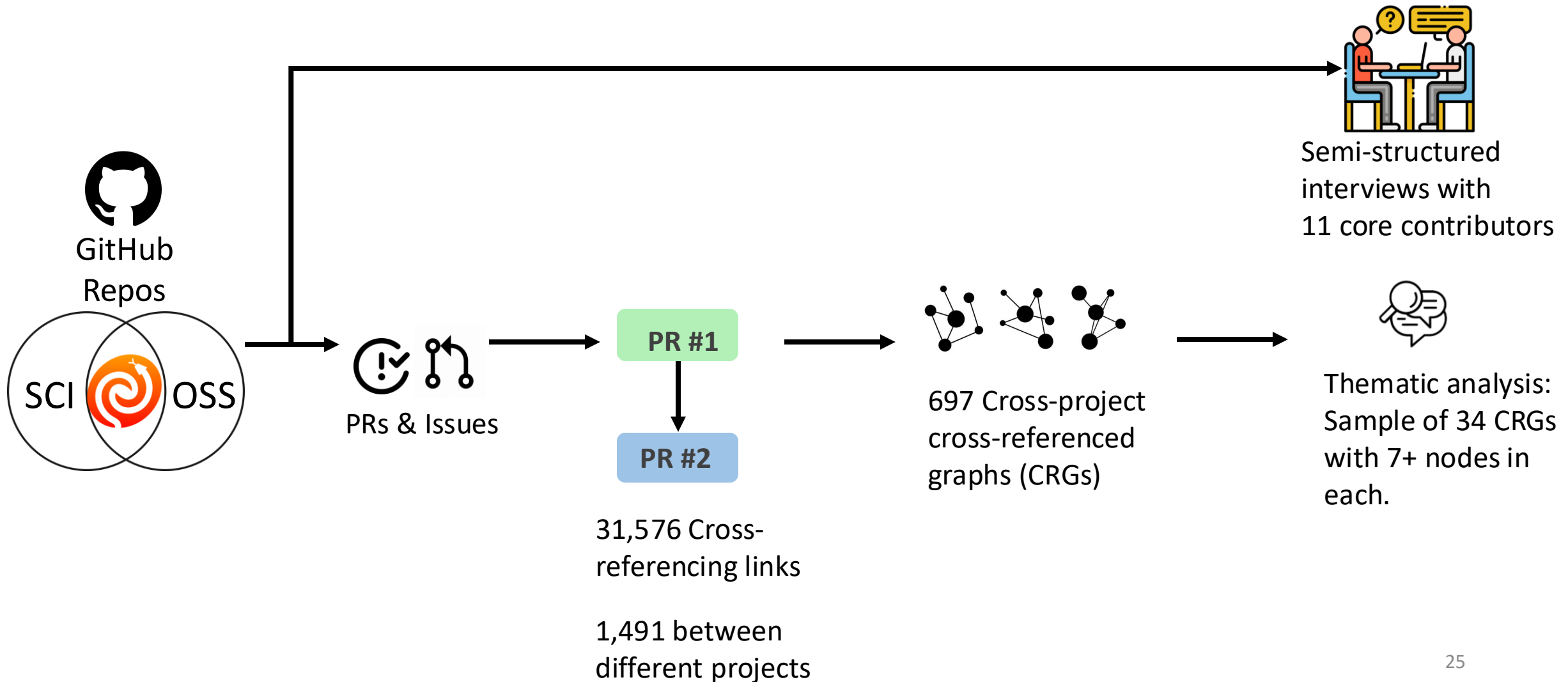
Astroquery
#1339



Poliastro
#364

Poliastro
#605

RQ2 Method: Cross-project Collaboration



RQ2: Collaboration Intentions



Intention		Occur. (ratio)	Example
Change coordination due to code dependencies		39.63%	Reporting bugs
Knowledge sharing		26.5%	Discussing solutions for Astronomy specific problems (e.g., design for “coordinate system reference frames”)
Coordinating shared functionalities		25.14%	Migrating features between projects
Centralized infrastructure management		8.74%	Batch update of shared testing tools
Coordinating during release cycles			Gather feedback on new features

RO2: Collaboration Intentions

■ Add ecliptic frames and transforms #3749

 Merged eteq merged 15 commits into `astropy:master` from `eteq:add-ecliptic-frames`

 Conversation 26  Commits 15  Checks 0  Files changed 8



eteq commented on May 3, 2015



Member ...

This closes [#1430](#) by adding frames (and transforms into the rest of the coordinate graph) for ecliptic coordinates.

As implemented I before merging it coordinates, bec such coordinates [@mtbannister](#), @ you can commen



astrojuanlu commented on Aug 31, 2017

I *suspect* (but cannot prove) that the NEOS lookup database of JPL uses heliocentric *mean* ecliptic coordinates:

<http://ssd.jpl.nasa.gov/sbdb.cgi?sstr=2003122>

We had some discussion about it in poliaastro, which is summarized here:

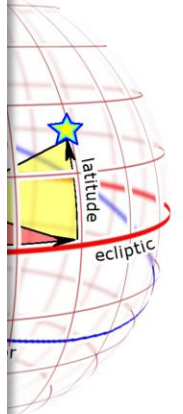
<https://github.com/poliastro/poliastro/wiki/Reference-frames#small-bodies-coordinates>



imple

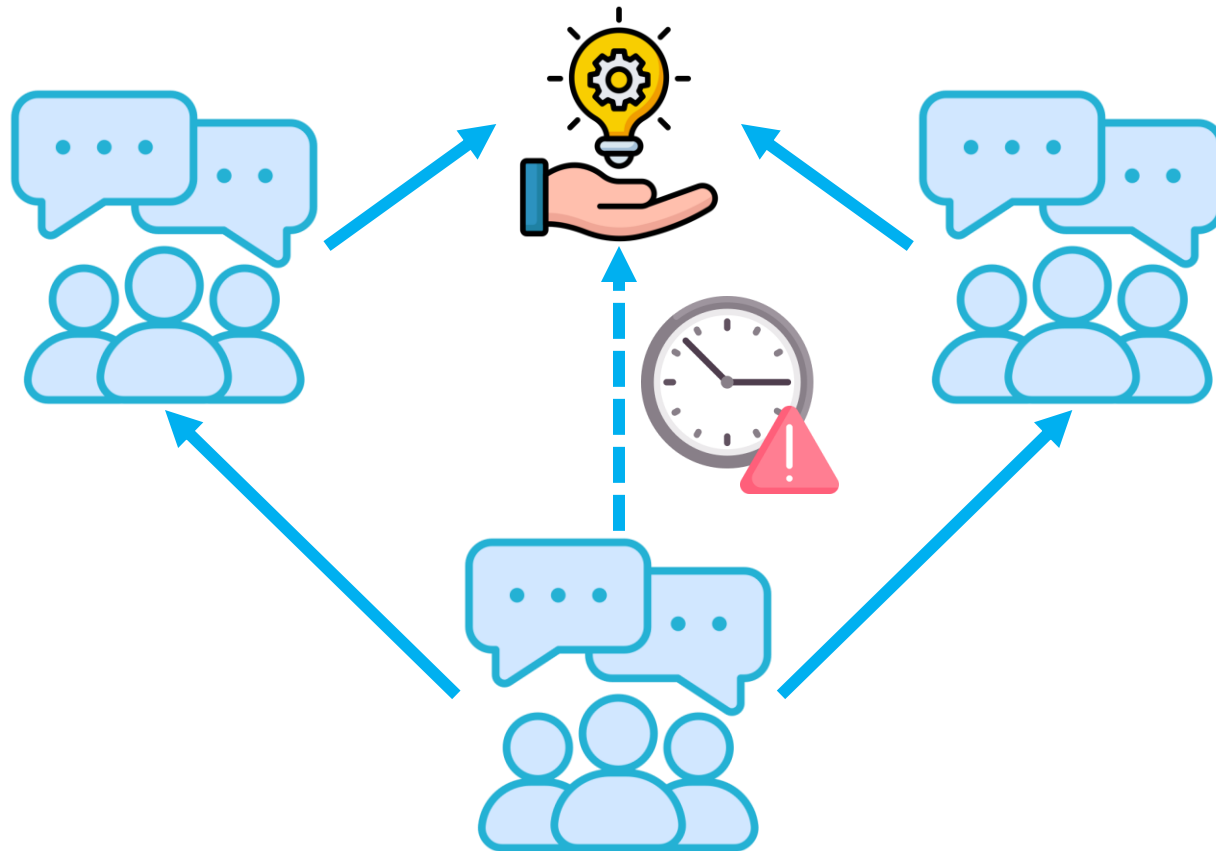
orting bugs

Discussing solutions for Astronomy specific problems (e.g., design for “coordinate system



RQ2: Collaboration Challenges

Information among scattered issue threads



RQ2: Collaboration Challenges

Information among scattered issue threads



Fitting/Modeling/Sigma Clipping Crashes with NaNs, whereas before it just gave a warning #1345

✓ Closed

orifox opened this issue on May 13, 2022 · 21 comments · Fixed by #1350



orifox commented on May 13, 2022 · edited by pllim

I'm working on PSF Fitting using Photutils. My script previously worked (astropy == 5.0.4) worked. I upgraded astropy, and now it fails. Although I'm not exactly sure why. Error message below. Nothing has changed.

Now (astropy == 5.2.dev75+ga3c27114b):

Assignees

No one assigned

Labels

bug

Projects

photutils
1345

RQ2: Collaboration Challenges

Information among scattered issue threads



Fitting/Modeling/Sigma Clipping Crashes with NaNs, whereas before it just gave a warning #1345

Closed orifox opened this issue on May 13, 2022 · 21 comments · Fixed by #1350



orifox commented on May 13, 2022 · edited by pilim

I'm working on PSF Fitting using Photutils. My script previously worked (astropy == 5.0.4) worked. I upgraded astropy, and now it fails. Although I'm not exactly sure why. Error message below. Nothing has changed.

Now (astropy == 5.2.dev75+ga3c27114b):

Assignees

No one assigned

Labels

bug

Projects

larrybradley commented on May 16, 2022 · edited

Member



IMHO, I think the sensible thing to do here is have the fitting code automatically remove the non-finite values (with a warning). I see that @astrofrog proposed exactly that 7 years ago ([astropy/astropy#3575](#)), but apparently it never went anywhere.



photutils
1345



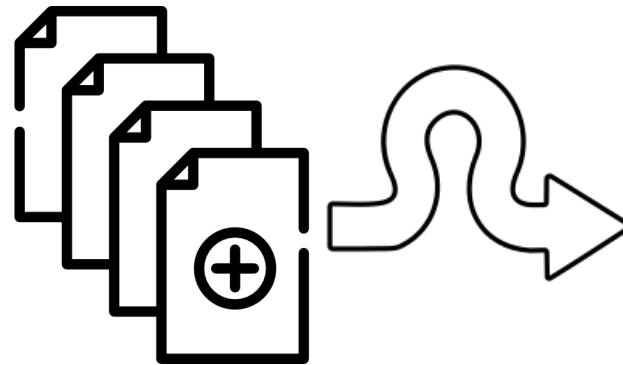
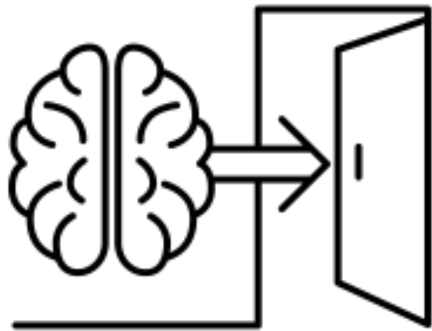
astropy
3575

$\Delta t = 7$ years!

RQ2: Collaboration Challenges



- Information among scattered issue threads
 - Knowledge loss If not for long-term contributors
 - Difficult to keep track across projects within the ecosystem
- Fragmented, duplicated implementations, and sub-optimal domain-specific solutions
- Delayed responses from downstream packages



Takeaway: Within & Cross-project Collaboration

RQ1 Within-project

- Reduced communication barrier
- Different mindset tensions on task prioritization
- Academic seniority influences decision making

RQ2 Cross-project

- Cross-project collaboration goes beyond just dependency-related issues
- Knowledge sharing of existing expertise
- Hard to manage scattered knowledge



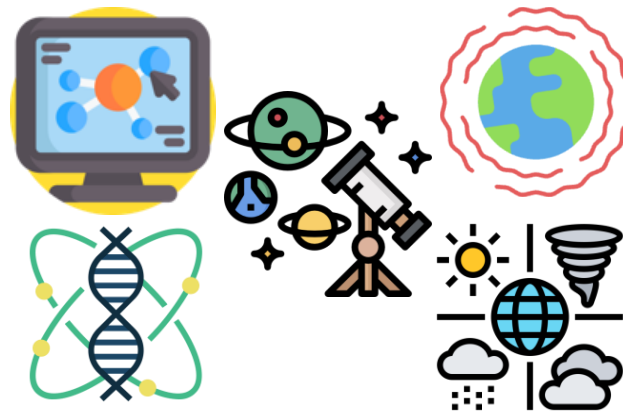
Open questions for future study

- Understanding of expertise-tasks matching on sci-OSS
- Tooling or practices to support balancing tension?

- Strategies for connecting communities
- Tools to better provide overview of the ecosystem

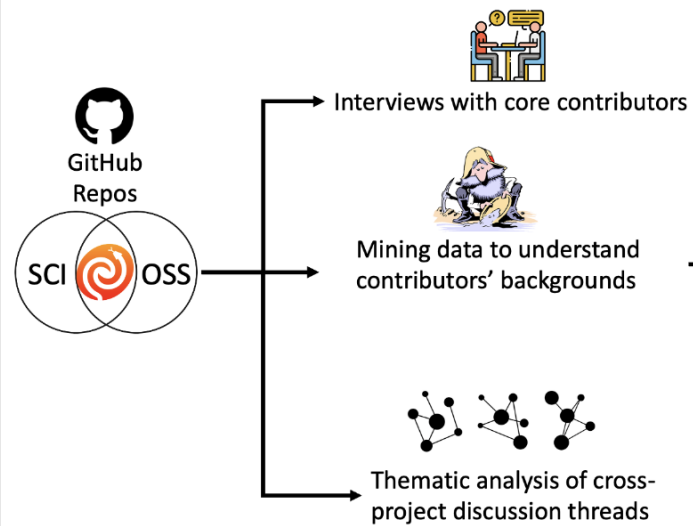
What's Next?

- What are the perspectives from short-term and disengaged contributors?
- Beyond the case study in Astronomy, does the observation generalize to other scientific domains?
- With the recent advances in LLM for coding tasks, does it help scientists with software development tasks?



Collaboration Challenges and Opportunities in Developing Scientific Open-Source Software Ecosystems: A Case Study on Astropy

Method Overview



Takeaway: Within & Cross-project Collaboration

RQ1 Within-project

- Reduced communication barrier
- Different mindset tensions on task prioritization
- Academic seniority influences decision making

Open questions for future study

- Understanding of expertise-tasks matching on sci-OSS
- Tooling or practices to support balancing tension?

RQ1 Within-project

- Cross-project collaboration goes beyond just dependency-related issues
- Knowledge sharing of existing expertise
- Hard to manage scattered knowledge



← **Contact me &
read our paper!**

Jiayi Sun
jiayisaria.sun@mail.utoronto.ca