Reflections on our experience moving to open source pan-pharma code collaboration on clinical trial reporting tools

James Black, Associate Director, Insights Engineering
Data and Statistical Sciences, Roche Pharma
james.black.jb2@roche.com

May 1st - 4th | PhUSE US Connect 2022 | Atlanta

What does open-source look like in late-stage pharma?

We are well on track to transition to an R based (*multi-lingual??*) backbone for clinical trial reporting

Open source software (OSS) is becoming common for internally resourced and funded tooling

How we used to interact with OSS

Using packages like 

Pushing out statistical packages into the ether

Occasional collaboration and often only ad-hoc

The new normal

Officially **resourcing and securing internal funding** for teams to work on open source R packages

More of **our talent is naturally flowing to**, or starting, ad-hoc open source projects

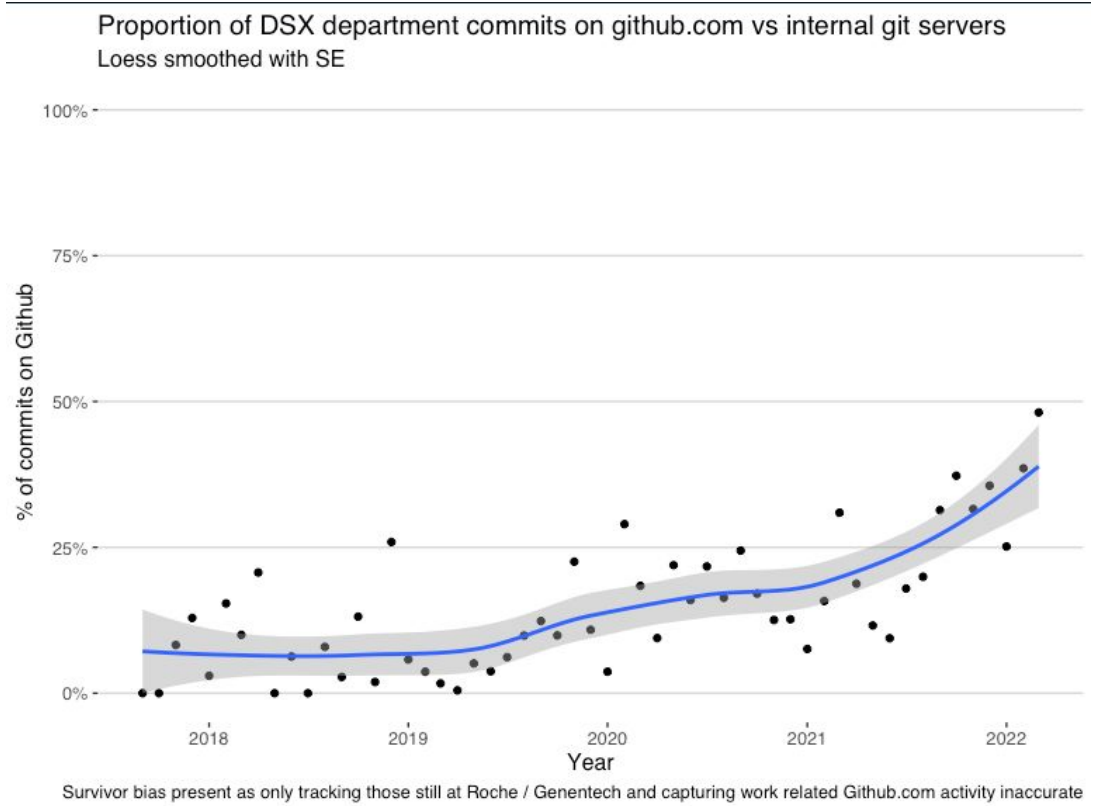
Co-creating packages that are core to workflows

Increase in the release of packages, including formal collaborations appearing on statistical packages (e.g. )

Disclaimer

The following data is based on the Github and Gitlab APIs, and has not been cleaned. Numbers are indicative rather than accurate.

We are gradually doing more on GitHub



Survivor bias present as only tracking those still at Roche / Genentech and capturing work related Github.com activity inaccurate

Four repositories to represent 4 types of projects

Part of Roche's clinical reporting strategy

8

10 yrs

25

1.2yrs

77

5yrs

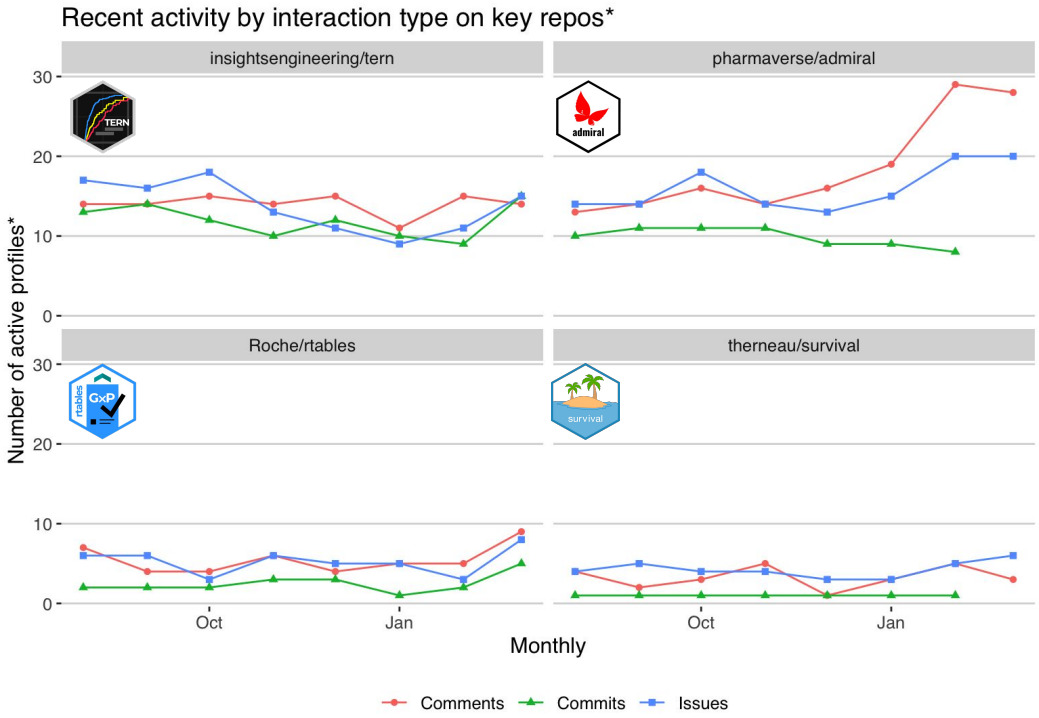
21

4yrs

Open source ‘health’ is about people and communities

Devs ≈ Reporters, active but plateaued package

Small active package



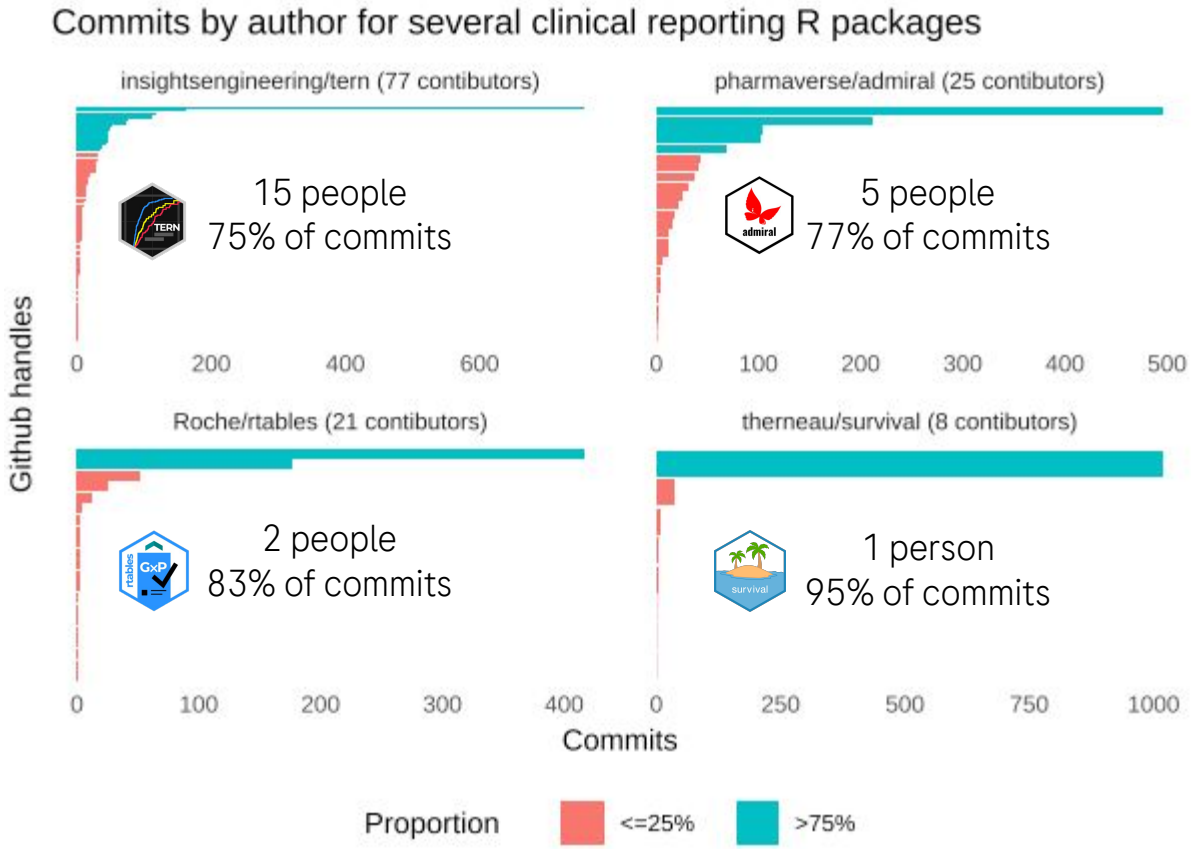
*Github profiles used where possible, otherwise unique emails where commits made do not match github profiles

Devs < Reporters, growing, active project

Small stable package

What have we learnt?

People can sometimes mean *a person*...

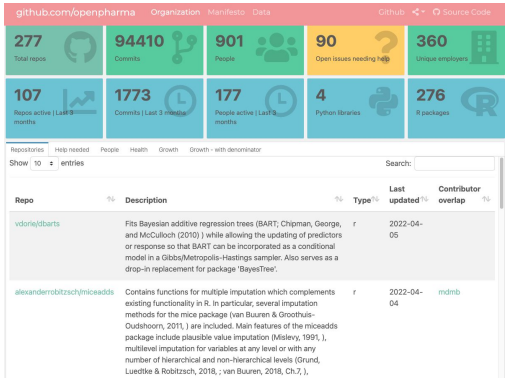


*Github profiles used were possible, otherwise unique emails where commits made do not match github profiles

Moving from publication to collaboration is non-trivial

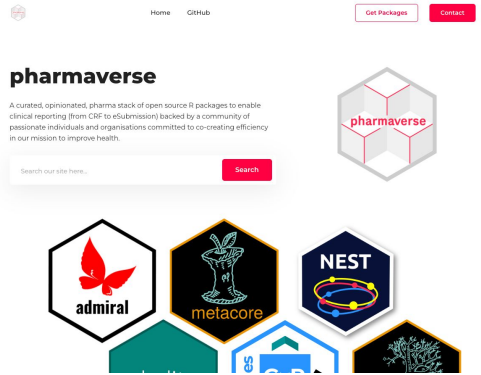
But there are tools and communities to help!

`openpharma.github.io`
Pulls and surfaces metadata daily on 270+ R/Python packages relevant to Late-Stage Pharma on Github.
A substantial overhaul of openpharma metadata started this month



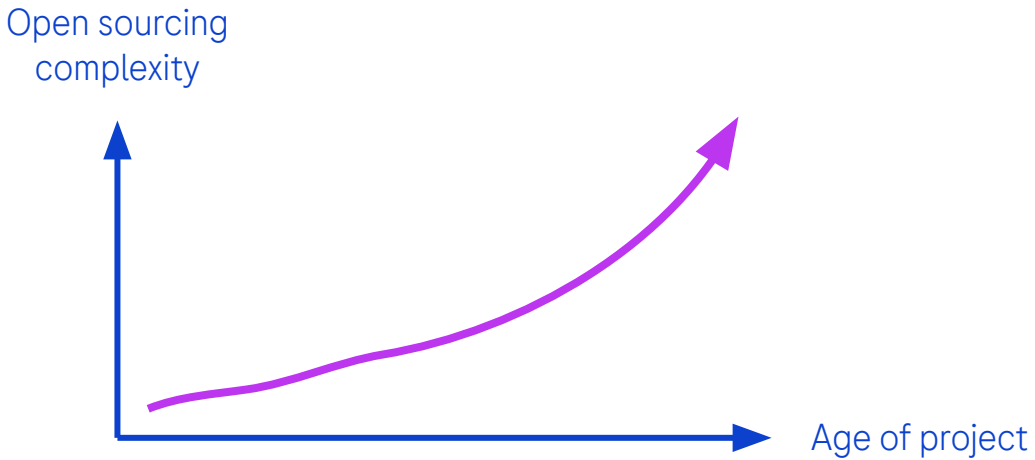
package metadata

`pharmaverse.org`
Opinionated and curated window into R packages used in clinical reporting.



Release early!

IP complexity when open sourcing has an exponential relationship with maturity of project



- ✅ New project
- ✅ Clear defined scope
- ✅ Free of patient data
- ✅ Post-competitive IP
- ✅ Useful outside Roche
- 🚀 Don’t wait to open source it!*

* After formal IP release via internal processes

Lawyer up!

Dependency licences

My code

```
my_cool_function <- function(x){
  y <- dependency::helper(x)
  y + 1
}
```

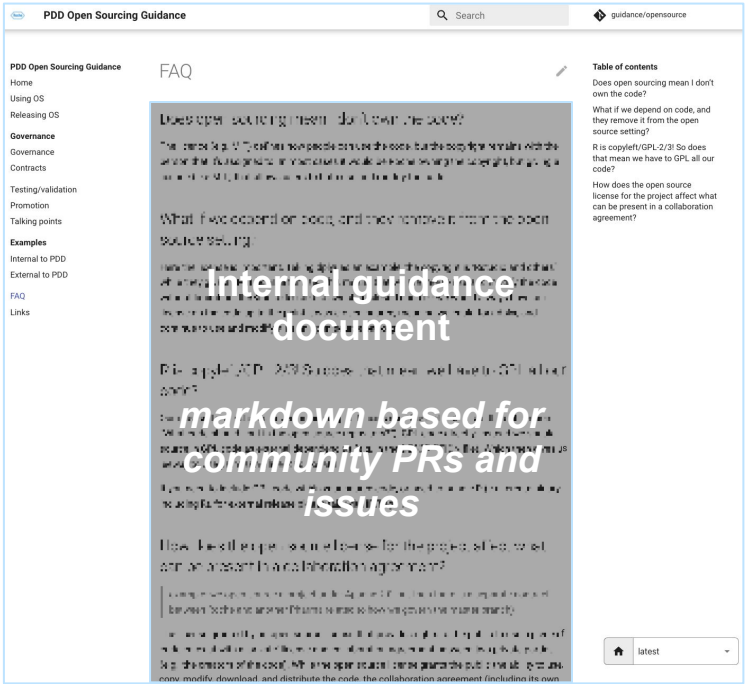
Sharing code that references dependencies???

Compiled propriety product??

From the Roche Data & Statistical Sciences experience, one of our most important resources navigating open source has been a dedicated IP lead from legal that understands research software.

Open source guidance can be inner sourced

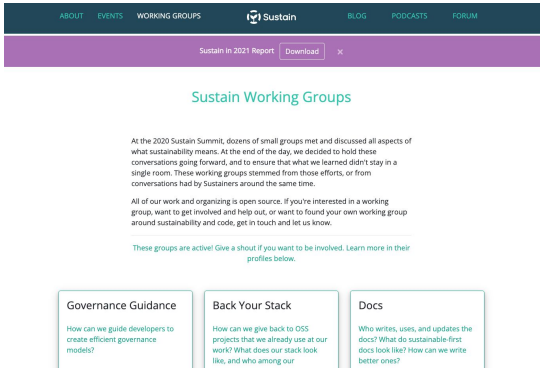
- Initial discussions internally showed a variety of opinions on what we open source, and developing a common set of classifications helps to protect our IP
- Partnering with Legal/IP, we created an internal guidance based on MKDocs, and continue to update as new questions arise.



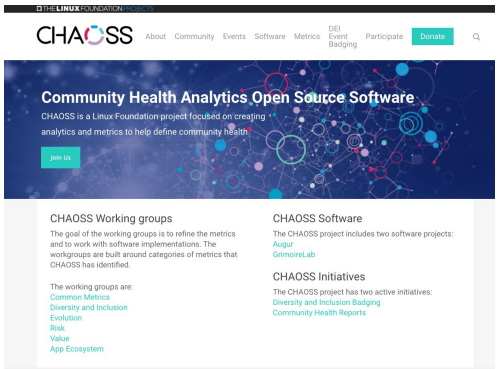
Resources are plentiful

Some example sources the Roche guidance builds on (and references)

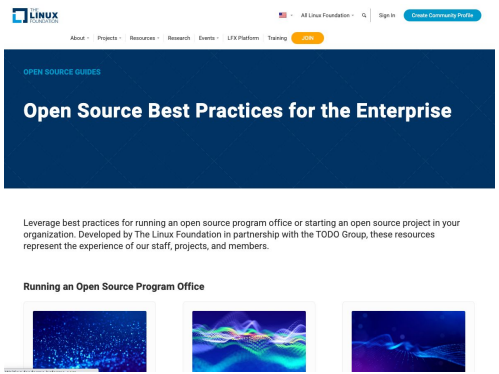
`sustainoss.org`
Focus on making OSS sustainable, but still covers topics like licences



`chaoss.community`
Working groups and tools focussed on OSS health



`linuxfoundation.org/resources/open-source-guides/`
A comprehensive guide



Collaboration contracts are a WIP

Do they really need to be?

- Our early contracts were modified research agreements, which is a fundamentally different type of engagement
- Arising IP (generated together in the project) is a simpler topic than existing IP
- It's important to stress permissive licences allow you to fork the code
- Licences dictate how the code can be used and modified, so a focus has been on lighter contracts focus on governance of the main branch
- Bi-lateral contracts don't reflect the fact companies may step in, and out, of wanting to govern an open source projects main branch

Conclusions, gaps and future discussions

Summary

- Open source collaborations are diverse collaborations with different types of engagement
- Open sourcing early helps ensure you have simpler discussion ‘this is the IP can we co-create’, rather than ‘this is the IP we will release and merge with your IP’, and hopefully prevents duplication - and better products rather than diversity through arbitrary decisions.
- Tools exist (and are being improved!) to help understand the health and engagement on your open source projects
- Licences can fundamentally change how your project can be used
- Partnering with Legal/IP, and users, to capture learnings in robust guidance - and help to evolve collaboration contracts
- The PhUSE **End-to-End Open-source Collaboration Guidance** working group hopes to help collect and share learnings and guidance across companies

Links

- Roche Pharma Development (late stage clinical reporting) notes on R code collaboration between companies: **`codecollaboration.org`**
- **`pharmaverse.org`** (opinionated stack for clinical reporting with R)
- **`openpharma.github.io`** (metadata on any Pharma R/Python packages added to it's tracker)
- Jeoren Ooms talk on R package open source health

More at this conference

- Tue 13:30-17:00; **workshop**; `pharmaverse` workshop on building open source e2e clinical reporting tools collaboratively
- Wed 11:15-11:30; **talk**; Pan-pharma code collaboration successes and new horizons

Doing now what patients need next