

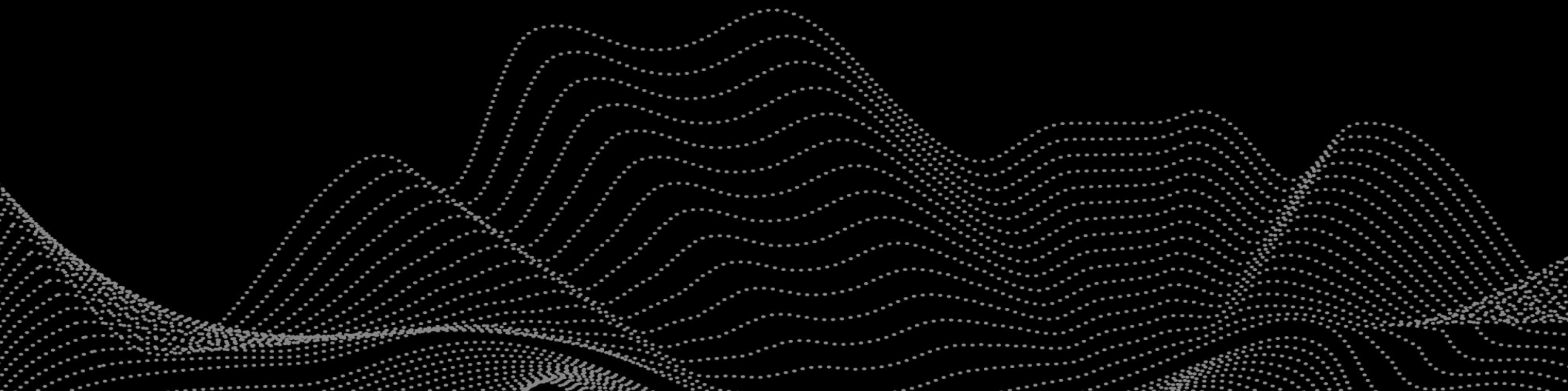


A research institution framework for publishing open code to enable reproducible science

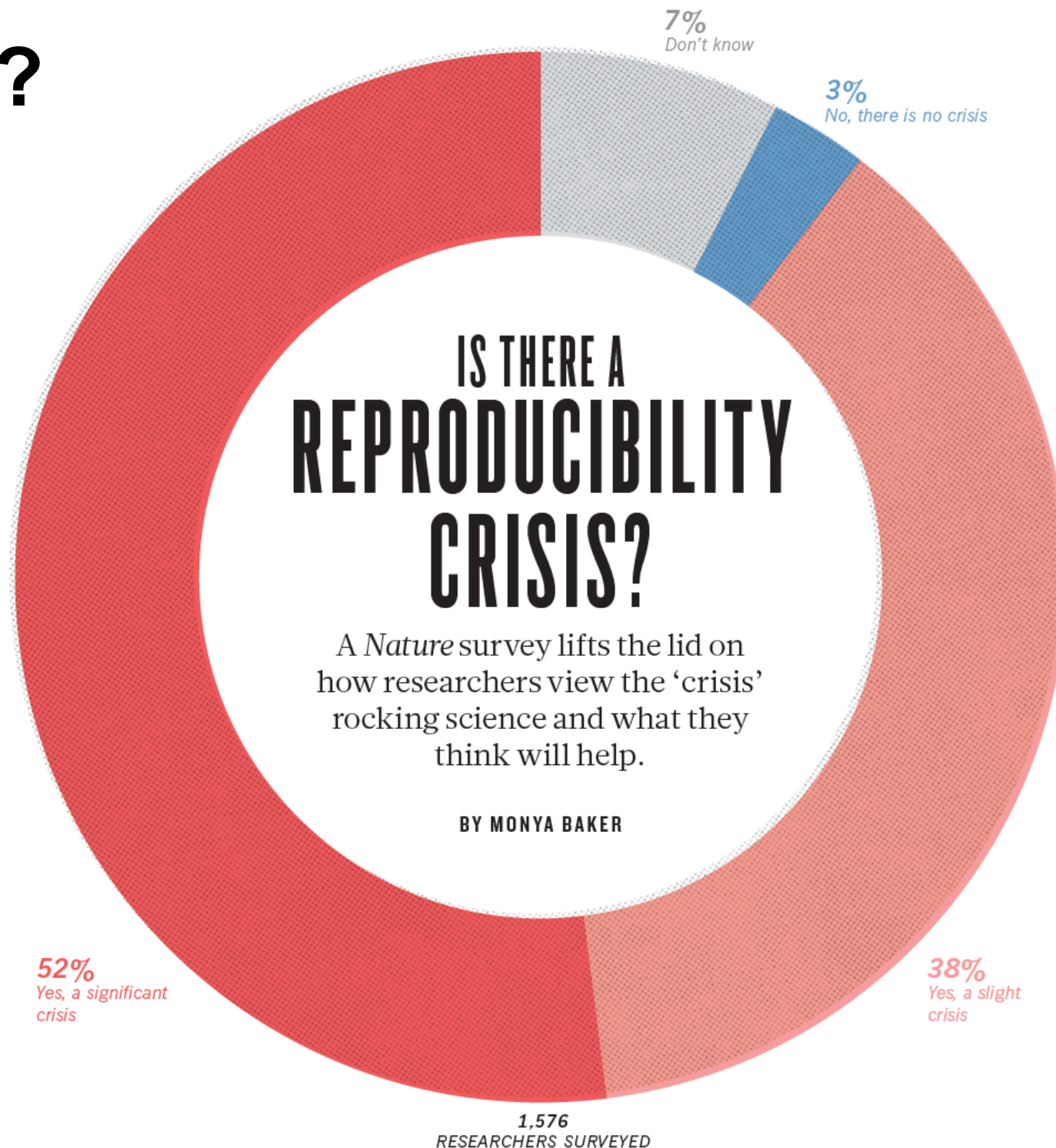
Tom Etherington

etheringtont@landcareresearch.co.nz

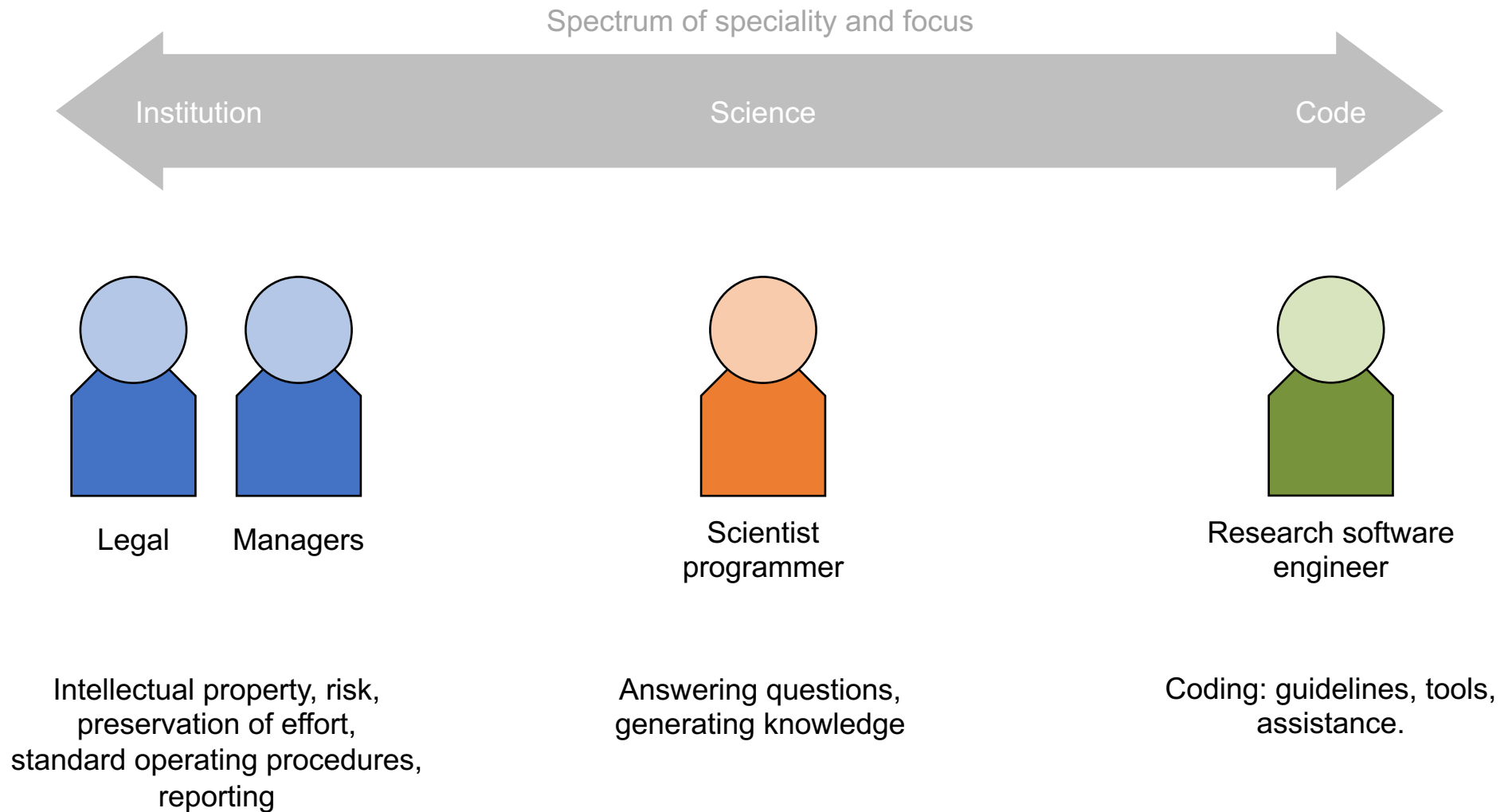
 [@tretherington](https://twitter.com/tretherington)



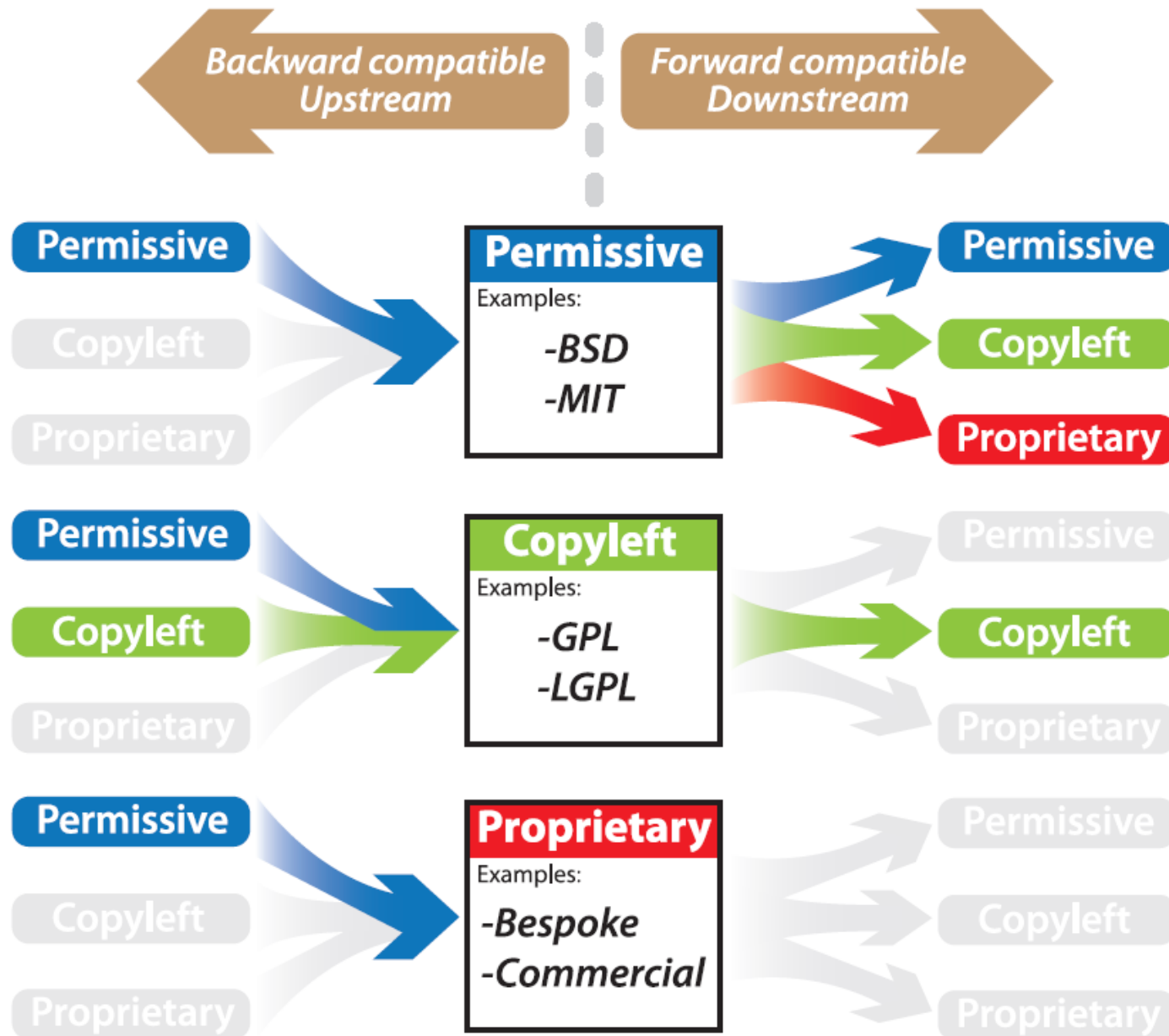
Why?



Who will be involved?

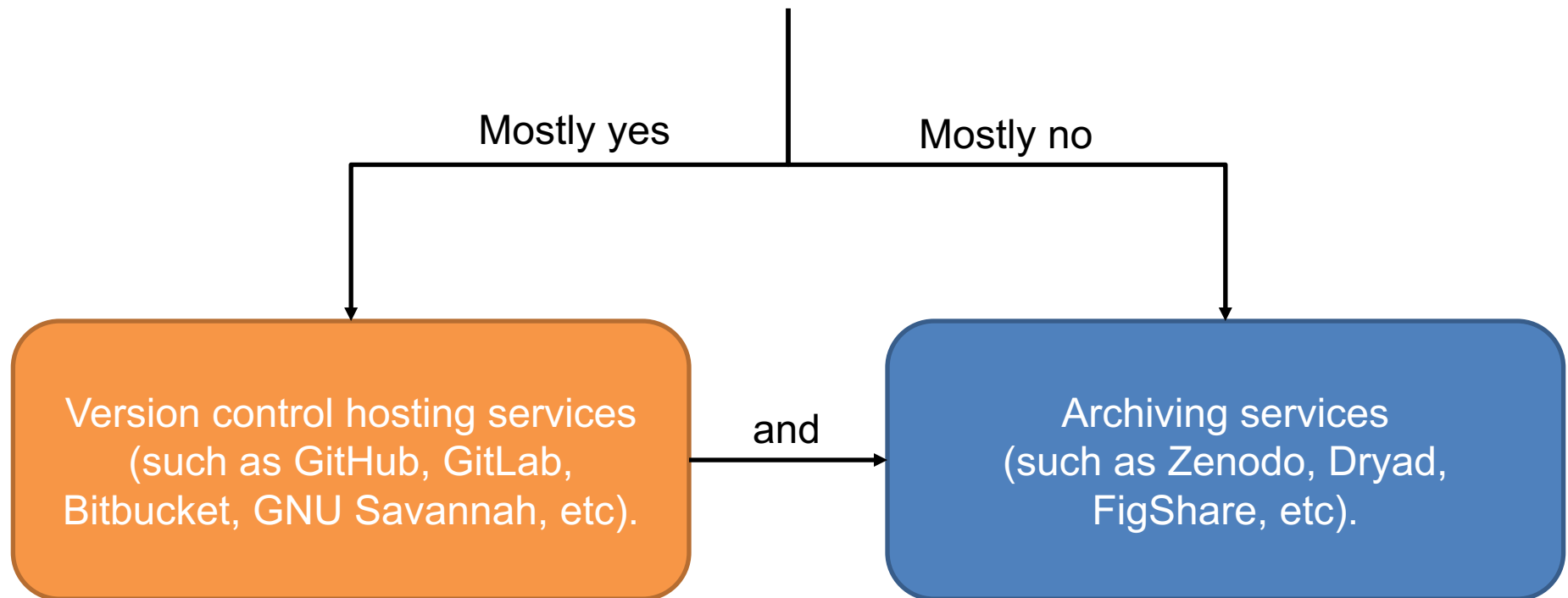


How should code be licensed?



Where should code be published?

- Have you chosen a permissive or copyleft licence?
- Does your code come with instructions, examples, and tests?
- Would someone else find your code useful, and be able to use it easily?
- Your code doesn't link to closed-source or local data or bespoke data formats?
 - Can you install your code on a computer?
 - Are you hoping the code will be developed further?
- Are you hoping people will let you know about issues and errors?

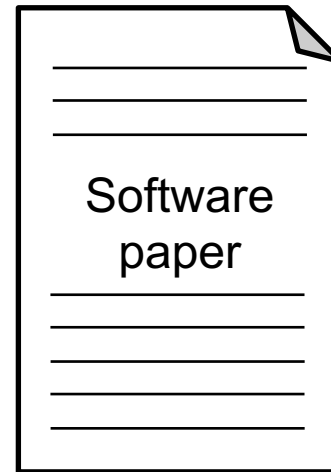


How to get credit?

CITATION file or README instructions



(from archiving service)

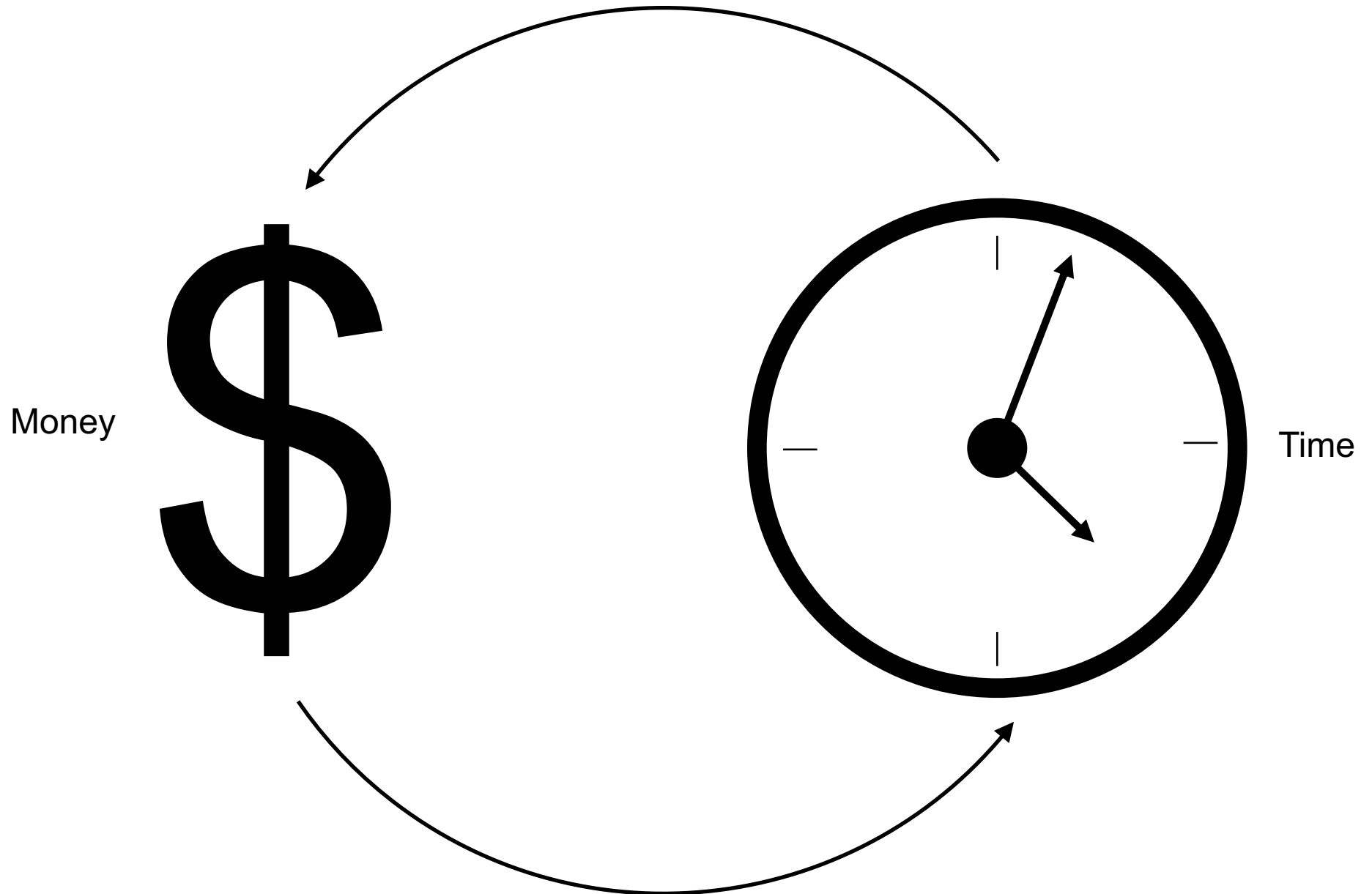


(Software Sustainability Institute
maintains a list)

What standards?

- README file (including: project overview, installation requirements, setup instructions, dependencies),
- LICENSE file (permissive or copyleft),
- citation instructions (either included in the README or as a citation file),
- example data and script,
- documentation embedded within functions,
- good coding practice (such as: commented, indented, white space, logical variable names, function definition),
- sensitive information removed (including: usernames, passwords, application programming interface (API) keys, full paths to files on network drives),
- version control history removed.

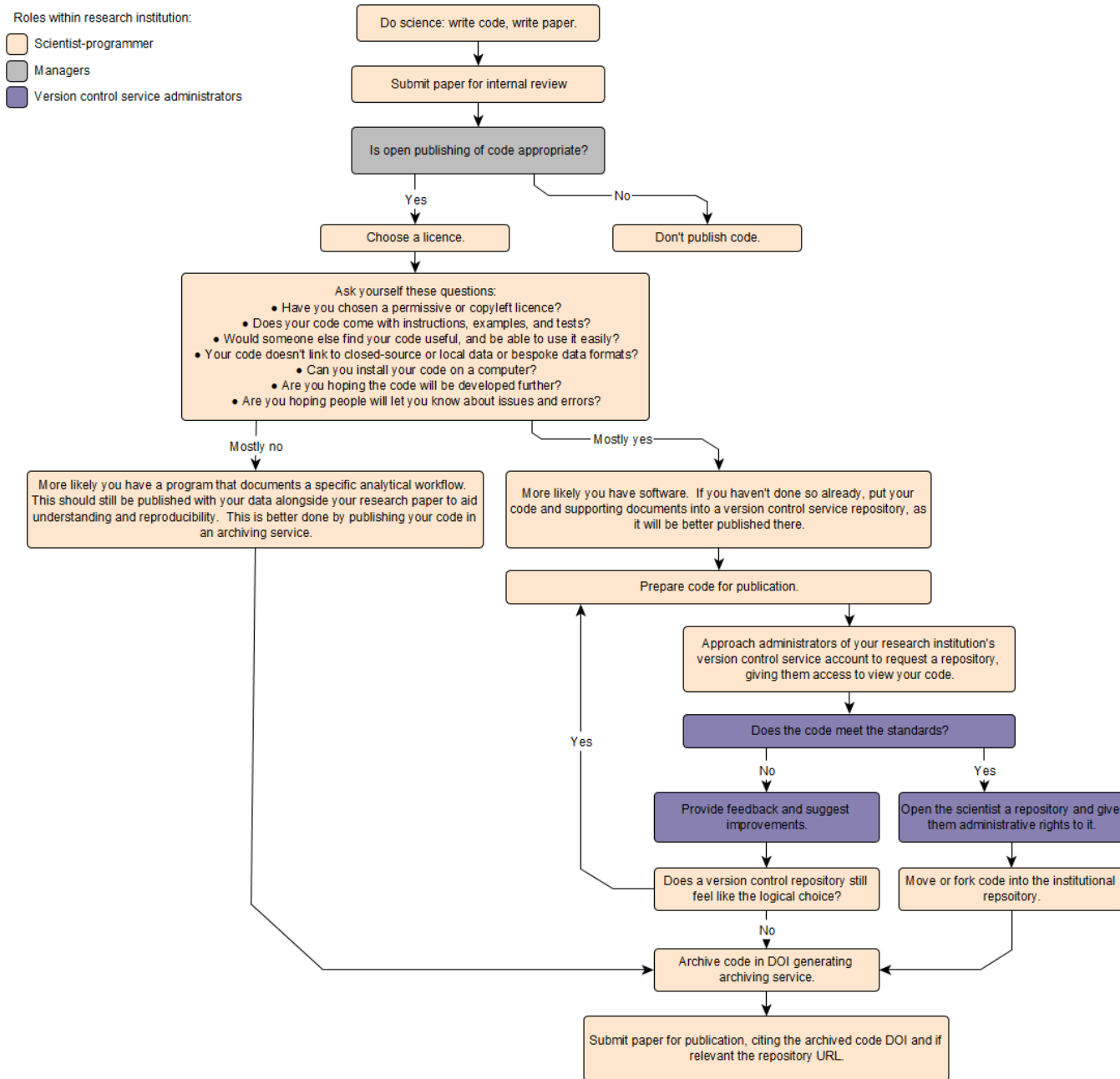
What costs?




Pros and cons

| Pros: | Cons: |
|--|--|
| <ul style="list-style-type: none">• enables reporting to funders, | <ul style="list-style-type: none">• services such as version control hosting and code review procedures need to be administered, |
| <ul style="list-style-type: none">• better outreach for the scientist and the institution, | <ul style="list-style-type: none">• scientists like to have freedom and may prefer to use another platform or like to publish code personally, |
| <ul style="list-style-type: none">• stimulates collaboration with other researchers, | <ul style="list-style-type: none">• ongoing institutional costs associated with version control hosting and archiving services, |
| <ul style="list-style-type: none">• quality control and transparency of science, | <ul style="list-style-type: none">• increased project costs to factor in staff time to make open code well documented and supported, |
| <ul style="list-style-type: none">• greater visibility of code projects compared to personal accounts, | <ul style="list-style-type: none">• accounts and repositories could become abandoned if nobody is using them. |
| <ul style="list-style-type: none">• provides repository exemplars useful for other scientists, | |
| <ul style="list-style-type: none">• other research institutions are already doing this, | |
| <ul style="list-style-type: none">• provides long-term support or at least access of published code, | |
| <ul style="list-style-type: none">• access and continuity of institutional knowledge if the person responsible for a repository leaves the research institution. | |

A framework for publishing



https://github.com/manaakiwhenua



Manaaki Whenua – Landcare Research
<https://www.landcareresearch.co.nz/>

Repositories 3

Packages

People 7


Teams

Projects

Settings

Pinned repositories

Customize pinned repositories

 **manaakiwhenua-manifesto**

A manifesto and code publishing framework for the Manaaki Whenua - Landcare Research GitHub account

● TeX

Find a repository...

Type: All

Language: All

New

virtualNicheR
virtualNicheR: creating virtual fundamental and realised niches

r ecology niche biogeography

ecological-simulations niche-modelling

ecological-niche-modelling

R MIT 0 0 1 0

Updated 21 days ago

pycrown
PyCrown - Fast raster-based individual tree segmentation for LiDAR data

python tree lidar segmentation numba

crowns

Python GPL-3.0 0 0 1 0

Updated on Jun 4

manaakiwhenua-manifesto
A manifesto and code publishing framework for the Manaaki Whenua - Landcare Research GitHub account

markdown opensource code publishing

reproducible-science reproducible-paper opensci

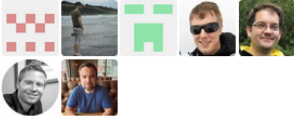
TeX CC-BY-4.0 0 0 2 0

Updated on May 30

Top languages

● Python ● R ● TeX

People 7



Invite someone

1 A research institution framework for publishing
2 open code to enable reproducible science

3 Thomas R. Etherington Ben Jolly Jan Zörner
4 Nick Spencer

5 **Abstract**

6 Reproducible science is greatly aided by open publishing of scientific computer
7 code. There are also many institutional benefits for encouraging the publication of
8 scientific code, but there are also institutional considerations around intellectual
9 property and risk. We discuss questions around scientific code publishing from
10 the perspective of a research organisation asking: who will be involved, how
11 should code be licensed, where should code be published, how to get credit,
12 what standards, and what costs? In reviewing advice and evidence relevant to
13 these questions we propose a research institution framework for publishing open
14 scientific code to enable reproducible science.

15 **Keywords:** archiving, code, version control, open science, programming



Manaaki Whenua
Landcare Research

Thank you!

Tom Etherington

etheringtont@landcareresearch.co.nz

 @tretherington

But...

Issues identified

1. We don't really know all possible uses.
2. Outside pressures.
3. We want to encourage people to open up code, and not put bureaucratic obstacles in their way.
4. Do we try to 'quality control' and 'organise' or just open up completely – will this create a 'noisy' or 'messy' set of repositories?
5. No one wants to be, or has the time to be, a 'gatekeeper'.