Computational Medicine Relay Teams

Steve Harris

2023-01-24

|  |  |
| --- | --- |
|  | A proposal and handbook for delivering data science work for the Computational Medicine theme. |

# Summary

We propose running two data science anchor teams over the first three years of of BRC 4. Each anchor team is composed of a pair of two skilled researchers: one with a background in analytical methods in data science (statistics, machine learning, epidemiology etc.); the other with a background in research software engineering. This anchor pair would be joined by members of an academic group to deliver a pre-defined component of project ‘in-flight’ within an allocated time window (typically 3-4 months). The anchor pair would then rotate to anchor the next academic group. Outputs are attributed to the academic group owning the project with suitable acknowledgement of the contribution from the rotating anchor pair.

The two pairs (4 people) would be managed by a project (relay) coordinator. The metaphorical relay baton is the academic project: the anchor is typically the fastest runner in a relay team. Speed takes preparation, and the anchor pair are efficient both because they bring the right skills, and because the ‘track’ has been well prepared. The relay coordinator is reponsible for this preparation. This key project management role scopes and prepares each project, ensures that regulatory and other requirements are in place before the anchor pair starts, and coordinates the daily delivery during the engagement. Once the anchor leg has been run, the analytical product or application is handed back to the academic team who are responsible for generating impact (e.g. publication, grant application, or implementation and evaluation).

|  | Anchor Pod | Relay co-ordinator | Academic Group |
| --- | --- | --- | --- |
|  | No involvement | Scheduling and coordination with the theme oversigh board | Ongoing workLonger term approvals and governanceProject proposalSignal intention to engage |
| **Approval milestone** |  |  | Project ‘go’ criteria defined |
| Preparation1-3 monthsValue definition | Approvals & TrainingDetailed project scoping and planningApproximately 10 person days available (5 per person) including exploratory scoping ‘spikes’ | Coordinate approvals and trainingDetailed project scoping and planning | Deliver approvals and trainingSpecify project aims and define impact |
| **Go milestone** | Sign off on feasibility | Sign-off on preparation | Sign-off planned scope |
| Engagement3-6 monthsValue generation | Full time commitment to project (less 10 person days preparation time for next project) | Coordinate daily check-ins and weekly review and planningProgress report to oversight board | Agreed time supporting project Attend daily check-insWeekly review and planning |
| **Hand-off milestone** |  | Sign off against planned scope | Sign-off handover |
| Value extraction | Minor support role only | Impact report to oversight board | Manuscript submissionGrant applicationApplication evaluation |

The programme will be over seen by the Computational Medicine Theme Board who would prioritise projects, and agree the time commitment for each engagement. The objective would be to deliver 3 projects per team per year.

## Notes

This delivery orientated approach to spending resource mirrors successful approaches in the software industry. The team’s capability is designed to support as many of the compuational medicine theme’s objectives as possible. A back of the envelope calculation assuming £80k/year per anchor pair member, two pairs, and one ‘relay’ co-ordinator would cost around £400k/year, or £1.2m for the proposed 3 year initial life-cycle. The recruitment process would be staggered, and there would be stage gates where the oversight board would review the delivery model before releasing further funds.[[1]](#footnote-1)

Assuming 3 projects per year (e.g. two 3 month, and one 4-6 month engagement) then this investment would return 15 projects over the funding period allowing one project each year to fail. This should provide opportunity to deliver on a wide range of the planned BRC objectives, and largely in the first years of the BRC giving time for the impact stories to be built.

### Benefits

1. A more efficient approach to hiring coordinated centrally and run once per team rather than once per project.
2. A more efficient approach to engagement with projects by deliberately curtailing the duration of the work whilst investing in planning and scoping
3. Throwing ‘mud at the wall’: predicting which projects will be successful is difficult. This approach allows meaningful evaluation of a wide range of projects, and opportunity to re-invest (interally or externally) in those showing the greatest promise.
4. An efficient division of labour: the academic group focuses on innovation, design, and impact (manuscript writing, grant applications), and the anchor pairs focus on the technical delivery task.

### Risks

1. A harmful effect on potential career structure for the pod members.

* Researchers who chose to adopt this role might not be seen as on a career track to become and independent researcher
* **Mitigation or rebuttal**
  1. Attribution at a middle author level would be expected
  2. These would be suitable bridging posts for researchers who needed experience and were not yet suitable for full independence
  3. Some individuals would enjoy the freedom to concentrate on project delivery rather than manuscript preparation and grant writing

1. Team composition is excessive for the task

* Projects might not need both a data scientist (modellers) and a software engineer.
* **Mitigation or rebuttal**
  1. Smaller projects are not suitable for this process and should be screened out at the planning stage
  2. The decision to deliver these projects using a ‘pair’ is forcing function to that helps build clarity, and improve quality. Individuals working alone are not required to communicate externally until a much later stage. Working in a small team externalises decisions and communication. This ought to improve quality, engage a wider audience, and permit supportive oversight in a timely manner.
  3. We have previously had good results pair the engineers and modellers. The former enjoy learnig methods from the latter, and the latter enjoy learning an approach to reproducibility and reliability from the latter.

1. Team composition is inadequate for the task

* Projects might need more than a data scientist and a software engineer
* **Mitigation or rebuttal**
  1. Projects would not be approved if they could not deliver with the resource proposed
  2. The academic project team would be expected to contribute their own expertise, and make a clear and synchronised commitment of time
  3. Where the academic project team could justify additional specific requirements, resource would be available to support time (salary) from external collaborators

1. To give those individuals being hired job security, we propose hiring individuals for 1 year with a planned extension to 3 years assuming the model is working well. [↑](#footnote-ref-1)