Learning Curve Initiative [Letter v0.1]

Trainee collective

Abstract An ongoing community letter for a trainee-led initiative to promote sharing otherwise-unpublishable data in computational biophysics

1 Why this initiative?

A group of trainees came together on Twitter and brainstormed possible ways to share negative results in a way that trainees showcase their effort and data, get feedback from the community and help flatten the learning curve for the next generation of computational biophysicists. **Learning Curve Initiative** will create a platform for computational biophysics trainees to share otherwise-unpublished data, common pitfalls and worked-out problems with the community. Our ultimate hope is to promote scientific discussion, protocol and data sharing practices between trainees.

Short-term goals:

1. Form a community consensus on the definition of otherwise-publishable data

2. Once the consensus is made, prepare a checklist of minimum criteria for the trainees to follow a consistent style

3. Form an advisory board: Faculty members from the community will give advice and mentor the organizing committee

4. Send out an open invitation to find more volunteers to serve in the organizing committee

5. Plan a webinar: The webinar will host talks presented by trainees. Trainees will submit an abstract in order to be chosen to present their otherwise-unpublished data.

The advisory board will be kept in the loop during this entire process.

Long term goals:

1. Organize the first webinar. If all goes well, come up with a plan to turn the webinar into a tradition. An organizing committee formed by trainees can hold these webinars regularly. The organizing committee members can train the upcoming year's members to ensure the continuity of the initiative.

2. Plan a potential publication (such as a LiveCoMS Lessons Learned article) by grouping the otherwise-unpublished presented in the webinar.

3. Talk to professional societies (such as BPS, ACS, APS and ASBMB) to organize a poster session for trainees presenting otherwise-unpublished data.

4. Generate a platform for trainees to openly talk about hardships. This could be virtual coffee chats, a blog, gathering/dinner at BPS, webinars, etc.

5. Promote this model for other research communities to use.

2 Discussion

The trainee group is requesting feedback from the community on (1) the overall model outlined earlier, (2) the definition of inconclusive data (next page), and (3) a checklist for trainees to identify the inconclusive data they may have and a minimum criteria expected for a related presentation /contribution (next page).

Following the LiveCoMS protocol, we plan to use the Github issue tracker for feedback, suggestions, and further discussions.

1. Feedback or discussion on the overall model: here

2. Feedback or Discussion on the definition and name for otherwise-unpublished data: here

3. Feedback or Discussion on the minimum criteria checklist: here

4. General discussion on the initiative: here

DEFINITION OF OTHERWISE-UNPUBLISHED DATA

Otherwise-unpublished data must belong to at least one of the following categories:

- ☐ The data do not support the initial hypothesis or expected outcome after careful and reproducible meticulous scientific research and hence the data will not be included in a traditional publication. Yet the data are generated by a robust protocol/analysis method that other people can learn from.
- ☐ The data clearly demonstrate a reproducible problem. This problem can be something that (1) the trainee does not have a manageable solution for; (2) the trainee needs feedback from the community; (3) the trainee wants to increase awareness for a phenomenon or an atypical pitfall.
- ☐ The data clearly demonstrate a solved problem. Even though the problem is fixed, the trainee believes that they have a reproducible and easy-to-explain way of demonstrating the problem along with some generalizable steps/suggestions to identify and fix such problems.

Checklist for a minimum criteria expected for a related presentation /contribution ★ Work in progress ★

INITIAL HYPOTHESIS

- ☐ An initial testable hypothesis backed up by strong rationale is formed (based on literature review, prior knowledge, research gap identified, etc.)
- ☐ A detailed, careful experiment design is formed to test the hypothesis with statistical rigor. (Included independent replicates, following best practices in the community, …)
- ☐ Benchmarks for testing the validity of results is set up a priori.

ANALYSING THE DATA

- ☐ Sanity checks have been done to the best of your abilities. Code/method review, debugging, testing..etc. Independent replicates
- ☐ Identified or tried different analysis schemes when the initial scheme failed the anticipated results.
- ☐ Additional testing scheme was added to check for a potential flaw in the design
- ☐ Considered alternative hypotheses and tested them

INTERPRETING THE DATA

- ☐ You have not done any "cherry-picking" of data. Analysis schemes are done on data without any bias
- ☐ Your analysis indicates the rejection of your initial hypothesis with statistical rigor. Or, you see an absence of anticipated features or characteristics that helped you formulate your hypothesis. Including simple, semiquantitative checks that describe them.

REPORTING THE DATA

- ☐ Honest and complete reporting of all the attempted tests and rationale
- ☐ Report alternative hypothesis tested if any. Discuss such possibilities.
- ☐ Discuss the potential flaws, pitfalls, and other features of the method that was used.