

CW20 - 2020-03-31

Group I - DSR9-CW20

Speed Blog

Please use the area below to draft the speed blog. Ideally, your speed blog will be a written piece with a title, picture (max 2, the main picture should be scalable to 400px wide by 300px high) and contain some links and references. In terms of word count, you are aiming for between 500 and 1200 words.

Some prompts for speed blogging:

You could take a conclusion first oriented approach to your blog and cover:

- *What are the five most important things learned during this discussion*

You could also include:

- *What are the problems, and are there solutions?*
- *What further work could be done, and who should do it?*
- *Are there any useful resources that people should know about?*

These could include all or some of:

- *Why is this issue important?*
- *Summary of the discussion topic*
- *Recommendations*
- *Other significant points outside the main topic of discussion*

[BLOG STARTS]

Bootstrapping a development team during the time of crisis

Raniere Silva, Malvika Sharan, Colin Sauze, Yo Yehudi, Robert Haines, Claire Wyatt

This blogpost is written at the Collaboration Workshop 2020. Authors' names are arranged in no particular order.

Introduction

As researchers, scientists, or techies we look for ways to use our skills to solve pressing, current issues. We have responded to the [COVID-19](#) pandemic with a sudden surge of hackathons, data modelling projects, task forces, and working groups. This has resulted in an unprecedented desire to pull together, which is surely an ideal way to move forward. Nevertheless, forming effective collaborations is a challenging process, even in the best of times. So, how can we ensure that the teams and research communities we form during the time of urgency are built on the foundation of excellent science and community practices? To help researchers address this issue, especially when starting a project during the COVID-19 pandemic, we have compiled recommendations drawn from our experience as a community and technical experts. We hope that this post will prompt people to think critically not just about the technicalities of the work, but also the social environment in which it is being conducted.



10 recommendations for managing a nascent team or community

What If you started a project with a thought like, “I have this great idea that I want to try on COVID-19 data!”? There is nothing to worry about if you’re the only one working on it, but the moment you share your idea with others in your network inviting them to join you - you become responsible for making people feel included in your project.

As a project lead, you want to set up a welcoming and inclusive environment and create the first set of visions and goals for your collaborators. You cannot assume that everyone you collaborate with knows what is expected of them when they start to work with others on your project. Therefore it’s important to set the right expectations from the beginning for your community, even though you might not have planned on having one (see the additional background in [Malvika’s talk from CW20](#)).

In the best case scenario, you would have already planned for community collaboration in your research project, but if that was not the case, then the checklist provided below will help you in making this process thoughtful and structured. The practices listed here are derived from and limited by the experiences of the authors who participate in several successful Open Research communities and lead community-driven projects such as [The Carpentries](#), [Mozilla Open Leaders](#), [The Turing Way](#) and [Open Life Science](#), hence they may need to be adapted for projects that are very different in nature.

1. Choose a communication platform:

- When leading an open project, use collaborative and open platforms such as [GitHub](#).
- Evaluate if you need any real-time communications such as a text chat system like [Slack](#) and [Riot/Matrix](#), or if a mailing list or forum will be sufficient
- Consider a separate internal communication platform for your community members and an external one for showing what you’ve done to the rest of the world. A Twitter account or a simple non-technical website can be useful external platforms.
- Make sure that there is a very low barrier to joining your platform. See [The Carpentries](#) and [The Turing Way](#) for examples.

2. Provide a project summary file:

- The first document in your project should be a project summary file, which on GitHub repository will be a [README.md file](#).
- This will provide basic information about your project so that people can evaluate why your project will be interesting for them. Here is [a template](#) by the GitHub user [PurpleBooth](#).
- In this file, include what your project vision and goals are, why the project is useful, what the possible milestones in the project are, how a contributor or user can get started, who they can reach out to for help, and what is currently missing in the project in terms of stakeholders, skills, or scope.
- You can use emojis, GIFs, videos or your personal narrative to make your project relatable. See [The Atom project](#) for example.

3. Select a Code of conduct:

- Add an [Open Source Project Code of Conduct](#) to your project. Don’t use it as a token, put intentional effort into it.

- When using GitHub, you can add a “CODE_OF_CONDUCT.md” file in your GitHub repository.
- List the expected and unacceptable behaviours, describe the reporting and enforcement process, explicitly define the scope and use an inclusive tone ([see instructions here](#)).
- Whenever you update your code of conduct, invite comments from your members to ensure that their concerns are addressed. This can be done on [GitHub issues](#), or [Pull Requests](#).

4. Provide contribution guidelines and interaction pathways:

- A thoughtful guideline helps people decide which pathway they can choose to contribute to your project, or if at all they want to be in your community.
- Make sure that your community interactions and different pathways to contribute are open, inclusive and clearly stated. If people can't figure out how to contribute they will drop off without help. Here is [a template provided](#) by the GitHub user [PurpleBooth](#).
- Value different types of contributions. Coding projects are not only about code, therefore list documentation and other management skills as well.
- You can use [persona and pathway exercise](#) to brainstorm who could be your possible community members.

5. Create a basic management/leadership structure:

- A leadership structure in an open project should aim to empower others and develop agency and accountability in your community.
- You can start by listing different tasks within your project and inviting your members to lead those tasks.
- Provide appropriate incentives and acknowledgement for contributions made by your community members.
- Create opportunities for members to share some leadership responsibilities with you in the project.
- When inviting suggestions and ideas from the community, provide the first set of plans where your community can develop from.

6. Provide contact details wherever useful:

- Clarifying responsibilities for different members will allow people to reach out to the right person with any query.
- Add details of the designated contact persons for technical problems, leadership questions or any report on Code of Conduct.
- This will be particularly useful if something needs immediate resolution.

7. Identify failed approaches, and stop:

- Development happens in an iterative manner. Revisit your plans and ideas at regular intervals and involve your members in the process.
- Check if there are parallel developments or multiple approaches that can be changed. Importantly, recognise what isn't working for your project.
- Fail fast, fail informatively. Document and share why you failed and how you change your project or approaches going forward.
- For Open Research communities you can maintain transparency when

discussing failures and successes but refrain from singling out or blaming others.

8. Have documentation and dissemination plans for your project:

- Communication is important. With new members coming all the time, it is important that they are able to find the information they need without asking you.
- Investing in the documentation of plans will free you from repetitive questions regarding past decisions or the decision making process your project uses.
- A good place to store the documentation is wiki or platforms like GitHub where information can be updated by your community members democratically.
- To disseminate outputs of your project, you want something permanent that can be shared and cited, for example, by using their digital object identifier (DOI). [Figsare](#) and [Zenodo](#) are good examples of platforms that can provide you with DOI for all your shareable data.

9. Address technical issues:

- Make sure that you have plans in place for people who want to contribute to your project but might deviate from your original goals very fast without supervision or guidance.
- If a certain skill or practice is required for your project, you should be able to point people to relevant resources so that they can engage with your project effectively. Provide short video or online tutorials on skills that your collaborators might need.
- It's important to promote practices such as code review and testing to ensure the reproducibility of your work.
- Reach out to other communities with specific expertise to save effort and time that can be invested on other tasks.
- Encourage data and code sharing with the use of appropriate licenses, but also make ethical decisions around data privacy.

10. Acknowledge differences and value them:

- Make sure that you invite perspectives from your members who can help you maintain heterogeneity and a fair distribution in leadership, methods, dataset and solutions that can be reproduced for different scenarios or across different platforms.
- Start by identifying who your stakeholders are. Are they given opportunities to be heard? If crucial voices and viewpoints are missing in your project, how can you invite them in? Talk to an expert (clinician, epidemiologist, etc.) if you are not one yourself.
- Always remember that different stakeholders can introduce new but critical and often unexpected viewpoints and help you avoid incorrect or ineffective assumptions.

Conclusion

Keeping a team engaged is a continuous activity that requires regular checking in and evaluation. Based on each evaluation we can change and improve our workflow and community practices. During this unprecedented time, while we navigate through the COVID-19 pandemic

situation and apply our collective knowledge to tackle the problem at hand, we don't have the benefit of time to carry out such evaluation. Therefore it is even more important to be intentional when creating teams and communities and apply the tried and tested best practices from the beginning. We hope that the practices recommended in this post will give a good starting point for building successful collaboration through your projects.

---NOTES from the CW20 discussion---

COVID-19 research: what are the challenges?

Participants

Please list the participants here (please use full names).

- Yo Yehudi: Chair
- Raniere Silva: Note-taking
- Caroline Jay:
- Malvika Sharan: Note-taking
- Robert Haines
- Colin Sauze
- Claire Wyatt

Notes from the discussion

Please use the area below to capture notes from the discussion session.

Tips on discussion sessions and Speed Blogging can be found here -

<https://www.software.ac.uk/speed-blogging-and-tips-writing-speed-blog-post>

- Yo: Software engineer in Uni Cambridge, Ph.D. in Manchester
- Raniere: A Masters student in Brazil, prev community manager of SSI
- Caroline: Department of Computer Science in Manchester. Supervisor of Yo. Interested because she was involved with discussions with the Alan Turing Institute.
- Malvika: Community Manager at the Turing Way at the Alan Turing Institute. Co-leading Open Life Science with Yo.
- Robert: Head of Research IT. Have a team of 50 people that provide research platforms and research software engineering. He was asked to participate in COVID-19. But is busy with other projects that are paying the bills.
- Colin: RSE in Wales (Supercomputing Wales), want to support Ramp initiative
- Claire: SSI based in Southampton. Community Manager for RSE Community. Lots of initiatives going on.

Topic: Malvika has been talking with a lot of people about using public data for the first time to understand COVID-19. [Missed something here] - People may not be aware of how the data are obtained and whether ethics were taken into account with regards to this.

People are running against the clock. Can people use the data? Due GDPR. Is the data good to be used for modeling? For example, is it a good representation of the population? Or is the data bias for any reason. For example, only "rich" neighborhoods have report cases.

Some countries, e.g. Singapore and South Korea, has used apps to decide if people could go out or not.

NHS is releasing a reporting app.

Most of the projects are volunteering and not paid research. What happens with ethical approval?

Some data is hard to download, e.g. Influenza. You don't have API. The current dataset is not machine-readable. You don't have an open license. Maybe cultural issues?

Is one week enough to do modeling?

<https://www.kaggle.com/c/covid19-global-forecasting-week-2>

Questions:

- How can we navigate ethical research when there is a race with time?
- Can we replace an in-person diagnosis with an app-based/contactless diagnosis?
- The concern with the huge volume of data that is being produced at this point - technical challenges.
- How to provide information and privacy protection when using technology while working from home? - developers and companies are in the place of power where they can misuse information that people have no control over.
- Even in healthcare, since we need to understand this situation - is it possible that the patients data are also not handled with the highest level of privacy?

Apps

- Evergreen / linked with NHS (UK based)
- China - http://www.xinhuanet.com/english/2020-02/10/c_138770415.htm
- Kings College COVID19 Symptom Tracker <https://covid.joinzoe.com/app>
-

By product of COVID - e.g. red tape being removed in terms of data confidentiality.

Super-fast research in the Coronavirus pandemic - what could possibly go wrong?

- Under sharing vs. oversharing issue: Oversharing can save lives - but can also breach the privacy of people.
- We can't treat data with the highest level of protection because the situation demands solution (vaccine, trend, cure etc.)
- The speed research can also lead to missed opportunities and accuracy - and above all lack of best practices

Ecology people have to use models for migration to COVID-19. We don't have any literature to back this up.

Culture - people want to pitch in perhaps whether or not we have useful or sensible skills to apply.

Researcher's ego

- Everyone wants to get involved - they might not be making the same degree of decision making as they would in other situation
- We need to understand and accept that everything can't be turned into a phone app.

Collaborating vs. competing

- People can still help each other, by actually doing what they are good at:
 - Supercomputing people can offer resources to researchers rather than using their resources in a misinformed way
 - Should community managers facilitate the use of COVID-19 data by researchers? Or promote their participation in volunteering actions?
 - This a great opportunity for collaborative working if we can figure out what our actual role is.
 - We need to enable people who want to get involved in.
 - As long as people accept that they might not be most suitable for one task

Role of the community manager. How do we organize efforts? It is hard to explain to people that their certain contribution is not beneficial - instead, we can empower them by assigning a task that matches their expertise

Live inside the bubble. Lots of people aren't aware of the problems of "bad" code.

Dangerous of anonymous contributions. You can use your anonymization to trolling. You don't need to be anonymous to send false information/fake news these days. For example, you can write you one message on WhatsApp and forward it anonymously.

[BLOG WAS HERE - but now moved to the top]

Part 2

LONG VERSION OF LAST 2 RECOMMENDATIONS. MAYBE WORTH DISCUSSING IF THIS CAN GO OUT AS SEPARATE BLOGS.

Address technical issues

Make sure that you also have plans in place for people who want to contribute to your project but might deviate from your original goals very fast without supervision or guidance. If a certain skill or practice is required for your project, you should be able to point people to relevant resources so that they can engage with your project effectively.

- **Provide short and concise tutorials**

In the time of crisis, we work on what is important right now, which might mean that we may overlook what is important in the long term. For example, we might want to test several algorithms on our COVID data but don't pay attention to recording the outcome systematically in a central platform that others access. Offering training or pre-recorded videos on recommended practices that are less than 2 hours long can enable your community members to work using a standard workflow, or take over some tasks from others.

- **Testing is important**

To err is human! And when working under pressure, they might be more frequent. Test your codes and encourage your community to review and test each other's code. Thankfully, developers are able not only to write code to solve problems but also write code to test if the code they write is correct. You can also set up a continuous integration environment to help automate testing in your workflow.

- **Reproducibility is even more important**

Great thing for less involved team members to do is constantly test the reproducibility of any code/environment. Do this from the start and it won't be a surprise later when it doesn't work on somebody else's computer.

- **Reach out to the experts (especially when dealing with legacy code)**

Reach out to other communities with specific expertise to save effort and time that can be invested on other tasks. For example, a lot of the scientific knowledge is built on top of results from FORTRAN, C and Java code that isn't maintained any longer and, probably, isn't documented. Finding someone with knowledge and experience of the legacy code to answer questions that other developers have will be a huge time saver.

- **Share code (and data) earlier**

When running against the clock, it is important to not waste time doing the same thing twice. Developers must share their code in a public version controlled repository (like GitHub and GitLab) and coordinate who is working on what feature or fix. Choose an appropriate [open source license](#) for your data, scripts and software.

- **Take note of the privacy issues**

Ask yourself, how can people who need to access this data get to it. How they can re-use and share the data appropriately. Choose a [relevant licence](#) ensuring protection of information that is sensitive such as movement and location data, personal health

issues, contact information, names, date of birth, and personal addresses. Avoid gathering personal information that is not necessary or breaches confidentiality.

Acknowledge differences and value them

When people collaborate with others, no matter if they are from your network or complete strangers, you need to take into account that everyone will have different backgrounds, different cultures, different time zones, different levels of skills and very different lived experiences. For example, when talking about COVID-19 related projects, you can expect that some people will join you because they want to know what kind of data on COVID-19 exist, some will come with an intention to be part of a group while learning some skills, some will simply hang out to watch what you are doing, whereas some people will join you because they have relatives who are suffering from COVID-19 infection and they want to speed up the diagnostic process. Welcome their varying perspectives into your project.

As an individual we may think that everyone has the same working conditions as us, such as sanitation, work setup at home or resources, whereas the reality is far from that. Our colleagues even in the same city have different living situations, caring responsibilities, medical conditions and resources available to them. Don't assume that everyone will have equal access or capacity to participate in your project. To create opportunities for everyone, create task lists with varying requirements of time and resources (like hardware, software, reference materials or internet bandwidth) so that people can choose where to contribute based on resources available to them.

Offer access to resources to ensure that everyone has the opportunity to gain a common understanding. For example, if your project idea is based on a recent peer reviewed paper on COVID-19 that is behind the paywall, members from low income institutes who don't hold subscription to that journal will not be able to participate unless given access to the information discussed in that paper.

Make sure that you invite perspectives from your members who can help you maintain heterogeneity and fair distribution in leadership, methods, dataset and solutions that can be reproduced for different scenarios or across different platforms. For example, when making recommendations based on the selected COVID-19 dataset, think if they will apply to both developed and developing countries, or if they will hold true for rich and poorer households alike. If not, what you can change, and who can you invite to help improve your project.

Think carefully about who your stakeholders are. Are they given opportunities to be heard? If crucial voices and viewpoints are missing in your project, how can you invite them in? Talk to an expert (clinician, epidemiologist etc) if you are not one yourself. Always remember that different stakeholders can introduce new but critical and often unexpected viewpoints and help you avoid incorrect or ineffective assumptions.