The RECON COVID-19 challenge: leveraging the R community to improve COVID-19 analytics resources

The Problem

The pandemic of COVID-19 is arguably one of the largest public health emergencies since the 1918 influenza pandemic, and poses considerable challenges to health services and public health agencies worldwide. Over the last decade, the response to large-scale epidemics such as the 2009 influenza pandemic [1], Middle-East Respiratory Syndrome Coronavirus (MERS-Cov, [2]), Ebola in West Africa [3,4] or in Eastern Democratic Republic of the Congo (DRC, [5]) has been increasingly data-driven, resulting in the emergence of outbreak analytics[6] as a new branch of data science aiming to inform the response to epidemics in real time.

Since its creation in 2016, the R Epidemics Consortium (RECON; www.repidemicsconsortium.org) has been at the forefront of free, open-source outbreak analytics using the R software, having developed a large number of packages for infectious disease epidemiology and setting references in terms of scientific and technical standards. RECON is also committed to education and to building bridges between R-developer and epidemiologists, and regularly runs short courses and hackathons worldwide. Despite these efforts and substantial progress made over the years (e.g. [7–11]), many gaps remain to be filled in terms of R software for epidemic analysis and modelling. These gaps include, for instance, tools for handling epidemiological data (including the handling of geographical data), estimation of key epidemiological delays accounting for lags in reporting, inference of under-reporting, standardised procedures for handling case forecasting, and tools for predicting bed occupancy based on admission data (present, and predicted) and duration of hospitalisation.

Currently, these gaps are to a large extent being addressed independently by different academic groups, public health agencies and non-governmental organisations (NGO) throughout the world on an *ad hoc* basis. Moreover, these groups being typically over-worked, many of these developments are sub-standard, using poor coding practices, with scarce documentation and often no automated testing nor code reviews, and with little code sharing beyond immediate collaborators. This approach is not only inefficient as it involves tremendous duplication of efforts, it is also prone to generating errors which could in turn invalidate subsequent analyses, with a potential for misleading information fed by modellers and analysts to policy makers. It also fails to tap existing resources, namely the large community of R users and developers willing to volunteer their time and skills to address some of these issues, and generally become part of the worldwide response to COVID-19.

The Plan

The aim of this proposal is to federate the infectious disease modelling, the epidemiology and the R communities to improve analytics resources for the response to COVID-19. This will be achieved by creating the *RECON COVID-19 challenge*, a website providing a platform which will centralise, curate and update R development tasks relevant to the response to COVID-19. Not

unlike other initiatives such as the Open Street Map Tasking Manager (tasks.hotosm.org), this platform will allow potential contributors to quickly identify outstanding tasks submitted by groups involved in the response to COVID-19 and will ensure that developments follow the highest scientific and technical standards.

These tasks will include code reviews, *ad hoc* or automated testing, the development of new features in existing packages, and the development of new R packages. They will not be limited to new developments, but will also include reviewing, refactoring, re-writing and packaging of existing code developed by COVID-19 responders, to ensure adherence to high coding standards and good coding practices. Each item will provide sufficient context for potential contributors to understand objectives and key deliverables, ideal timelines, and the potential impact on the response to COVID-19. Links to related projects, github repositories and issues will be provided, as well as contacts of the focal point (the 'author' of the task). Each task will be dated and labelled using keywords to facilitate searches. A community-driven scoring system will be used to rate each task according to two metrics: the *priority*, which will measure the importance and urgency of the task, and *difficulty*, which will reflect the perceived amount of work needed to complete the task. A simple search engine will enable searches by dates of creation, keywords, authors, and *priority/difficulty* scores.

Ahead of the initial release of the website, a *project manager* will be appointed to help organise the collection and curation of tasks. They will contribute to defining processes for handling the submission of new tasks, involving a panel of expert statisticians, modellers and epidemiologists across multiple institutes, already involved in analyses and forecasts for the response to COVID-19. The project manager will also maintain and curate the website for the first weeks post-release, and maintain contact with the main task-issuing groups to monitor progress. They will be responsible for advertising the project on relevant social media platforms and forums, and generally maintaining communication with the community. Finally, they will be coordinating the creation of a CRAN task view for epidemiology, which will provide an exhaustive account of existing R resources for analysing epidemics.

While this project is aimed at leveraging R tools for helping to respond to COVID-19, we expect that it will lead to long-lasting developments of partnerships between the R and epidemiological communities, and that the resources developed will become key assets for supporting outbreak responses well beyond this pandemic.

Deliverables and timeline
Project duration: 15th April 2020 - 18th September 2020

Creation of the website

- 15th April - 1st May: recruitment of a consultant web developer
- 20th May: beta release of the website for internal testing
- 1st June: first public release of the website
- 1st July: second release of the website

Project coordination

- 1st May - 15th May: recruitment of a project coordinator; assembly of an expert panel by the RECON board
- 1st May - 18th September: collect and curate tasks, ensure communication with community
- 15th May: early stage assessment (including RECON board and expert panel)
- 15th June: mid stage assessment (including RECON board and expert panel), including technical review ahead of the second release of the website
- 1st July: creation of a CRAN task view for epidemiology
- 1st September: final assessment

The Team

This proposal is led by Thibaut Jombart on behalf of RECON, and will be managed by the Executive Board of the organisation. TJ is an Associate Professor in Outbreak Analytics at the London School of Hygiene and Tropical Medicine, a Senior Lecturer in Genetic Analysis at Imperial College London, a core member of the UK Public Health Rapid Support Team, and the founder and current president of RECON. He has co-authored more than 70 publications in statistics and epidemics analysis, 15+ R packages, and has first-hand experience of emergency outbreak responses, most recently as a WHO consultant deployed to North Kivu, Eastern DRC, as lead analyst for the Ebola outbreak response (25 weeks on the ground across 3 missions).

RECON was created in 2016 in the aftermath of a hackathon for epidemics analysis (hackout3.ropensci.org) co-organised by Imperial College London (TJ) and rOpenSci (Karthik Ram). In September 2018, RECON became a Non-Governmental Organisation registered as a not-for-profit, incorporated association in France (registration number: W751246083). With 12 packages on CRAN and about twenty others in development, it is currently the single largest initiative for the development of free R packages for outbreak analysis. RECON is also extensively involved in training, through the development of a platform providing open-access lectures and practicals (reconlearn.org), and has run free or sponsored short courses in various places including Bénin, Bulgaria, Colombia, Malta, South Africa, Sweden and the USA, with major public health partners such as the US Centre for Disease Control and Prevention, the World Health Organization, and the European Centres for Disease Prevention and Control. In collaboration with Doctors Without Borders, RECON has created a series of R resources for field epidemiologists through the R4epis initiative (r4epis.netlify.com).

Recently, RECON has been at the forefront of the response to the Ebola outbreak in Eastern DRC, having contributed a large fraction of the R tools for in outbreak analytics pipeline of the Emergency Operation Centre (in Béni, then Goma, DRC), and deployed 2 members to Goma as analysts through the Global Outbreak Alert and Response Network (GOARN: extranet.who.int/goarn). Members of the executive board and the broader membership of the organisation include epidemiologists, statisticians, modellers and developers from the major infectious disease modelling groups and public health institutions worldwide. Because of our

extensive track record in the development of R tools for infectious disease epidemiology, our experience in emergency outbreak response, and our deep links with the field epidemiology community, we believe we are in an excellent position to carry the proposed project forward and ensure its success.

How Can The ISC Help

1. The formation of an ISC working group to coordinate the project work described above, and to serve as a focal point for coordinating contributions from the R Community.
2. $23,300 of funding broken down into the following items:
    - web developer consultant fees for creating the website (initial release): $5,000
    - domain name registration: $100
    - zoom pro licenses for conferencing / coordination: $200
    - consultant fees for a project coordinator: $16,000 (budgeted at 10h per week for 20 weeks, at a rate of $80 per hour)
    - consultant fees for maintaining the website (second release): $2,000

In addition to financial support, we would require administrative support from the ISC to help with human resources aspects of the project, ideally including the issuance of contracts. We also believe we could greatly benefit from additional support of the ISC in terms of general advice on community strengthening and infrastructure development, technical advice, and networking with the wider R community.

Dissemination

The website itself will be a tool for different communities (epidemiologists, analysts, outbreak modellers, developers) to share skills and information. Its success will crucially depend on our ability to engage the members of the respective communities. We will build on our extensive network amongst field epidemiologists, infectious disease analysts, public health agencies and humanitarian organisations to engage the infectious disease epidemiology community. With support from ISC, we believe we will be able to reach out to the wider R community, and engage members and followers of e.g. RStudio, rOpenSci, and Rladies.

New packages will be hosted on github and stable versions released on CRAN. We will offer hosting on the RECON github organisation (github.com/reconhub/), although it will not be mandatory. Each RECON package released on CRAN has its own dedicated website, providing documentation and tutorials (e.g. https://www.repidemicsconsortium.org/incidence/). Complementary training material will be hosted on RECON learn (www.reconlearn.org). In addition, Greg Martin (RECON communication coordinator) will contribute videos on his global health R analytics Youtube channel (62.7K subscribers) to advertise the project, and further illustrate positive outcomes once new tools have been released.

References

1. Fraser, C. *et al.* Pandemic potential of a strain of influenza A (H1N1): early findings. *Science* **324**, 1557–1561 (2009).
2. Cauchemez, S. *et al.* Middle East respiratory syndrome coronavirus: quantification of the extent of the epidemic, surveillance biases, and transmissibility. *Lancet Infect. Dis.* **14**, 50–56 (2014).
3. WHO Ebola Response Team. Ebola virus disease in West Africa--the first 9 months of the epidemic and forward projections. *N. Engl. J. Med.* **371**, 1481–1495 (2014).
4. WHO Ebola Response Team *et al.* West African Ebola epidemic after one year--slowing but not yet under control. *N. Engl. J. Med.* **372**, 584–587 (2015).
5. Jombart, T. *et al.* The cost of insecurity: from flare-up to control of a major Ebola virus disease hotspot during the outbreak in the Democratic Republic of the Congo, 2019. *Eurosurveillance* vol. 25 (2020).
6. Polonsky, J. A. *et al.* Outbreak analytics: a developing data science for informing the response to emerging pathogens. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **374**, 20180276 (2019).
7. Cori, A. EpiEstim: a package to estimate time varying reproduction numbers from epidemic curves. *R package version* 1–1 (2013).
8. Thompson, R. N. *et al.* Improved inference of time-varying reproduction numbers during infectious disease outbreaks. *Epidemics* **29**, 100356 (2019).
9. Kamvar, Z. N., Cai, J., Pulliam, J. R. C., Schumacher, J. & Jombart, T. Epidemic curves made easy using the R package *incidence*. *F1000Res.* **8**, 139 (2019).
10. Nagraj, V. P. *et al.* epicontacts: Handling, visualisation and analysis of epidemiological contacts. *F1000Res.* **7**, (2018).
11. Moraga, P. *et al.* epiflows: an R package for risk assessment of travel-related spread of disease. *F1000Res.* **7**, (2018).