

## Application name

Unifying, harmonising, and establishing sustainable governance for the epinowcast/epiforecasts ecosystem of infectious disease modelling tools

## Project details

*Please indicate if you are expecting to apply for a Large Award of up to 2 years duration and up to £500,000 budget or a Small Award of up to 1 year duration and up to £150,000 budget. Full budget details will be required at the full proposal stage.*

Large Award: 2 years duration, up to £500,000 budget

## Proposal summary / scope of work

*0 / 500 words Please provide a short summary of the work being proposed. Please note that this summary may be publicly published - do not include any confidential information.*

Despite widespread usage by public health agencies and researchers, our infectious disease modelling tools are fragmented across two organisations with inconsistent interfaces, overlapping functionality, and distributed maintenance burden, creating barriers to adoption and inefficiencies. These tools, including EpiNow2, epinowcast, scoringutils, primarycensored, and epidist, are currently used by major public health agencies including UKHSA, WHO, CDC, and ECDC for situational awareness of infectious diseases, as well as by numerous research groups. Collectively they have been downloaded 100,000+ times with 200+ stars on GitHub and have been cited in over 100 academic papers.

Many of these packages were developed rapidly during the COVID-19 pandemic, prioritising utility over long-term sustainability. Users now struggle to discover and apply appropriate methods, while developers face significant technical debt and duplicated maintenance efforts. The ecosystem lacks clear governance and contribution pathways that would allow it onboard new developers and evolve to match user needs. Without intervention, these challenges will limit the impact and sustainability of these valuable tools.

This project will unify, harmonise, and establish sustainable governance for the epinowcast/epiforecasts ecosystem of infectious disease modelling tools. We will systematically address the current challenges through the following key activities:

1. Creating a single entry point for R packages across epinowcast and epiforecasts organisations to create a coherent ecosystem with clear signposts for users.
2. Seeking user and developer feedback on the divergent user interfaces across packages to enable standardisation.
3. Establishing formal governance structures for the unified ecosystem to support sustainable maintenance, facilitate decision making, and enable broader community participation.
4. Reducing code duplication across packages to improve interoperability and reduce technical debt.

5. Implementing shared components and common backends across packages to enable future integration with next-generation probabilistic programming frameworks such as Turing.jl and create sustainable pathways for maintenance.
6. Improving contribution guides and onboarding materials to expand our contributor pool and distribute maintenance responsibility across a wider community.
7. Creating infrastructure that facilitates communication across the ecosystem to support community-driven packages and diverse contributors.
8. Establishing a unified roadmap for the ecosystem to ensure ongoing relevance and funding opportunities.
9. Running two targeted workshops: one focused on onboarding new developers to contribute to the ecosystem, and another dedicated to training new users in effectively applying our tools for infectious disease modelling.

An example specific activity is the development of primarycensored (2,500+ downloads) to handle key biases in delay distributions. This module is integrated into epidist and will soon be integrated into EpiNow2, and epinowcast leading to reduced duplication of effort and improved functionality. Our project will identify additional shared components like distribution interfaces, latent models, and preprocessing steps that can be implemented once and reused across packages, reducing maintenance burden, improving functionality, and enabling greater collaboration.

Through the activities in this project, we will provide public health agencies and researchers with a more coherent, maintainable set of tools for infectious disease modelling, enabling more effective response to both routine surveillance and emerging threats across diverse global settings.

## Categories of work

- ☒ Technical
- ☒ Community
- ☒ Documentation
- ☒ Training
- ☒ Governance

*Please check all categories of activities that apply to the work you are requesting funding for. Click on the help text (?) for further details of each category. Please note that these categories may be publicly published.*

## Project Team

*0 / 250 words Briefly describe why your team is the appropriate one to carry out these activities. You will be asked to provide details of the Project Lead and Co-Leads on the Team Details page, along with their connection to the software.*

Our team created and maintains the tools this project seeks to harmonise, giving us unique understanding of both the code and communities that rely on these tools.

Dr. Sam Abbott will lead the project at 100% FTE. As founder of the epinowcast community and developer of key packages across both organisations, Sam has translated methodological innovations into accessible tools used by UKHSA, CDC, and ECDC. He has built the epinowcast community to include over 50 researchers globally, demonstrating his ability to foster inclusive collaboration.

Professor Sebastian Funk will provide strategic guidance at a fractional FTE. As Director of CMMID at LSHTM and a Wellcome Senior Research Fellow, Sebastian brings leadership experience and public health networks. He has led multiple open-source software projects, including socialmixr and rbi.

We have established contribution pathways with guides, forums, and support channels across both ecosystems. Core tools like epinowcast and EpiNow2 each have 20+ contributors, showing our ability to convert users into community members. This demonstrates our capacity to scale the community further through improved governance and documentation.

The team will also include a senior research software engineer who will implement technical aspects of unification, interface standardisation, and code deduplication, delivering the technical improvements that form the foundation of this project.

By combining Sam's expertise in software development and community building with Sebastian's strategic vision and dedicated technical resources, we are positioned to create a sustainable ecosystem serving current and future needs of the infectious disease modelling community.

## **Benefit to UK research**

*0 / 250 words Briefly describe the expected value of the proposed work to UK research. You may wish to describe the fields of research or types of research method the software is used in, and describe the benefits to a particular community. You may also want to give examples of what research is enabled by the software.*

This project will deliver significant value to UK research by harmonising critical infectious disease modelling tools currently used by UKHSA, academic groups, and organisations. We need to address the fragmentation of infectious disease modelling tools across our organisations so that users can discover, apply, and contribute to appropriate methods.

UK public health agencies currently use these tools to support outbreak response and surveillance for infectious diseases including COVID-19, influenza, and norovirus. The harmonised ecosystem will enable more robust outbreak response capabilities for emerging threats and facilitate research into reporting delays and biases in surveillance systems.

Researchers use our tools for methodological comparison and validation. For example, scoringutils is widely used for evaluating forecast models, EpiNow2 has been used for evaluating methods like EpiFusion (Judge et al. 2024) and viral load based approaches (Hay et al. 2021), and epinowcast has been used to evaluate methods for nowcasting (Mellor et al. 2024). This cross-method validation is crucial for advancing epidemiological research.

Other examples of research enabled by this ecosystem include real-time situational awareness during outbreaks and robust estimation of reporting delays in surveillance systems. By simplifying the handling of complex surveillance data, our tools make sophisticated modelling more accessible, allowing researchers to focus more on epidemiological questions rather than data challenges.

The project aligns with key UK research priorities around pandemic preparedness, health data science infrastructure, and methodological innovation. By reducing duplication and inconsistencies between tools, we will enable more efficient knowledge transfer between research groups and institutions.

## **Landscape analysis**

*0 / 250 words Briefly describe the other software (either proprietary or open source) that is primarily used by the research community addressed by the work in this proposal. Summarise, to the best of your knowledge, how the software in this proposal compares to these other software in terms of user base size, usage, and maturity. Describe, if appropriate, how the other software interacts with the software in this proposal.*

Our core packages have progressed beyond the prototype stage with multiple releases. EpiNow2, scoringutils, and primarycensored are on CRAN, while epinowcast and epidist have multiple stable releases despite not being on CRAN due to dependencies. All packages have multiple contributors (5+ for smaller packages, 20+ for core tools), 1-5 years of active development, and are past version 1.0. Downloads vary between 2000 and 50,000. All packages are used by multiple public health agencies and research groups.

For real-time estimation, EpiEstim has a significant user base, offering simpler functionality but with methodological limitations. Its user base tends to be less sophisticated and it is less widely used at public health agencies than our tools. Epidemia provides comprehensive features but has seen dormant development for two years. EpiMap has limited traction despite integrating population flows and again is dormant.

For delay distribution estimation, our packages improve upon existing options like fitdistrplus (high downloads but biased methods for outbreak data) and discrete (no active development for 5+ years and biased methods).

For forecast evaluation, scoringRules provides the statistical foundation that scoringutils builds upon, but lacks epidemiological focus. No other significant tools exist specifically for forecast evaluation in infectious disease modelling.

Our work differs from initiatives like Epiverse-TRACE by prioritising methodological innovation and integration for advanced users while maintaining accessibility. Our tools demonstrate higher user engagement metrics and more external contributors.

## **Measure of impact (optional)**

*0 / 250 words Briefly describe how you would measure the impact of the proposed work. You may have existing measures that you use, or you may propose new measures. For the Expression of Interest we*

*are looking to understand how you would approach this, and what you think is important to measure. In the full application, you will be asked to define specific measures.*

We will measure impact across four key dimensions with specific metrics for each:

1. Technical improvements:
  - Increased test coverage
  - Decreased resolution time for critical issues
  - Number of shared components implemented
  - Successful implementation of standardised interfaces
2. User adoption:
  - Downloads
  - GitHub engagement (GitHub stars, issues, and pull requests)
  - Academic citations
  - Public health agency adoption. We would track this based on personal communication with the agency and/or publicly available information.
3. Community health:
  - Number of active contributors. We would track this based on GitHub activity using the API.
  - First-time contributor conversion rate. Measured based on the number of first-time contributors who go on to make a second contribution.
  - Community forum engagement. Active members based on page visits and posts.
  - Contributors spanning multiple packages. We would track this based on GitHub activity using the API and aim for an increase in contributors working on multiple packages.
4. Governance and sustainability:
  - Number of community members securing independent funding to contribute to the ecosystem.
  - Number of community-driven packages integrated into ecosystem.

We will assess these metrics both overall and by package where it makes sense to do so. These metrics will be tracked, with baseline measurements established at project start. We already collect many of these metrics in an ad-hoc manner, but this project will establish a unified process.