CS&S

Co-Designing the Foundations of a Climate Sensitive Infectious Disease Community of Practice

Breakout group discussion at Cape Town, South Africa during 2023 Co-Design Workshop. Photo by: Lihlumelo Hlumie

wellcome

Report Authors

ANGELA OKUNE, CODE FOR SCIENCE & SOCIETY

Director of Programs at CS&S. She is a social scientist of science with a PhD in Anthropology from the University of California, Irvine. She was formerly the Research Director at Nairobi's first innovation hub and Network Coordinator for the Open and Collaborative Science in Development Network.

MILIAKU NWABUEZE, CODE FOR SCIENCE & SOCIETY

Senior Programs Manager at CS&S. She holds an MFA in Transdisciplinary Design from Parsons at The New School and a BA in Economics and Gender, Sexuality, & Feminist Studies. Her experience in tech spans designing software interfaces for international banking systems and writing code for community-based mapping projects.

MICHELLE BARKER, RESEARCH SOFTWARE ALLIANCE

Co-founder and Director of Research Software Alliance. She holds a PhD in sociology with extensive expertise in open science, research software, and research infrastructure. Michelle is a former Director of the Australian Research Data Commons.

KARTHIK RAM, THE ROPENSCI PROJECT | US RESEARCH SOFTWARE SUSTAINABILITY INSTITUTE

Co-founder and Director of The rOpenSci Project and lead at US Research Software Sustainability Institute. He is a research scientist at UC Berkeley's Institute for Data Science and Berkeley Initiative for Global Change Biology. He graduated with a PhD in Ecology and Evolution from UC Davis.

GEMMA TURON, ERSILIA OPEN SOURCE INITIATIVE

Co-founder and CEO of Ersilia. Trained as a molecular biologist, Gemma completed a PhD in colorectal cancer and stem cells at IRB Barcelona. She focuses on building and expanding a user-friendly platform where AI models can be deployed so researchers worldwide can access them.

SUSAN WINKS, H3D FOUNDATION

Head of Research Operations and Business Development at H3D Foundation. Susan holds a PhD in Organic Chemistry and an MBA. She brings experience as a drug discovery scientist leading research projects.

REPORT REVIEWERS

Laura Ación, Metadocencia
Kristen Aiemjoy, UC Davis | Mahidol University
Váleri N. Vásquez, Stanford University
Rayya El Zein, Code for Science & Society
Danielle Robinson, Code for Science & Society

REPORT DESIGN

Huston Malande, Skyline Design

Table of Contents

Executive Summary

Changes in global and regional climate patterns are exacerbating existing health inequities and introducing novel challenges. A particularly concerning ramification of climate variation and change is its impact on the spread of infectious diseases, many of which are climate sensitive. The term Climate Sensitive Infectious Disease (CSID) refers to these infectious diseases whose transmission and spread are influenced by changes and variations in climate and weather. In response to growing awareness about these issues, as well as advances in technologies such as artificial intelligence and machine learning, there has been an expansion of digital tools to better understand and predict the impacts of near-term and long-term shifts in climate on disease spread. If implemented well, such tools have the potential to support governments, grassroots organizations, and individuals to be better equipped and have meaningful impact on health policy.

A Community of Practice (CoP) is a group of individuals who come together over shared interests, have comparable levels of domain knowledge or expertise, and interact often enough to develop a shared understanding of challenges and opportunities (Wenger 2000). In the field of open-source software, CoPs are increasingly recognized as a key piece of ensuring software sustainability and resilience. In early 2023, Code for Science & Society (CS&S) was commissioned to understand the existing communities of practice in the CSID space. This report shares insights from an initial 6-month landscaping of the field and makes recommendations for the convening of the emergent CSID community.

Findings

1. There is demand for a community of practice focused on CSID open-source software tools

There is notable demand for a CSID CoP that draws individuals and their project communities together to grow shared understandings of challenges and opportunities related to CSID tools including data and their standards, software, models, and policy. Since CSID tools require multiple areas of specialization—climate modeling, infectious disease modeling, research software engineering, end-user engagement—a CoP that draws together diverse experts from across these fields is seen as integral for supporting successful CSID software tools. While numerous knowledge-exchange initiatives exist at the intersection of health and environment, there remains a clear gap for a CoP that focuses explicitly on shared learning for tool and model creators. Beyond individual projects, there is not a current community that seamlessly integrates CSID's topical focus with software tool development, inclusive conversations on data, technical infrastructure, and proactive end-user engagement. Through an in-person convening of 45 diverse individuals with an interest in CSID software tools, we found great interest for a new CoP that would connect disease modelers with those who build software tools and climate change researchers to improve the effectiveness and usability of CSID tools. Delegates in attendance expressed their interest in being part of such a CoP and were keen to develop and lead it.

2. This CoP will need social and operational infrastructure to be sustainable

Sustainable CoPs depend on social and operational infrastructure to thrive and grow (Sethi 2017; Ram 2023a). This includes functional governance that adapts with the CoP's evolution; clear pathways to leadership that can help foster new talent and dynamism; reliable financial and operational infrastructure for long-term stability; and community engagement that promotes collaboration and the advancement of shared areas of interest. An emerging CSID CoP must foreground effective decision-making, culture-building, and operational strategies from its inception to avoid common pitfalls that often challenge new communities.

3. This CoP must proactively address barriers to participation and access

38% of the 37 tools used for CSID modeling, as identified by Ryan et al. (2023), were developed in the USA or UK, despite the primary intended users being in the global South. Persistent barriers to community participation such as a heavy reliance on volunteered labor and the inequalities that can result; asymmetrical power dynamics resulting from unequal funding distribution; and uneven access to data and technical infrastructure are some of the pressing challenges that a CSID CoP will need to consider to create an inclusive community. This work is international and interdisciplinary. To build a community that can effectively engage across multiple axes of difference, a CSID CoP must foreground discussions of governance, labor, and barriers to funding, access, and participation. This landscaping work has ignited a community ready to meet this great challenge.

Summary of Recommendations for the Development of the CSID CoP

Our findings underscore that power asymmetries—across regions, identities, and areas of expertise, among other categories of difference—significantly shape aspects of CSID research and tool development. We recommend the following approach to address these issues:

- **Integrate equity considerations into all facets of CoP activities.**
- **Offer differentiated value for CoP members.**
- **Embrace adaptability and anticipate the evolving needs of the CSID community over time.**

Suggested Domain

We recommend a new CSID CoP serve as a home for CSID modeling projects and decision-making tools to grow shared understandings of challenges and opportunities related to CSID tools including data and their standards, software, models, and end-user engagement.

Requires Interdisciplinary Community

The CSID CoP community will need to be comprised of an interdisciplinary set of communities representing expertise in climate science, infectious disease modeling, data science, and software engineering, among others. Key categories of actors to engage moving forward will include:

1. CSID researchers focused on modeling and methods (such as epidemiologists, climate scientists and ecologists).
   - data and software specialists interested in CSID issues (such as research software engineers, data scientists, UX/UI designers).
   - end-user communities and public health practitioners keen to engage on the development of software tools (such as public health decision makers, citizen scientists); and
   - funders supportive of the work.

Community Activities

A CSID CoP is well placed to advance:

A) Sharing best practices for meaningful engagement with CSID tool end users.
B) Bringing existing CSID data, models, software resources together for a global community, and providing a platform to develop new tools that takes advantage of the rapidly evolving technology landscape.
C) Providing thought leadership for the growing field of CSID.
D) Consolidating relevant job, grant, and training opportunities for CSID researchers and tool makers, and nurturing CSID leadership amongst early career researchers in LMICs.

As part of the next phase of work, the COP scope, value-addition for differentiated members, and activities should be further nuanced. We offer prompts, a suggested 24-month timeline and further details in the full report.

Project Overview and Introduction

Climate change, an indisputable and stark reality of our time, impacts human health. Rising temperatures, more frequent and intense heat waves, altered precipitation patterns, and escalating severity of extreme events like storms and floods cause physical injury and death. These changes also indirectly trigger issues like malnutrition, mental health issues, and an upsurge in various infectious diseases. A concerning ramification of climate change is its impact on infectious diseases, many of which are climate sensitive.

What is Climate Sensitive Infectious Disease (CSID)? How does Open-Source Software relate to CSID?

Pathogens and their vectors, such as mosquitoes and ticks, have specific climate needs for survival and spread. Climate changes can alter these conditions, affecting disease distribution and transmission in humans and animals. For example, a modeling study by Colón-González et al. (2021) found that the population at risk of malaria and dengue diseases might increase by up to 4.7 additional billion people by 2070 relative to 1970–99. Therefore, it is crucial to understand these climate-infection connections and integrate climate resilience into our health systems for a healthier future. The term “Climate Sensitive Infectious Disease” is used to describe infectious diseases whose transmission and spread are directly influenced by changes and variations in climate and weather. These include mosquito-/vector-borne diseases as well as respiratory pathogens and water-borne diseases.

In response to growing awareness about CSID as well as advances in technology such as artificial intelligence and machine learning, there has been an expansion of digital tools, such as climate-informed early-warning systems, to better understand and predict the impact of near-term and long-term shifts in climate on disease transmission. If implemented well, such tools have the potential to support governments, grassroots organizations, and individuals to proactively respond. However, to date, these tools and related practices have been unequally distributed, decentralized, and primarily developed and directed by those based outside of regions most affected by CSID.

The Wellcome Trust has articulated concern about the far-reaching implications of climate change on human health and begun to make significant investments towards supporting a transdisciplinary field of “Climate Sensitive Infectious Disease” (CSID). In 2022, Wellcome Trust commissioned a report by the Inter-American Institute for Global Change Research (IAI) that identified technology gaps in global climate-sensitive infectious disease preparedness. The IAI reviewed 9,500 pieces of published work and identified 37 (named) digital tools for CSID modeling (Ryan et al. 2023). The report highlighted several important considerations for the further development of the field of CSID, especially the importance of fostering leadership and building capacity among those most affected by CSID; a need to invest in understudied diseases beyond the vector-borne; and that building technology alone is not enough.[1]

Further, the report authors noted a significant accessibility gap; while several of the available models had freely accessible model outputs, they did not have code repositories, making it challenging to reproduce or use them (IAI 2022; Ryan et al. 2023). Therefore, it was recommended that future funding ensure validated models and research can be readily translated into automated packaged tools. As a result of this recommendation, in early 2023, Wellcome Trust announced the awarding of £22.7 million to 24 research teams across 12 countries for developing open CSID digital tools.

Wellcome Trust and others have recognized that by sharing code and methodologies openly, researchers can verify results and build upon existing work, fostering trust, and advancing scientific knowledge. The accessibility of open source empowers researchers to more efficiently build on others’ work to improve the reach and impact of scientific research.

[1] The three points highlighted are taken from those highlighted in a Wellcome blog about grants for new digital CSID tools written by Bilal Mateen (2022).

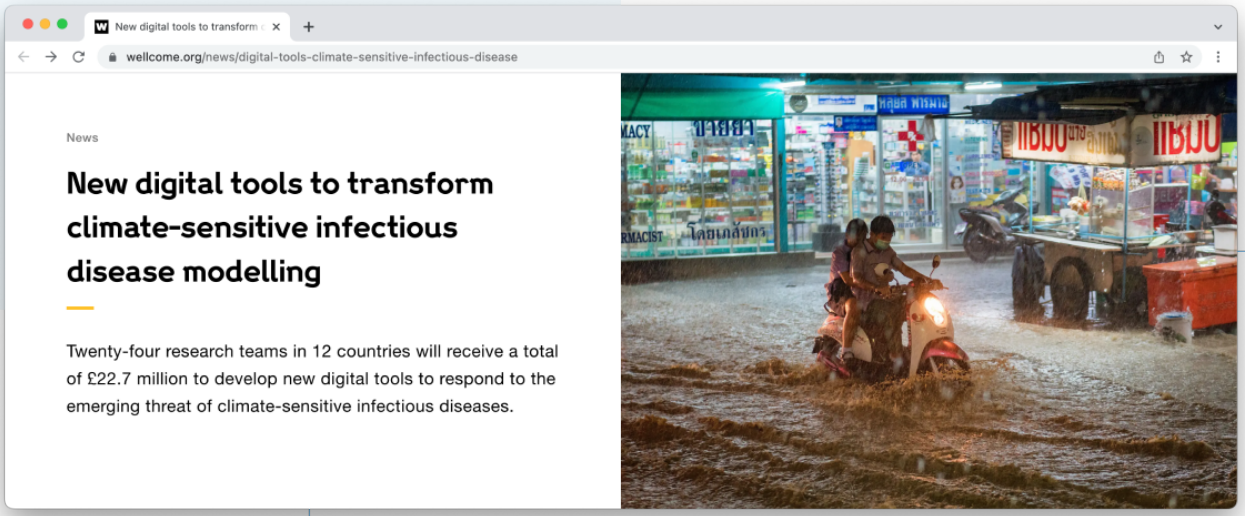


Figure 1. Image of Wellcome Trust blog post announcing investment in new digital tools. Source: wellcome.org/news/digital-tools-climate-sensitive-infectious-disease

In addition to further support of CSID tools, the IAI report also found that there was a need for connecting disease modelers with those who build software tools and climate change researchers to improve the effectiveness and usability of the tools. Thus, in 2022, Wellcome opened an expression of interest for a convening partner to support the development of a broader Community of Practice, extending beyond Wellcome grantees, to effectively combat the challenges faced by the CSID field and its tools, and build more collective knowledge. In early 2023, Code for Science & Society (CS&S) was commissioned to understand the existing communities of practice in the CSID space and opportunities for a community of practice for this field.

What is a Community of Practice and how does it relate to CSID software?

Today, the concept of a Community of Practice (CoP) is broadly understood as a group of individuals who come together over shared interests, have comparable levels of domain knowledge or expertise, and interact often enough to develop a shared understanding of challenges and opportunities (Wenger 2000).

In the field of open-source software, CoPs are increasingly recognized as a key piece of ensuring software sustainability. If a software tool is to be sustained over time, mechanisms for continuous software development, integration, and deployment by a decentralized and open network of contributors are needed.[2]

[2] Open-source software ideas often spring from research paper/project analyses, with some analysis code evolving into prototype research software for community visibility, feedback, and buy-in. These prototypes, often minimally documented, tested, and not necessarily engineered for optimal performance, are made accessible via platforms like GitHub. When these tools find a significant fit in the research infrastructure, either by integrating into ongoing projects or by fostering a surrounding community, questions of long-term sustainability arise. At this stage, the open-source tool, whether it is software, models, or methods, is made public but still needs further transformation before evolving into an ecosystem.

These mechanisms are best established and kept lively by growing a community of practice around a tool. An emergent example is the Epinowcast community which builds and maintains the epinowcast R package and other related packages for real-time analysis of infectious diseases.

In other words, CoPs can enable the transformation from an open-source tool to a self-sustaining community (Ram 2023a) that enables the ongoing collaborative, asynchronous development of an open-source tool designed to be publicly accessible, modifiable, and distributable by anyone under an open-source licensing model (National Science Foundation 2022).

Beyond specific software tools, CoPs can also form around shared domains of interest. For example, CGIAR's crop modeling community of practice focuses on parametrizing interactions within and among the main drivers of cropping system.

Communities of practice are voluntary so what makes them successful over time is their ability to generate enough excitement, relevance, and value to attract and engage members (Wenger, McDermott, and Snyder 2002). Given that the domains of climate science, infectious disease modeling, data science and software engineering are highly specialized, it is improbable for one individual to possess comprehensive expertise in all fields. This circumstance underscores the importance of a CoP in bridging these specialized fields and fostering effective collaboration for improved CSID tools.

Research Framework and Methodology

This research included three distinct aspects:

- First, conducting primary research with researchers and modelers working on CSID to understand how they are currently organized and the existing communities of practice with whom they engage.
- Second, learning from existing open-source and scientific software communities of practice through primary and secondary research to understand key features of a sustainable software community of practice and lessons learned from existing CoPs.
- Third, hosting a co-design convening to discuss and begin to build the foundation for a CSID CoP that the emergent community members would like to see take shape moving forward.

Project Aims and Key Research Questions

This project aimed to co-design and enable the foundations for a sustainable and robust CSID community of practice through engaged research processes. The methodology was framed around four key research questions:

1. What community-building mechanisms have existing CoPs for open-source software leveraged? What has worked and what has not?
2. What models of community-led research are more responsive to local (research) needs?
3. What mechanisms exist for the sharing of data, code, and resources across sectors in Low- and Middle-Income Countries (LMICs)? Which mechanisms appear to be most effective?
4. What practices and structures have been used by open-source software communities to enable collaborative work at regional/intra-regional scales?

Methodology

A collaborative approach is fundamental to CS&S's work as we catalyze new and creative partnerships and alliances to tackle the complex issues of today. We engaged with researchers and open-source software advocates working in or interested in the nexus of climate science, infectious disease, and open-source software to understand the existing communities of practice. The research importantly was a collaborative endeavor with experts in CSID, data science, and scientific research software community-building; see Appendix 1 and 2 for a summary of the contributors to and collaborations undergirding this project. See Figure 2 below for a summary workflow of the overall landscaping.

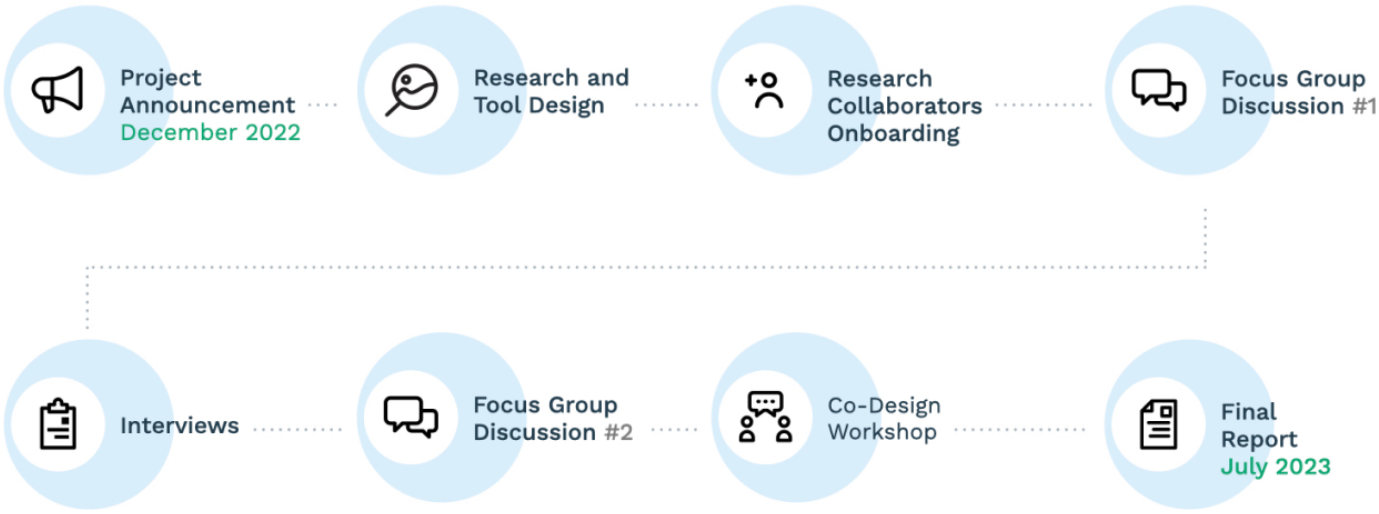


Figure 2. Summary of the landscaping project workflow.

The research team conducted 50 one-on-one consultative meetings and interviews (mostly virtually via Zoom and a few in-person) and two virtual focus group discussions (of 3-4 participants each, one in Spanish and one in English). We spoke with people located in 20 countries and 6 continents including Brazil, Bangladesh, UK, Thailand, Cameroon, Kenya, Uganda, Cyprus, the USA, and the Philippines. Interviewees were asked about CSID community initiatives already in existence, to identify the best way to support an emerging CSID community. Find the full research instruments used in Appendix 4.

Data and Analysis

Data was gathered using primary and secondary methods; see Figure 3 below for a summary.

Open Source and Scientific Software CoPs	CSID Modelers and Researchers	Software End-users and Practitioners
Methods: • Secondary literature review • Consultative meetings • Interviews • Workshop	Methods: • Interviews • Focus group discussions • Consultative meetings • Workshop	Methods: • Secondary literature review • Consultative meetings • Workshop

Figure 3. Summary visualization of the data gathered for the landscaping project.

Data Protection

Compliance with UK GDPR requirements was confirmed in collaboration with the Wellcome Trust legal and data protection teams. Prior to interviewing, the project purpose and scope of involvement were explained to participants. We obtained written informed consent from all focus group discussions and interview participants.

In addition to the qualitative methods named above, we used a Participatory Action Research framework to design and convene an in-person workshop in May 2023 in Cape Town, South Africa, to better understand the needs of the emergent community and to begin to establish a strong, collaborative foundation for the CoP. Find further methodological details in Appendix 3.

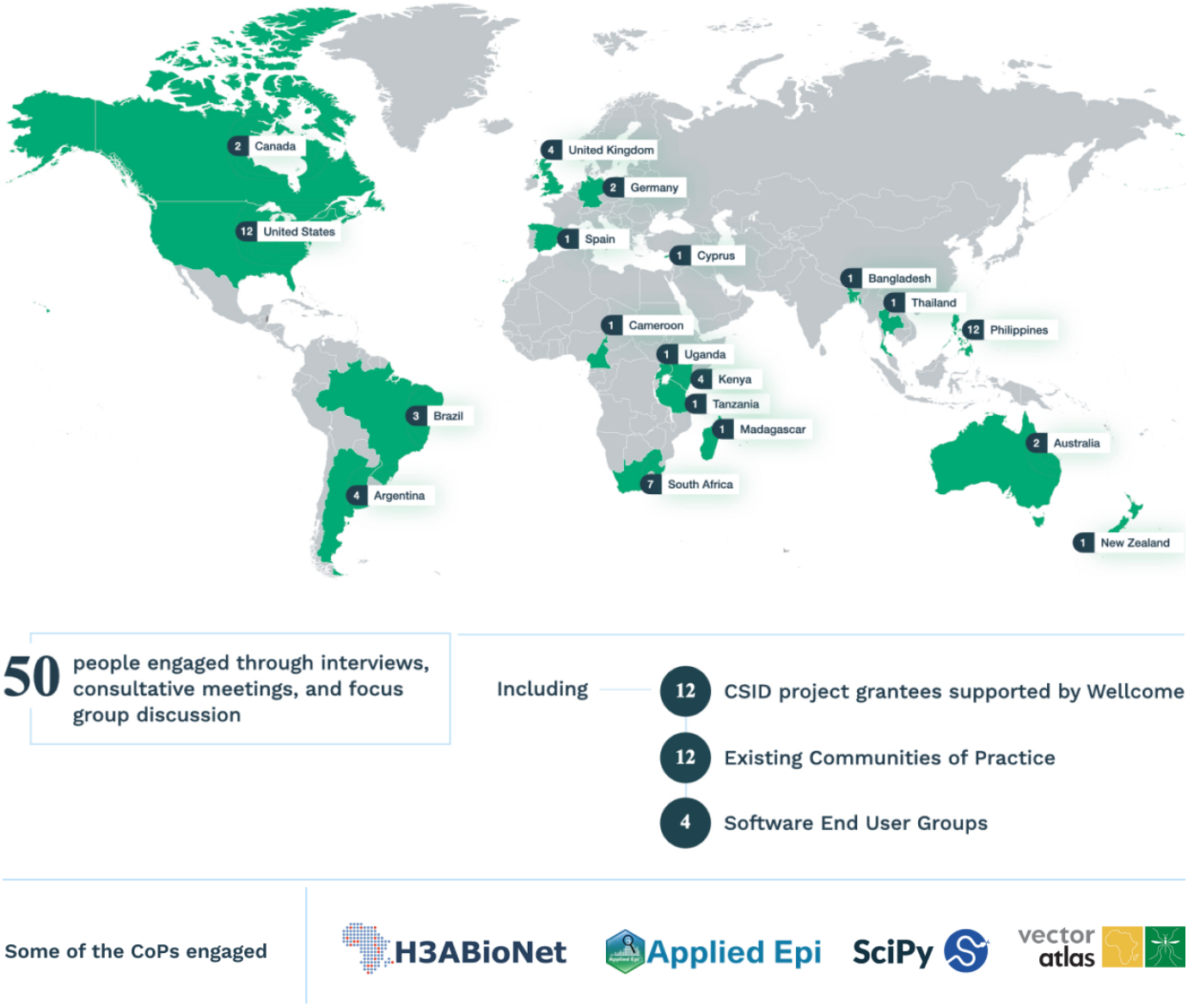


Figure 4. Illustration of the geographic location for research interlocutors engaged as part of this project.

Findings

There are different groups involved in CSID research, with heavy overlap across groups and most affiliated with academic institutions. Those who consider themselves CSID researchers are primarily from fields including epidemiology, ecology, and public health, among others. Some of these researchers engage in the development of modeling tools and research software; however, they often lack up-to-date software engineering practices. Research software engineers form another significant group within this community, instrumental in crafting modeling tools, including open-source software.[3] Climate scientists form a third group that is imperative for this community, however, it has proven more challenging to draw these stakeholders into the conversation.

1. There Is Demand for a New Community of Practice (CoP) Focused on CSID Open-Source Software Tools

Our landscape analysis intended to focus on the various CoPs related to CSID modeling and software development. However, we quickly learned that many of the relevant actors have yet to work together, resulting in a landscape filled with individual persons, lab groups, and independent projects, represented by the numerous small bubbles in Figure 5 below. Existing CSID CoPs were small-scale and many primarily focused on their project community.

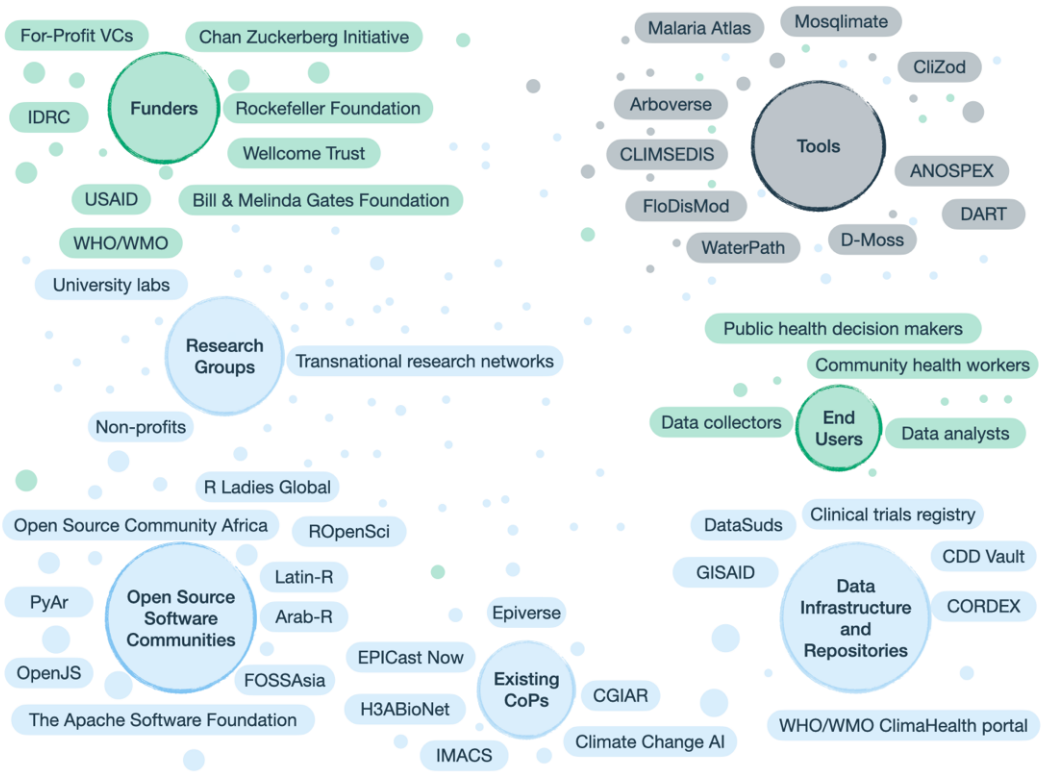


Figure 5. Visual representation of existing CSID communities of practice and related stakeholders. The organizations and types of end-users listed in this diagram are illustrative, and not exhaustive. The many small dots illustrate the plethora of individual research groups and tools that are currently unconnected.

[3] Several organizations, including Research Software Alliance (ReSA) and Journal of Open Source Software (JOSS), are dedicated to addressing the challenge of Research Software Engineers (RSEs) having trouble gaining recognition for their work. While the CSID CoP should partner and work with these groups, the CSID community should maintain its focus on supporting and building connections amongst those developing CSID software (disease and climate modelers and the RSEs they employ), and those using it (end-users).

Perhaps because of this siloed landscape, the participants invited to the co-design meeting were eager to play an active role in shaping the evolution of a new CSID CoP. The co-design participants were interested in a CoP that draws together various existing CSID initiatives to grow shared understandings of challenges and opportunities related to CSID tools including data and their standards, software, models, and end-user engagement. Attendees demonstrated a collective willingness to not merely participate in, but to also guide the development and future direction of this new community.

> *"My initial expectations for the workshop were to participate in a constructive exchange of ideas, learn from the experiences of other organizations, and identify opportunities for collaboration on the themes of digital equity, data governance.*
>
> *I'm delighted to say that these expectations were far exceeded. The workshop succeeded in bringing together individuals from different backgrounds and expertise and facilitated dynamic and productive discussions."*
>
> *— Post-workshop survey respondent*

A post-workshop feedback survey[4] conducted among the participants surfaced that a CSID CoP holds significant value to respondents. Already, after an initial two-day convening, 12 out of 16 post-workshop survey participants agree or strongly agree that they feel part of a broader community working on software tools to understand climate change and disease patterns. Many expressed interest in continuing their engagement with the group as long as it adds value to their everyday work.

There has also been great excitement for a CSID CoP to be formed as expressed during consultative meetings with potential future collaborators representing various CSID-relevant organizations including funding bodies, community groups, and scholarly partners. This affirmation of interest and commitment amongst members and potential partners bodes well for the prospective growth and success of the community of practice.

A CoP focused on growing shared understandings and work related to CSID models, software, data, and end-user engagement could help galvanize the efforts of the CSID modeling community which otherwise would remain isolated individual efforts.

Figure 6. CSID Co-Design Workshop attendees ideating next activities for the CoP. Photo by Lihlumelo Hlumie.

[4] Find further insights from the post-workshop survey in Appendix 8.

Key Issues of Concern within the CSID Community

Based on interviews and extended discussions with co-design workshop participants, several sector issues emerged that the CSID community is grappling with and might tackle in a CoP:

Lack of Support for Early Career and Marginalized Groups to Access Data, Computing Infrastructure, and Mentorship Support

Relationships were highlighted as key not only to data access but also to ensuring CSID models are used. A workshop participant highlighted the importance of considering why the model is needed in the first place, to what purpose the data will be used, and how to close the feedback loop to enable communication thereafter. “If data and model uptake are based on social relations,” another workshop participant asked, “how can new entrants to CSID—small organizations or young Primary Investigators that are just starting their careers and may not have a strong network—be supported to have access that is equitable and enables them to participate?” Building processes that enable those who have access to support newcomers also gain access was viewed as a starting point to addressing this issue.

Local Engagement Across the Tool Creation Process

The importance of understanding the local contexts within which CSID actors are operating was a key point made with regards to understanding CSID software end-users. Rather than simply pushing the uptake of existing CSID tools to end-users, workshop participants encouraged each other to listen for users’ own areas of concerns related to CSID. This was referred to as thinking about the “demand-side” for the tools rather than just “supply-side.” While early involvement of local stakeholders was deemed beneficial to the development of a successful CSID tool, respondents acknowledged that this is difficult and can often slow down the speed at which things can happen. There was prolonged discussion during the workshop about disaggregating the “end user” category to identify the diverse skills, interests, and needs of the different end-users. It was recommended that the right individual champions who understands the issues and can influence decision-making and resource allocation be identified and engaged throughout the process of development of CSID tools (not just at the end).

Responsible Handover of CSID Models and Tools

Workshop participants discussed the need for more responsible handover of CSID tools and models from those who develop them to those who use them, noting that there must be improved ways to pass on the data and model context when sharing the data/model because divorcing the context from the original creation can at times lead to dangerous misunderstandings of the limits of the model/data. A representative from a County Government Hospital mentioned the importance of understanding the confidence and knowledge of different decision-makers to ensure that any necessary additional training can be offered as part of the handover of the model/tool to end-users.

Evidence Standards and Communicating Uncertainty

There was discussion about the match-up of different uncertainties between climate models and epidemiological models and how different scientists think and incorporate uncertainty across scenarios, across models, and within models. Questions were raised about the most appropriate sets of uncertainties or samples to feed from one model into another and the matching of spatial and temporal scales, which informs which uncertainties are propagated from one model to another. The workshop underlined that continuing these discussions could help the CoP support the community on how to better handle scientific uncertainty and develop evidence standards / mathematical models for reuse as well as communicating model uncertainties to end-users.

Find conference proceedings and further insights surfaced during the workshop in Appendix 8.

Figure 7. Fishbowl Session on Day 1 of the CSID Co-Design Workshop. Photo by Lihlumelo Hlumie.

In addition to learning about many individuals, labs, and projects, we identified broader knowledge-exchange initiatives at the nexus of public health and the environment. There are existing CoPs that support the development of CSID at large without an explicit focus on software tools. Some of these communities include: GeoHealth Community of Practice; The International Society for Neglected Tropical Diseases; and the Institute of Malaria and Climate Solutions hosted by Malaria No More. A more comprehensive list of existing groups and projects are relevant is listed in Appendix 6 and a subsection of these may be highly relevant to a CSID CoP.

Community structures of these CoPs ranged from very formal governance models such as the Open Modelling Foundation to much more informal networks such as Sisonke Biotik which reach individuals outside the traditional academic circle, giving opportunities to those without official affiliations.[5] Paradoxically, one interviewee noted the challenges can arise when a community has strong governance and licensing structures - which would seem valuable to transparency and sustainability - but can then result in participants being concerned about issues such as the ultimate ownership of their research.

The research indicated limited interaction between various CoPs, suggesting they operated in silos. Recognizing this, one participant stressed the need for platforms that share information across these communities, for example on upcoming events and training relevant for the field of CSID at large.

Current CSID Community-Building Mechanisms (Research Question 1)

Current community-building mechanisms across CSID communities largely rely on common channels through which academics interact, such as:

- **Conferences** (e.g., Models of Infectious Disease Agent Study (MIDAS) annual conference),
- **Regular online events or meetups** (e.g., Epiverse TRACE seminar series and The International Society for Neglected Tropical Diseases' topical webinars),
- **Email lists** (e.g., British Ecological Society Special Interest Group in Quantitative Ecology),
- **Code hosting platforms** (e.g., GitHub used by Applied Epi), and
- **Real-time messaging platforms** (e.g., Mattermost, Discord, and Slack are among the most commonly used).

[5] See further details about governance structures that have been used to enable work at different scales in Appendix 11 (Research Question 4).

Many of the most significant interactions outside of laboratory or departmental contexts occur with collaborative partners formed via funded projects. These collaborations often conclude when the project reaches its completion and the accompanying funding ceases.[6]

TIME-BOUND TRAININGS

Beyond the knowledge exchange mechanisms mentioned, there are other initiatives that aim to enhance capabilities relevant to CSID via focused, time-limited training programs. For example, there are various annual trainings run by the International Clinics on Infectious Disease Dynamics and Data Program (ICI3D), which include a clinic on Software Engineering for the Applied Mathematical Sciences (SEAMS); a clinic on Dynamical Approaches to Infectious Disease Data (DAID); and a clinic on Meaningful Modeling of Epidemiological Data (MMED). These clinics range from 1 to 2 weeks and focus on topics such as infectious disease dynamics and foundations of modeling; principles and practices of software engineering; and the use of data in understanding infectious disease dynamics.

Interview respondents saw training as a successful strategy in cultivating CoPs, especially beneficial to Early Career Researchers (ECRs) seeking to enhance their skills. Potential training opportunities included peer-to-peer learning, research rotations, workshops, and in-depth talks. CSID researchers can find support in field-specific training programs such as Applied Epi, which offers epidemiological model training to field researchers, and the H3D Foundation, focusing on capacity building for infectious disease drug discovery in Africa.

EARLY-CAREER RESEARCHER OPPORTUNITIES

Many CoPs were seeking more members from Latin America and/or Africa, with one participant noting that since their CoP consisted only of organizations, geographical diversity was even more difficult to achieve in its membership.

Opportunities for early-career researchers to allow them to pursue careers in their countries of origin appear important for moving CSID research forward. Notable initiatives such as the H3ABioNet, a Pan-African initiative spanning 16 countries, successfully train researchers through a community approach across 28 nodes. However, such consortia require substantial, sustained funding and are challenging to replicate.

FELLOWSHIPS

Fellowships outside academia, akin to the Mozilla Open Leaders program, a model adopted by organizations such as Open Life Sciences (OLS), Open Hardware Makers (OHM), and the Software Sustainability Institute (SSI) are currently non-existent in CSID software. A “champions” program model could also have significant potential to nurture a diverse range of community leaders and contributors, as exemplified by rOpenSci and Openscapes.

CONFERENCES AND JOURNALS

Some of the relevant annual conferences identified for the CSID community include the American Society of Tropical Medicine and Hygiene (ASTMH) annual meeting; Joint International Tropical Medicine Meeting (JITMM); and Epidemics: International Conference on Infectious Disease Dynamics. A shared publishing venue can be another important field- and scientific community-building mechanism. A sample of relevant journals for the CSID community include Emerging Infectious Diseases, Environmental Modelling and Software, and Lancet Planetary Health. A more extensive list of relevant conferences and journals can be found in Appendix 5.

[6] While some CoPs emerge from such funded research projects, often these research projects are not designed to serve as peer-to-peer spaces for collective knowledge building and sharing, but rather designed as project-managed initiatives with clear directives and external deliverables.

Networking and community building, facilitated through consortiums, open-source communities, or personal connections, are essential to establishing successful CSID projects. However, because of the geographic and disciplinary diversity of the CSID research community, there is currently not a clear conference that most actors attend or one journal in which most CSID researchers publish. The ASTMH meeting was the most frequently referenced meeting, but several interviewees noted that they are not able to attend due to the high travel costs and visa restrictions associated with attending a conference in the US in person. Community-led events, such as those sponsored by the CS&S Event Fund, though smaller and less prestigious, are more accessible to local scientists and may be important for initiating new communities of practice in the field.

New models for more distributed and hybrid or virtual events are also being established and appear to pave a path for more equitable ways for gathering international communities, especially important given the climate crisis and the contribution of air travel to the crisis (Chasi and Heleta 2022).

There are few spaces that facilitate regular engagement with non-academic stakeholder groups such as industry software developers, policy makers or clinicians. Some notable examples include those organized by national or regional bodies such as the Pan American Health Organization (PAHO). Our research indicated that while CSID researchers engage with different academic (and sometimes non-academic) communities, sharing insights for collaboration on open code and software tools specific to CSID modeling is not a focus. A new CSID CoP would bridge the gap between developers, experts, and end users (such as policymakers and physicians), creating the necessary infrastructure for a holistic, effective community. It also promises to aid in transitioning tools from research labs to end users, increasing accessibility and impact.

2. A CSID CoP Will Need Social and Operational Infrastructure to Be Sustainable

In the following segment, we characterize existing relevant CoPs and highlight key features of a sustainable CoP.

CoPs can be understood using three characteristics: domains (such as particular diseases or a function such as modeling), community development (such as peer-to-peer learning or annual conferences), and practices of shared interest (such as facilitating networks or creating new sharing best practices) (Edmonton Regional Learning Consortium 2016).

In terms of domains, the largest number of Open-Source Software (OSS) CoPs target the development of a specific software project. For example, the ESMValTool is a community diagnostic and performance metrics tool for routine evaluation of Earth system models in the Coupled Model Intercomparison Project (CMIP) around which a community has developed to build and maintain the tool. Other types of CoPs are focused around areas such as:

- Programming languages, e.g., rOpenSci, pyOpenSci
- Research disciplines, e.g., Astropy, CoMSES Net
- Software platforms, e.g., Google Developer Groups, Galaxy Community Hub
- Developer demographics, e.g., Arab-R, Latin-R

CoPs have different ways to develop community, including:

- Use of an open code repository platform for shared development of an OSS project, e.g., GitHub
- Regular meetups, e.g., R-Ladies Nairobi (and R-Ladies Global)
- Conferences, e.g., SciPy Latam
- Discussion groups, e.g., Google Developers Group Lagos
- Podcasts, e.g., Sustain

- Sharing best practices, e.g., PyCon US Maintainers Summit discusses and fosters best practices on how to maintain and develop sustainable projects and thriving communities.
- Code contributions, e.g., Open Source Community Africa (OSCA) aims to increase the rate of credible contributions by African software developers, designers, writers, etc., to open-source projects both locally and globally.
- Impact, e.g., OpenForum Academy Symposium covers questions relating to the social, political and economic impact of open source.

Key Features of a Sustainable Software Community of Practice

The establishment of a sustainable software Community of Practice (CoP) rests on several critical foundations that nurture its growth and evolution. In particular:

FUNCTIONAL GOVERNANCE

A balanced governance system is crucial - neither too informal nor too rigidly structured, but efficient and capable of driving community actions. Governance models can and should evolve in accordance with a community's needs. For instance, early-stage projects may not require formal governance, but as they gain traction—signified by a growing user base or increased funding streams—more structured governance can become beneficial (Ram 2023b).

PATHWAYS TO LEADERSHIP

A sustainable CoP must establish pathways to new leadership for continuous growth and evolution of a community. Such pathways can include mentorship programs, rotational roles, or leadership training, which are designed to support the growth of future leaders and ensure a steady flow of diverse talent into leadership positions, keeping the community dynamic and adaptable over time.

FINANCIAL AND OPERATIONAL INFRASTRUCTURE

Effective mechanisms to receive and manage funds are crucial to the functioning and sustainability of any community. This could involve establishing an organizational structure that allows for receiving donations, grants, or other forms of funding, whether as a non-profit entity, a sponsored project under an existing organization, or via other legal frameworks. Furthermore, implementing transparent and accountable financial management practices, such as regular financial reporting and audits, can enhance trust within the community and among external supporters. Such financial stability and transparency can attract further investment, fostering the growth and long-term sustainability of the community. Establishing strong operational infrastructure also enables community members and topical experts to focus on content and programming.

CLEAR SCOPE

Establishing a clear scope and mandate for a CoP is essential to its functioning and coherence. A clear scope helps to align members around a common purpose, specifying what the community aims to achieve and what activities or issues fall within its purview. Scope should be established by and regularly reflected upon by community members. Equally important is delineating what lies outside the CoP's realm of responsibility or interest, which can help to prevent dilution of efforts and potential misunderstandings. This understanding essentially provides a roadmap for the collective journey, articulating what the community is working towards and the boundaries of their shared endeavor.

CREATING A VALUE-ADDED EXPERIENCE

While the sharing of knowledge is a foundational aspect of a CoP, its long-term sustainability often relies on providing added value beyond this core function. This can take the form of new funding opportunities, which can enhance individual and collective capacity to pursue shared interests. Equally important are incentives that align with the individual evaluation criteria in members' respective professional or social environments. For example, offering recognition or opportunities that enhance career advancement, such as exchange fellowships, can provide motivation to actively contribute to the CoP. Hence, creating a rich, value-added experience for members is crucial in fostering a robust and enduring CoP.

INVESTMENT IN COMMUNITY'S SOCIAL INFRASTRUCTURE

The sustenance of a vibrant CoP hinges significantly on the recognition and consistent investment in social infrastructure, that is, broadly speaking, the elements of community that act as a conduit to bring people together and create affordances for social connection (Klinenberg 2018). This involves roles such as a community manager, who ensures smooth coordination, active engagement, and conflict resolution within the community. Furthermore, regular communal social events can provide platforms for networking, learning, and fostering a shared sense of identity and purpose. Additionally, maintaining an online platform for continuous dialogue allows the community to collaborate, share insights, and build on each other's work in real-time, thereby fostering a dynamic and interactive CoP. This investment in social infrastructure plays a pivotal role in the growth, cohesion, and longevity of the community.

3. A CSID CoP Must Proactively Address Barriers to Participation and Access

Expanding upon the previous section, which detailed key characteristics of a sustainable software CoP, this section focuses on key challenges that will need to be addressed by the new CSID CoP including persistent barriers to community participation; uneven access to funding; and unequal access to data and infrastructure. Research Question 2 sought to understand existing models of community-led science that are more responsive to local needs. But despite a stated desire by many of the CSID researchers to produce work responsive to local needs, we did not identify strong models for local engagement beyond the leveraging of pre-existing social relations built by the Primary Investigator of a project.

Persistent Barriers to Community Participation

OSS projects require extensive maintenance from diverse community members with different skills (Eghbal 2020). With a project's expansion, technical demands evolve as well as an increasing need for non-technical skills such as user adoption advocacy, training, and onboarding (Goble 2022). Such community expansion necessitates greater investment in community engagement focusing on increasing participation and handling issues of motivation, retention, and contribution barriers.[7]

Despite growing investment in the social infrastructure of OSS communities and increasing attention to issues of diversity and inclusion, barriers to participation persist in OSS communities (Kaur, Kaur Chahal, and Saini 2022) and there are still notable differences in who participates in these communities and how (Dunbar-Hester 2020). Many OSS communities, like other areas in science and technology, continue to be led by white men from the U.S., Canada, and Western Europe. But, as biostatistician Laura Ación (Research Scientist, Universidad de Buenos Aires) and colleagues have articulated, there remains a promise of scientific open-source development that offers people "not only access but also agency as first-class participants and co-creators to people from all nations," (Ación, Peña-Castellanos, and Pérez 2022). The question of how open science movements can resist reproducing long-standing scientific systems of extractivism and instead actively dismantle such legacy structures is part of ongoing discussion and praxis (Chan et al. 2019). The following are some of the persistent barriers to more diverse community participation in OSS projects:

---

[7] Models such as the Center for Scientific Collaboration and Community Engagement (CSCCE) Community Participation Model are useful in describing different methods of member engagement that are utilized by many CoPs to increase participation, and CSCCE Community Profiles provide snapshots of a range of scientific communities, enabling comparison across areas such as activities, opportunities and challenges. These show that an increasing number of OSS CoPs now recognize the need for continued investment in social infrastructure. rOpenSci provides an example of a community that recognizes the need for social infrastructure, hiring their first Community Manager in 2016, and diversifying their stakeholder participation modes. See Appendix 9 for further details on rOpenSci.

HEAVY RELIANCE ON VOLUNTEER LABOR

OSS projects are often undertaken by a small group of centralized experienced maintainers (Littauer et al. 2021) who are typically poorly compensated, or work for free (Eghbal 2016; 2020). In the focus group discussion conducted in Spanish, a respondent emphasized this point, noting the heavy reliance on volunteer labor and the inequality it can create within communities, where those with the means or time to contribute can do so, inadvertently excluding others. The group concluded that if the workforce consists only of volunteers, true inclusion and diversity are not achieved and easily becomes a "burnout machine" in the long run.

LACK OF CLEAR EXPECTATIONS AND BENEFITS

The topic of payment for CoP labor and participation is a complex and nuanced topic that also emerged during the Cape Town workshop and should be an ongoing topic of discussion as the CoP establishes. Short of making everything highly transactional, which is not desirable or sustainable outside of a business context, volunteers are key to a CoP. But in order that CoP participation is not experienced as extractive, expectations need to be clearly stated and the community structured so that those who do contribute experience clear benefits stemming from their participation. These benefits might include being a co-author on a published scientific paper, a shifted perspective of one's work or field, or connecting with a new collaborator.

LACK OF EQUITY IN EVERYDAY PRACTICES, DESPITE STATED VALUES

If volunteers do not derive value from their participation, then despite the appearance of being a grassroots community, the Spanish focus group respondent noted that there can be an undercurrent of exploitation. This can gradually erode the community, especially when community leaders use discourse of progress, inclusion, and diversity, but such values are not actually practiced in the community. This point–that CoPs that do not genuinely put their stated values into practice are not actually sustainable–emerged throughout the co-design workshop, emphasizing the need to weave equity into every decision about the CSID CoP, from determining the technical infrastructure to put shared code on to how to fill leadership functions of the community.

MORE IMMEDIATE ISSUES TO ADDRESS

Another important point that arose in the focus group discussion in Spanish and during the Cape Town workshop was that it can be difficult to agree on long-term planning related to climate change and disease because immediate problems such as hunger and homelessness take precedence. A participant noted that the stark reality of these present social issues in her region in Latin America, make it challenging to discuss larger, long-term issues like the governance of communities of practice.

> ***ES:*** *"Entonces, a ver, no estamos pensando en la gobernanza de comunidades que trabajan en cambio climático, ni de otras cosas. Tenemos gente que se está muriendo de hambre. O sea, tengo una persona viviendo en la puerta de mi casa y vivo en Capital Federal. Tenemos problemas mucho más graves que la gobernanza de comunidades de práctica, ¿no es verdad?"*
>
> —
>
> ***EN:*** *"So you see, we are not thinking about the governance of climate change communities, nor other things. We have people who are starving to death. I mean, I have a person sleeping at my front door and I am in the capital city. We have a lot of much more pressing problems than community of practice governance, don't we?*

Case Study: Building and Sustaining rOpenSci as a Community of Practice

(A summary of the detailed case study found in Appendix 9)

Founded in 2011, rOpenSci is a non-profit that provides resources and a community for scientists and research software engineers using R programming. From its initial domain of creating R software tools, it has evolved alongside the R ecosystem, transitioning to supporting individuals in creating their software by establishing standards, offering educational materials, and providing technical support.

rOpenSci's CoP evolution exemplifies the co-evolution of domain, community, and practice. The domain, originally focused on open-source R software development, morphed in response to the broader R ecosystem's dynamics. The community is a diverse group of R package users and developers, and the practice comprises activities like software peer review, community calls, and proposing new standards for evaluation.

Key turning points include the hiring of a community manager in 2016, which influenced community practices, moving away from a focus on specific tools and technical approaches to more meta-topics. This in turn brought visibility to new aspects of rOpenSci's domain and helped lead to the expansion of peer review to include statistical software and Spanish-language submissions. The hiring of a new community manager —i.e., investing in rOpenSci's social infrastructure—had important positive knock-on effects on both the domain and community. This constant evolution, however, has led to some members struggling to define rOpenSci's mission, revealing the need for tailored content for specific sub-groups in the CoP to support diverse participation levels and inclusivity. The rOpenSci CoP showcases the importance of responsiveness to ecosystem dynamics, and clear communication of community purpose for sustaining a successful CoP.

Uneven Access to Funding and Mechanisms to Receive Funding

There are a range of funding challenges to OSS CoPs, including lack of funding, and lack of mechanisms to receive funding. Funding takes quite varied forms, including the following six types in Figure 8 below.

Even if funding is obtained to support some of the work, OSS projects may not have a mechanism for receiving income.

Figure 8. Different types of funding for open-source projects. Source: Goble 2022.

Given the difficulties associated with setting up a U.S. non-profit tax designation as a 501(c)(3), the immense overhead that comes with administration, and the challenges of receiving US-funding as a non-US entity, many projects have opted for other models (Ritvo, Hessekiel, and Bavitz 2017). For example, one project solved their long-term maintenance and funding challenge for their research software by founding the OpenBioSim Community Interest Company (Woods 2022). Other communities join existing non-profits as fiscally sponsored projects or set up other corporate structures (Benefit Corp, S-Corp, LLC).[8] Fiscal sponsors include NumFOCUS and Open Collective who support open-source and grassroots projects. Among fiscal sponsors, there are marked differences and a range of distinct services offered, with each organization tailoring its support and resources to best meet the unique needs and goals of the communities they serve.

It is important to recognize that funding flows unevenly around the community and often can exacerbate global North - global South inequalities. “There is always a shortage of money to do everything,” a focus group participant mentioned in Spanish. “But in the North, there is more money, so I have been climbing to the international level, because then there is a little bit more funding.” While this dynamic at present may be inescapable because of current capital distribution globally, it goes against the development of local networks and capacity when global North actors remain necessary intermediaries to connect the global South. Asymmetrical power distribution within a project team is often exacerbated by which primary organization receives the funding from external sources and how the money is distributed between collaborating organizations.

Several interview respondents noted that it was common for a CoP to form because of a funded research project. Despite offering various benefits like building networks, sharing knowledge and methodologies, and identifying collaborative opportunities, these project-based groupings often restrict participation to project-funded members and focus on delivering a research outcome. One participant highlighted that such CoPs can inadvertently exclude parts of the community, particularly those formed through competitive bidding, where only winners join initially.

Unequal Access to Data and Sociotechnical Infrastructure

The digital divide, referring to unequal access to information and communication technologies, is a significant issue across all sectors, including CSID research. Such a divide is illustrated by the fact that 38% of the 37 tools used for CSID modeling, as identified in the IAI report, were developed in the USA or UK, despite the primary intended users being in the global South.

Since CSID modeling is a field that merges two specializations—climate and health—robust data sharing protocols and infrastructure are crucial to the success of CSID software tools. However, mismatches in the timescales of infectious disease epidemiology studies and climate change models (Metcalf et al. 2017)—that is to say, the data are collected at different scales—coupled with limited collaboration between climate scientists and infectious disease researchers during project conceptualization, among other things, create challenges to building impactful interdisciplinary modeling tools. A CSID CoP will need to play a coordination role to help prevent data fragmentation, promote comprehensive dataset collection, and encourage ethical data collection and reuse across disciplines and projects.

Importantly, the CoP must consider historical and ongoing practices of data extraction that lead to certain marginalized groups feeling exploited by data collection and less able to benefit from data sharing. In a 2021 State of Open Data report, over 4,000 scientists worldwide reported their main reasons for not sharing data were tied to concerns about its misuse (43%) and not receiving appropriate acknowledgement (39%) (Digital Science et al. 2021). These issues are

[8] A good example of the latter is Julia Lang, the open-source data science language. It has a fiscally sponsored open-source project, and a separate Julia Computing business wing (https://juliahub.com). The Julia programming language community is a good example for further study as it has established standards for both software and community interactions/norms as well as governance structures spearheaded by volunteers (initially) and substantive collaborations with multiple domain-specific sectors (e.g., pharma, energy) across both private and public institutions. As an important and growing scientific programming language, the language bears consideration as software specifics are further discussed within the CSID CoP.

particularly worrying to scientists in LMICs with 45% concerned about misuse of their data (Research and Goodey 2021). In 2018, an interview administered to 100 scientists from the Southern Africa Network for Biosciences (SANBio) identified the major concern regarding data sharing was “having other researchers take my results.” Observational data and existing literature suggest that these concerns arise from neocolonial approaches in global health (Khan et al. 2021; Serwadda et al. 2018) and the fact that, for a variety of reasons, research often takes longer to be completed in LMICs, providing better resourced scientists with the chance to publish more quickly for their own ends (Bezuidenhout and Chakauya 2018).

A "helicopter research" approach where High-Income Country organizations use LMIC partners for data collection but offer minimal recognition, compensation, and training (“Nature Addresses Helicopter Research and Ethics Dumping” 2022) has thankfully begun to be widely critiqued. But a recent study found that almost 30% of Global Health journal articles still do not cite local scientists (Ghani et al. 2021) promoting a "data accessibility division,” whereby pockets of researchers are created, some with extensive data sources and others, the individuals without well-established networks or that are not part of large consortia, without. Such divisions make it particularly difficult for younger scientists and/or those in Low-Income Countries (as opposed to Lower Middle- and Upper Middle-Income Countries) to start their own research programs. Additional challenges shape the contours of CSID data sharing such as inadequate data management and sharing plans and insufficient governmental support in LMICs.[9] Find further detailed information about existing data sharing infrastructures, open data benefits and limitations; and open-source software considerations in LMICs in Appendix 10 (Research Question 3).

[9] Most health data in LMICs is collected in non-electronic formats or basic electronic systems, presenting archival risks like metadata loss, versioning issues, typographical errors at transcription, and difficulties in sharing, which is primarily done via email attachments. The absence of structured databases limits data reuse and combination for greater value. Furthermore, clinical trial data use is often restricted to the trial by patient consent forms, and while deidentified patient data can be shared, it often lacks critical geolocation and temporal data points necessary to meaningfully link to climate data. When data is shared, it is usually governed by a Data Use Agreement which is highly specific and restricted to particular researchers. On the other hand, climate data, mostly collected by governmental agencies, requires continuous monitoring. Sub-Saharan Africa has the lowest investment in weather and climate information services (Georgeson, Maslin, and Poessinouw 2017; Manteaw et al. 2022) and while international support is advancing public data collection and digitization of historical climate records (Kaspar et al. 2022) accessing these datasets can be challenging, even for local researchers. Climate datasets present a further challenge; their size can reach up to petabytes of information, a volume difficult to host and manage by an average computer workstation.

Recommendations and Conclusions

Based on the findings of the study, in this section we outline key recommendations for Wellcome Trust as it looks to establish a new community of practice.

Our research found that 1) there is demand for a new CoP focused on CSID open-source software tools; 2) a sustainable CoP requires investment in social and organizational infrastructure; 3) a CSID CoP must proactively address barriers to participation and access.

These findings underscore that power asymmetries—across regions, identities, and areas of expertise, among other categories of difference—shape CSID research and tool development, including collaborative relationships, data access, and model utilization. However, these issues cannot be relegated to one group to solve. There is a pressing need to establish new relationships and community mechanisms for engaging across different backgrounds and perspectives. These insights play a critical role in informing the recommended approach for further developing a CSID CoP.

Recommended Approach

Integrate Equity Considerations into All Facets of CoP Activities

Addressing disparities in the production of global scientific knowledge has been articulated as a cornerstone principle for this CoP. During the first six months, we recommend the CSID CoP host community-wide discussions that tackle different aspects of how to integrate equity considerations[10] into all facets of the group's activities. Discussions should enable reflection on group power dynamics and practice divergence and articulation of implicit norms and habits. Outcomes from the discussions may include the development of a community decision-making guide or a community agreement on the culture desired within the CSID CoP including core values and guiding principles. We recommend follow-up discussions be scheduled at regular intervals to ensure these values remain vibrant and prominent.

Offer Differentiated Value for CoP Members

A successful CSID CoP needs to be interdisciplinary by design, drawing in people working in different work environments including non-profits, governments, and universities. These different work environments each have their own incentive structures. For example, in today's academic environment the maintenance of code and software is still not rewarded, while publishing an academic paper in a high impact factor journal is. Thus, the CSID CoP will need to take such incentive structures into account to ensure participation for CoP members offers value to its different members.[11] To achieve this end, during the first 6 months, we recommend the CSID CoP map out the various roles within the CoP and anticipated tasks and benefits. Instead of expecting all CoP members to engage equally (realistically, a futile goal), we suggest that the community make the different levels of engagement explicit,

[10] As part of this discussion, the term "equity" should be further unpacked. We observed in our research that "equity" often referred to racial and geographic axes of difference and that there was great concern about who was and was not able to participate in the CSID CoP. Notably other categories of difference seemed less salient for the group, for example socioeconomic, sexuality, and linguistic difference. A variety of barriers were identified that prevent people from certain backgrounds and jobs from fully participating in CoPs. For example, it was noted that those who had to fund themselves through "soft money" that is, fundraise for their own salaries, could not participate in the same way in volunteer-run community networks as those who might have more stable income sources who have more time to donate.

[11] For example, for many academics, long-term career stability offered by paths like tenure will outweigh the allure of short-term financial incentives for participation that the CoP might be able to provide (e.g., an honorarium for serving on a board of advisors or as a mentor). Consequently, incentive design should be strategic and grounded in a deep understanding of the systemic motivations that pervade the institutions where CoP participants work. This might look like ensuring that some of the CoP activities include opportunities for academic publishing or developing programming to connect potential collaborators for project grant applications.

give each group of stakeholders[12] a title and set of rights and responsibilities, and intentionally design transparent pathways for people to move in and out of different roles.

Embrace Adaptability and Anticipate the Evolving Needs of the CSID Community Over Time

We have learned from Open-Source Software communities that recognizing the evolving needs of a community is paramount to creating a resilient and responsive CoP. Interviews with open-source software initiatives revealed they were hampered by excessively bureaucratic and heavy governance structures. Such structures proved not only ill-suited to their actual needs, but also inflexible to accommodate changes, ultimately leading to their failure. It is vital to establish the right level of structure that corresponds to the current state of the community. Long-term decisions should be made with caution, prioritizing flexibility and adaptable infrastructure over rigid, predetermined systems. To this end, we suggest an interim legal structure be established initially to allow for longer community discussions about the long-term appropriate governance structure for the CoP. We recommend a fiscal sponsorship arrangement or other similar structure that will enable the community to focus on topical content rather than the administrative set-up for a new CoP.

Recommended Community Design

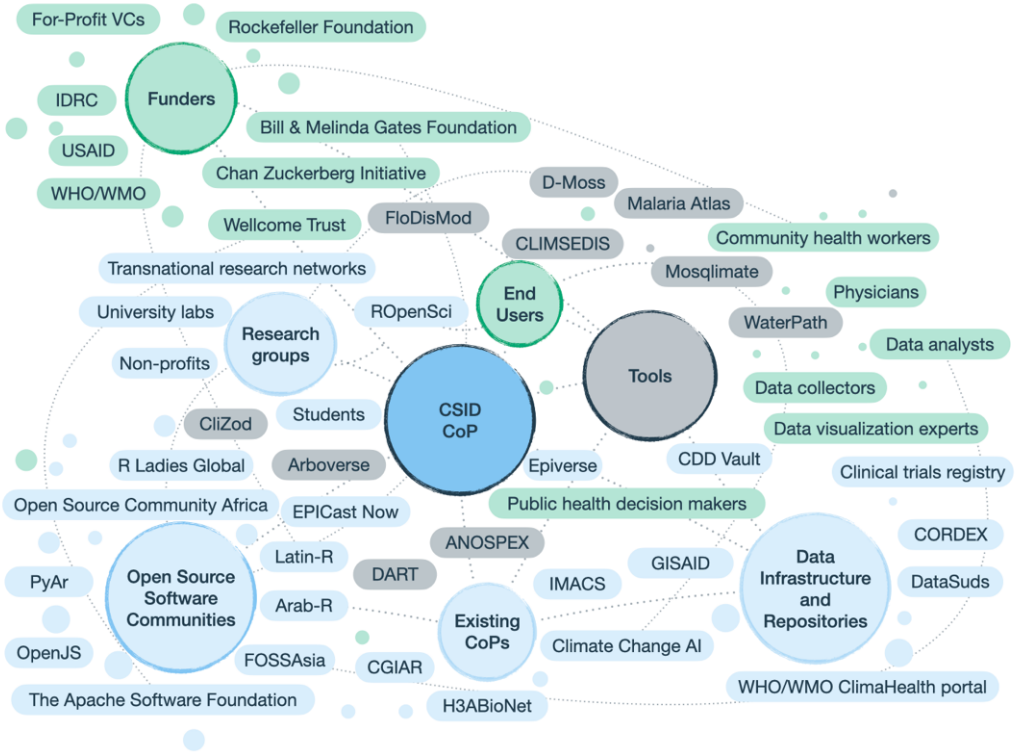


Figure 9. A diagram illustrating the CSID ecosystem with the new proposed CSID CoP creating new connections across different existing communities, groups, and projects. The organizations and end-user categories listed in this diagram are illustrative and not exhaustive. The previously small, unconnected dots illustrating the plethora of individual, unconnected research groups and tools from Figure 5 have been replaced with increased engagement both within and across stakeholder groups illustrated by connected lines. In this conceptualization, the CoP serves as a facilitator and catalyst for connections.

[12] Some possible roles to be discussed by the community could include an executive committee; working group lead; working group participant; and general member.

As noted earlier in this report, there are three core structural elements to a CoP: i) a domain of focus, ii) a community of practitioners at various levels of expertise and iii) a practice or set of activities the group does together to support and advance their learning. Based on the landscaping research, we propose the following design elements for the CSID CoP:

Domain of Focus: Climate-Sensitive Infectious Disease Open-Source Tools

There is a growing field of CSID research and accompanying tools. The proposed CSID CoP would be a home for CSID modeling projects and decision-making tools to grow shared understandings of challenges and opportunities related to CSID tools including data and their standards, software, models, and end-user engagement.

As part of the next phase of work, the CSID CoP scope should be further nuanced by collectively answering questions such as:

- What diseases are in scope/out of scope?
- What detailed methods will be used to solidify that topical scope?
- What types of software tools (simulation, decision, forecasting) are most needed, and are these different for different diseases?
- What policy structures – local, regional, national, international, or combination thereof – need to be engaged, and are these different for different diseases and/or countries?

Key Stakeholders to Engage

The CSID CoP community will need to be comprised of an interdisciplinary set of communities. Key categories of actors to engage moving forward will include:

- CSID researchers focused on modeling and methods (such as epidemiologists, climate scientists and ecologists).
- data and software specialists interested in CSID issues (such as research software engineers, data scientists, UX/UI designers).
- end-user communities and public health practitioners keen to engage on the development of software tools (such as public health decision makers, citizen scientists); and
- funders supportive of the work.

In this initial phase of work, we found it challenging to connect with climate scientists who could be interested in joining a CoP on Climate Sensitive Infectious Disease modeling and tools. This may be explained by the fact that many who currently identify as CSID researchers come from health and disease training rather than from backgrounds in climate science. In many cases, the PI developing the software for a CSID tool is from a health rather than from a climate background. This is not to say that climate scientists are not interested or relevant for a CSID CoP, quite the contrary. The one climate scientist who participated in the co-design workshop mentioned that she found it very relevant and hoped others would participate in the future. She highlighted that it was important to draw in climate scientists early on as part of the conceptualization of a CSID project and not just for the data at the end. It will be imperative to establish specific channels to identify, engage and draw in climate scientists interested in exploring this interdisciplinary area of work.

A key first step could be to attend climate science conferences to establish relations with those trained in climate science that could be interested in the interdisciplinary work of CSID modeling.

This challenge also brings up an important issue regarding prevailing understandings of who constitutes a CSID researcher. Given that this is a relatively new and emerging field, as well as a new label, there are many researchers who would not immediately self-identify as a CSID researcher. Furthermore, the field has remained dominated and led by academics from certain fields, reinforcing long-standing expectations about who the "experts" are and perpetuating conventional knowledge hierarchies. If the CoP seeks to meaningfully include those from marginalized communities, regular assessment of who is considered a CSID researcher and software end-user will be necessary.

be considered. As depicted in Figure 10, we borrow an "oyster" model of community engagement studied by digital humanities scholars (Fenlon et al. 2022), which illustrates expanding circles of potential allied groups. A bidirectional arrow signifies the dynamic nature of community interactions, and the wavy edges symbolize the fluidity and dynamism of communities across different strata. These diverse communities have distinct motivations and reasons for participation. There is a natural and necessary movement in and out of various levels of engagement.

An integral part of the next phase of our work will be a deeper engagement with these groups to more thoroughly comprehend their individual motivations and interests in contributing to the community. Over time, engagement metrics and other measures of demographic, domain-specific, or salient policy reach could also be incorporated as part of an inward-looking CSID CoP data collection and impact assessment over time.

In the visual below, an arrow represents the pull from outer strata to the core of the CoP, which is explicitly part of the sustainability planning: how to "reel" communities and their members closer to the core of the project, to increase their investment and their role in sustaining the CoP. At the same time, the arrow is bidirectional, to acknowledge that there is a natural and necessary movement in and out of various levels of engagement.

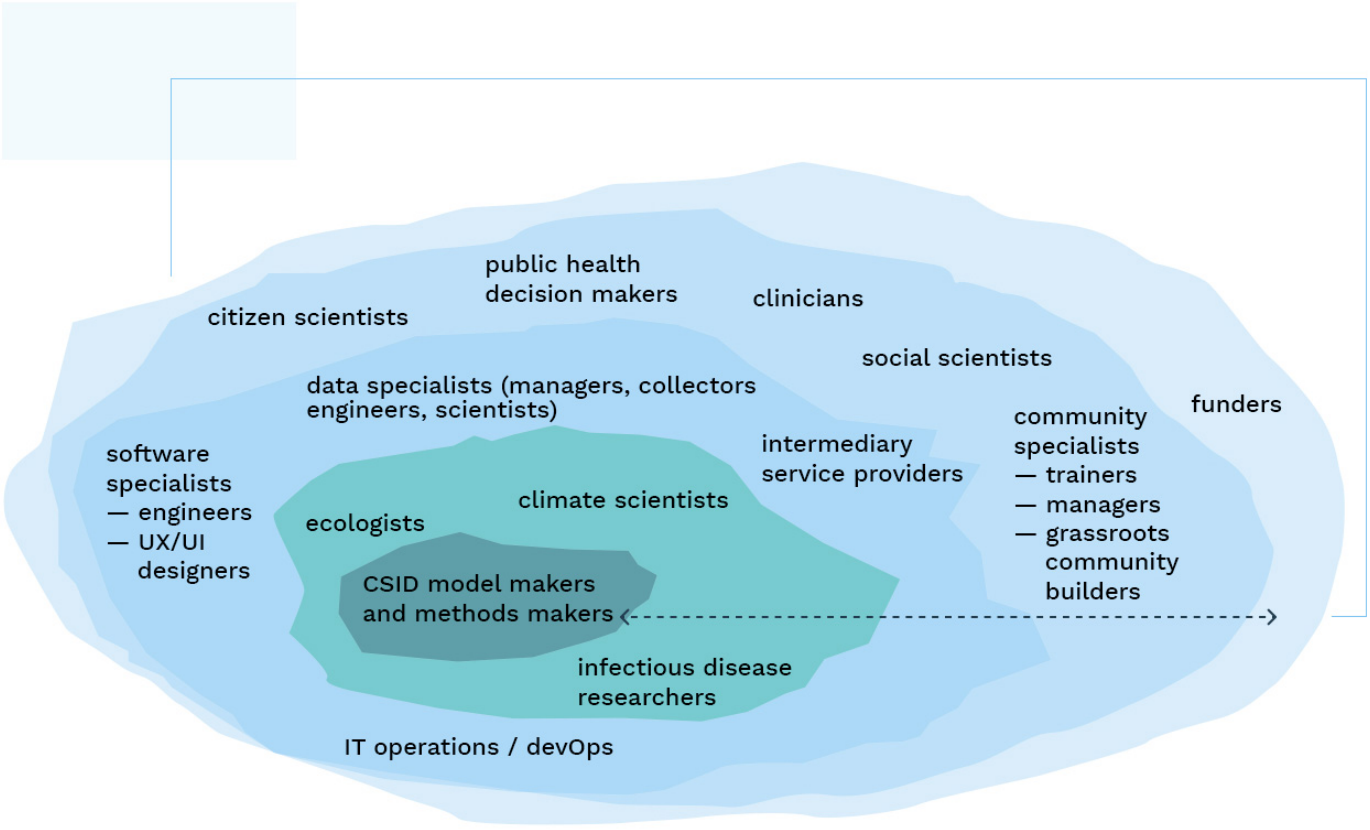


Figure 10. An "oyster" model of community engagement (Fenlon et al. 2022), which illustrates expanding circles of potential allied groups for the CSID CoP. Note the bidirectional arrow signaling shifts and movements between the groups. The "CSID model and methods makers" located in the core are comprised of interdisciplinary teams that work towards CSID model/methods/tool making; they will also have subject matter expertise in some of the fields in the expanding circles.

These CSID actors possess diverse needs and motivations for participation. These differences will result in varying degrees of involvement, and each group will play unique roles and make distinct contributions. Therefore, the engagement strategies and activities must be tailored to meet varying needs.

Community Activities

Some of the areas that a CSID CoP is well placed to advance include:

a) Sharing best practices for meaningful engagement with CSID tool end users. Workshop participant noted that such engagement is critical, not optional, for improved CSID tools, but difficult to do alone (and often outside of the expertise of the tool or model-makers).
b) Bringing existing CSID data, models, software resources together for a global community, and providing a platform to develop new tools that takes advantage of the rapidly evolving technology landscape.
c) Providing thought leadership for the growing field of CSID.
d) Consolidating relevant job, grant, and training opportunities for CSID researchers and tool makers, and nurturing CSID leadership amongst early career researchers in LMICs.

Thus far, an email list has been established for the group as well as a database for continuing to comprehensively identify who key stakeholders and participants of a CSID CoP should be. Further communications infrastructure including a brand kit and public-facing website should be established as part of the next phase of work.

Recommendations from the CSID Co-Design Workshop

As mentioned in the recommended approach above, we expect the community's specific activities to evolve and adjust in response to changing circumstances. The initial activities proposed emerged during the Co-Design Workshop and are the result of engaging discussions with workshop delegates, many of whom are keen to get started. Establishing procedures for regular evaluation and adjustment of activities is crucial to creating a resilient and adaptable CoP, guaranteeing they remain responsive to the fluctuating needs of the CSID community.

NURTURE CSID LEADERSHIP

- Develop a clear, collaborative articulation of what equity means in the context of a CSID community and sketch a roadmap for how equity is woven into all aspects of the community including its processes, activities, trainings, outreach, etc.
- Establish a governance structure for a CSID community that takes equity and intersectional diversity into account.
- Build a mentorship program.
    - While the exact mechanism should be further developed collectively in the next phase, this program will likely include horizontal and vertical mentorship, for example, a student/technical intern-funded exchange program in different organizations of the community. Important not to reproduce problematic dynamics of sending “global South” representatives to “global North” to learn supposed “best practices.”
- Sketch a roadmap towards a community-led fund that supports the growth of the field.
    - The fund might support things such as: funding for server time for scientists in LMICs; development of CSID generalizable tools/methods; rescuing/maintaining of tools and databases; workshops/ training to support local/regional CSID community growth.

ADVANCE CSID RESEARCH

- Build upon the IAI report findings (2021) to further understand the CSID tools/methods that are available and new techniques to develop a CSID software best practices guide.
- Explore the methodological challenges inherent in blending infectious disease and climate modeling towards the development of effective and integrative modeling approaches (i.e., exploration of generalizable methods and tools).
- Develop a more nuanced categorization schema of tools.
- Share effort on big climate data ingestion/wrangling to be used across different disease modeling projects.
- Write co-authored journal publications on various CSID topics.
    - E.g., uncertainties across climate and disease scenarios, across models, and within models, and which of those are the most appropriate sets of uncertainties or samples to feed from one model into another.
    - E.g., challenges of matching both spatial and temporal scales which informs which uncertainties are propagated from one model to another.

SUPPORT ADOPTION AND MAINTENANCE OF CSID TOOLS

- Establish outreach and engagement with end users (e.g., policy makers; data analysts; citizen scientists).
- Cultivate connections with software communities to grow capacity for research software engineering.
- Support projects with structured/certified code review.
- Develop a process for understanding the impact of existing tools and conduct a postmortem of the "dead" software tools to surface lessons learned.
- Build community health analytics for open-source software metrics to monitor CSID projects at scale so the CoP can offer support to projects potentially in trouble.

CULTIVATE CONNECTIONS

- Partner with other relevant CoPs to host community events/trainings/seminars.[13]
- Host annual CSID CoP gathering to facilitate knowledge exchange and further community connection.
- Develop a database of CSID-relevant databases, projects, community members.
- Draw in funding partners. Several funders such as The Rockefeller Foundation, Bill & Melinda Gates Foundation, the International Development Research Centre, and the U.S. National Institutes of Health are supporting relevant CSID software work (see for example, a funding call by the National Institute of Allergy and Infectious Diseases to support Exploratory Data Science Methods and Algorithm Development in Infectious and Immune-mediated Diseases). It will be important moving forward to engage with relevant funding bodies early on to discuss opportunities for partnership.

Allied communities working on similar topics will also be important to engage moving forward such as: GeoHealth Community of Practice; The International Society for Neglected Tropical Diseases; Institute of Malaria and Climate Solutions hosted by Malaria No More, and, among others, CAFÉ: a new Community of Practice funded by the US National Institutes of Health (NIH) to expand work on Climate Change and Health. Engagement with such allies might take the form of co-taught summer school courses, jointly hosted training webinars, or co-located events. RSEs working on CSID software can find programming language-specific support through existing communities such as SciPy, which develops an open-source library for scientific and technical computing in Python, or rOpenSci which provides high-quality R packages that lower barriers to working with local and remote scientific data sources.

Other organizations that can help to support the development of CSID OSS tools include Outreachy, a program that offers stipends to university students and underrepresented groups to work on open-source projects, and The Carpentries, which provides training and workshops to help researchers and scientists acquire foundational computational and data skills.

We have not yet substantially explored what partnering with for-profit initiatives adopting open-source practices in their business models might look like and what role they could play in a CSID CoP, as participants and funders. There may be benefits to engaging with for-profit ventures to help translate the impact of the developed tools from academia and research to field projects. This topic is something to be discussed collectively once the new CoP structure is established.

[13] A significant role of the CSID CoP will be to promote engagement with other fields, geographies, and stakeholders beyond the realm of CSID software and modeling. For instance, the CoP should actively participate in and track discussions in the Open-Source Software community about software maintenance and sustainability. Regular training sessions conducted by groups like Climate Change AI and International Clinics on Infectious Disease Dynamics and Data Program (ICI3D) may also present collaborative opportunities.

Timeline of Next Steps

While many of the details will be revised and iterated on as the community comes into greater formation, we suggest a general timeline for the CSID CoP to establish over the next 24 months.

First 6 months

1. Establish an interim legal structure that will enable the CoP to send/receive funds, hire staff, and other necessary administrative tasks for the community.
2. Host community-wide discussions that tackle different aspects of how to integrate equity considerations into all facets of the group's activities. Identify specific structures, outputs, and new processes to develop based on insights from these discussions.
3. Develop consensus on a clear scope for the CSID CoP with community members (see “Domain of Focus” above for some prompt questions to use) and map the levels of possible community engagement. Identify the anticipated tasks required for the CoP’s work and link these to benefits for different participants. Name the roles and establish a set of rights and responsibilities. Design transparent pathways for people to move in and out of different roles.
4. Hire full-time staff member to support the CoP with establishing and maintaining the necessary social and operational infrastructure.
5. Begin developing the communications and outreach infrastructure including a brand kit, website, social media, and newsletter.
6. Determine location and priorities for 2024 annual gathering and establish event planning committee.

6 – 12 months

7. Establish working structures and groups as informed by the community discussions on equity. Possible working groups may include governance; training and capacity; fundraising; outreach; and CSID tools and models.
8. Host first working group meetings and establish cadence and 6-month work plan for each group.
9. Host 2024 annual gathering.
10. Launch new website and public announcement of the CSID CoP.

13 – 24 months

11. Reflect on the utility of the working group model and, if appropriate, develop next work plan for each group.
12. Continue to outreach with relevant CSID CoPs, funding partners, end-user stakeholders.
13. Determine location and priorities for 2025 annual gathering and establish event planning committee 6 months prior to the gathering.
14. Host reflective community discussions on what has been learned thus far as a CoP and iterate on the community roles and structures. Determine next steps towards desired long-term governance structure.
15. Public launch of CSID CoP working group outputs (ideas for outputs that have been suggested include a database of relevant CSID databases; database of CSID actors; climate data ingestion tool; co-authored journal publication on challenges on the emerging CSID field, or papers on the creation of generalizable CSID methodologies).
16. Host 2025 annual gathering.

Beyond 24 months

16. Transition from interim legal structure to the community-determined longer-term governance structure.
17. Reflect on learnings from the first two years to articulate longer-term vision of the CSID CoP.
18. Secure additional funding support to sustain the CSID CoP.

Conclusion

Growing a sustainable CSID CoP will require significant effort, resourcing, and coordination. Today's CSID actors seek to be part of something that will chart new ways of fostering collaborations between scientists that avoid reproducing the usual power and knowledge hierarchies. CS&S sees promise in the strength of collective action and unique alignment in the CSID community today. There is significant potential to seize this moment and change scientific and software collaboration for the better, while advancing the study of climate sensitive infectious disease—one of humanity's most pressing and existential challenges. The right coordination and lightly structured approach to community-building can catalyze a ripple effect of new relationships, tools, opportunities, and learnings across diverse groups, as we already have begun to observe from early efforts at convening this community.

This convening work holds important promise to address the heavy silos that prevent transformative scientific work and restrain widespread impact of scientific tools and software. Our climate is changing, and with this change, the patterns of infectious disease are shifting. Changing patterns of climate-sensitive diseases will impact us all and will be felt most harshly by vulnerable and marginalized populations. A diverse CoP is therefore needed to address challenges in CSID and to focus and mitigate impacts on vulnerable populations. This landscaping work has ignited a community ready to meet this great challenge. By fostering mutual learning, we can expand and accelerate the research needed to understand this change and mitigate risks to humanity.

ACKNOWLEDGMENTS

We extend our gratitude to all involved in this research. Thank you to the external members of the research team who provided steadfast support throughout the project, even when the scope was broad and tasks occasionally ambiguous. Big appreciation to the various members of the CS&S team for not only learning about a new field, but also conducting research and organizing a comprehensive workshop within a tight six-month timeframe.

Our gratitude extends to the Wellcome team for their consistent feedback and support during bi-weekly check-ins. Lastly, we want to acknowledge all the interlocutors and especially the CSID potential allied organizations who were willing to engage in discussions and begin strategizing collaborative partnerships from the outset. The dedication and collective efforts of all parties made this project an exciting reality.

Bibliography

Ación, Laura, Gonzalo Peña-Castellanos, and Fernando Pérez. 2022. “Open Source and Open Science in Latin America,” September. https://doi.org/10.5281/ZENODO.8092090.

Bezuidenhout, Louise, and Ereck Chakauya. 2018. “Hidden Concerns of Sharing Research Data by Low/Middle-Income Country Scientists.” *Global Bioethics = Problemi Di Bioetica* 29 (1): 39–54.

Chan, Leslie, Angela Okune, Rebecca Hillyer, Denisse Albornoz, and Alejandro Posada, eds. 2019. *Contextualizing Openness: Situating Open Science*. Ottawa: University of Ottawa Press. https://www.idrc.ca/en/book/contextualizing-openness-situating-open-science.

Chasi, Samia, and Savo Heleta. 2022. “Towards More Sustainable, Equitable and Just Internationalisation Practices: The Case of Internationalisation Conferences.” *Journal of Studies in International Education* 0 (0): 10283153221139924. https://doi.org/10.1177/10283153221139924.

Colón-González, Felipe J, Maquins Odhiambo Sewe, Adrian M Tompkins, Henrik Sjödin, Alejandro Casallas, Joacim Rocklöv, Cyril Caminade, and Rachel Lowe. 2021. “Projecting the Risk of Mosquito-Borne Diseases in a Warmer and More Populated World: A Multi-Model, Multi-Scenario Intercomparison Modelling Study.” *The Lancet Planetary Health* 5 (7): e404–14. https://doi.org/10.1016/S2542-5196(21)00132-7.

Digital Science, Natasha Simons, Greg Goodey, Megan Hardeman, Connie Clare, Sara Gonzales, Damon Strange, and others. 2021. “The State of Open Data 2021.” 2021. https://doi.org/10.6084/m9.figshare.17061347.v1.

Dunbar-Hester, Christina. 2020. *Hacking Diversity: The Politics of Inclusion in Open Technology Cultures*. Princeton Studies in Culture and Technology. Princeton, New Jersey: Princeton University Press.

Edmonton Regional Learning Consortium. 2016. “What Is a Community of Practice?” *Community of Practice* (blog). 2016. https://www.communityofpractice.ca/background/what-is-a-community-of-practice/.

Eghbal, Nadia. 2016. “Roads and Bridges: The Unseen Labor Behind Our Digital Infrastructure.” Ford Foundation. https://www.fordfoundation.org/work/learning/research-reports/roads-and-bridges-the-unseen-labor-behind-our-digital-infrastructure/.

———. 2020. *Working in Public: The Making and Maintenance of Open Source Software*. First Edition. San Francisco: Stripe Press.

Georgeson, Lucien, Mark Maslin, and Martyn Poessinouw. 2017. “Global Disparity in the Supply of Commercial Weather and Climate Information Services.” *Science Advances* 3 (5). https://doi.org/10.1126/sciadv.1602632.

Ghani, Marium, Rowan Hurrell, Avelino C. Verceles, Michael T. McCurdy, and Alfred Papali. 2021. “Geographic, Subject, and Authorship Trends among LMIC-Based Scientific Publications in High-Impact Global Health and General Medicine Journals: A 30-Month Bibliometric Analysis.” *Journal of Epidemiology and Global Health* 11 (1): 92–97.

Goble, Carole. 2022. “Research Software Sustainability Takes a Village,” November. https://doi.org/10.5281/ZENODO.7304596.

IAI, Inter-American Institute for Global Change Research: 2022. “Landscape Mapping of Software Tools for Climate-Sensitive Infectious Disease Modelling.” https://wellcome.org/reports/landscape-mapping-software-tools-climate-sensitive-infectious-disease-modelling.

Kaspar, Frank, Axel Andersson, Markus Ziese, and Rainer Hollmann. 2022. “Contributions to the Improvement of Climate Data Availability and Quality for Sub-Saharan Africa.” *Frontiers in Climate* 3. https://doi.org/10.3389/fclim.2021.815043.

Kaur, Rajdeep, Kuljit Kaur Chahal, and Munish Saini. 2022. “Understanding Community Participation and Engagement in Open Source Software Projects: A Systematic Mapping Study.” *Journal of King Saud*

University - Computer and Information Sciences 34 (7): 4607–25. https://doi.org/10.1016/j.jksuci.2020.10.020.
Khan, Mishal, Seye Abimbola, Tammam Aloudat, Emanuele Capobianco, Sarah Hawkes, and Afifah Rahman-Shepherd. 2021. “Decolonising Global Health in 2021: A Roadmap to Move from Rhetoric to Reform.” *BMJ Global Health* 6 (3). https://doi.org/10.1136/bmjgh-2021-005604.
Klinenberg, Eric. 2018. *Palaces for the People: How Social Infrastructure Can Help Fight Inequality, Polarization, and the Decline of Civic Life*. First Edition. New York: Crown.
Littauer, Richard, Benjamin Nickolls, John Hill, and Martin Wright. 2021. “Sustain 2021.” 2021. https://drive.google.com/drive/folders/1aw6keDVHOEaSuGFRHcIxNPhJfCwzP4I_.
Manteaw, Bob O., Antwi-Bosiako Amoah, Bernard Ayittah, and Kirk B. Enu. 2022. “Climate-Informed Decision-Making in Data-Poor Environments: Managing Climate Risk Through Citizen Science Networks.” *Frontiers in Climate* 4.
Metcalf, C. Jessica E., Katharine S. Walter, Amy Wesolowski, Caroline O. Buckee, Elena Shevliakova, Andrew J. Tatem, William R. Boos, Daniel M. Weinberger, and Virginia E. Pitzer. 2017. “Identifying Climate Drivers of Infectious Disease Dynamics: Recent Advances and Challenges Ahead.” *Proceedings of the Royal Society B: Biological Sciences* 284 (1860): 20170901. https://doi.org/10.1098/rspb.2017.0901.
National Science Foundation. 2022. “Frequently Asked Questions (FAQs) for the NSF Pathways to Enable Open-Source Ecosystems (POSE) Program (NSF 22-572).” https://www.nsf.gov/pubs/2022/nsf22062/nsf22062.jsp.
“Nature Addresses Helicopter Research and Ethics Dumping.” 2022. *Nature* 606 (7912): 7–7. https://doi.org/10.1038/d41586-022-01423-6.
Ram, Karthik. 2023a. “How CoPs Are Formed and Sustained around Open Source Ecosystems.” https://doi.org/10.5281/zenodo.8165092.
———. 2023b. “Key Considerations to Take into Account for Governance Models of an Open Project given Size, Age, Maturity Level, Resources, and Total Addressable Community.” Zenodo. https://doi.org/10.5281/zenodo.8165102.
Research, Nature, and Greg Goodey. 2021. “State of Open Data Survey 2021 Additional Resources.” figshare. https://doi.org/10.6084/M9.FIGSHARE.17081231.
Ritvo, Dalia, Kira Hessekiel, and Christopher Bavitz. 2017. “Challenges & Opportunities Concerning Corporate Formation, Nonprofit Status, & Governance for Open Source Projects.” https://doi.org/10.2139/ssrn.2937956.
Ryan, Sadie J, Catherine A Lippi, Talia Caplan, Avriel Diaz, Willy Dunbar, Shruti Grover, Simon Johnson, et al. 2023. “The Current Landscape of Software Tools for the Climate-Sensitive Infectious Disease Modelling Community.” *The Lancet Planetary Health* 7 (6): e527–36. https://doi.org/10.1016/S2542-5196(23)00056-6.
Savini, Lara, Susanna Tora, Alessio Di Lorenzo, Daniela Cioci, Federica Monaco, Andrea Polci, Massimiliano Orsini, Paolo Calistri, and Annamaria Conte. 2018. “A Web Geographic Information System to Share Data and Explorative Analysis Tools: The Application to West Nile Disease in the Mediterranean Basin.” Edited by Ulrich Melcher. *PLOS ONE* 13 (6): e0196429. https://doi.org/10.1371/journal.pone.0196429.
Serwadda, David, Paul Ndebele, M. Kate Grabowski, Francis Bajunirwe, and Rhoda K. Wanyenze. 2018. “Open Data Sharing and the Global South—Who Benefits?” *Science* 359 (6376): 642–43. https://doi.org/10.1126/science.aap8395.
Sethi, Akanksha. 2017. “Learning in Communities: Understanding Communities of Practice in the Development Sector. Communities of Practice in Development: A Relic of the Past or Sign of the Future?” *Knowledge Management for Development Journal* 13 (3): 4–21.
Wenger, Etienne. 2000. “Communities of Practice and Social Learning Systems.” *Organization* 7 (2): 225–46.
Wenger, Etienne, Richard A. McDermott, and William Snyder. 2002. *Cultivating Communities of Practice: A Guide to Managing Knowledge*. Boston, Mass: Harvard Business School Press.
Woods, Christopher, dir. 2022. *How You Could Possibly Fund Maintenance for Your Software*. https://www.youtube.com/watch?v=JuCBQH7pcxU.

Glossary

COP

This abbreviation stands for Community of Practice. A Community of Practice often looks like a group of people who share a common interest, profession, or passion and actively engage in sharing ideas and experiences to learn from each other and improve their skills. A CoP is different from a network in the sense that it focuses on a substantive topic; it is not just a set of relationships. A CoP is distinct from a work team in that the shared learning and interest of its members keep it together. It is defined by knowledge rather than by an individual task and exists because participation has value to its members. Finally, a CoP is differentiated from other communities since its members are more likely to share a common profession or work situation and therefore share interest in developing common practices.

CSID

This abbreviation stands for Climate Sensitive Infectious Disease, which are infectious diseases whose transmission and spread are influenced by changes and variations in climate and weather. Factors such as temperature, precipitation, humidity, and extreme weather events can impact the lifecycle of pathogens or vectors, such as mosquitoes or ticks, that transmit these diseases. While mosquito-/vector-borne diseases are commonly considered CSID, other diseases such as respiratory pathogens and water-borne diseases can also be considered CSID. The study and modeling of these diseases often require an interdisciplinary approach combining epidemiology, climatology, and other fields.

INFECTIOUS DISEASE MODELING

A scientific technique used to understand the spread of infectious diseases in populations. It involves the use of mathematical models and statistical analysis to predict disease dynamics and inform public health interventions.

LOW- AND MIDDLE-INCOME COUNTRIES (LMICS)

LMIC is a term commonly used by international organizations and development agencies to categorize countries based on their income levels and economic development. The World Bank classifies countries based on their Gross National Income (GNI) per capita to determine their income group; LMIC countries have relatively lower average income levels compared to “High-Income Countries” but are higher than “Low-Income Countries.”

OPEN DATA

Open data refers to data that is made freely available and accessible to the public without restrictions on its use, reuse, or redistribution. The concept of open data is rooted in the principles of transparency, accountability, and sharing of information for the benefit of society as a whole. When data is considered “open,” it means that anyone can access, download, modify, and share it, often without the need for special permissions or licenses.

OPEN SCIENCE

Open science refers to the practice of making scientific research, data, methodologies, and findings openly accessible to the public. It promotes transparency, collaboration, and the sharing of knowledge among researchers, institutions, and the wider community.

OPEN-SOURCE SOFTWARE (OSS)

Software that is released with a license allowing anyone to view, use, modify, and distribute the software's source code. This encourages collaboration and community-driven development.

SCIENTIFIC SOFTWARE

Any software application or tool that is designed to aid scientific analysis, data processing, simulation, or modeling. This type of software is used by scientists and researchers across various scientific domains, such as physics, chemistry, biology, the social sciences, astronomy, and engineering. The primary focus of scientific software is to provide solutions to scientific problems and facilitate data-driven research. Scientific software is a specific subset of research software that is explicitly focused on scientific analysis, simulation, or modeling, while research software encompasses a broader range of tools and applications used in various research disciplines.

TOOL

A software tool is a computer program or application designed to perform specific tasks or functions to assist users in accomplishing their goals. These tools are created to simplify, automate, or enhance various processes and activities related to computing, data manipulation, or problem-solving. In the context of CSID modeling, some of the existing tools include early warning systems like D-MOSS which aims to give beneficiaries several months advance warning of likely outbreaks of dengue fever; and disease surveillance and monitoring tools like the Disease Monitoring Dashboard, which is a web-based application in which data on West Nile Virus occurrence in Europe and the Mediterranean basin are collected from multiple sources and displayed on interactive maps (Savini et al. 2018).

Code for Science & Society envisions a future where open research, data, and technology effectively work towards equitable distribution of resources, knowledge, and power. We create impact towards our vision through practical & strategic support, funding & resources, and research & programming.

**Code for Science & Society, 3439 SE Hawthorne Blvd.
Portland, Oregon, 97214
USA
codeforsociety.org**

Wellcome supports science to solve the urgent health challenges facing everyone. We support discovery research into life, health and wellbeing, and we're taking on three worldwide health challenges: mental health, global heating and infectious diseases.

**Wellcome Trust, 215 Euston Road, London NW1 2BE, United Kingdom
T +44 (0)20 7611 8888, E contact@wellcome.org, wellcome.org**

Appendices

Appendix 1: Contributor Roles[1]

	Contributor Role 1	Contributor Role 2	Contributor Role 3	Contributor Role 4	Contributor Role 5	Contributor Role 6
Laura Ación	Investigation	Research interlocutor	Writing – review & editing			
Kristen Aiemjoy	Research interlocutor	Writing – review & editing				
Rebecca Asher	Conceptualization	Resources	Supervision			
Whitney Bagge	Research interlocutor					
Michelle Barker	Conceptualization	Investigation	Writing – original draft	Writing – review & editing		
Michael Barton	Research interlocutor					
Anna Carnegie	Research interlocutor					
Flávio Codeço Coelho	Research interlocutor					
Hardip Dhaliwal	Project administration					
William De Souza	Research interlocutor					
Rayya El Zein	Conceptualization	Funding acquisition	Writing – review & editing			
Kamil Erguler	Research interlocutor					
Michelle Evans	Research interlocutor					
Isabel Fletcher	Conceptualization	Resources	Writing – review & editing	Supervision		
Marios-Eleftherios Fokaefs	Research interlocutor					
Nick Golding	Research interlocutor					
Joe Hand	Project administration					
Manzoor Ahmed Hanifi	Research interlocutor					
Emily Jesper-Mir	Conceptualization	Resources	Supervision			

[1] Categorization schema from https://credit.niso.org/ with the addition of “research interlocutor” category. Interlocutors were asked as part of formal interviews if they would like to be named or not in the attribution of work. Many additional interlocutors including workshop participants were also integral to this research but have not been explicitly named.

Rachel Lowe	Research interlocutor					
Tim Lucas	Research interlocutor					
Aigerim Massabayeva	Data curation					
Huston Malande	Visualization					
Emma Mendelsohn	Research interlocutor					
Kerrie Mengersen	Research interlocutor					
Page Metcalf	Project administration					
Nhia Moua	Project administration					
Miliaku Nwabueze	Conceptualization	Project administration	Software	Writing – review & editing		
Angela Okune	Conceptualization	Methodology	Project administration	Funding acquisition	Writing – original draft	Writing – review & editing
Karthik Ram	Investigation	Writing – original draft	Writing – review & editing			
Ryan Rising	Project administration	Data curation				
Danielle Robinson	Conceptualization	Funding acquisition	Supervision	Writing – review & editing		
Noam Ross	Research interlocutor					
Sheetal Silal	Research interlocutor					
Jaspreet Turner	Resources					
Gemma Turon	Investigation	Writing – original draft	Writing – review & editing			
Emilie Vallee	Research interlocutor					
Váleri Vásquez	Research interlocutor	Writing – review & editing				
Daniel Villela	Research interlocutor					
Susan Winks	Investigation	Writing – original draft	Writing – review & editing			

Appendix 2: Project Collaborators

Project Collaborator (alphabetical)	Affiliated Organizations	Areas of Expertise	Input to the Project
Dr. Laura Ación	MetaDocencia, CONICET-Universidad de Buenos Aires	health data science, health artificial intelligence, responsible use of data, community building, capacity building	Conceptualized and organized Focus Group Discussion in Spanish; reviewed final report
Dr. Kristen Aiemjoy	UC Davis, USA Mahidol University, Thailand	Infectious disease epidemiology, surveillance, seroepidemiology, R open-source tools	Offered insight on the field of infectious disease surveillance, neglected tropical disease, open-source software tools; reviewed final report
Dr. Michelle Barker	Research Software Alliance (ReSA)	research software, research infrastructure, open science, collaboration, syste m change	Conducted 10 of the research interviews; secondary research; developed reports of research learnings; offered insight on research software engineering communities; reviewed final report
Dr. Karthik Ram	rOpenSci, US Research Software Sustainability Institute	Open-source software, sustainability, reproducible research, open science, global change biology	Offered insight on open-source software communities of practice; conducted secondary research and documentation of OSS learnings; reviewed final report
Ryan Rising	EcoSocial Action	ecological design, facilitation, governance, community organizing, ecosocial direct action	Facilitated in-person workshop in Cape Town
Dr. Gemma Turon	Ersilia	infectious diseases, artificial intelligence, machine learning, python, open source	Offered insight on open science and Africa-based CoPs; conducted secondary research and documentation of open data learnings; reviewed final report
Dr. Váleri Vásquez	UC Berkeley / Stanford	environmental health, ecology, policy, optimization, simulation, open-source scientific software, genetic-based intervention technologies	Offered insight on field of CSID; reviewed final report
Dr. Susan Winks	H3D Foundation	infectious disease, drug discovery, scientific governance, pan-African CoPs, end-user training	Offered insight on open science and Africa-based CoPs; conducted secondary research and documentation of open data learnings; reviewed final report

Appendix 3: Methodological Details

Sampling and Interview Approach

The team began by reviewing of existing initiatives, groups, events, publications, and tools related to climate science, infectious disease, and/or open-source/scientific software communities. The following criteria was used to identify and select appropriate stakeholders for conducting research interviews and discussions:

- For inclusion in a CSID stakeholder list:
    - Is this an individual, group, project, event, publication venue, tool that is related to climate science, public health, and/or infectious disease?
    - Is this an existing community of open-source software/scientific software?

From the stakeholder list, an interview list was curated that represented a diverse cross-section of individuals based on gender, career stage, country, disease focus, expertise, and institution type. As we neared theoretical saturation,[2] we prioritized leaders of existing CoPs rather than individual researchers.

The initial interview participants were largely composed of recent Wellcome-funded grantee teams after which snowball sampling[3] was used to reach new communities of practice that might be relevant. The professional networks of the CS&S research team members were leveraged to reach relevant individuals and groups outside of the Wellcome Trust network that are working on climate science, infectious disease research and modeling, and open-source and scientific software. While it was outside the scope of work for this phase of the project, future research during the next phase should include interviews and discussions with CSID policy makers as software end users, as well as funders interested in supporting CSID software.

Data Protection

Compliance with UK GDPR requirements was confirmed in collaboration with the Wellcome Trust legal and data protection teams. Prior to interviewing, the project purpose and scope of involvement were explained to participants. We obtained written informed consent from all focus group discussions and interview participants.

Data Analysis

The research team generated field notes based on observations during focus group discussions and interviews. Following the interviews and discussions in which at least two members of the research team were present, a discussion of emerging insights and notable findings was held between the research team. Insights gained following interviews that were conducted by one member of the team were shared in a collaborative document of emerging insights.

---

[2] Theoretical saturation is often used in qualitative research as a criterion for judging when to stop sampling (Barney G Glaser and Anselm L Strauss, *The Discovery of Grounded Theory Strategies for Qualitative Research* (Chicago: Aldine Pub., 1973), http://search.ebscohost.com/login.aspx?direct=true&scope=site&db=nlebk&db=nlabk&AN=580315.)

[3] B. L. Berg, *Qualitative Research Methods for the Social Sciences*, 7th ed. (Boston, MA: Pearson Education, 2009).

In addition to the qualitative methods named above, two additional approaches were leveraged to gain further insight on the needs of a new CSID CoP:

- First, a member of the research team, Karthik Ram, ran a quantitative analysis of the usage of CSID open source/scientific modeling software by popularity and disease coverage which provided a complementary insight into the collections of software used to develop CSID models.[4]
- Second, we used a Participatory Action Research framework to design and convene an in-person workshop in Cape Town, South Africa, to better understand the needs of the emergent community and to begin to establish a strong, collaborative foundation for the CoP. The conceptual framework and decisions behind the workshop invitee list and agenda (find the agenda in Appendix 5) were informed by the IAI report and insights from interviews. The invitee list for the workshop was iteratively developed and confirmed concurrently with the interviews and consultative meetings. CS&S sought to invite not only those who already saw themselves as doing CSID work, but also those who could bring lessons and learnings from other relevant parallel spaces. We leveraged Wellcome and CS&S networks to invite infectious disease modelers, climate scientists, research software engineers, data scientists, founders and leads of other communities of practice, and potential end users of CSID software.

[4] Karthik Ram, “A Quantitative Analysis of the Usage of CSID Open Source/Scientific Modeling Software by Popularity, Disease Coverage, and Geography,” July 2023, https://doi.org/10.5281/zenodo.8165083.

Appendix 4: Research Instruments Used

1:1 Interview Guide

This interview guide was preceded by the sharing of an Interview Consent Form, which was emailed to participants in advance for review and signing. The consent form included details about the data protection and confidentiality protocols that would be followed for this research including that the discussion would be recorded and transcribed for data analysis purposes and that the video recording and transcription would not be published or shared.

Introductory Remarks

Background

1. Can you tell me a bit more about your organization(s)?
2. What is your role and involvement with the climate science infectious disease modeling community? Which infectious disease/s do you focus on?

[IF THEY LEAD A CSID COP]

3. Who constitutes your community?
4. How does your community come together (i.e., What community building mechanisms do you use?)
5. What is the problem your group is addressing?
6. What mechanisms do you use for the sharing of data, software, and resources?
7. Who do you see as missing in your collective? What sectors are missing? What countries are missing?
8. How do you encourage collaboration? What prevents collaboration?

On relevant CSID work

9. Which other projects, organizations, communities, initiatives, etc., do you engage with in this field?
10. What is the brief history behind your project? (i.e., why are they doing this work? What are their motivating factors?)

Community Engagements

11. What are key groups (such as networks, scholarly societies, not-for-profits) do you work with as part of your everyday work?
12. Which policy makers and funders are involved in these efforts and how does their influence affect the community?
13. How do you/your team connect with your end users (policy makers/public health officials/etc.)?
14. Which journals, conferences/workshops do you find most relevant for you? Why?

On challenges and opportunities

15. What are the major issues that you face in your everyday work? How do you want to better engage with others (who/in what ways)?
16. What gaps do you see in the CSID field? Who is missing from discussions? What do you wish there was more of?

17. What are major challenges faced by the CSID communities that you are part of (such as resourcing, governance, infrastructure, sustainability)?

Futures

18. What future work do you see yourself/your group engaging in? What kinds of partnerships are you looking for?

19. Is there anything we didn't get a chance to talk about that you would want to mention?

20. Do you have other people or groups that you think I should talk to about CSID?

English Focus Group Discussion Guide

Introductory Remarks

1. Introductions

a. Please introduce yourself, your organization, and your role
b. Which disease/s, geographic areas, disciplines, etc., does your work focus on?

2. Existing Communities

We're interested in understanding the types of communities and mechanisms that facilitate stakeholder engagement for the climate science infectious disease modeling community. Think about the communities that you're involved with. These could be formal or informal, and include organizations, teams, working groups, events, LinkedIn groups, etc.

a. Which communities do you find most useful for meeting other researchers and practitioners? Can you tell us more about this/these communities?

1. *Probes:*
    1. *Where do they meet? How frequently? Why do they work? What kinds of CSID folks are there? Who is missing?*
    2. *What are some of the missed opportunities of this community? How would you like to see it improved?*

b. Of those communities that you think have been most active with researchers, why do you think they have been able to get such strong engagement?

c. By a show of hands, how many of you have been part of communities that have:
    1. shared data
    2. collaborated or provided input on analysis
    3. co-developed research software
    4. offered training or mentoring
    5. offered funding or help in accessing funding

d. Of these possibilities, which ones might you be most interested in seeing develop as a new community?

e. What is the composition of each of these communities, e.g., local team, institutional, regional, national, international; focused on a discipline, programming language, function

(such as software development, malaria, open modeling, training), etc.?

f. What challenges do you face in your work that you would benefit from engagement with others to solve? What sort of activities could be helpful?

Closing remarks

If time permits, then these questions will also be discussed:
a. What are the challenges faced by these communities and mechanisms? E.g., how effective are these at enabling stakeholders' interaction, are there issues related to resourcing, governance, infrastructure, sustainability, etc.?

Spanish Focus Group Discussion Guide

This discussion was run entirely in Spanish. CS&S used a different guide to tailor the conversation for a different audience to better understand regional context.

1. Presentaciones
a. Para comenzar, les voy a pedir que se presenten con su nombre, la organización a la que pertenecen y el rol que ocupan.
b. ¿En qué temas se enfoca su trabajo?

2. Comunidades existentes
a. Comunidades:

- ¿Qué comunidades (formales o informales) encontrás **más útiles** para reunirte con otros profesionales? ¿Podés contarnos más de estas comunidades?

- Les invito a pensar en aquellas comunidades que son **más activas** con las personas que investigan, ¿por qué creen que consiguen un compromiso activo y fuerte con su comunidad?
    - Otras:
        - ¿Qué hacen para fortalecer y mantener ese compromiso, esa actividad e intercambio constante?
        - ¿Qué incentivos/propuestas/beneficios ofrecen? Ejemplo?

b. Grupos de investigación. Si pensamos en los grupos de investigación de los que fuiste o sos parte, ¿cómo lograron comprometer y trabajar con otros usuarios y comunidades? ¿Qué funcionó y qué no?

3. Importancia del contexto
¿Qué crees que es único, distintivo, especial del trabajo que hacés desde/en contextos latinoamericanos?

4. Conexiones y redes deseables en el futuro

- HOY: ¿Cómo son las conexiones y redes que tenés hoy? ¿Tienen escala local, alcance regional o internacional? ¿Cómo las podrías definir?
- FUTURO: ¿Qué tipos de conexiones o redes te gustaría construir? ¿Por qué motivos?

- BARRERAS:

    - Según tu opinión, ¿qué dificulta crear esas redes? ¿Por qué? ¿Motivos?
    - Pensando en las particularidades de Argentina o (si se animan y conocen) de nuestra Región, ¿qué otras barreras identifican para crear conexiones? Motivos: lenguaje, tiempo, falta de conexiones, recursos, financiamiento.
- **Ejercicio.** Les voy a pedir que piensen en los desafíos que enfrentan en su trabajo. Desafíos que pueden ser de todo tipo. Y les voy a dar 2 minutos para que los anoten en un papel o en drive si son menos analógicos que yo. ¿Listo? Ahora, les voy a pedir que vuelvan a leerlos y resalten aquellos (y sólo aquellos) desafíos que podrían resolverse trabajando con otros, trazando nuevas alianzas.
    - Minuto para compartir desafíos que podrían resolverse trabajando con otros.
    - ¿A través de qué acciones y estrategias se les ocurre que podrían resolverse?
- **Experiencia**. ¿Conocen alguna comunidad o iniciativa exitosa que lo haya logrado escalar y crecer a partir de nuevas conexiones o alianzas? ¿De qué se trata? ¿Por qué la consideran exitosa?

5. Datos y uso compartido

Por último, vamos a abordar un tema que es de interés y se espera que la nueva comunidad de práctica que CSS busca apoyar trabaje: **el intercambio de datos y la infraestructura**

- Según tu opinión, ¿cuál es el mayor desafío o preocupación existente a la hora de compartir datos en Argentina? (Si se sienten cómodos/as: ¿Y en LATAM?
    - ¿Trabajás con data software: querés agregar algún desafío o preocupación a los identificados?
- ¿Cuáles son las formas y razones existentes para compartir datos, software y recursos entre sectores en Argentina/América Latina?

- Antes de terminar, les ofrecemos un **espacio de reflexión** para abordar algún tema extra que quieras compartirnos o incluso la posibilidad de ampliar alguna de tus respuestas, realizarnos alguna sugerencia o comentario

Para cerrar. Muchas gracias por participar y sus aportes. La conversación fue muy productiva y nutrirá directamente a nuestra comunidad. Vamos a compartir el informe final de aprendizajes y recomendaciones cuando se publique en agosto, así como los próximos pasos para ver cómo seguimos en los meses siguientes. ¡Gracias!

(English translation of the Spanish FGD)

1. Introductions

a. Let's go around and please introduce yourself, your organization, and your role. What topics/subjects does your work focus on?

2. Existing Communities

a. Which communities (formal or informal) do you find most useful for meeting other researchers and practitioners? Can you tell us more about this/these communities?

b. Of those communities that you think have been most active with researchers, why do you think they have been able to get such strong engagement?

b. How have existing research groups you have been part of engaged or worked with end-users and other community groups? What worked and what didn't?

3. Importance of Place
What do you think is unique about doing the work you do from/in Latin American contexts?

3. **Desired Future Connections**
a. Are your connections and networks more local, regional collaborations or international connections/collaborations? What kinds of new connections would you like to make and why? What do you think is a barrier to creating those new connections?

b. What challenges do you face in your work that you would benefit from engagement with others to solve? What sort of activities could be helpful?

6. Data and Sharing
We expect that the topic of data sharing and infrastructure will be a big part of the new community of practice that CS&S is looking to support. What are existing ways and reasons for sharing data, software and resources across sectors in LatAm? What are some of the biggest challenges and concerns when it comes to sharing of data in LatAm?

Closing remarks

Appendix 5: Relevant CSID Conferences and Journals

Conferences

Some of the relevant annual conferences identified for the CSID community include:

- American Society of Tropical Medicine and Hygiene (ASTMH) annual meeting
- Joint International Tropical Medicine Meeting (JITMM)
- Prince Mahidol Award Conference (PMAC)
- British Ecological Society and Ecological Society of Australia annual conferences
- CASCON by IBM-Canada Advanced Studies
- Elsevier's Impact of Environmental Changes on Infectious Diseases (IECDI) in 2015 and 2017 (and canceled in 2020)
- Epidemics: International Conference on Infectious Disease Dynamics
- IEEE/ACM International Conference on Software Engineering (ICSE)
- National research software engineering association events, such as that of de-RSE in Germany

Journals

Another important field- and scientific community-building mechanism can be a shared publishing venue. A sample of relevant journals for the CSID community include:

- BMC Infectious Diseases
- Emerging Infectious Diseases
- Ecological Modelling
- Environmental Modelling and Software
- Epidemics
- Frontiers
- International Journal of Epidemiology
- Journal of Open Source Software
- Journal of Royal Society Interface
- Lancet Planetary Health
- Lancet Microbe
- American Journal of Tropical Medicine and Hygiene
- New Zealand Veterinary Journal (for One Health)
- PLOS Global Public Health
- PLOS Neglected Tropical Diseases
- Proceedings of the National Academy of Sciences

Appendix 6: Relevant Tools, Projects & Organizations

As CS&S has conducted the landscaping, in addition to the various tools already surveyed by IAI, we have encountered various tools, projects, and organizations that may be relevant for the CSID software community. Find an initial list below which we plan to continue to grow and maintain as a public database during the next phase of work.

CSID Relevant Projects and Organizations	URL
Tech & Data Projects + Communities	
EpiModel	https://www.epimodel.org/
Framework for Reconstructing Epidemiological Dynamics (FRED)	https://fred.publichealth.pitt.edu/
United States National Center for Atmospheric Research (NCAR) Climate Inspector	https://ral.ucar.edu/pressroom/features/new-climate-change-tool-for-gis-climate-inspector
Climate-Sensitive Infectious Diseases Modeling (CSIDM) Toolbox	
Eclipse Spatio-Temporal Epidemiological Modeler (STEM)	https://projects.eclipse.org/projects/technology.stem
Malaria Atlas	https://malariaatlas.org/people/
REPICON	https://www.repidemicsconsortium.org/
Vector Byte	https://www.vectorbyte.org/
Vector Atlas	http://vectoratlas.icipe.org/
EPICast Now	https://www.epinowcast.org/
EpiVERSE Trace	https://data.org/initiatives/epiverse/
Centro Interdisciplinario de Estudios en Ciencia, Tecnología e Innovación (CIECTI) - “Gestión epidemiológica basada en inteligencia artificial y ciencia de datos” (ARPHAI)	http://www.ciecti.org.ar/arphai/
Ecological Forecasting Initiative	https://ecoforecast.org/about/
ROpenSci	https://ropensci.org/
Climate Change AI (CCAI)	https://www.climatechange.ai/about
Software Sustainability Institute	https://www.software.ac.uk/
Global Alliance for Genomics and Health (GA4GH)	https://www.ga4gh.org/
Open Microscopy Environment	https://www.openmicroscopy.org/
Foundation for Public Code	https://publiccode.net/
Open Source Community Africa	https://oscafrica.org/
OpenJS	https://openjsf.org/
FOSSAsia	https://fossasia.org/
The Apache Software Foundation	https://www.apache.org/
Planetary Health Alliance	https://www.planetaryhealthalliance.org/

R Ladies Global	https://rladies.org
H3ABioNet	www.h3abionet.org
Sisonke Biotik	https://www.sisonkebiotik.africa/
Bioinformatics Hub of Kenya Initiative	https://bhki.org/
Netlify	https://liibre.netlify.app/
Open Life Science	https://openlifesci.org/ols-7/projects-participants/
Rio Abierto	https://pulsante.org/wp-content/uploads/2020/07/Pulsante_rio_abierto.pdf
RSE Asia	https://rse-asia.github.io/RSE_Asia/
R-Ladies Nairobi	https://www.rladies.org
R-Ladies Argentina	https://www.rladies.org/
R-Ladies Mumbai	https://www.rladies.org/
PyAr	https://pypi.org/project/pyar/
Python Namibia	https://www.instagram.com/_pynam/?hl=en
Google Developers Group Nigeria	https://gdg.community.dev/gdg-lagos/
Google Developer Groups GDG Jakarta	https://gdg.community.dev/gdg-jakarta/
Arab-R	https://arabr.github.io/
Latin-R	https://latin-r.com/en
Carpentries MENA	https://carpentries.org/
SciPy LatAm	https://github.com/scipy-latinamerica
Journal of Open Source Software	https://joss.theoj.org/
OSGeo - Open Source Geospatial	https://wiki.osgeo.org/wiki/Main_Page
IDExtremes project	https://wellcome.org/news/new-digital-tools-use-climate-data-better-predict-and-prepare-infectious-diseases-outbreaks
IDAlert	https://cordis.europa.eu/project/id/101057554
Info Dengue	https://info.dengue.mat.br/
D-Moss	https://www.hrwallingford.com/projects/d-moss-dengue-forecasting-model-satellite-based-system
E4Warning	https://twitter.com/e4warning?s=20&t=gHprZcU2OXMuCNSOdNGg3Q
Global Biodiversity Information Facility (GBIF)	https://www.gbif.org/
VectorMap	https://vectormap.si.edu/
Applied-EPI	https://appliedepi.org/people2/
CORDEX	https://cordex.org/data-access/how-to-access-the-data/

Global Adaptation Mapping Initiative	https://globaladaptation.github.io/index.html
INDEPTH Data Repository	https://www.indepth-ishare.org/index.php/home
DataSuds	https://dataverse.ird.fr/
KWTRP	https://dataverse.harvard.edu/dataverse/kwtrp
Clinical Trials Community	https://www.ctc.africa/map
Global Initiative on Sharing Avian Influenza Data (GISAID)	https://gisaid.org/
Africa Open Science Hardware	https://africaosh.com/
MboaLab	https://mboalab.net/
Gathering for Open Science Hardware	https://openhardware.science/
NumFOCUS	https://numfocus.org/
SustainOSS	https://sustainoss.org/
Society of RSE	https://society-rse.org/
US-RSE	https://us-rse.org/
R for EPI	https://www.r4epi.com/
Research Centers, Scholarly Societies, Topical Networks & Non-Profits	
Open Modelling Foundation (OMF)	https://www.openmodelingfoundation.org/
Bioinformatics Research Network	https://www.bioresnet.org/
Center for Infectious Disease Modeling and Analysis (Yale Public Health)	https://ysph.yale.edu/cidma/
The Inter-Sectoral Impact Model Intercomparison Project, hosted by the Potsdam Institute for Climate Impact Research (PIK)	https://www.isimip.org/
Harvard Global Health Institute, Harvard University	https://globalhealth.harvard.edu/domains/cc-health/
Emerging Pathogens Institute (EPI), University of Florida	https://www.epi.ufl.edu/about/
WHO Collaborating Centre on Early Warning Systems for Malaria and other Climate Sensitive Diseases	https://apps.who.int/whocc/Detail.aspx?9a7R6L/IwK+2ZdhEjI+HRA==
Centre for Research in Infectious Diseases (CRID)	https://crid-cam.net/about-our-company/
EcoHealth Alliance	https://www.ecohealthalliance.org
Inter-American Institute for Global Change Research (IAI)	https://www.iai.int/en/#aboutus
Global Consortium on Climate and Health Education	https://www.publichealth.columbia.edu/research/global-consortium-climate-and-health-education/about
Healthy Environments and Lives	https://heal2021.com.au/heal-network/
British Ecological Society SIG in quantitative ecology	https://www.britishecologicalsociety.org/membership-community/special-interest-groups/quantitative-ecology/

Australian Ecological Society SIG in quantitative ecology	https://www.ecolsoc.org.au/category/research-chapters/quantitative-ecology-research-chapter/
International Society for NTDs (ISNTD)	https://www.isntd.org/
Society for Modelling and Simulation International	https://scs.org/
MSSANZ - Modelling and Simulation Society of Australia and New Zealand	https://www.mssanz.org.au/
Centre for Mathematical Modeling at LSHTM	https://www.lshtm.ac.uk/research/centres/centre-mathematical-modelling-infectious-diseases
Center for Scientific Collaboration and Community Engagement	https://www.cscce.org/
Centre on Climate Change and Planetary Health (with Infectious Diseases Lab) at the London School of Hygiene & Tropical Medicine (LSHTM)	https://www.lshtm.ac.uk/research/centres/centre-climate-change-and-planetary-health
Global Health Resilience Group, Barcelona Supercomputing Centre	https://www.bsc.es/discover-bsc/organisation/research-departments/global-health-resilience
Verena/Viral Emergence	https://www.viralemergence.org/
Land Body Ecologies	https://www.landbodyecologies.com/
Lancet Countdown Europe	https://www.lancetcountdown.org/europe/
Institute for Malaria and Climate Solutions (IMACS)	https://www.malarianomore.org/our-impact/international-programs/forecasting-healthy-futures/#:~:text=The%20Institute%20for%20Malaria%20and,in%20malaria%20control%20programs%20worldwide.
SAMRC	https://www.samrc.ac.za/
EHRU -SAMRC intramural unit	https://www.samrc.ac.za/intramural-research-units/EnvironmentHealth
CAIR, Centre for Artificial Intelligence Research	https://www.cair.org.za/about
ACDI, African Climate and Development Initiative	http://www.acdi.uct.ac.za/acdi-research
Modelling and Simulation Hub, Africa	http://www.masha.uct.ac.za/masha/who_are_we
African Population and Health Research Center	https://aphrc.org/
Mali ICER	https://www.niaid.nih.gov/about/mali-icer-program
RedUAS	Red Universitaria de Salud y Ambiente (RedUAS), https://reduas.com.ar/
ISGlobal	https://www.isglobal.org/
DS-I Africa	https://dsi-africa.org/
APBioNet	https://www.apbionet.org/
Environmental Data Science Innovation & Inclusion Lab (ESIIL)	https://esiil.org/center
Government(s) and Funder-Led Initiatives	
Africa Center for Disease Control	https://africacdc.org

National Institute of Environmental Health Sciences	https://www.niehs.nih.gov
National Programme on Climate Change & Human Health (NPCCHH)	https://ncdc.gov.in/index1.php?lang=1&level=1&sublinkid=876&lid=660
Data Science for Health Discovery and Innovation in Africa	https://dsi-africa.org/
CLIMA	https://climaupch.com/
World Health Organization's Hub for Epidemic and Pandemic Intelligence	https://www.who.int/news/item/01-09-2021-who-germany-open-hub-for-pandemic-and-epidemic-intelligence-in-berlin
U.S. CDC's Center for Forecasting and Outbreak Analytics	https://www.cdc.gov/media/releases/2021/p0818-disease-forecasting-center.html
Global Immunological Observatory	https://elifesciences.org/articles/58989
Global Virome Project	https://www.globalviromeproject.org/
NIAID CREID Network	https://creid-network.org/
USAID's PREDICT program	https://p2.predict.global/
USAID's STOP Spillover program	https://stopspillover.org/
USAID's DEEP VZN program	https://www.usaid.gov/news-information/press-releases/oct-5-2021-usaid-announces-new-125-million-project-detect-unknown-viruses
WHO/WMO ClimaHealth portal	https://www.who.int/news/item/31-10-2022-who-and-wmo-launch-a-new-knowledge-platform-for-climate-and-health
GOARN - global outbreaks analytical resource networks	https://goarn.who.int/?page=1part of WHO outbreak cells
European Centre for Disease Prevention and Control	https://www.ecdc.europa.eu/en
Caribbean Public Health Agency (CARPHA)	https://carpha.org
Pan American Health Organization (PAHO)	https://www.paho.org/en
National Institute of Allergy and Infectious Disease (NIAID)	https://www.niaid.nih.gov/about/mission-planning-overview
National Centers for Environmental Information	https://www.ncei.noaa.gov/about
Relevant Funders	
Wellcome Trust	https://wellcome.org/
Chan Zuckerberg Initiative	https://chanzuckerberg.com/
Sloan Foundation	https://sloan.org/
Bill and Melinda Gates Foundation	https://www.google.com/search?q=bill+melinda+gates+foundation&oq=bill+melinda+gates+&sourceid=chrome&ie=UTF-8
Oswaldo Cruz Foundation	https://en.wikipedia.org/wiki/Oswaldo_Cruz_Foundation
International Development Research Centre (IDRC)	https://www.idrc.ca/en/initiative/climate-adaptation-and-resilience-clare

Rockefeller Foundation	https://www.rockefellerfoundation.org/news/the-rockefeller-foundation-and-world-health-organization-announce-partnership-to-expand-global-pandemic-preparedness-in-era-of-climate-change/
Robert Wood Johnson Foundation	https://www.rwjf.org/
U.S. National Institutes of Health	https://www.nih.gov/
Science for Africa Foundation	https://scienceforafrica.foundation/
Adjacent CSID Allies and Potential Collaborators	
Research Software Alliance	https://www.researchsoft.org/
Research Data Alliance	https://www.rd-alliance.org/
African Leaders Malaria Alliance	https://alma2030.org/
Asia Pacific Leaders Malaria Alliance Secretariat (APLMA)	https://www.aplma.org/
Allen Institute for AI	https://allenai.org
TranslatESciences	https://translatesciences.com/
OpenUK	https://openuk.uk/
Intergovernmental Software Collaborative	https://beeckcenter.georgetown.edu/projects/intergovernmental-software-collaborative/
ILDA	https://idatosabiertos.org/en/
PLOS Climate	https://journals.plos.org/climate/
Intrahealth International	https://www.intrahealth.org/
Drugs for Neglected Diseases initiative (DNDi)	https://dndi.org/news/2022/dndi-at-world-conference-francophone-science-journalists/
Robert Stempel College of Public Health & Social Work, Florida International University	https://stempel.fiu.edu/index.html

Appendix 7: Workshop Agenda

Developed by the Code for Science and Society facilitation team for the May 2023 convening to co-design of a Climate-Sensitive Infectious Disease Open Source Software Community of Practice

Day 1: Tuesday May 16, 2023

Before Breakfast: Pre-arrival COVID-19 Test

Included in your welcome bag provided at hotel check-in is a COVID-19 rapid test. Please self-administer as per our COVID-19 policies. It takes 15 minutes to read results. Any participant who receives a positive result should notify organizers via email (angela@codeforsociety.org) or text +1 808-358-8563.

7:00 - 8:30 Breakfast provided by The Vineyard @ The Square Restaurant in the hotel

8:30 - 9:00 Registration and Arrival to the Conference Centre Summerhouse

Please find your way to the conference space "Summerhouse" on the ground floor (Level 1) to allow enough time to receive your name badge and sign photo consent form at registration. We will begin our first session together promptly at 9:00 hrs.

9:00 - 9:20 Introduction and Context Setting

9:20 - 9:40 How We Want to Engage with Each Other

9:40 - 10:15 Introducing Our Stories and Why We're Here

10:15 - 10:45 Tea and Conversation @ *The Foyer outside Conference Centre Summerhouse*

10:45 - 12:00 A CSID Roundtable Conversation, featuring you

This conversation will bring participants into a roundtable discussion for short amounts of time to start revealing some of the information in the room: sustaining open-source scientific software, building collectivist and community-based solutions, developing networks at scale, training the trainers, end-user experience on the digitization of community health systems, data sharing infrastructures, open data initiatives and working with government extension offices, and other aspects of digital equity, data governance, and end-user co-creation. **No preparation is necessary.** This is meant to exhibit summary expertise you already carry regularly.

12:00 - 13:30 Lunch provided by The Vineyard @ The Square Restaurant

13:30 - 14:30 Design Process Introduction & Landscape Analysis

The facilitation team will share our process for the coming days and together we will review the landscape of existing CSID / Open-Source Software Communities.

14:30 - 16:00 Embodied Reflective Workshop

16:00 - 18:00 Free Time or Self-Organized Group Outing in Cape Town

18:30 Group Dinner provided at GOLD Restaurant, 15 Bennett Street, Green Point.

Shuttles will be provided to and from the restaurant in Cape Town.

Day 2: Wednesday, May 17, 2023

Before Breakfast: Pre-arrival COVID-19 Test

Please allow enough time in your morning routine to self-administer a COVID-19 rapid test prior to breakfast as per our COVID-19 policies (noting that it takes 15 minutes to read results). Any participant who receives a positive result should notify organizers via email (angela@codeforsociety.org) or text +1 808-358-8563.

7:00 - 8:30 Breakfast provided by The Vineyard

8:30 - 9:00 Arrival to The Workshop Space (Conference Centre Summerhouse)

9:00 - 9:15 Review of Day 1, and Day 2 Logistics

9:15 - 10:45 World Cafe Theatre to Design the Community of Practice

An embodied design-led research format to understand what participants want the community of practice to look like and synthesize different ideas that emerge.
Proposed Explorations:

- *What does an Open CSID Community of Practice look like to you?
- What can you contribute to and receive from a Community of Practice?
- What's in-scope and out-of-scope for a Community of Practice to handle?*

10:45 - 11:00 Tea and Conversation at The Foyer outside conference center

11:00 - 11:30 Scheduling Emergent Conversations for the Afternoon

11:30 - 13:00 Lunch provided by The Vineyard at Morii fine dining restaurant

13:00 - 14:00 Breakout Discussions and Activities - Session 1

There will be multiple simultaneous tracks to choose from of both pre-planned and emergent participant-generated conversations that will include sustaining a community

of practice and what functional governance might look like, key features of impactful open-source software, and an interactive workshop where participants co-design software using speculative design. Participants are invited to host conversations. Discussions and Activities will be listed here as they are scheduled.

List of discussions and activities will be generated in a dynamic agenda on Notion.

14:00 - 14:15 Tea Break

14:15 - 15:15 Breakout Activities and Discussions - Session 2

There will be multiple simultaneous tracks to choose from of both pre-planned and emergent participant-generated conversations that will include transnational data sharing issues and potentials, identifying CSID education and training needs, and speculative co-design of software.

List of discussions and activities will be generated in a dynamic agenda on Notion.

15:00 - 16:00 ****Virtual Attendance via Zoom for those unable to attend in person: register.***

For those attending virtually, please enter the Zoom Room at 15:00 hrs (South Africa Standard Time) via the link to be included here soon. Virtual attendees will be able to listen to report backs from the Breakout Discussions, and will be able to send in notes, suggestions, and discussion offerings via this page.

15:15 - 16:00 Plenary Report Backs on Breakout and Open Space Discussions
This will include virtual participation for those not able to attend to listen.

16:00 - 16:30 Prepare for Happy Hour and Meet at Shuttles
Shuttles provided by Facilitation Team

17:00 - 19:00 Happy Hour in the Sky Bar at Grand Daddy Boutique Hotel, 38 Long Street
Limited drinks and appetizers will be provided
Dinner not provided; please inform Alida if you need transport back to the hotel.

Day 3: Thursday, May 18, 2023

Before Breakfast: Pre-arrival COVID-19 Test

Please allow enough time in your morning routine to self-administer a COVID-19 rapid test prior to breakfast as per our COVID-19 policies (noting that it takes 15 minutes to read results). Any participant who receives a positive result should notify organizers via email (angela@codeforsociety.org) or text +1 808-358-8563.

7:30 - 8:30 Breakfast provided by The Vineyard

8:30 - 9:00 Arrival to Workshop Space

9:00 - 9:20 Review of Day 2, and Day 3 Logistics

9:20 - 11:15 Envisioning Our Path Forward Together

Consensus Building and Agreement Setting; Identifying What's Needed to Compose and Maintain our Community and Identifying Next Steps and Action Items; Setting Expectations for Follow Up and Communication Going Forward.

11:15 - 11:30 Tea Break

11:30 - 12:30 Cultivating Joy and Engagement as a Community

13:00 - 14:00 Lunch provided by The Vineyard

Workshop Ends after lunch. Check-out by the next day, Friday, May 19th, at 11 hrs.

Appendix 8: CSID Workshop Proceedings and Participant Feedback

1) Workshop Proceedings

Figure 1. Participants at the 2023 CSID Co-Design Workshop. Photo by Lihlumelo Hlumie.

Day 1

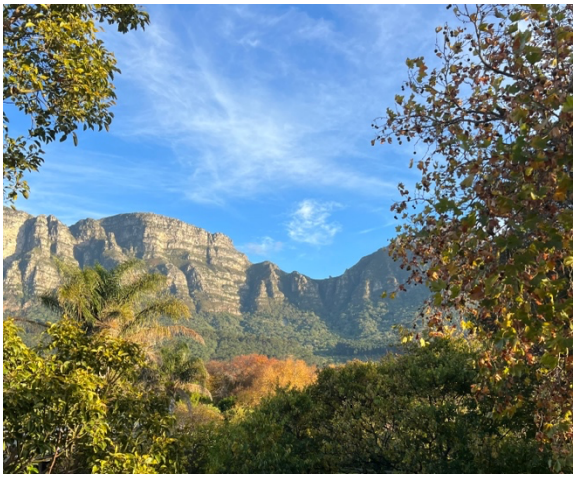


Figure 2. CSID Co-Design Workshop took place at the foot of Table Mountain in Cape Town, South Africa. Photo by Angela Okune.

Day 1 of the program aimed to build new relationships and set the context for discussion including a sense of who was in the room and the expertise they brought.

Day 1 conversations opened by establishing Community Agreements, which were phrased as "How Do We Want to Engage with Each Other." This session encouraged participants to be creative and political. The initial uncertainty was broken when someone noted that "we've never done this before" and the conversation quickly moved from "this is weird" to "how about this!" and then "here's something important." This exercise encouraged participants to have agency over their own space and community rapidly and brought in issues of equity and horizontalization without the facilitators having to.

The following were the community agreements determined by the participants:

- Build psychological safety for and with one another
- Recognize the barriers and restrictions we face for collaboration and participation
- Be present as mutual respect
- Imagine non-judgement, seed kind, co-supportive culture
- Practicing and being human in these AI times
- Noticing normative hierarchies and trying to practice flattening them
- Step up; step back
- Room for a range of emotions
- Create open space for one another to join conversation
- Spell out acronyms and define jargon
- Seek understanding: ask questions
- Be Present: notes will be shared by facilitators
- Build from the bottom-up taking into account one-another's divergent lived experiences. Remember that we all have different cultural pathways to decisions.

Figure 3. Participants begin to get to know each other during warm-up activities on Day 1. Photo by Lihlumelo Hlumie.

A CSID Roundtable Conversation, featuring you

Figure 4. During the fishbowl activity, participants sat in the outer circle and listened intently to the conversation in the inner circle. People were invited into the inner circle to share their thoughts and began to also self-nominate to join. Photo by Lihlumelo Hlumie.

A "fishbowl" session was hosted where a dynamic, changing circle on the inside had an organic conversation that was being actively watched and listened to and at times interjected by an outer circle of participants. The aim of this activity was to offer participants a sense of why we are all in the room and the different backgrounds and areas of expertise that everyone brought.

- The conversation opened with a discussion on software sustainability. The founder of ROpenSci mentioned that documentation of best practices for software development have significantly improved in the past 10 years, making it easier to create sustainable

software but there is no need to start at the very beginning with a question of sustainability but rather to make sure there is appropriate demand and interest in the product before beginning to ensure that it is sustainable.

- Others joined the circle and the conversation turned to questions of data access and ethical data practices. If a model is only as good as its data, how do you ensure you are accessing good data?

- One member of the group mentioned that there are several examples of opening up data that have gone badly. Data access, she emphasized, is fundamentally about relationships. That is also how you enable your models to be used. She underlined it is important to consider why you need the models in the first place, to what purpose will the data be used, and how to close the feedback loop to enable communication thereafter. Data needs to be procured in a way that is respectful to the original terms of data collection in the field and understood within its context and the individuals it represents, rather than solely focusing on as numbers and statistics.

- A question was raised, if data is based on social relations, for small organizations or young Primary Investigators that are just starting their careers and may not have a strong network, how can such new entrants be supported to have access that is equitable and enables them to participate? This was answered that it is about building processes that enable those who have access to support newcomers to also gain access.

- A new voice joined the conversation and reiterated the importance of getting feedback early and understanding the needs of the specific community for whom a software product is being developed. He made the distinction between the client (the actual user or decision-maker) and the ultimate beneficiary (the general public) and explained that targeting the right client is crucial, especially in cases where there may be a subscription model or financial considerations. This person mentioned that decision-making power may lie with planning and finance departments rather than the line ministry responsible for a specific domain so he advised that it is important to understand the decision-making dynamics within government and finding the right individual champions who understand the issue and can influence resource allocation within the government.

- Importance of having a good understanding of your local setting in which you're trying to engage your user. The facilitator emphasized here another benefit of a Community of Practice is the ability to tap into a collective network and knowledge base of diverse end users. By coming together, individuals can leverage the connections that different members of the community have across different local contexts.

- The conversation turned to the imbalance between the abundance of data in some parts of the world and the lack of data in others, making it challenging to build robust models. A climate scientist added to the conversation to point out that contrary to popular perceptions about climate data as not having ethical issues, climate scientists, like public health scientists, face complexities when working with different types of data, including biases and limitations that need to be carefully considered. A software engineer noted that metadata, or data about the data, and context building are crucial for understanding

and utilizing datasets effectively. Synthetic data and geospatial data were raised as possible useful alternatives when original data is limited or inaccessible, providing options for analysis and stakeholder engagement.

- An open data expert raised the importance of establishing feedback loops in data systems, specifically illustrated through the example of the DHIS (District Health Information Service) in Bangladesh. Despite the system being widely used and collecting data at the facility level, there were instances where the data did not make its way back to inform decision-making and service delivery. Encouraging the flow of data back to the individuals and communities it pertains to requires intentional efforts and can be achieved through operational practices rather than solely relying on policy.

- Another health policy expert shared that feedback loops are essential to maximize the value of government data, as many datasets lack a clear purpose or understanding of their potential value. Creating value from data often involves aligning it with specific organizational needs and generating insights that resonate with stakeholders, but it can be challenging to establish meaningful data utilization without a clear understanding of the organization's operational context and objectives.

- A representative from a local county government hospital mentioned understanding the confidence and knowledge of different decision-makers, including healthcare workers, government officials, and end-users, is crucial in effective data communication. Feedback language should be simplified and tailored to the specific audience to ensure comprehension and utilization of the data. When communicating with higher-level decision-makers, it is important to incorporate a political angle and language that resonates with them to garner support and buy-in.

Figure 5. Dinner on Day 1 included a lesson in djembe drumming and a 14-course African menu. Photo by Angela Okune.

The fishbowl activity was an embodied way to practice the kinds of advice and conversations a supportive community of practice might be able to provide. People practiced self-managing during fishbowl, leaving the circle and making space for others without having to be asked to; always leaving an open seat and not filling any space open to platform oneself; not hogging the mic, etc.

The afternoon of Day 1 closed with a grounding presentation by Code for Science & Society Director of Programs, Dr. Angela Okune who laid out some of the reasons for the gathering and the work ahead. Ample time was given for groups to rest and recoup before an evening dinner and show at Gold Restaurant.

Day 2
Day 2 leveraged the relationships and conversations built on Day 1 as a springboard for deepening conversation about the ideal design for a CSID community of practice. We began with a World Café session where participants sat at different tables to answer the following prompts:

- *What difference will having the various people we discussed be involved make?*
- *What activities can connect these different people together? What or who will this impact and why do we want that?*
- *What's out-of-scope for this CoP? What are its bounds that guard and define what it will focus on?*

One person anchored the conversation and remained at the table while others scattered around the room for each question. Through these conversations, we collectively surfaced the elements of the CoP design, namely the who, what, and why. (See the main report for the summarized findings from these conversations).

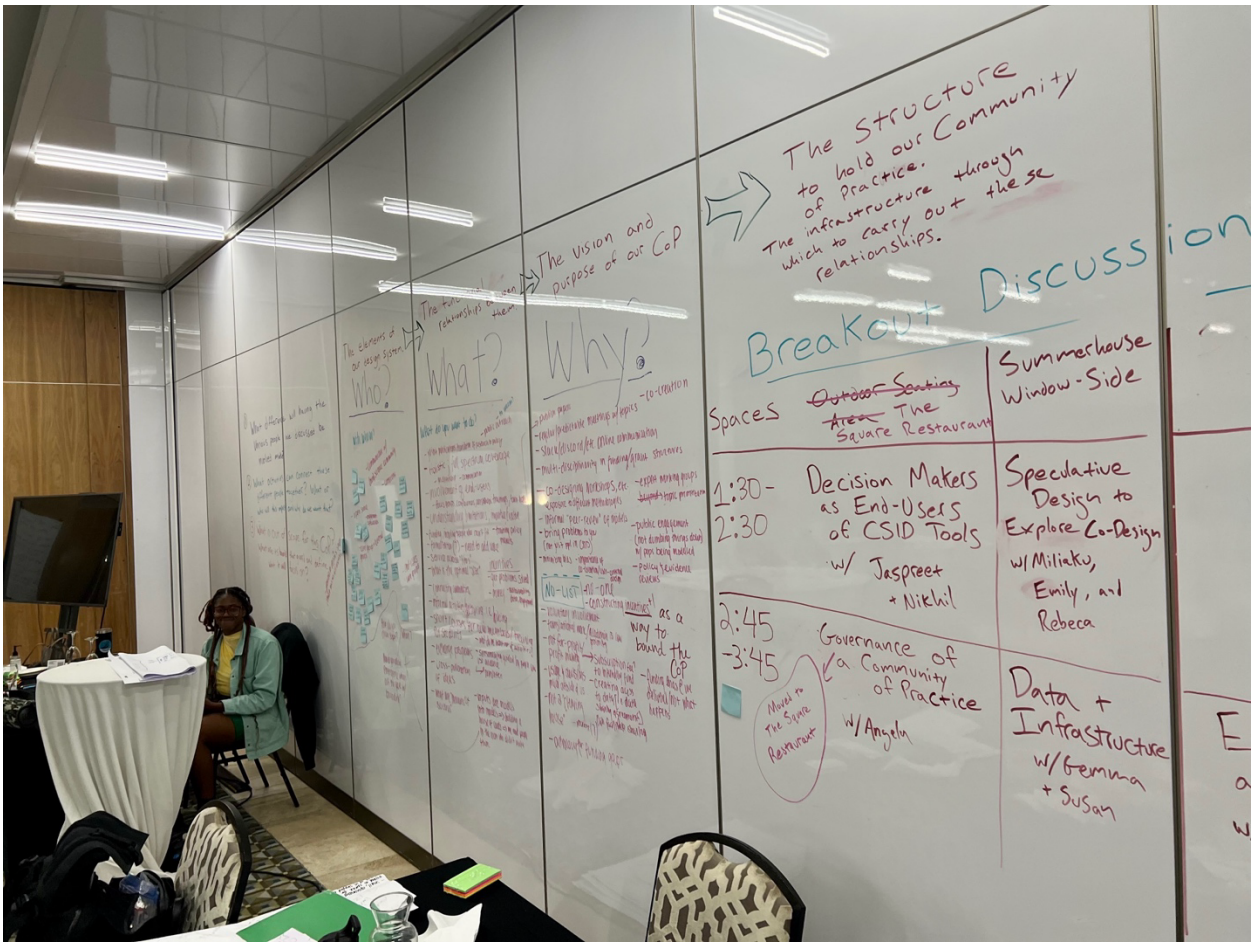


Figure 6. CS&S Sr. Program Manager Miliaku Nwabueze in front of a whiteboard of notes from the World Cafe on Day 2. Photo by Angela Okune.

Following lunch, we scheduled emergent breakout conversations. These breakout groups covered the following topics, several which were self-proposed by participants:

Figure 7. Images from different breakout group discussions. Photos by Lihlumelo Hlumie.

- Decision Makers as End-Users of CSID Tools
- Governance Structures of a CSID CoP
- Using Speculative Design to Explore Co-Design
- Data and Infrastructure
- Equity in a Community of Practice
- Modeling Uncertainty and Interfaces between Climate and Epidemiological Models

Following the breakout discussions, groups reported back and virtual participants from teams who could not attend were invited to listen in.

We closed Day 2 with a happy hour event in downtown Cape Town to connect CSID workshop participants and also draw in new and old collaborators and partners working in open science, open source and research software, social justice, and related fields in Cape Town.

Figure 8. Participants mingle at the CSID happy hour. Photo by Lihlumelo Hlumie.

Day 3

Figure 9. Participants warm up on Day 3 with an icebreaker that has everyone in smiles and high-fives. Photo by Lihlumelo Hlumie.

Figure 10. Whiteboarding next steps for the CSID CoP. Photo by Lihlumelo Hlumie.

Day 3 sought to sketch out an articulation of what work the participants may want to start moving forward on. We began with a light ice breaker before diving into articulating what specific activities the CoP would start with and who and when such activities should begin. These activities are represented in the full report.
We closed our time together cultivating joy and engagement as a community with a clap circle.

Figure 11. We closed with an energizing and joyful clap circle with five words from everyone to summarize their time together. Photo by Lihlumelo Hlumie.

2) Key Development Considerations for Those Building CSID Software Tools

From the workshop, several key development considerations emerged that the group thought were important for those building CSID software tools to take into account:

A. Know Your User(s)

Understanding the diverse nature of government is crucial when expecting uptake of models or technology. Differentiate between the client (decision-maker) and the ultimate beneficiary (general public) when targeting stakeholders. Building relationships and trust with government officials is key to accessing data and driving change. Investing time and energy in building relationships and finding champions within the government is necessary for success. Getting feedback early and understanding the specific community's needs is essential.

B. Data Ethics are Important for Good Science

Important to consider how the data was originally collected and under what terms sharing can be done. How can we share data as openly as possible without being unethical and recognizing there are unequal benefits to data sharing. It is important to understand data context beyond the currently documented meta-data. A key benefit of a CSID CoP could be the ability to discuss and engage with experts from different domains of expertise to understand the right context for data use.

C. Consider Your Data Feedback Loops

Consider the purpose of data usage and closing the feedback loop for better communication, acknowledging data as more than just numbers and statistics. There is a need for intentional efforts to ensure data flows back to inform decision-making and service delivery. Many datasets lack a clear purpose so generating insights that are aligned with a user's organizational needs and operational context is crucial to unlock the value of the data and tool.

D. Gritty Minimum Viable Tool vs End Users Getting Burned

During the workshop, some participants mentioned the importance of "test early, fail early," meaning that you should put your software out there as early as possible to determine its usefulness and potential adoption outside of one research group. However, in other parts of the workshop, it was raised that when the software does not work, software users may not be keen to use it again. You could lose their confidence in the model or tool, and they may not trust the technology again. This tension – when and how to release a software tool to users – is one to continue to discuss and around which greater best practices are needed.

3) Post-Workshop Participant Feedback

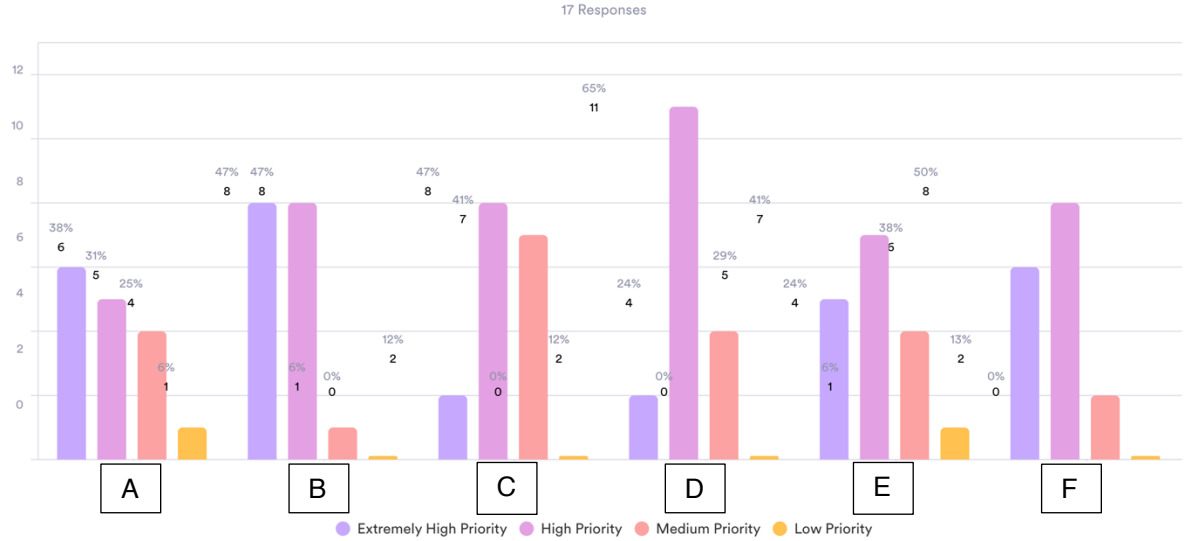
A survey was sent to workshop participants following the event and 17 responses were received.

What were your initial expectations of the event? Were your expectations met? Please explain why or why not.
My expectation was to engage with various professionals such as researchers (human, animal, and environmental), clinicians, modelers, academicians, policymakers etc. to understand their contribution to modeling, the challenges they go through, and how these are addressed. I also got contacts to tap into whenever I have a modeling project. Yes, my expectation was met.
My expectations were to learn, understand the needs of the community, and explore pathways of ensuring a community is designed to be inclusive and impactful. Yes, the expectations were fully met. Quite enlightened.
Before attending the meeting, I was afraid that the meeting would be too technical with computer experts. Organizers made a good attempt to include scientists from different fields and the discussions were productive. Still the group needs more experts from public health, environmental microbiology, and social science.
No expectations. Meeting was interesting as a means of meeting people, but I had hoped to heat more about what each group is doing to assist in developing networks and relationships.
My expectation was to learn from the CSID community, to exchange ideas and to get a start of the art in the CSID space.
I primarily thought of this as a networking event and the structure of it allowed for that. It was more prescriptive in terms spending the time of formalized community building exercises than I expected.
This was a very well-organized meeting and so much needed for the interdisciplinary work everyone is currently doing. I met all my expectations and am looking forward to continuing my engagement with the group. Making sure that it does add to the daily research needs of the group would be the true motive for everyone to stay engaged.
My initial expectations for the workshop were to participate in a constructive exchange of ideas, learn from the experiences of other organizations, and identify opportunities for collaboration on the themes of digital equity, data governance. I'm delighted to say that these expectations were far exceeded. The workshop succeeded in bringing together individuals from different backgrounds and expertise and facilitated dynamic and productive discussions.
I expected to have a clear way forward and to know what is expected of me and what role I'll be playing in the group. No, they were not met but it was interesting to be part of the planning /design process.
I didn't know what to expect from the meeting as it was new to me but I had a feeling that it was going to be something exciting and useful for research community. The pre-meeting, kind of, helped me to understand the concepts a bit and also got to know more people. I enjoyed the workshop very much. Although the outcomes were not very tangible, I felt I learned a lot and I keep thinking how to participate in these communities when I do research. It was also inspiring to hear from others as well.
Yes. I was expecting this event to be a place for networking and collaboration. It was perfectly set up for this opportunity.
My initial expectations were to meet people close to my research interest and with my expertise, but I also meet other researchers and professionals from other knowledge areas not so closely related to mine and that open my mind to new ideas on my own investigation line. It was more than expected.

This was a fantastic event. Excellent planning and stellar attendees. The meeting was run very well, with many excellent opportunities for interactive activities and feedback.
I expected to learn what people are doing in their CSID projects and perspectives/recommendations on how to promote the adoption and sustainability of developed CSID software for disease prevention. My expectations were mostly met. I wish I would hear more perspectives from the governmental sectors related to the deployment and sustainability of the CSID tools.
My expectations were to meet other grantees and learn more about CoPs. My expectations were fully met.
I had no expectations. The programming was great but felt a little forced/contrived and not that natural. Not that it's bad but it felt very top-down. Because the unconference part was also forced into a little window in terms of generating topics and time to discuss, it felt short of a better outcome. Next time I'd suggest giving people a head's up beforehand, providing time to post topics for everyone to see and digest, and then collate/combine.

Given the conversations hosted in Cape Town, how would you prioritize the benefits of participating in a CSID Community of Practice?

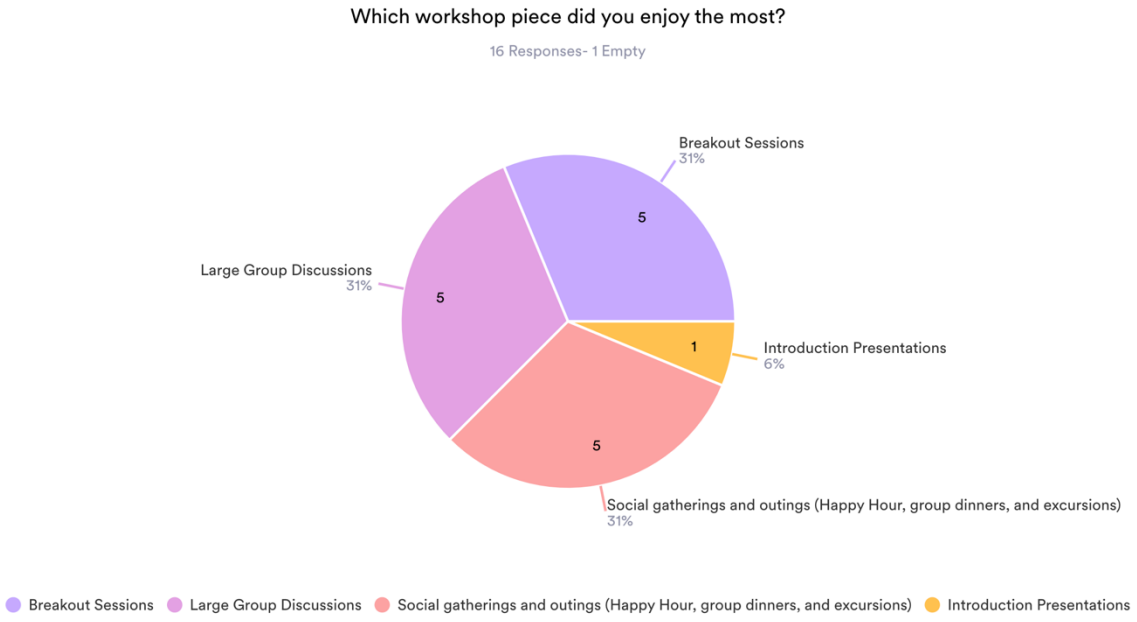
17 Responses



A = Funding for my project
B = Meeting potential collaborators
C = Connections with software end users in various contexts
D = Secondary and tertiary connections across the wider network
E = Opportunities for horizontal and vertical mentorship and exchange
F = Knowledge sharing about relevant events and opportunities

Any suggestions for next event topics, activities, or locations?
An update on what specific aspect of our work has changed as a result of the CSID workshop
A deep dive into one of the components highlighted on the board, especially with the teams interested in those items.
Developing a centralized platform for sharing resources and integrating ideas/modelling approaches/data resources.
Review of key software and methods adopted by attendees after Cape Town. Africa could still work for me.
I feel more a part of a community working on and researching the topic than specifically developing software tools.
A demonstration/ workshop for common methods would be a good meeting/recurring meeting we can plan. Location: finding a location that has the least visa restrictions is necessary. Sometimes is ignored how much preparation, time, and effort is put in to arrange visas. Also, the group was not representative of geographies around the globe. I understand it is hard to find, but some solutions could be Indonesia, Hong Kong, Singapore, Maldives, etc.

Topics: Sessions dedicated to exploring different sources of climate, satellite data and how to access them. A workshop on the application of machine learning techniques in the field of climate-sensitive diseases. Activities: Hands-on coding sessions where participants can work together to develop or improve models. Small group discussions to share specific challenges and possible solutions.
I think we need to narrow down what exactly it is we want to do and how and where do we start.
Case studies of climate sensitive research, challenges and key factors for success.
I would be equally interested in a conference type of gathering where individuals present their work about climate impacts on diseases
No idea. Maybe some place in Asia?
Loved the Capetown location!
- Topic: Lesson learned on success/failures of developing and using CSID software tools in disease prevention - Activities: presentation, interactive discussion on problems and solutions, and social gathering. - Location: An Asian country in the South or South-East region
maybe add some lightning talks (max 5 minutes) by members so that we can learn more about what's going on in each group
- More unconference time. - More people writing down the discussions (on shared google docs). Have a paid staff member edit these so that at the end, we leave with extensive notes from which to build and make progress.



On a scale of 1-5, how much more do you feel a part of a broader community of people working on software tools to understand climate change and disease patterns?

17 Responses

Best Response

4

41%
Percentage

17
Responses

Data	Response	%
4	7	41%
3	5	29%
5	5	29%
1	0	0%
2	0	0%

Would you recommend coming to our next gathering to a friend or colleague? Why or why not?
Yes, to ensure that everyone is onboard for greater support and buy in of the CSID COP
Absolutely, it is a very inclusive environment.
Yes. I will ask my collaborators also to be the part of this community
Not sure...depends on the focus of the event.
Yes, I definitely would. For networking and touching base with the wider community.
Yes, there's an excellent group of people involved in this so that's a great asset.
Yes, particularly my postdoc.
Yes, I would definitely recommend your next event to a colleague or friend. The level of discussion and the opportunity to share experiences and knowledge were very rewarding at this workshop. The interactive format also enabled in-depth discussions on key topics and opportunities for collaboration. In addition, we are working internally with other colleagues in our research team on the PRIDE-C (PRedicting Infectious Diseases via the Environment and Climate) project, which added an extra dimension of relevance and interest for us. In short, the experience was highly beneficial, and we believe it would be for others in our field of work too.
Yes, I would. I'm sure they will learn a lot.
Definitely. I think just learning things by talking and sharing with people is what we sometimes need. The atmosphere was set up excellently for this workshop. Well done to the team!
Yes. It was different from our normal workshops, and I enjoyed every activity
Yes. Because is a very good experience.
Yes
Yes, I would do recommend this to my colleagues who are experts on this field.
Yes. It is a unique opportunity for making new connections and expanding our world view.
Absolutely!

Appendix 9: Lessons and Challenges from Building and Sustaining rOpenSci as a Community of Practice

By: Karthik Ram, The rOpenSci Project and University of California, Berkeley. Berkeley, CA 94720 karthik.ram@gmail.com. January 2023

Introduction

rOpenSci is a non-profit organization founded in 2011 by Karthik Ram, Scott Chamberlain, and Carl Boettiger. It provides community support, standards, and infrastructure for scientists and research software engineers working in the R programming language to develop, maintain, and publish high-quality open-source scientific software. rOpenSci has evolved from a software development organization to a community support organization over the past ten years. In this whitepaper, I describe the evolution of the community, explore whether rOpenSci can be described as a true community of practice (hereafter CoP), and conclude with some of the challenges and opportunities of sustaining this effort.

The organization was initially founded to create R software tools to provide access to scientific data sources when infrastructure and resources for creating and disseminating such software packages were sparse. Scientists largely used their bespoke code to access data sources if programmatic tools were used at all. rOpenSci staff developed several widely-used tools for data access, especially in the ecological and phylogenetic scientific fields. As community members started contributing their own tools, rOpenSci became a clearinghouse for such packages. Over time the general R ecosystem became richer and more accessible, and it became easier for people to develop their packages, so rOpenSci's focus shifted from providing users with a different tool for every data source to helping people create their own software by developing standards, educational materials, technical support and scaffolded access to a community of peers.

rOpenSci's current activities and community are like an onion's layers, with a rigorous yet collegial open software peer review system at its core (Figure 1). This is a transparent, constructive, non-adversarial, and open review process that combines academic manuscript review with aspects of code review (Ram et al. 2019). Roles include editor, author, and reviewer. Each review is a discussion in a single thread publicly accessible on the GitHub platform. Briefly, an author of an R package chooses to submit it for review, the editor-in-chief (a rotating role) determines whether the package is in scope, triggers automated checks, approves (or declines) the submission, assigns one of many associate editors, and that handling editor invites two reviewers from a database of volunteers or solicits reviewers from the broader community. After reviewers are assigned, the community manager invites package author(s) and reviewers to join rOpenSci's invitation-only Slack workspace where they can get help or share knowledge in technical or domain-specific areas. Once the package passes review, the author retains maintainership and copyright while the package is transferred into the rOpenSci GitHub organization, marked with a "peer-reviewed" badge, and rOpenSci branding is applied to the documentation. This helps make the package more discoverable and gives the author more visibility. Software peer review would not be possible without the time and effort of volunteers. A reviewer takes an average of nine hours and a median of five hours to review an R package [5],

[5] https://ropensci.org/blog/2018/05/03/onboarding-is-work/

and many people review three or more packages. For their ongoing commitment, associate editors receive an annual honorarium.

Beyond software peer review, rOpenSci has developed projects and programs (Table 1) to support and promote its mission and the work of contributors and people wanting to engage with and learn from them. Work is carried out using multiple channels to facilitate different types of communication among participants inside and outside the community in many time zones worldwide. The rOpenSci community is a self-identifying group composed of users and developers of R software tools who, together, contribute to the technical and social infrastructure for open and reproducible research. These are people who use, cite, and share use cases for rOpenSci packages, attend or present in our Community Calls, write a post for our blog, participate in a rOpenSci unconf, or in our research-domain focused communities, ask or answer questions in our fora, report problems, recommend or implement solutions, contribute to, or maintain a package, or participate in rOpenSci software peer review as a reviewer, author or editor.

Figure 1

Table 1. rOpenSci programs and projects

Program	Purpose
Packages	>350 curated staff- and community-contributed tools

Software peer review	system for open peer review of community-contributed R packages, supported by documentation of good practices
Use cases	community-contributed examples of applications of rOpenSci packages or resources
R-universe	infrastructure for publication and discovery of research software in R with integrated measuring and monitoring tools
Community calls	virtual 1-hr multi-speaker presentations and discussion on technical and community topics, quarterly, open to anyone
Blog	narrative or technical posts by staff or community members
Social coworking and office hours	virtual 2-hr sessions for getting work done alongside staff, R users, and developers, monthly, open to anyone
Unconferences	in-person annual hackathon (discontinued)

Supporting public documentation - dev guide (https://devguide.ropensci.org/), stats review guide (https://stats-devguide.ropensci.org/), blog guide (https://blogguide.ropensci.org/), and a contributing guide (https://contributing.ropensci.org/).

Does rOpenSci fit the description of a COP?

There is a growing need for CoPs in a range of contexts in STEM as scientists and scientifically trained professionals need to deepen their skills in specific areas in an ongoing manner. This can range from teaching science communication to developing software skills and preparing for newly emerging roles such as research data manager. Training alone is rarely sufficient to gain proficiency in a new “craft” such as creating code or building a community. Belonging to a learning community of peers who can learn together is essential.

While general definitions of a CoP circle around the idea that it’s a place where individuals learn how to advance a skill or craft through interaction with one another, three core structural elements are at the heart of CoP theory: i) a ***domain*** of focus, ii) a ***community*** of practitioners at various levels of expertise and iii) a ***practice*** or set of activities the group does together to support and advance their learning.

These elements provide a useful way to assess why CoPs succeed or fail. A ***domain*** may be too broad and potential members cannot see the direct relevance of the topic to their own work, or too narrow and outside the scope of their interest. Defining who needs to be present within the ***community*** of a community of practice can also be challenging. If membership in the

community is too homogenous it will not represent the diversity of perspectives needed to truly advance learning. Too many novices in the domain and not enough individuals with expertise may result in unsatisfactory learning together or put greater pressure on a community manager to schedule learning activities. Too many experts may result in different challenges - from busy schedules and lack of incentive to engage, to issues of competition and unwillingness to share intellectual property. Finally, in terms of ***practice***, many proto-CoPs state a desire to "share knowledge and best practices" in a specific domain but fail to create adequate programming and/or forums in which to nurture that knowledge - hoping instead that members will spontaneously determine the ways in which they wish to engage together. This can result in a lack of clarity about how to participate or fragmented attention across different activities.

Domain

In its current configuration, rOpenSci has a clearly defined *domain*: the development of well-documented open-source R software for research, statistical modeling, and general-purpose data science. Key to the formation of this community in 2011 was the decision to focus on the centralized development of software packages in specific areas that aligned with the expertise of the founders and early community members: biodiversity data science, and scientometrics (Boettiger and Chamberlain, n.d.; Ram et al. 2019). This avoided over-extending the founders and diluting community member contributions across multiple domains. As the general R ecosystem became richer and it became easier for people to develop their own packages, rOpenSci's domain shifted in 2015 to helping people create their own software by developing and documenting standards and a software peer review system to support that. The next iteration of this in 2020 was to expand into a new subject area where R is relevant: statistical software peer review. This gradual shifting of domain shown by rOpenSci may be relevant to other longer-standing CoPs that are able to be responsive to changing dynamics across their broader ecosystem, and thus continue to sustain member participation.

Community

The *community* that rOpenSci is focused on is a subset of the people who use or develop R packages *for open and reproducible research* in any sector, including academia, government, industry, or non-profits. The community consists of people with diverse motivations (often tied to their professional roles and/or levels of R expertise) and it follows that their participation in the community will also take a variety of forms. Some members are focused on having their package reviewed or on building their skills by reviewing a package, while others aspire to change the broader research culture to be more open and embrace reproducible research practices.

One consequence of this growth that emerged through conversations with the community is that many people inside and outside the rOpenSci community are not able to clearly define the mission of rOpenSci, even though most said that they thought of themselves as members of the community and specifically appreciated how welcoming it is. This confusion likely stems from the evolution of the domain that rOpenSci focuses on as many people think of rOpenSci as primarily a producer of R packages and are less aware of the newer emphasis on software peer review.

Many self-identifying rOpenSci members did not name rOpenSci as their primary community - often focusing on the R community more broadly. This is not uncommon in an evolving ecosystem of overlapping communities and can be seen in other CoPs too, such as those focused on open science more generally, where members may belong to local, regional/national, and domain-specific CoPs at the same time. We increasingly believe that successfully convening a CoP requires understanding and communicating its position within a broader ecosystem of other

communities so that its outputs are discoverable and shareable between communities. This will also protect members of multiple communities from being burned out by multiple requests to make similar contributions across the ecosystem.

Finally, while nurturing a sense of belonging and being clear about the intended membership and purpose of a CoP is crucial, it does not guarantee that all individuals feel included and able to participate. Specific programming may still be needed to support diverse participation, including from newcomers and/or members of historically under-represented groups and this need may become more obvious as a community grows. As it continues to grow, rOpenSci is now planning a champions program to increase the diversity of contributors to its software peer review process.

Practice

The *practice* that rOpenSci community members undertake together is to evolve a shared understanding of what good software (and supporting infrastructure) looks like - including how we evaluate it. This ultimately results in making the whole research software ecosystem more robust and reproducible. This practice includes the shared resources that members produce (e.g., packages, reviews, documentation, and more) and the activities that they engage in together (e.g., community calls, Q&A, Slack discussions, software peer review). Not every member needs to engage in the practice in the same way. For some members, attending a community call to learn about a specific output produced by other community members may be sufficient. For others, the practice looks like participation in the software peer review process or proposing new standards for evaluation.

The specific activities that the core team supports create multiple opportunities for members to engage in practice together and to move into deeper levels of engagement with the practice should they choose to. (Table 1. rOpenSci programs and projects). These activities have evolved over time, as the domain of focus for rOpenSci has evolved and also as the location for practice has shifted to online-only spaces. For example: from 2011 -2014 the practice included the co-development of packages among a small group of experts. In 2014 - 2019 this shifted with the addition of unconferences, which supported in-person trust-building and led to an active Slack workspace. In 2020 - 2021 rOpensci's activities were all virtual - with 50- 150 people attending each community call and the introduction of new social coworking and office hours.

How have domain, community, and practice co-evolved in the rOpenSci CoP?

The initial, clearly defined and limited scope of rOpenSci's domain allowed a small community of active members to congregate around it. The practice continued to evolve with the actions of members, such as proposing key packages that should be developed, or developing and contributing them to the rOpenSci repository to make them more discoverable, rather than keeping them in their own individual collections. As the domain has evolved with the broader R ecosystem, so has the community and the practice. Defined roles in software peer review were and continue to be an explicit prompt for the growth and evolution of the community and the practice.

The hiring of a community manager in 2016 influenced the practice of the community which in turn had knock-on effects on the domain and community. Pre-2016, community calls focussed on specific tools or technical approaches, but post-2016 some addressed more meta-topics like "How do I Create a Code of Conduct for my Event/Lab/Codebase?", "How to ask questions so they get answered. Possibly by yourself!", or "Involving Multilingual Communities". In 2017, a new series

of blog posts provided a platform for people whose packages had passed software review or who had reviewed a package, to write about their experiences. This brought visibility to the practice of software peer review, new aspects of rOpenSci's domain, and helped those contributors identify more strongly as members of the community.

Recently the practice of software peer review has expanded to include the review and development of standards and guidelines for *statistical* software, as well as Spanish-language submission and review. Currently, the domain is expanding further into developing infrastructure for individuals and organizations to curate and share their own R packages and documentation. It remains to be seen how these recent expansions will influence community membership and how their contributions will influence the practice.

Challenges and Opportunities

A big strength of rOpenSci has been anticipating and adapting to the needs of our community to sustain member participation. If CS&S were to support CSID communities to build out CoPs, it would be important to help efforts, especially ones that have become established, be responsive to changing dynamics across the ecosystem and shift resources and priorities accordingly.

One of rOpenSci's biggest challenges, outside of financial sustainability, is that many members are not clearly able to define the mission. Part of this struggle is that many still perceive rOpenSci as a producer of tools and are not aware of the newer services. This issue is particularly acute among experts, who no longer need our tools and find it challenging to engage with the project beyond serving as a well-wisher. For some, peer review is not as rewarding or too time-consuming. This ties into the community aspect of the CoP, where programming needs to address every skill level. Specific programming may also be needed to support diverse participation, including from newcomers and/or members of historically under-represented groups and this need may become more obvious as a community grows.

Conclusion

While research software creation is easier than it was a decade ago, there are challenges in of the research software lifecycle such as evaluation, sustainability, and discovery. These areas are underdeveloped compared to equivalent products in research papers or commercial software. rOpenSci's model - a community of researcher-developers creating federated software, via mutual support and a core of technical, social, and educational infrastructure - is well suited to address these challenges and scale to other communities and products.

References

Boettiger, C., and S. Chamberlain. n.d. "Building Software, Building Community: Lessons from the rOpenSci Project." *Journal of Atomic and Molecular Physics*. https://doi.org/10.5334/jors.bu/print.

Ram, Karthik, Carl Boettiger, Scott Chamberlain, Noam Ross, Maëlle Salmon, and Stefanie Butland. 2019. "A Community of Practice Around Peer Review for Long-Term Research Software Sustainability." *Computing in Science & Engineering* 21 (2): 59–65.

Salmon, Maëlle, and Karthik Ram. 2021. "The R Developer Community Does Have a Strong Software Engineering Culture." *The R Journal* 13 (2): 673.

Appendix 10: Contexts for Sharing Data in Low- and Middle-Income Countries (LMICs)

By: Gemma Turon and Susan Winks. Summarized by Angela Okune (any errors that result from summarizing are my own). April 2023.

Existing Data Sharing Infrastructures

To facilitate data sharing, a good option is to leverage existing software services, including well-established web servers (e.g., AWS, Google Cloud, Azure) but these typically require a high level of expertise or are locked behind paywalls not affordable to many researchers in resource constrained settings. As a solution, the Grand Challenges Africa Drug Discovery Accelerator Network, supported by the University of Cape Town's H3D, provides a Collaborative Drug Discovery (CDD) Vault for researchers to organize, visualize, and share biological and chemical data, emphasizing the importance of structured databases for effective drug discovery research. Although valuable, reliance on product donations from companies is unsustainable; hence, robust data management plans using open-source technology are essential. Newly emerging programs like the NIH Common Fund's Harnessing Data Science for Health Discovery and Innovation in Africa (DS-I Africa) and the Genomic Research Approach for Diversity and Optimizing Therapeutics (GRADIENT, jointly funded by GSK and Novartis) prioritize data sharing infrastructure, with DS-I Africa using the eLwazi platform for open data sharing and GRADIENT requiring a data management plan. MASHA is also developing an internal data platform that will allow individual data users to confidentiality share their data with MASHA and obtain MASHA's analysis through it as well. For non-consortia researchers, "on-demand" data access model, where scientists can apply to be given access to specific data collections like the GISAID database of genomic data for respiratory viruses, may be a viable data sharing framework. Numerous free databases and search engines, like Google's dataset search engine, have emerged, though most are developed and maintained in High-Income Countries with limited resources focused on Lower Middle Income Countries. Examples of databases maintained in LMICs include INDEPTH, DataSuds, KWTRP, and AHRI.

Open Data Benefits and Limitations

Open Data promotes transparency, accountability, and collaboration by providing free and unrestricted access to data, playing a critical role in advancing CSID modeling. However, it must ensure empowerment and fairness in research, particularly in LMICs, without perpetuating existing scientific knowledge production inequalities. Different organizations adopt varying data sharing approaches, ranging from fully open (DS-I Africa, GRADIENT) to limited due to privacy and IP concerns (drug-discovery consortiums, MASHA modeling group). Overcoming technical barriers to data sharing and interoperability requires a cultural shift for equitable data ownership and retribution, preserving indigenous sovereignty, and giving back to the community where data has been collected. Advancements in Artificial Intelligence, such as privacy-preserving AI, synthetic dataset generation, and blockchain-based Trusted Execution Environments like Nautilus, can mitigate IP-sensitive data usage concerns, offer powerful additions to CSID modeling, and help overcome open data paradigm limitations.

In discussions about accessibility of data, it is also important to note that in current CSID modeling research, there exists a bias towards Malaria, Tuberculosis, and HIV, reflecting a funding and interest feedback loop due to historic resource allocation. This imbalance is beginning to be

addressed by organizations such as the African Research Universities Alliance (ARUA) with Centers of Excellence in 13 thematic areas including climate change, and the Lacuna Fund, which supports dataset collection for LMIC-focused issues. As the field evolves, it is critical to consider different strategies for model and tool development, recognizing the distinct requirements of "data saturated" and "data scarce" contexts.

Open-Source Software

Open-Source Software (OSS), where source code is released under a license that grants users the freedom to view, modify, distribute, and use the software freely, is recognized as important and particularly valuable to CSID researchers, especially those working in low-resource settings. Despite the advantages, OSS development faces several challenges. One of these is the potential decay of software due to the reliance on volunteer developers or grant-funded staff. Once core contributors leave or funding runs out, maintenance can suffer. In fact, about 41% of the code cited in scientific articles up to 2012 is no longer accessible,[6] demonstrating this concern.

Sharing software and models was noted as challenging by several interviewees, with reasons including the lack of standards to enable interoperability of these in new operating environments; and a lack of knowledge on best practices for development that can support reuse by others and reproducibility of research outcomes. Many researchers have the skills to do a small amount of software development, but their approaches are often ad hoc, reducing their ability to share with colleagues in a manner that is useful to others. Documentation also came up as a challenge, as it needs to be usable by stakeholders with a wide range of backgrounds, from clinicians to policy makers to researchers from other disciplines. One participant noted the need for infrastructure that enables research software and model developers to understand what already exists, to avoid repeated duplication of effort in a fragmented landscape, which projects like Epiverse TRACE are seeking to address.

Another issue that surfaced was that granting access to source code does not necessarily mean that models are usable by the broader community. The interpretation of model outcomes requires a nuanced understanding of the limitations of the data and the context of the model's development. While model repositories like Hugging Face and DL Hub offer free access to model predictions, their application in CSID modeling raises a key question: are field scientists, policymakers, and funders equipped to interpret and adapt these model outcomes to their use cases?

Different strategies have been adopted to address these issues. MASHA, for instance, provides ready-to-analyze outputs and policy recommendations rather than direct access to their models, ensuring that data is interpretable and useful for decision-makers. In comparison, the Ersilia Open Source Initiative offers free access to in-house developed and third-party OSS AI models for drug discovery in infectious diseases, coupled with user training, acknowledging the lack of AI expertise in LMICs and the risks of using AI without appropriate guidance.

In sum, effective CSID modeling in LMICs requires more than robust data sharing frameworks. It also needs sustainable funding, support for community building, and accessible frameworks for modelers to share their work with end-users and decision-makers.

---

[6] Mangul, Serghei, Thiago Mosqueiro, Richard J. Abdill, Dat Duong, Keith Mitchell, Varuni Sarwal, Brian Hill, et al. 2019. “Challenges and Recommendations to Improve the Installability and Archival Stability of Omics Computational Tools.” *PLOS Biology* 17 (6): e3000333. https://doi.org/10.1371/journal.pbio.3000333.

Appendix 11: Governance Structures to Enable Work at Different Scales

A myriad of governance structures, from highly centralized to largely decentralized, are currently utilized by various communities of practice to enable work at different scales. Each has different strengths and weaknesses, and the best choice often depends on the specific needs, characteristics, and goals of the group.

Top-Down: The most centralized model, frequently found in highly structured government and international agencies as well as in early-stage software projects where it is often referred to as the “Benevolent Dictator For Life” (BDFL) model, is where decision-making authority rests with a single person or a small group of individuals at the top. The leaders make decisions, and the rest of the organization follows. This model can lead to a lack of buy-in or engagement from those lower down in the hierarchy. Notable examples include the U.S. Centers for Disease Control and Prevention (CDC), the Pan American Health Organization (PAHO), and the Python Software Foundation operated this way until 2018 when the lead stepped down and the foundation adopted an elected board model.

Board-Led: This governance model involves a board of directors who oversee the organization's operations and make strategic decisions. Day-to-day management is often delegated to executive staff. This can be efficient and bring a range of perspectives to decision-making but can also risk disconnecting leadership from the broader community or group. Some examples of groups leveraging this model include The International Research Institute for Climate and Society (IRI) under the Earth Institute at Columbia University, which is governed by a board of directors and Climate and Health Alliance (CAHA), an Australian organization whose board of directors set the strategic direction.

Federated: A federated governance model, typically associated with decentralization, allows multiple entities to cooperate while preserving their distinct identities and autonomy. Communities like CGIAR, the Research Data Alliance (RDA), and H3ABioNet employ this model, achieving a balance between local autonomy and global interconnectedness. This encourages diverse participation, while fostering a collective identity and shared purpose. Although generally applied to communities comprising pre-existing organizations, a federated model can also work in a primarily individual-based community through autonomous working groups or committees, focusing on shared interests. The core principles remain constant: autonomy with affiliation, shared resources, consensus-based decision-making, and an equilibrium between centralized coordination and decentralized execution.