

Epiverse-TRACE Summit Report

Collaborative Software Development Ecosystem for Public Health Bogotá, Colombia, June 26 - July 7, 2023

Institutions involved in the Summit

Universidad de los Andes Pontificia Universidad Javeriana The London School of Hygiene & Tropical Medicine The Medical Research Council Unit The Gambia data.org











The views expressed herein do not, however, represent those of IDRC or its Board of Governors. This work was carried out thanks to a grant from the International Development Research Center (IDRC), Ottawa, Canada. Universidad de los Andes | Vigilada Mineducación | Reconocimiento como Universidad: Decreto 1297 del 30 de mayo de 1964. Reconocimiento personería jurídica: Resolución 28 del 23 de febrero de 1949 Minjusticia.

Funded by:





EPIVERSE TRACE-LAC

Principal Investigators

Catalina González-Uribe Juan M. Cordovéz Zulma M. Cucunubá **Scientific Coordination**

Natalia Niño-Machado Diana Fajardo Pulido

SUMMIT 2023 REPORT

Editorial Coordination | Miller Díaz-Valderrama | Natalia Niño-Machado| Diana Fajardo PulidoPhotographs | Laura Franco | David Gallo

Editorial design | Denís Toledo-Nieto

Graphic recording | Melissa Ávila Posada

Infographic | Maria Paula Forero

Support team

Mauricio Cortés-Cely Mabel Chavarro Hawyn Díaz Denís Toledo-Nieto Carol Buitrago Chávez

Contact information

trace@uniandes.edu.co tracelac@javeriana.edu.co https://medicina.javeriana.edu.co/trace-lac https://data.org/initiatives/epiverse/

* TRACE-LAC is a collaborative project co-led by Pontificia Universidad Javeriana and Universidad de Los Andes in Colombia for the region. Its aim is to enhance the response to infectious disease epidemics in Latin America and the Caribbean by providing robust data tools to inform decisionmaking effectively. TRACE-LAC is linked with the Epiverse-TRACE initiative of data.org, which is led by the London School of Hygiene and Tropical Medicine (LSHTM) in the United Kingdom, in collaboration with the Medical Research Council - The Gambia.

This work is part of the TRACE-LAC research project funded by the International Research Centre (IDRC) Ottawa, Canada. [109848-001-]. The funders had no role in study design, data collection and analysis, decision to publish or preparation of manuscripts, reports or working papers. The views expressed herein do not, however, represent those of IDRC or its Board of Governors.











Summit Participants

TRACE-LAC Team (Universidad de los Andes and Pontificia Universidad Javeriana)

Zulma M. Cucunubá (PI) Catalina González-Uribe (PI) Juan Manuel Cordovez (PI) Natalia Niño-Machado Diana Fajardo Pulido Hawyn Diaz Miller Diaz-Valderrama Mauricio Santos-Vega Juan D. Umaña Julian Otero Maria Camila Tavera-Cifuentes Juan Montenegro-Torres

José Bocanegra Andrés Moreno David Quevedo Geraldine Gómez Jaime Pavlich Nicolás Torres Santiago Loaiza Adriana Buitrago José M. Velasco Laura Gómez Magda C. Cepeda

TRACE Team (LSHTM and MRC Unit Team)

Adam Kucharski (PI) Rosalind Eggo (PI) Sebastian Funk (PI) Bubacarr Bah (PI) James Azam Bankole Ahadzie Anna Carnegie Abdoelnaser Degoot Pratik Gupte Joshua Lambert Karim Mané Nuredin Mohammed Carmen Tamayo Cuartero Andree Valle Campos Edwin van Leeuwen

Data.org members

Claudine Lim

Hugo Gruson









- Assumptions of packages Stechnical "As developer I want to Knew about users: - Advantages to use packages for users. Plachicality/usability Want to Know about developers:

icability = cases in which it won 4 work ication ning to developers or users seedback - coers (















Table of Contents

Introduction	. 8
 Software development for emergency preparedness as a global endeavor: issues for discussion	10
2. Building Teams and People's Careers for Scientific Software Development	13
3. Building a community of users: scale and context	20
Communities of users are organized as collaborative and adaptive groups	23
References2	24
SUMMIT GRAPHIC REPORT	25
TRACE: Protocol key elements2	25
Day 1. Collaborative software development in context	26
Day 2. Software development2	28
Day 3. Building RSE & Data Science Teams2	29
Day 4. Building a community of users	31



Introduction

Over the last two years, a group of researchers interested in data science tools for public health and epidemiology have been working together on the research project "Tools for Response, Analytics and Control of Epidemics – TRACE". This initiative aims to change how analytics are used in the global infectious disease response, moving from inflexible tools and ad-hoc collaboration to integrated, generalizable, and scalable communitydriven software. TRACE focuses on building a high-quality, open-source, and transparent analytics toolkit to inform decision-makers in response to epidemics and grow an engaged users community. With the support from The Rockefeller Foundation and Wellcome, the project began at the London School of Hygiene & Tropical Medicine (LSHTM) in the United Kingdom under the leadership of Adam Kucharski, Rosalind Eggo, and Sebastian Funk and in collaboration with the Medical Research Council Unit in the Gambia with Bubacarr Bah. In 2022, with the support of the Canadian International Development Research Centre- IDRC, the project was extended to Latin America to establish TRACE-LAC, led by Zulma Cucunuba at Pontificia Universidad Javeriana, Catalina González-Uribe and Juan Manuel Cordovez at Universidad de los Andes in Colombia.

In addition, TRACE is part of the data.org initiative "Epiverse". This global collaborative initiative aims to develop a trustworthy data analysis ecosystem dedicated to getting ahead of the next public health crisis. Epiverse convenes a global consortium to design and build open software tools enabling privacy-preserving distributed analysis of data to power pandemic response. (https://data.org/initiatives/epiverse/).

Between the 26th of June and the 7th of July of 2023, 44 members of Epiverse-TRACE got together in Bogota (Colombia) to share their experiences and strengthen collaboration to develop an ecosystem of software for public health. As an extension of the summit, between the 30th of June and the 7th of July, the "TRACE-ing R Issues" session took place to strengthen package development. During this week, software development teams met in person to address issues in co-design, planning of new functions for future integration, and, most importantly, enhancing collaboration among developers within

Epiverse. This report consolidates the main discussions and conclusions of both meetings. It is organized around three main topics of discussion: (1) collaborative software development (2) assembling Research Software Engineers (RSE) and Data Science teams and (3) building communities of users. Each topic had a key aim and a set of questions orienting the summit activities.

Epiverse

TRACELAC



Infographic: Main topics of Summit 2023



Epiverse TRACE and TRACE-LAC team in Summit 2023, Bogotá -COL

Collaborative Software Development Ecosystem for Public Health

SUMMIT 2023 - COLOMBIA Bogotá, 26th -29th of June 9

Universidad de los Andes

JAVERIANA



Universidad de los Andes

1. Collaborative software development for emergency preparedness as a global endeavor: issues for discussion

Aim and activities

Aim: To achieve a set of agreements on good practices in collaborative software development.

During the event software development teams socialized their needs and approaches for package development, they also discussed topics on quality, transparency, and interoperability within the tools and activities that Epiverse-TRACE leads in the context of epidemic preparedness.

Main questions

- How to ensure the quality of the tools and validation of the methods?
- How to communicate to the users the risks, uncertainty, assumptions, and biases within the tool's design?
- How to promote reproducibility, integration, and cooperation among multiple scales, actors, and tools in co-design processes?
- Which challenges can be identified in the implementation of the Epiverse-TRACE pipeline in your context (ie. user community, country, team)?
- How to build long-lasting, adaptable, and evolving tools in the context of changing problems and technologies?
- How to build data science teams for epidemics analysis?

Methodologies applied: World Cafe¹ and discussion groups.

¹ On the world café as format for discussion groups consult: <u>World Cafe Method | The World Cafe</u>







Producing public health data tools with global reach - such as those we are developing in the Epiverse-TRACE initiative - requires agreements on important aspects of software development such as quality, interoperability, and sustainability. In addition, software development for emergency preparedness requires focusing on users' needs when discussing each aspect mentioned.

Being these considerations at the heart of the <u>values</u> and <u>blueprints</u> of the Epiverse-TRACE community, some experiences and reflections were shared to reach the following conclusions:

QUALITY is a difficult feature to define and to track. It involves user feedback, technical maintenance, and reproducibility of the methods. Measures to assess the quality of the packages require further development, and for this purpose, an expert committee could be helpful, cross-validation of methods among data scientists could be a strategy for quality checks. Also, code reviews need to be increased and improved by implementing internal reviews on the documentation and training materials to avoid biases.

Although internal validation could be useful for assessing the quality of the Epiverse-TRACE tools, external reviews should be taken into account, including their different conceptions of quality. **Users and developers may have different understandings of what** *quality* **means** and they all must be considered. Epiverse-TRACE packages should be creative about **explaining the usages of their functions but also the misusages, biases, and assumptions**. This includes more warnings on the input data and the analysis outputs, training material for different situations/cases, and the use of explicit and clear language regarding identified risks, usages, and limitations. INTEROPERABILITY must be pursued to share information between different levels, organizations, and systems. Users (individuals and organizations) are at the core of interoperability, where flexibility and trust become important challenges to consider.

Epiverse

TRACELAC

The development of context-specific tools might lead to very specific needs, capacities, and limitations. Conversely, the production of "universal" packages might lead to decontextualized developments in which users' needs and local infrastructure are disregarded. Both broad and narrow tools may fail. It is essential then, to remember that the scale of the context does shape the pertinence of the analyses and that the goal of developing epidemiological tools includes different challenges in their implementation for either local or global spheres.

Translation is needed among different scales, actors, and contexts. Multiple versions of the same package (ie. console-based version, interface/app-based version, online version, etc.) can be helpful for the translation of the methods into decision-making users and different communities of practice. Aspects such as how users visualize and interpret data, or how they understand it within an organization must be considered.

Co-creation is an essential dimension of translation as a tool for understanding users' needs and producing relevant tools. Open source tools promote feedback and collaboration from users, however, it is insufficient to consider that a package is co-designed with the users if a need-based development is not implemented, including social sciences approaches. In addition, reproducibility is not limited to code availability: it includes data governance, structure, and usage in different settings and the standardization of practices for data analysis.

• The **ENGAGEMENT** of stakeholders and the **SUSTAINABILITY** of the packages developed by the Epiverse-TRACE initiative are two sides of the same coin. Making long-lasting, adaptable, and evolving tools depends on knowledge sharing and capacity building. On one hand, a closer relationship with stakeholders and users can improve the ability of Epiverse-TRACE to shape its development products and anticipate novel features needed in its tools to address emerging needs.

On the other hand, technical sustainability is closely related to the development of human talent. Attracting and retaining talent should be a priority for the initiative with strategies such as training opportunities, career development paths,





interdisciplinary learning, partnership with stakeholders, and long-term visibility. It is urgent to develop an onboarding process so new members will have a better understanding of the development processes, the structure of the teams, the expertise required, the standards of quality and the collaborative work that Epiverse-TRACE is working on.

Recommendations and some work for the future

- Epiverse-TRACE will adopt responsible practices to design correct usages of the packages including cases of use, misusage warnings, and compliance standards.
- Usability barriers will be documented, and communicated within Epiverse-TRACE teams so awareness of implementation challenges and capacity building can be identified and prioritized.
- A platform for in-house and external training on Epiverse-TRACE R packages is needed.
- Needs identification and co-design processes require trust building between technical and social science teams, and stakeholders.









Universidad de Ios Andes Colombia

2. Building Teams and People's Careers for Scientific Software Development

Aim and activities

Aim: To provide insights and recommendations for organizations and leaders involved in building and managing inclusive and diverse RSE and Data Science teams.

During the event, we considered how Epiverse-TRACE has built its teams, what skills are needed, what kind of expertise will answer our research questions in software development and training, and how these teams can be inclusive and diverse. This information contributed to the development of effective strategies and practices to foster collaboration, diversity, and professional growth within these teams.

Main questions

- What roles does each team have?
- What roles are missing in the teams and why?
- Is it necessary to think of new roles for scientific software development (SSD) within the teams?
- How do you think the Epiverse-TRACE initiative has impacted your career development?
- What new challenges (as opportunities), skills, and experiences do you think Epiverse-TRACE can offer you to boost your career development further?

Methodologies applied: designing roles, mapping collaborations, and discussion groups.







Collaborative software development involves more than coding, it involves thinking about the people involved in this process, their needs, their career development expectations, and the creation of fertile environments for interdisciplinarity. Thus, defining roles and mechanisms for interdisciplinary collaboration, building opportunities and for people's professional growth are key for RSE and data science teams working in academic and researchoriented settings targeting social impact.

 ROLES are a good point of departure for this conversation since members of the Epiverse-TRACE can perform more than one. The diversity and particularity of contexts lead to different definitions of roles and responsibilities within the teams. Additionally, roles are perceived as flexible activities that different members of the teams adopt to conduct a user-centered software approach. Therefore, defining roles and their interaction is an ongoing and still unfinished need, depending on each team's challenges.

For scientific software development (SSD) teams, the search for the "unicorn" -that exceptional person with all the skills needed, eg. programming, domain expertise, data management, and leadership- is as rare as in any software development project. It could be even harder to find, as the conjunction of domain expertise with different data science and programming-related skills demands a significant investment of time in training for a single person. Instead of trying to find the needle in the haystack, the potential of SSD relies on interdisciplinary collaboration and cross-fertilization.

A variety of roles sustain SSD as a feasible enterprise in which not only code-related roles or those focused on particular domain expertise are



Figure 1. Roles and gender distribution

Collaborative Software Development Ecosystem for Public Health SUMMIT 2023 - COLOMBIA Bogotá, 26th -29th of June 15

necessary. Other expertise oriented to fulfill scientific research requirements, knowing users' needs, understanding the socio-technical systems in which tools are developed, translation, communication, and networking appear as useful for SSD. Thus, **Data Science must be reimagined as a much wider field that includes all the people, roles, and expertise needed to build an innovative ecosystem.**

Epiverse

TRACELAC

Capacity building is therefore necessary for harvesting successful collaborative work, establishing diverse teams (some work must be done for promoting more gender-balanced teams in the performance of different roles *-see figure 1-*), and boosting the professional careers of those engaged in SSD.

SKILLS can be developed through knowledge sharing among the members of Epiverse-TRACE who come from different backgrounds and disciplines. It is noticed that interdisciplinarity is one of the greatest strengths of the initiative and one of the main motivations of the TRACE members to remain in the project. The TRACE members from LAC, UK, and the Gambia teams see their projects as a way to learn new skills, transform their learning paths, and a place for career development.

There are different CHALLENGES and **OPPORTUNITIES** for people's growth for the Epiverse-TRACE teams. For example, communication between professionals from different backgrounds has been demanding. However, members perceive multidisciplinarity as way а of transforming people's ideas and acquiring new skills and questions. An example of this is the incorporation of a gender perspective within SSD that some members found as a new issue to look at and a new skill to be developed (see figure 2).



Figure 2. Skills required or to be strengthen

Regarding interdisciplinarity, the transition between domain expertise and RSE is seen as difficult. Learning "soft skills" (such as *Leadership*, *Communication*, and *Multidisciplinary work*) is identified as vital for improving collaborative work. At the same time, training in statistics quality analysis, data ethics, and domain expertise in Public Health, epidemiology, and disease ecology is seen as an ongoing need.

Universidad de los Andes

IAVERIANA

Despite language barriers, the encounter of cultural differences that promotes a global initiative that meets heterogeneous contexts constitutes an exciting opportunity for people's growth and learning. In this sense, Epiverse-TRACE is expected to build stronger collaborations at national and international levels and find new academic opportunities and research orientation (mentoring being highlighted by Principal Investigators (PIs) as an important way of boosting career development).

Epiverse

TRACELAC



 MAPPING COLLABORATIONS The graph for Epiverse-TRACE showcases collaboration within different teams, revealing a complex network of interactions at the heart of collective efforts. This mapping initiative aims to address the question: How do we work together? It does so by visually depicting team connections and exploring diverse methodologies. Through ego-sociographs mapping, the teams could comprehensively explore the dynamic landscape shaped by roles within inclusive and diverse teams dedicated to SSD.

The map offers a dynamic snapshot, uncovering the interplay of workflows crucial for the overall collaboration networks in scientific software development within Epiverse-TRACE. This visual representation (see figure 3) The map offers a dynamic snapshot, uncovering the interplay of workflows crucial for the overall collaboration networks in scientific software development within Epiverse-TRACE. This visual representation (see figure 3) illuminates the central role individuals, who simultaneously fulfill diverse roles, play within each team. Consequently, it underscores the significance of diversity and interdisciplinarity in fostering collaboration and empowering team members within Scientific Software Development teams².



² A publication on roles and skills in RSE and data science teams is currently under construction, for additional information please get in touch with the TRACE-LAC research team.



Figure 3. The Epiverse-TRACE as a collaboration network

FUTURE is a special concern for different members of the initiative. Staff retention and a list
of specialized professionals that need to be built up are a concern. If unicorns are hard to find,
building strong teams for SSD is a long-term process, and engaging the human talent
constitutes a challenge in itself, for initiatives such as the Epiverse-TRACE. How can
people be engaged in SSD for a longer period in such conditions?.

For those who work in SSD **the social impact of their work, remains as the most prominent expectation for working in the Epiverse-TRACE**. The sustainability and funding of the initiative (increasing its lifetime), then become fundamental pillars for capacity building.





Recommendations and some work for the future

- Social science, communication and translation roles (those related to improving the interaction between developers, users, and other stakeholders) are seen as lacking and necessary by most of the teams.
- As a space for diverse interaction, the Epiverse-TRACE could enhance the learning and development of those skills needed for better -responsible and fluid- interaction among team members and users.
- The task of chasing unicorns is fruitless, as its very enunciation suggests. Instead, building diverse, interdisciplinary and inclusive teams is essential to the success of SSD.
- Face-to-face activities in which different members of the Epiverse-TRACE (from different places and backgrounds) can encourage and influence each other's thinking is essential to the initiative.









Universidad de Ios Andes Colombia

3. Building a community of users: scale and context

20

Aim and activities

Aim: To generate a comprehensive understanding of the Epiverse-TRACE community of users, laying the groundwork for building strong connections, overcoming obstacles, and collectively shaping a vibrant and inclusive community on a global scale.

Each team examined how to integrate stakeholders into a comprehensive, worldwide community of Epiverse-TRACE users. The main topics were creating robust links, tackling impediments, and cooperatively forming an energetic, inclusive community. Additionally, the teams reviewed the evaluation framework.

Main questions

- 1. How would you define the Epiverse-TRACE community (situate yourself in the work you do specifically or in the general work of TRACE)? What characteristics do you imagine that community should have?
- 2. What are the characteristics of current or potential users of the Epiverse-TRACE tools (e.g. packages, training material, qualitative findings, etc)?
- 3. Is there any consideration in terms of gender, ethnicity, and/or geographical background that you consider relevant when thinking about the concept of "community" for the TRACE project?
- 4. Identify stakeholders (key actors and organizations) that are or will be relevant to your work. Who or which organizations should be/or are already part of the Epiverse-TRACE community?

Methodologies applied: mapping stakeholders, discussion groups, and presentations.





Building a community of users is a different challenge for each of the TRACE teams since their definition of stakeholders covers a wide range of types of actors. Some of the UK members, for example, were interested in actors that operate at national levels or at the global scale (e.g. international organizations).

The TRACE-LAC members, on the other hand, highlighted as relevant actors located in specific cities and municipalities.



Figure 4. Impact scale of organizations with global reach and those with impact on specific territories



For TRACE-LAC, local actors are those who work with data from municipalities and towns. **The connection between scales needs to be characterized (fig 4)**. The potential of the Epiverse-TRACE relies on the wide relations that can be established between organizations with global reach and those with impact on specific territories, thanks to the strategic and diverse placement of its members.

USERS are heterogeneous. Two types of users have been identified, those who already acquired the necessary skills to use the Epiverse-TRACE tools and those who need training.

One particular user niche for the TRACE initiative is located in academia: scholars and people with technical knowledge. However, decision-makers at different levels should be included. This could be challenging, considering language barriers, different problems with infrastructure, and available resources at more local scales like those that face the TRACE-LAC project. Thus, even the needs of "non-users"³ (Oudshoorn & Pinch, 2003; Wyatt, 2014) should be included in the development process, by offering alternative ways for interacting with the Epiverse-TRACE tools and concerns.

Epiverse

TRACELAC

The list of actual and potential users and stakeholders involves various organizations and groups. More work needs to be done to include actors such as industries, MOOCs providers, Cloud services providers, funding organizations, different universities, Non-Governmental Organizations (NGOs), Independent researchers, and especially **Climate and environmental organizations that could be a good (and indirect) niche interested in the development, sustainability and improvement of the Epiverse-TRACE tools.**

For this reason, diverse expertise and different levels of it are necessary to be considered: different skills configure different users of the training materials and the libraries. **Strategies such as defining profiles of archetypical learners or mapping actors and stakeholders are found useful tools for characterizing actual and potential users**. However, characterization must provide an initial approach to users that must be sensitive to how they change; the risk of over-characterization is just around the corner.

Communities of users are organized as collaborative and adaptive groups

It is not possible to foresee how they will interact in the future and scenarios for their active participation must be facilitated. This could be a challenge if context is not considered (for this reason, characterization is vital): tools must face and overcome language barriers (translation to different languages), and accessibility issues (think about offline use of the tools) and it is necessary to include different people in their testing processes.



Universidad de los Andes

JAVERIANA

³ All those for whom the use or learning of programming languages represents a barrier, but who can benefit from user-friendly versions of the tools, mediated by interfaces.





Recommendations and some work for the future

Building **TRUST** to establish **COLLABORATIVE RELATIONSHIPS** with identified organizations is a key element for future efforts. Building trust also includes understanding potential fears and concerns public health organizations might have regarding data science. Some of these concerns might be related to how to guarantee an ethical and responsible use of data, potential misinterpretation of models and their results in the public arena, fear of including new methodologies and sources of evidence in organizational contexts with a long tradition of using their procedures and analysis, and concerns on technology replacing activities done by humans.

Users' needs should be at the heart of SSD. Continuous interaction and communication with them is essential to achieve the common goal of *enhancing Tools for the Response, Analytics, and Control of epidemics* and for building. sustainability.



Collaborative Software Development Ecosystem for Public Health

SUMMIT 2023 - COLOMBIA Bogotá, 26th -29th of June References

Oudshoorn, N., & Pinch, T. (2003). How Users and Non-Users Matter. The Co-Construction of Users and Technology. Cambridge, MA.

Wyatt, S. (2014, April). Bringing users and non-users into being across methods and disciplines. In Refusing, Limiting, Departing, In: CHI 2014 Workshop Considering Why We Should Study Technology Non-use, Toronto. http://nonuse.jedbrubaker. com/wpcontent/uploads/2014/03/Wyatt_Toronto_April_2014. pdf (last accessed 03/07/2015)





Universidad de [A] los Andes











25

SUMMIT GRAPHIC REPORT

TRACE: Protocol key elements







Day 1. Collaborative software development in context

Day 1 aimed to reflect on the human resources and infrastructure characteristics that the team in the UK, the Gambia, and Colombia





27







Day 2. Software development

Day two aimed to provide a progress update of the TRACE pillars, a socialization of the needs and approaches that packages are taking into account in their development, and a discussion on quality, transparency, and interoperability within the tools and activities that Epiverse-TRACE leads in the context of epidemic preparedness.





Universidad de Ios Andes Colombia

Day 3. Building RSE & Data Science Teams

Day three aimed to provide insights and recommendations for organizations and leaders involved in building and managing inclusive and diverse RSE and Data Science teams. This information can contribute to the development of effective strategies and practices that foster collaboration, diversity, and professional growth within these teams.





Collaborative Software Development Ecosystem for Public Health Bogotá, 26th

SUMMIT 2023 - COLOMBIA Bogotá, 26th -29th of June





Day 4. Building a community of users

Day four aimed to generate a comprehensive understanding of the Epiverse-TRACE community of users, laying the groundwork for building strong connections, overcoming obstacles, and collectively shaping a vibrant and inclusive community.









ENHANCING TOOLS FOR RESPONSE, ANALYTICS AND CONTROL OF EPIDEMICS IN LATIN AMERICA AND THE CARIBBEAN.



Funded by:









The views expressed herein do not, however, represent those of IDRC or its Board of Governors. This work was carried out thanks to a grant from the International Development Research Center (IDRC), Ottawa, Canada. Universidad de los Andes | Vigilada Mineducación | Reconocimiento como Universidad: Decreto 1297 del 30 de mayo de 1964. Reconocimiento personería jurídica: Resolución 28 del 23 de febrero de 1949 Minjusticia.