<R eSA>
Research Software Alliance

# The Research Software Community Landscape in the Global South

**Author: Paula Andrea Martinez**

**DOI: 10.5281/zenodo.7179892**

**2022, October 10**

Please city this work as:

Martinez, Paula Andrea. (2022, October 10). The Research Software Community Landscape in the Global South. Zenodo. https://doi.org/10.5281/zenodo.7179892

Martinez, Paula Andrea. (2022). Research Software Communities Global South [Data set]. Zenodo. https://doi.org/10.5281/zenodo.7179807

Martinez, Paula Andrea. (2022). Research Software Funders Global South [Data set]. Zenodo. https://doi.org/10.5281/zenodo.7179867

# Introduction

The Research Software Alliance's (ReSA) mission is to bring research software communities together to collaborate on the advancement of research software. Given the ReSA mission, it is important to understand the landscape of communities involved with research software. In 2020, ReSA completed an initial exercise to scope the international research software community landscape. This work was reported by ReSA's Software Landscape Analysis task force via a blog post. The majority of the communities in the previous analysis represented the global north. To improve the extent of this landscape analysis, ReSA announced a paid opportunity for short-term contractors located in the global south to collect data on communities and funders in their region in early 2022. This document describes how the work was undertaken, a summary of findings, the gaps and opportunities perceived by the data collectors and some highlights. This work identified 126 organisations and communities and 62 funder bodies that support research software in the global south. Their main activities are connecting people, training, and networking, and support through research grants.

# Background

## *Regions*

The landscape analysis of research software communities in the global south addressed three regions: Africa, South America, and Asia. A total of twenty-two countries were reported on by the data collectors:

- Africa (9): Benin, Botswana, Cameroon, Democratic Republic of the Congo (DRC), Ghana, Kenya, Namibia, Nigeria, South Africa
- South America (4): Argentina, Bolivia, Brazil, Colombia
- Asia (9): Bangladesh, Bhutan, China, India, Indonesia, Nepal, Pakistan, Saudi Arabia, Sri Lanka

## *Data Collectors*

This new work was completed thanks to twelve consultants who conducted the investigation process and collected data from their own regions:

- Africa (4): Anthony Oko-Isu, Jessica Upani, Caleb Kibet, Narcisse Mbunzama
- Asia (5): Abdullah Shams Bin Tariq, Saranjeet Kaur Bhogal, Jyoti Bhogal, Arslan Sheikh, Batool Almarzouq
- South America (4): Nicolás Wolovick, Paula Andrea Martinez[1], Renato Augusto Corrêa dos Santos, Stephannie Jimenez Gacha

[1] Paula Martinez was not a casual contractor, but works for ReSA as in kind support from ARDC

## *How was this work undertaken?*

ReSA contracted twelve consultants from the global south to provide details of research software initiatives, events, funders, and a short contextual narrative from their region. Each of these data collectors received the same task, to provide information from their region divided into four categories 1) Communities, 2) Events, 3) Funders, and 4) Contextual narrative. Each data collector provided information about one or more countries, depending on their expertise. The investigation work was carried out in January and February 2022 and summarised in this report, which accompanies two dataset collections.

To differentiate from the first landscape analysis in 2020, the following two new collections have a tag name "V2". The communities listed by each data collector were all collated and curated in the Research Software Communities Global South spreadsheet. The funders listed were all collated and curated in the Research Software Funders Global South spreadsheet. It was determined not to report the identified events, as these were few and often past one-off events. A contextual narrative accompanied each region, which included personal views. These are summarised in three sections of this report: Gaps, Opportunities and Highlights.

# Summary of findings

## *Narratives*

The data collectors' contextual narratives agree in the following statements. Most, if not all, agreed with the need for local communities in their region, whether they existed or not. Many of them have also coincided with the view that since artificial intelligence and data science have become critical international topics, coding is generally agreed to be an essential skill. Another statement in accord with many data collectors is that Open Science communities typically involve researchers and introduce best practices in research, including best practices in coding and developing more efficient research software. A few of these communities actively participate in training, including topics of research software production, sharing, versioning and documenting.

## *Research Software Communities Global South*

- 126 regional initiatives, communities, training providers including 8 global
- For example, communities relevant in multiple countries include:
    - Google developers groups
    - R-Ladies groups
    - Grassroots open science communities
    - The Carpentries subcommunities
    - Accredited training providers
    - Local/institutional communities

## *Research Software Funders Global South*

- 62 funders in the global south, including 16 global

The global south funders curated list can be classified in three areas:

- US/ Europe-based initiatives supporting countries in the global south or globally
- Federal or other government funding
- Philanthropic funds

Since the concept of Research Software Engineers (RSEs) has not yet developed in many countries within the global south, the funding landscape seems narrow. The main funding source supporting Research Software's sustainability are US/Europe-based initiatives, such as Python Software Foundation, The Wellcome Trust and Chan Zuckerberg Initiative (CZI).

Federal or other government funding seems to be the second major source of funding in all investigated countries in this report. This includes Argentina (e.g. Ministry of Science and Technology), Botswana (e.g The Botswana Innovation Fund established by the government of Botswana). Brazil (e.g. National Council for Scientific and Technological), China (e.g. National Natural Science Foundation of China, Ministry of Science and Technology of the People's Republic of China), Colombia (e.g. Ministerio de Tecnologías de la información y de las comunicaciones, Ministerio de ciencia), Indonesia (e.g. The Indonesian Science Fund), Japan (e.g. Japan Science and Technology Agency), Kenya (e.g. Kenya Education Network Trust), Nepal (e.g. Nepal Academy of Science and Technology), Nigeria (e.g. National Information Technology Development Agency, Federal Government TETFund research grant), Pakistan (e.g. Ignite National Technology Fund), South Africa (e.g. National Research Foundation South Africa), Sri Lanka (e.g. National Research Council of Sri Lanka, National Science Foundation), Saudi Arabia (e.g. King Abdulaziz City for Science and Technology). However, none of these is explicitly directed towards research software but generally oriented towards research and technology. This reflects the attention paid to the role of research software to support science and deliver research outcomes.

There are also successful programmes based in the US and Europe, but directed to local communities in the global south, including ‘USAID Awards Six Research Grants’ to strengthen US - Bhutan research partnerships to support research and capacity-building. Other programs like the one in Nigeria include Canada’s International Development Research Centre (IDRC) and Swedish International Development Agency (SIDA). IDRC and SIDA have jointly awarded a research grant to the University of Lagos as part of the four-year Artificial Intelligence for Development in Africa (AI4D) programme. This programme is dedicated to accelerating the realisation of Africa’s Artificial Intelligence innovations toward the achievement of continental and global developmental goals. Also, the UK-Nigeria Tech Hub designed to reduce youth unemployment by providing a bootcamp and access to educational resources. There is an excellent opportunity for similar programs directed toward middle-to-low-income countries

explicitly oriented towards research software. It also seems that limited foundations use private philanthropic funds, an example is the John Templeton Foundation.

# Gaps

This section is composed from the gaps identified by each country's contextual narrative, which included personal views. The statements are quoted and presented followed by the country name in brackets.

There was a general agreement that the concept of Research Software Engineers (RSEs) is not yet mature in the regions investigated in this report. "The main barrier to engagement is the lack of awareness of the value that RSEs provide to research institutions ... demonstrating this value could develop communities" (Kenya). "There is an evident absence of RSE groups in African Universities" (Kenya), with very few other countries listing examples of institutional RSE communities or related research software communities.

Data collectors mentioned a few times that they were unaware of specific research software communities in their region. In turn, they listed grassroots open science communities and local science government departments as relevant communities that support research software. The lack of specific research software communities reinforces the need to establish more visible and specific communities supporting research software. Subsequently, "there is a need to create and establish active RSE communities to push for new career pathways in the future" (Kenya).

Equally agreed was that research software culture change is needed (Bangladesh, Chad, Colombia, Ghana, DRC, Kenya, Pakistan, and Saudi Arabia). One of the examples of cultural change needs is initiatives to overcome the "lack of economic appraisal of research software, generating challenges in research software sustainability" (Saudi Arabia). Another gap mentioned about the slow maturation of research software communities is that the "typical large-scale data and software-related tasks are service-oriented" (Pakistan), hence commissioned and not ingrained in the budget of a research project.

Regarding funding, the consensus was that there is a lack of specific funding to support the development and maintenance of research software. However, "research software is an intrinsic part of projects that acquire funding through other science-related grants to solve societal issues, such as those in the biomedical sciences" (Kenya). On the other hand, the lack of funding in Argentina is exemplified by the following statement, "in the specific field of High-Performance Computing (HPC), the government does not invest in supercomputers initially, less would they invest in software developed in the country to run on those supercomputers". It is problematic to address concerns about how and when to support research software when some initial premises, such as hardware availability, are not yet in place. Similarly, a hurdle in Bangladesh is a "deficiency in prioritising funding to purchase institutional software licences".

With regard to training, "disciplines that rely on algorithms and computational workflows such as bioinformatics, astronomy or environmental modelling, do not formally introduce the best coding practices into the curriculum in most local institutions" (Saudi Arabia). Hence, "there is a need for training opportunities carried out by the relevant academic institutions to introduce best practices in coding" (Saudi Arabia). In addition to accessibility, the Francophone countries in Africa mentioned language barriers in access to resources, training and funding opportunities (Benin, Cameroon, the DCR). There was also a comment on the need for additional support for localisation (L10n[2]) and internationalisation (i18n) of open source research software and resources to engage non-English speaking communities and researchers (Saudi Arabia).

A final gap not raised from the narratives or identified from the list of communities is that aspects of interest to ReSA, such as preservation, citation, career paths, productivity, and research software sustainability, were not exemplified in this collection.

# Opportunities

This section is composed from a range of developing opportunities listed by the data collectors. This list can be summarised in national strategic policies to grow digital literacy, acquiring funding for software through other types of grants, and drivers for community growth. A few of those are quoted followed by the country name in brackets.

1. Training programs and workshops are organised by groups interested in statistical analysis, data science and open science (Bangladesh, Saudi Arabia, Argentina, India).
2. To get funds for research software, a recommendation is to contemplate a software component in the research project and submit proposals to the Ministry of Science and Technology, Information and Communications Technology (ICT) Division, University Grants Commission (Bangladesh).
3. Research software developed in the region (Congo, DCR, Gabon, Chad) have received international funding from the United Nations Development Programme and the United Nations specialised agency for information and communication technologies. The funded projects are related to digital technologies supporting human rights, health, agriculture, environment and sustainable development.
4. Given the importance of the interaction between the university and industry, efforts exist today to integrate and foster collaboration between them (Brazil).
5. Funding agencies could include software generated in different projects as outcomes to show research impact, and not only the software but also the source code and data (Brazil).
6. MisionTIC2022 is an initiative  for Colombian citizens to extend their opportunities to improve professional and academic development, including computational skills (Colombia). It started in 2022 and has 14,000 training places.

[2] Internationalisation vs Localisation (i18n vs l10n)
https://blog.mozilla.org/l10n/2011/12/14/i18n-vs-l10n-whats-the-diff/

7. Saudi Arabia's Vision 2030 is a strategic framework developed in 2016. It is a package of economic and social policies designed to free the kingdom from dependence on oil exports in pursuit of a knowledge-based economy (Saudi Arabia).
8. Emphasis on artificial intelligence, digital skills, and digital infrastructure are some of the key drivers for growth in areas of research software (Pakistan, Nigeria).
9. There might be an opportunity for international organisations such as the International Council of RSE Associations to leverage experience exchanges across south and north through invited speakers or to encourage collaborations on specific projects. (Not quoted).

# Highlights

This section presents a few selected statements from the narratives and also from the data collections that are worth exemplifying, as these are specific initiatives currently happening in the countries listed. These have been edited and the country is listed at the beginning of the dot point.

- (Various countries agreed on this statement) Data management plans and open science policies are helping raise the value of RSEs.
- (Various countries agreed on this statement) The need and the interest to run active RSE communities exist, and conversations are ongoing among some trailblazing stakeholders.
- Argentina: This generation embraces free/libre open source software (FLOSS), development tools, and web apps and services to make scientific software production more reliable.
- Colombia: The growth drivers for research software include machine learning, artificial intelligence, the Internet of Things (IoT) and general algorithms used across multiple research areas. Researchers and industry have revolutionised how they use software resources to facilitate or enhance their applications in the last five years.
- Kenya: The Pan-African Bioinformatics Network for the Human Heredity and Health in Africa (H3Africa) consortium (H3ABioNet) offers genomics data analysis, software development, and analysis workflows to support the H3Africa projects. H3ABioNet provides a platform for RSEs (though not using the specific title) to be hired and where their contributions are recognised. The founders of Research Software & Systems Engineers of Africa (RSSE Africa) are also part of this network. The H3ABioNet has nodes all over Africa, and their work has catalysed the application of best practices for software development, including FAIR principles[3] and open source.
- Kenya: Adopting best practices in the research life cycle, driven by Research Data Management policies, catalyses the need for RSEs.

[3] The FAIR guiding principles are Findable, Accessible, Interoperable, and Reusable. The acronym and principles were defined by Wilkinson et.al and published in March 2016 in the journal Scientific Data.

- Namibia: The Python Conference (PyCon) is a significant role player in the open source community for both Namibia and South Africa.
- Nigeria: Is the most populous country in Africa, with one of the largest populations of youth in the world who are entrepreneurial driven and eager to develop, learn or use innovative software to advance their career. The drivers of growth in research software development in Nigeria are the youthful population, the recent emphasis on entrepreneurship, and the increasing investment in technology from international funding agencies and the government. The growth of modern technology hubs in learning institutions and co-working spaces in Nigeria has also contributed to research software development and upskilling in using software in research.
- Saudi Arabia: The forum for open research in MENA (FORM) works toward establishing a cross-regional dialogue in the MENA region (the Middle East and North Africa) concerning the development of practical policies and actionable insights for open Science, including research software at an institutional level.
- Saudi Arabia: Mentorship from well-established groups in the global north could be instrumental in growing the local communities in the global south. The Research Software Alliance could be pivotal in highlighting isolated initiatives and linking them to mentorship and funding opportunities.

# What are the learnings?

The sections on Opportunities and Highlights demonstrate that a few initiatives can generate a big impact. For example, national strategic policies (in Colombia and Saudi Arabia) to grow digital literacy are a significant contribution to scholarship and improvement of recognition of research software and its authors.

With the anticipated growth drivers of data science and machine learning cited by multiple data collector's narratives, there is an excellent opportunity to improve capacity building by integrating foundational competencies and programming best practices earlier in the curricula to prepare for the future need for a skilled workforce. On the other hand, there is a growing need for institutions to collaborate closely with researchers who code to articulate needs from discipline-informed perspectives to address the gaps listed in the report.

Also, more awareness about funding opportunities, possibly from past recipients, could be broadly shared to help acquire funding for software through grants that may not be specific to software. Additionally, the narratives concurred that grants rarely addressed development versus maintenance costs as separate items in funding proposals; hence funders could also be more explicit in providing mechanisms that allow proposals to budget for those costs associated with the research software life cycle.

Coincidental to this analysis, the United Nations Educational, Scientific, and Cultural Organisation (UNESCO) recommendation on open science was approved in November 2022 by all countries included in this analysis, which are also member states. This recommendation is intended to embrace the culture and practice of open science, by changing institutional and

national policies to encourage open science, investing in needed technological infrastructure, educating people about open science, and spurring innovation throughout the research process. Member states are due to report back every four years on their progress, and there is interest in the coming changes.

As previously identified, many communities listed have broader interests than just research software; similarly, those presented in this collection support training for digital skills or offer networking opportunities for people interested in research software.

# Conclusion

This analysis shows target audiences and drivers of growth in the research software communities. The gaps, opportunities and highlights sections provide practical ideas for governments and funders to understand and better support the global south research community.

Communities around research software in the global south are growing, which shows a positive recognition of the importance of RSEs. It also reinforces the need to coordinate activities in different regions and be more inclusive internationally. Current collaborations exist to support efforts in the global south. For example, a few global south communities receive support from local and international institutions (e.g. RSSE Africa in partnership with SANBI and Talarify, RSE at SUN in South Africa in partnership with the UK SSI). There are also already some international mentorship initiatives, such as that of the RSE Society in the United Kingdom and the Open Life Sciences (OLS) Program to develop open science initiatives.

This analysis also provides new information on various research software funders. Although not all funders are specifically related to research software, those listed support the broader initiatives and accept research software components in their scope.

To continue to progress this work readers are invited to continue to add or make corrections to the ReSA list of research software communities global south and the research software funders global south. ReSA are also interested to hear from community members who would like to deepen the analysis in their region, consider involvement in other ReSA activities, including task forces, or engage in writing a landscape paper based on further analysis and work.

# Acknowledgements

This work was supported by the Wellcome Trust grant [222436/Z/21/Z].

# CRediT (Contributor Roles Taxonomy)

CRediT (Contributor Roles Taxonomy) is a high-level taxonomy, used to represent the roles typically played by contributors to research outputs.

Paula Andrea Martinez: Methodology, Validation, Data Curation, Writing - Original Draft, Writing - review and editing.
Michelle Barker: Funding acquisition, Conceptualization, Project administration
Investigation: See Data Collectors
Stephannie Jimenez: Data curation, review and editing
Batool Almarzouq: review and editing
Jessica Hardwicke: Project administration