Making forest data fair and open

Data on tropical forests are in high demand. But ground forest measurements are hard to sustain and the people who make them are extremely disadvantaged compared to those who use them. We propose a new approach to forest data that focuses on the needs of data originators, and ensures users and funders contribute properly.

Renato A. F. de Lima, Oliver L. Phillips, Alvaro Duque, J. Sebastian Tello, Stuart J. Davies, Alexandre Adalardo de Oliveira, Sandra Muller, Euridice N. Honorio Coronado, Emilio Vilanova, Aida Cuni-Sanchez, Timothy R. Baker, Casey M. Ryan, Agustina Malizia, Simon L. Lewis, Hans ter Steege, Joice Ferreira, Beatriz Schwantes Marimon, Hong Truong Luu, Gerard Imani, Luzmila Arroyo, Cecilia Blundo, David Kenfack, Moses N. Sainge, Bonaventure Sonké and Rodolfo Vásquez

It is a truth universally acknowledged that those in possession of time and good fortune must be in want of information. Nowhere is this more so than for tropical forests, which include the richest and most productive ecosystems on Earth. Information on tropical forest carbon and biodiversity, and how these are changing, is immensely valuable, and many different stakeholders wish to use data on tropical and subtropical forests. These include scientists, governments, nongovernmental organizations and commercial interests, such as those extracting timber or selling carbon credits. Another crucial, often-ignored group are the local communities for whom forest information may help to assert their rights and conserve or restore their forests.

A widespread view is that to lead to better public outcomes it is necessary and sufficient for forest data to be open and 'Findable, Accessible, Interoperable, Reusable' (FAIR)[1,2]. There is indeed a powerful case. Open data — those that anyone can use and share without restrictions — can encourage transparency and reproducibility, foster innovation and be used more widely, thus translating into a greater public good (for example, https://creativecommons.org). Open biological collections and genetic sequences such as GBIF or GenBank have enabled species discovery, and open Earth observation data helps people to understand and monitor deforestation (for example, Global Forest Watch). But the perspectives of those who actually make the forest measurements are much less recognized[3], meaning that open and FAIR data can be extremely unfair indeed. We argue here that forest data policies and practices must be fair in the correct, linguistic use of the term — just and equitable.

In a world in which forest data origination — measuring, monitoring and sustaining forest science — is secured by large, long-term capital investment (such as through space missions and some officially supported national forest inventories), making all data open makes perfect sense. But where data origination depends on insecure funding and precarious employment conditions, top-down calls to make these data open can be deeply problematic[4]. Even when well-intentioned, such calls ignore the socioeconomic context of the places where the forest plots are located and how knowledge is created, entrenching the structural inequalities that characterize scientific research and collaboration among and within nations[5–9]. A recent review found scant evidence for open data ever lessening such inequalities[10]. Clearly, only a privileged part of the global community is currently able to exploit the potential of open forest data[11]. Meanwhile, some local communities are de facto owners of their forests and associated knowledge, so making information open — for example, the location of valuable species — may carry risks to themselves and their forests.

The challenge

The risks of open forest data exploitation are magnified by features of how forests are measured and who does the measuring. Generating long-term data on forest health and change involves physically measuring and identifying millions of trees. This means establishing, maintaining and revisiting plots, and curating records indefinitely. Trees are long-lived organisms so forests require decades of monitoring to properly infer change. Sustaining local observations for decades needs deep, long-term commitment to the unique but shifting combinations of people, institutions, regulations, interests and relationships that characterize each forest site. The challenge is enhanced by the great biodiversity of tropical forests. Measuring a single hectare of Amazon forest involves collecting and identifying up to ten times the number of tree species that are present in the UK's entire 24 million hectares. There are very few people with the skills to do this.

Long-term tropical-forest data measurements not only require effort and skill but also often carry risk and depend on some of the most disadvantaged actors in the global science community. Many forest workers (researchers, technicians, students, field assistants and local communities) lack basic job security, much less a career path, despite the long-term dedication that monitoring forests requires. In addition, many tropical forest workers may endure dangerous field conditions, with threats including kidnapping, armed insurgents, narcotraffickers, land-grabbers, infectious disease, snakebite, floods, fire, dangerous transport and gender-based violence. Besides these personal dangers, tropical scientists often lack the basic resources to measure and maintain their forest plots, let alone develop their research groups[8].

In contrast to the experiences of those monitoring forests on the ground, consider the context for satellite and aircraft-based measurements, which require ground-based data for validation. Space-based forest missions are expensive but are funded by public or private capital. Once in orbit, they stream data to analysts 'for free'. This requires relatively few people to sustain, and although the analysts' work is highly skilled, it carries little professional and physical risk and lacks commitment to place. Forest fieldwork is less capital-intensive, but needs sustained investment, is intensely human and carries substantial costs and risks. There are no automated collecting stations to help to identify and measure trees, so without the long-term dedication of many forest workers data collection simply stops.

The risks and costs involved in acquiring and sustaining ground forest data are persistently overlooked, ignored or regarded as externalities to be picked up by the forest

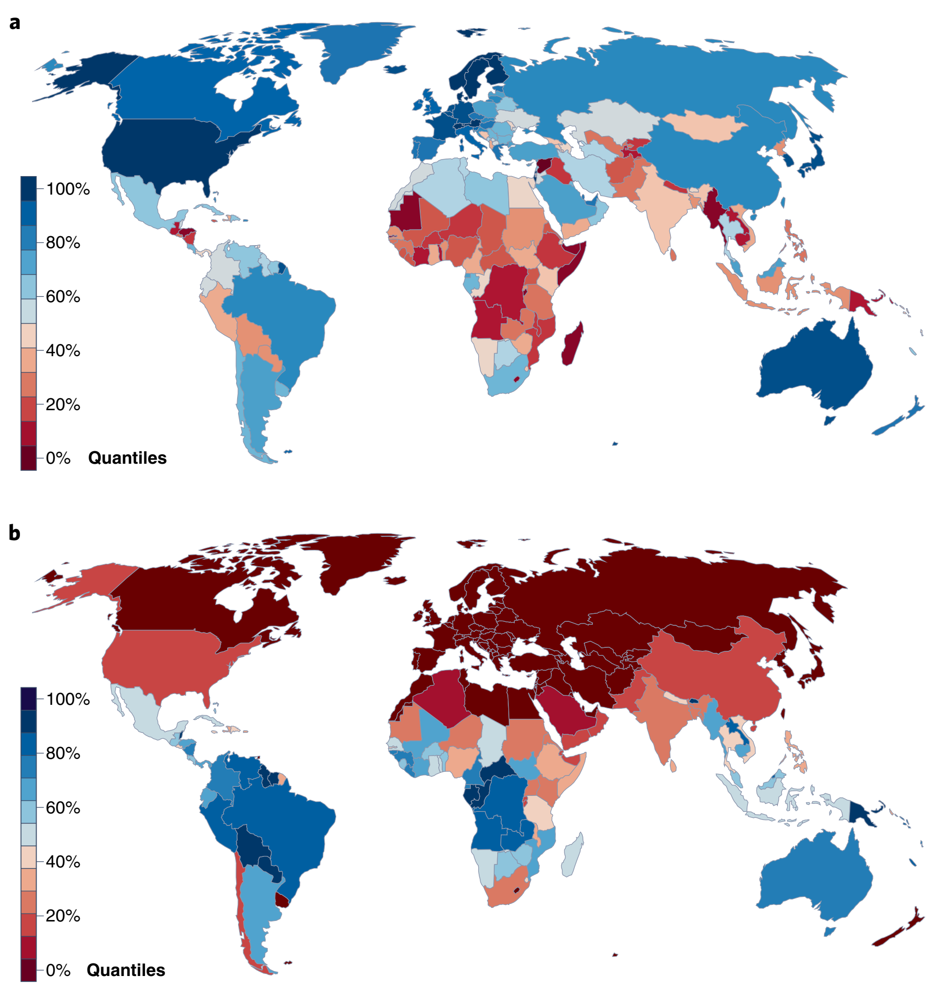

Fig. 1 | Global distributions of per capita gross domestic product and tropical forest area. a,b, The 2008–2018 national average gross domestic product per capita (**a**) and tropical forest area per capita (**b**). Countries are coloured according to position from lowest (dark red) to highest (dark blue) within each global distribution.

workers themselves. This is especially problematic because countries that hold the most tropical forests are among those least able to invest in science and development (Fig. 1, Supplementary Fig. 1). For example, monitoring the carbon balance of intact tropical moist forests has been estimated to cost US $7 million a year[12], easily exceeding present support. By contrast, the USA alone spends over $90 million annually on its national forest inventory[13]. So, many tropical forest data are collected by skilled people working with minimal funding, in challenging conditions and facing other constraints, including complex layers of rules, agreements and research permits. Given such huge disparities, it is hardly reasonable to expect this output to be served on an open plate to the world.

It is perhaps unsurprising that the most vocal proponents of making tropical and subtropical forest data open are often not those who actually measure and monitor them. Meanwhile, key beneficiaries include powerful publishers (usually with commercial interests), agencies and technology companies (often with commercial or political interests), and highly educated computer-savvy analysts wishing to integrate earth observation data with forest data (naturally with a career interest). Relatively few of these institutions and people are based in the tropics and subtropics. Fewer still are also data originators.

And so, for many data originators the present meaning of making tropical forest data 'open' is to transfer the hard-won output of their labours to more privileged individuals and institutions, and lose more of the limited control they have over their professional lives. Power flows from the originators to public agencies, private companies and data scientists, mainly in the Global North.

A way forward

Can this situation be changed to benefit data originators and users alike? We believe that the future of tropical forest data should be open, as is already the case for some biological data (for example, species records and DNA sequences), but realizing the potential scientific and societal gains of open data requires a radically new agreement among forest data originators, users and funders. It needs users and funders to explicitly acknowledge the power dynamics — and to do something practical about them.

To ensure the benefits of long-term on-the-ground forest data streams are fairly secured, we present here eight key recommendations based on an alternative model that focuses on the needs of the data originators and ensures users and funders contribute properly. In this model, the skills, careers and livelihoods of the originators are front-and-centre. An equitable and sustainable approach to measuring the world's forests therefore starts by recognizing the human challenge involved in long-term forest measurements. It puts people — not data — first. This means recognizing the true costs of forest data origination and supporting better-quality careers for those doing the fieldwork.

These include funding the direct and indirect costs of: (**1**) fieldwork and essential laboratory work, including herbaria support; (**2**) training, safe working practices and secure employment conditions for the professionals on whom forest data production depends; and (**3**) the overheads of institutions responsible for data delivery. Long-term support of integrated forest data management is also essential, beyond what GBIF and GenBank provide for species records and DNA sequences. Therefore, (**4**) covering the costs of coordinated data curation and database infrastructure needs to be standard[14]. Together, addressing these true costs will put funders in a position to ensure their support leads to open data releases and more open science.

Thus, agreements to make data open become explicitly tied to properly funded actions that sustain data origination and develop in-country capacity so that all benefit from the open data model[4,14]. Building local capacity and infrastructure will generate better science, which will lead to better forest management[15] and reduce inequalities between those producing and using data. It is not only fair to invest in data originators and their development, but it is also better for the ultimate goals we all want to achieve.

Meanwhile, we can help to build long-term tropical forest research on more equitable foundations. Authors and journals can support this by (**5**) embracing holistic definitions of authorship to include those involved in data collection and management and (**6**) ensuring results are communicated in the originators' languages. Because we share the same world but research investment capacity is highly uneven (Fig. 1), (**7**) international agreements and funding to support data origination, capacity building, stable and long-term careers are needed to empower subtropical and tropical institutions. Last but not least, it is essential to (**8**) develop deep, long-term and equitable collaborations, which should be the stated aim of funders, producers and users alike[8]. Global and national research networks have already emerged to help to originate, assemble and share forest data, while putting data originators in control of data management and access[12,16,17]. These initiatives inherit asymmetries in scientific research but can build bridges and develop the next generation of subtropical and tropical leaders by supporting them with data, tools, connectivity and opportunities to lead academic and applied outcomes.

Successful systemic change demands that we build partnerships across divides. Our shared need to secure a stable climate and protect biodiversity is becoming enshrined in global agreements and national laws, not only because the benefits are becoming ever clearer but also because the fundamental principle of differentiated obligations and contributions has been widely recognized. And just as the health of all depends on equitable global access to nutrition and medical resources, so the benefits of sharing data will flow much more easily when those who make forest measurements become truly valued. In short, for tropical forest data to be open, they first must be supported fairly.

Data availability

The data used to produce Fig. 1 are publicly available. Data on gross domestic product and population size are available from the World Bank at https://data.worldbank.org/indicator. Data on tropical forest area (closed, open or fragmented forests) were extracted from the Forests 2000 by Major Ecological Domains grid, provided by the Food and Agriculture Organization at http://www.fao.org/geonetwork. ❐

Renato A. F. de Lima[1,28] ✉, Oliver L. Phillips[2,28] ✉, Alvaro Duque[3], J. Sebastian Tello[4], Stuart J. Davies[5], Alexandre Adalardo de Oliveira[1], Sandra Muller[6], Euridice N. Honorio Coronado[7,8], Emilio Vilanova[9,10], Aida Cuni-Sanchez[11,12], Timothy R. Baker[2], Casey M. Ryan[13], Agustina Malizia[14], Simon L. Lewis[2,15], Hans ter Steege[16], Joice Ferreira[17], Beatriz Schwantes Marimon[18], Hong Truong Luu[19], Gerard Imani[20], Luzmila Arroyo[21,22], Cecilia Blundo[14], David Kenfack[5], Moses N. Sainge[23,24,25], Bonaventure Sonké[26] and Rodolfo Vásquez[27]

[1]Departamento de Ecologia, Instituto de Biociências, Universidade de São Paulo, São Paulo, Brazil. [2]School of Geography, University of Leeds, Leeds, UK. [3]Departamento de Ciencias Forestales, Universidad Nacional de Colombia, Medellín, Colombia. [4]Center for Conservation and Sustainable Development, Missouri Botanical Garden, St Louis, MO, USA. [5]Forest Global Earth Observatory (ForestGEO), Smithsonian Tropical Research Institute, Washington, DC, USA. [6]Departamento de Ecologia, Universidade Federal do Rio Grande do Sul, Porto Alegre, Brazil. [7]School of Geography and Sustainable Development, University of St Andrews, St Andrews, UK. [8]Instituto de Investigaciones de la Amazonía Peruana (IIAP), Iquitos, Peru. [9]Facultad de Ciencias Forestales y Ambientales, Universidad de los Andes, Mérida, Venezuela. [10]Wildlife Conservation Society (WCS), Bronx, New York, USA. [11]NORAGRIC Department, Norwegian University of Life Sciences, Ås, Norway. [12]Department of Environment and Geography, University of York, York, UK. [13]School of GeoSciences, University of Edinburgh, Edinburgh, UK. [14]Instituto de Ecología Regional, Universidad Nacional de Tucumán – Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Tucumán, Argentina. [15]Department of Geography, University College London, London, UK. [16]Tropical Botany, Naturalis Biodiversity Center, Leiden, The Netherlands. [17]Embrapa Amazônia Oriental, Belém, Brazil. [18]Programa de Pós-graduação em Ecologia e Conservação, Universidade do Estado de Mato Grosso, Nova Xavantina, Brazil. [19]Southern Institute of Ecology, Institute of Applied Materials Science, Vietnam Academy of Science and Technology, Ho Chi Minh City, Vietnam. [20]Biology Department, Faculty of Sciences, Universite Officielle de Bukavu, Bukavu, Democratic Republic of the Congo. [21]Facultad de Ciencias Agrícolas, Universidad Autónoma Gabriel René Moreno, Santa Cruz, Bolivia. [22]Museo de Historia Natural Noel Kempff Mercado, Santa Cruz, Bolivia. [23]Department of Biological Sciences, Fourth Bay College, University of Sierra Leone, Freetown, Sierra Leone. [24]Institute of International Education Scholar Rescue Fund (IIE-SRF), New York, NY, USA. [25]Department of Horticultural Sciences, Cape Peninsula University of Technology, Cape Town, South Africa. [26]Plant Systematics and Ecology Laboratory, Department of Biology, Higher Teachers' Training College, University of Yaoundé I, Yaoundé, Cameroon. [27]Jardín Botánico de Missouri, Oxapampa, Peru. [28]These authors contributed equally: Renato A. F. de Lima, Oliver L. Phillips.
✉*e-mail: raflima@usp.br; O.Phillips@leeds.ac.uk*

Published online: 11 April 2022
https://doi.org/10.1038/s41559-022-01738-7

References

1. Liang, J. & Gamarra, J. G. P. *Sci. Data* **7**, 424 (2020).
2. Wilkinson, M. D. et al. *Sci. Data* **3**, 160018 (2016).
3. Carroll, S. R. et al. *Data Sci. J.* **19**, 43 (2020).
4. Verhulst, S. G. & Young, A. *Open Data in Developing Economies: Toward Building an Evidence Base on What Works and How* (African Minds, 2017).
5. May, R. M. *Science* **275**, 793–796 (1997).
6. Bezuidenhout, L. M., Leonelli, S., Kelly, A. H. & Rappert, B. *Sci. Public Policy* **44**, 464–475 (2017).
7. Nuñez, M. A., Chiuffo, M. C., Pauchard, A. & Zenni, R. D. *Trends Ecol. Evol.* **36**, 766–769 (2021).
8. Armenteras, D. *Nat. Ecol. Evol.* **5**, 1193–1194 (2021).
9. Trisos, C. H., Auerbach, J. & Katti, M. *Nat. Ecol. Evol.* **5**, 1205–1212 (2021).
10. Davies, T. & Perini, F. *J. Comm. Inform.* **12**, 148–178 (2016).
11. Gurstein, M. B. *First Monday* **16**, https://doi.org/10.5210/fm.v16i2.3316 (2011).
12. ForestPlots.net et al. *Biol. Conserv.* **260**, 108849 (2021).
13. Castillo, P. S. B. & Alvarez, M. *Forest Inventory and Analysis Fiscal Year 2018 Business Report* (US Department of Agriculture, 2020)
14. Serwadda, D., Ndebele, P., Grabowski, M. K., Bajunirwe, F. & Wanyenze, R. K. *Science* **359**, 642–643 (2018).
15. Baker, T. R. et al. *Plants, People, Planet* **3**, 229–237 (2021).
16. Davies, S. J. et al. *Biol. Conserv.* **253**, 108907 (2021).
17. The SEOSAW partnership. *Plants, People, Planet* **3**, 249–267 (2021).

Acknowledgements

The authors thank T. Pennington, F. Costa and A. Vicentini for helpful discussions and comments on earlier versions of this manuscript. For support in making this possible, we thank the Royal Society (International Collaboration Award, ICA\R1\180100), the Leverhulme Trust (APX\R1\191094), NERC (NE/S011811/1 'ARBOLES'; NE/T01279X/1 'SECO'), São Paulo Research Foundation (FAPESP), National Council for Scientific and Technological Development (CNPq), Vietnam Ministry of Science and Technology (project contract 23/2020/HĐ-NĐT), the Global Land Programme, Missouri Botanical Garden, US National Science Foundation (Award 2020424) and the Peruvian Service for Natural Protected Areas (SERNANP).

Author contributions

O.L.P. drafted the initial version of the manuscript with subsequent inputs from R.A.F.d.L. R.A.F.d.L. produced the figures. All authors contributed to discussing the theme, made suggestions and approved the text, along with the following positionality statement. **Positionality statement:** We, the authors, are all forest researchers working across different disciplines and networks to better understand tropical and subtropical forests and support their sustainable future. Some come from the South and are based in the South, some come from the North and based in the North, and some from the South are based currently in the North. We recognize that labels such as North and South, while useful, simplify complex realities, and are to an extent limited. In our careers we have been originators in the sense of working in the field to establish, identify and/or re-measure forest and savannah plots. We have also worked to get funding to establish and sustain plots, support data management, grow collaborations and do science. Many of us have also benefitted in different ways, including publications, building research groups and achieving professional recognition. Although our experiences, career stages and backgrounds are diverse, we are united by our convictions that the production and the use of forest data must become more equitable, that many less-visible colleagues who contribute vital work need proper recognition, and that fair, long-term collaboration across geographical, socio-economic and cultural divides is essential to build the best outcomes for science and society.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41559-022-01738-7.

Peer review information *Nature Ecology & Evolution* thanks Rob Crystal-Ornelas and Mariana Chiuffo for their contribution to the peer review of this work.