

Software Heritage

# Strategic Software Insights report for Latin American research institutions

**Version 1.0 – 2026-10-24**

https://www.softwareheritage.org

List of the authors

| | |
|---|---|
| **Short title** | Strategic Software Insights for Latin American research institutions |
| **Full title** | Strategic Software Insights report for Latin American research institutions |
| **Author** | Roberto Di Cosmo (for the Software Heritage team) |
| **Date** | 2026-10-24 |
| **Contact** | Roberto Di Cosmo ⟨roberto@dicosmo.org⟩ |

# Contents

**1 What this report says, in five numbers** **3**
1.1 What follows from this . . . . . 3

**2 The largest projects** **5**
2.1 How much of each project comes from Latin America . . . . . 5

**3 Software Latin American researchers help build** **8**

**4 Which countries build software together** **9**

**5 What was counted, and how we know** **9**

**6 Where the software lives, and whose forge it is on** **11**

**7 Your country, and your institution** **13**
7.1 Which institutions . . . . . 14

**8 What would be lost, and is not** **16**
8.1 Can this software be cited . . . . . 16
8.2 Can this software be reused . . . . . 16

**9 What a ministry or a research council can do next** **17**
9.1 What this report does not support . . . . . 17
9.2 Where to send someone who asks how . . . . . 18

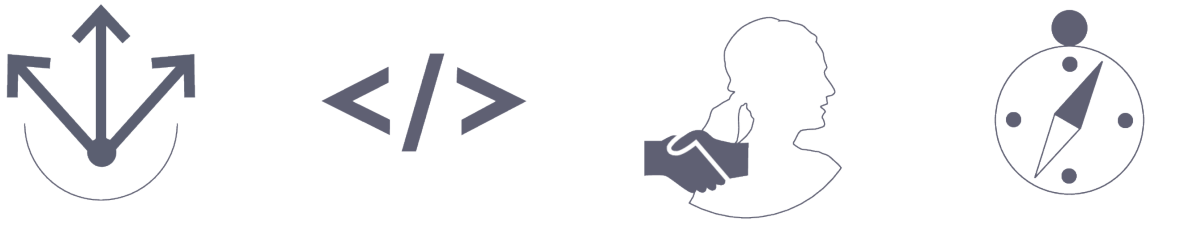
**10 About Software Heritage** **19**

**11 Appendix: the perimeter** **20**

# 1 What this report says, in five numbers

Software Heritage has collected, preserved and indexed the source code produced by Latin American research institutions. This edition counts projects with a contributor at one of the institutional domains listed in the appendix, which define the perimeter, less the exclusions described in point 3.

1. **620,095 projects** carry at least 1 contribution signed from an institutional address of the perimeter, across 26 countries.
2. **139,221 people** signed those contributions, and most of them write from a departmental or faculty subdomain rather than from the institution's main domain. A count built from top-level domains alone misses them.
3. **28,037 projects of the perimeter are coursework, templates, personal pages, test repositories or mailing-list archives, and are dropped rather than counted.** Of the 30,911 most active projects, judged one at a time, 51.8% fall in that category; the rest were identified by the course organisation or assignment template they belong to. This is the measurement an institution cannot make for itself, and no number above includes any of them.
4. **50,605 projects can no longer be found where they were published**, and exist today only because they were archived.
5. **3,580 projects received a contribution in the last twelve months.** The rest are not necessarily abandoned, but they are not moving.



**6230383**
**Contributions**

**620095**
**Software Projects**
**Active last 12 months: 3580**
**Still crawled: 569490**

**139221**
**Contributors**

**1971-06-21**
**2026-06-03**
**Time Span**

## 1.1 What follows from this

**Have the national and institutional forges archived.** The section on where the software lives names the platforms this code actually sits on. Software Heritage covers the large commercial platforms well and covers self-hosted national infrastructure only where someone has asked. A forge that is not archived contributes nothing to the figures above, and the code

on it is the code most specific to the country.

# 2 The largest projects

These are the projects of the perimeter with the most contributions from the institutions counted here, out of the 6,230,383 contributions the perimeter holds in total. Each bar is a project, not a repository: its forks and mirrors in the archive are counted once, under the name of the project itself.

| project | institutional contributions | institutional share % |
|---|---|---|
| https://github.com/optimusgisit/terralib | 6769 | 58.86 |
| https://github.com/AndresG25/Tesis_Maestria | 6353 | 99.97 |
| https://github.com/lua/lua | 5979 | 94.84 |
| https://github.com/eciis/web | 5290 | 64.63 |
| https://github.com/jdvelasq/techminer-api | 4884 | 82.32 |
| https://github.com/rsa-tools/rsat-code | 4842 | 8.47 |
| https://github.com/chrodriguez/meran | 4828 | 54.31 |
| https://github.com/TeleMidia/ginga | 3826 | 51.93 |
| https://github.com/RocketPy-Team/RocketPy | 3739 | 51.28 |
| https://github.com/NumCosmo/NumCosmo | 3271 | 52.16 |
| https://github.com/ifunam/salva | 3189 | 78.49 |
| https://bitbucket.org/inland-dgvm/inland | 3141 | 90.13 |
| https://github.com/amadeubarbosa/openbus-core | 2936 | 100.0 |
| https://github.com/LCAD-UFES/carmen_lcad | 2800 | 30.0 |
| https://github.com/TerraME/terrame | 2732 | 58.61 |
| https://gitlab.com/felipe.tellez/igac.git | 2715 | 45.57 |
| https://github.com/The-OpenROAD-Project/OpenROAD | 2602 | 5.36 |
| https://github.com/lauramilena3/MOSAIC | 2484 | 100.0 |
| https://github.com/TeleMidia/nclcomposer | 2407 | 89.45 |
| https://github.com/BioinformaticsFMRP/TCGAbiolinks | 2240 | 6.51 |

## 2.1 How much of each project comes from Latin America

The ranking above says how large each project is. It does not say how much of it the perimeter wrote. The next figure answers that: the same projects, in the same order, with each bar split into the contributions that came from an institutional address and the contributions that came from everyone else.

Read the percentage. Each bar is the contribution that came from an institutional address, and the figure beside it says what share of the whole project that is. A project at 3.7% is one the region participates in; a project at 100% is one the region carries. Those two call for different decisions, and the ranking alone cannot tell them apart.

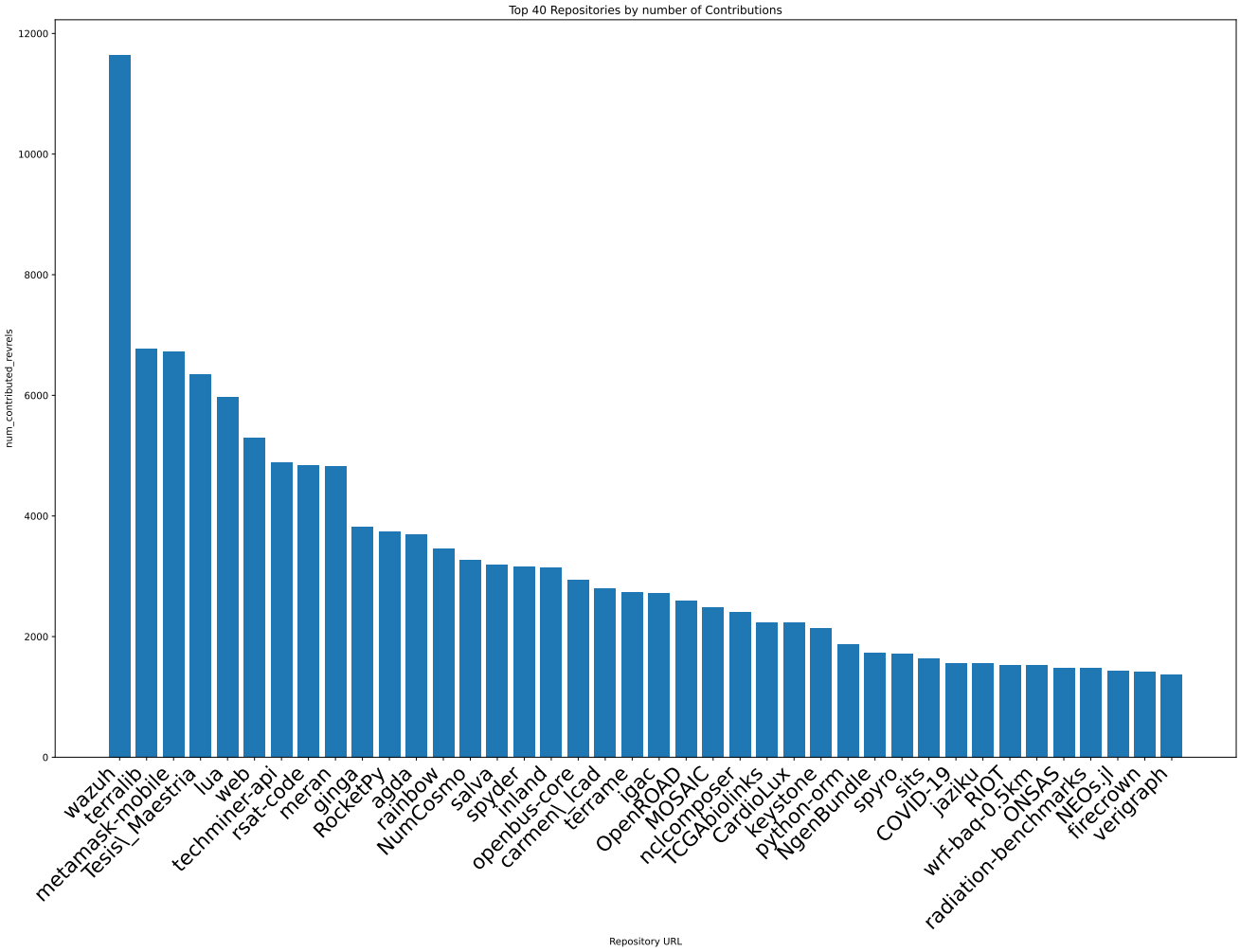



Figure 1: **The 40 projects with the most contributions from the perimeter.** Population: the 620,095 projects of the perimeter, measured in distinct revisions and releases.

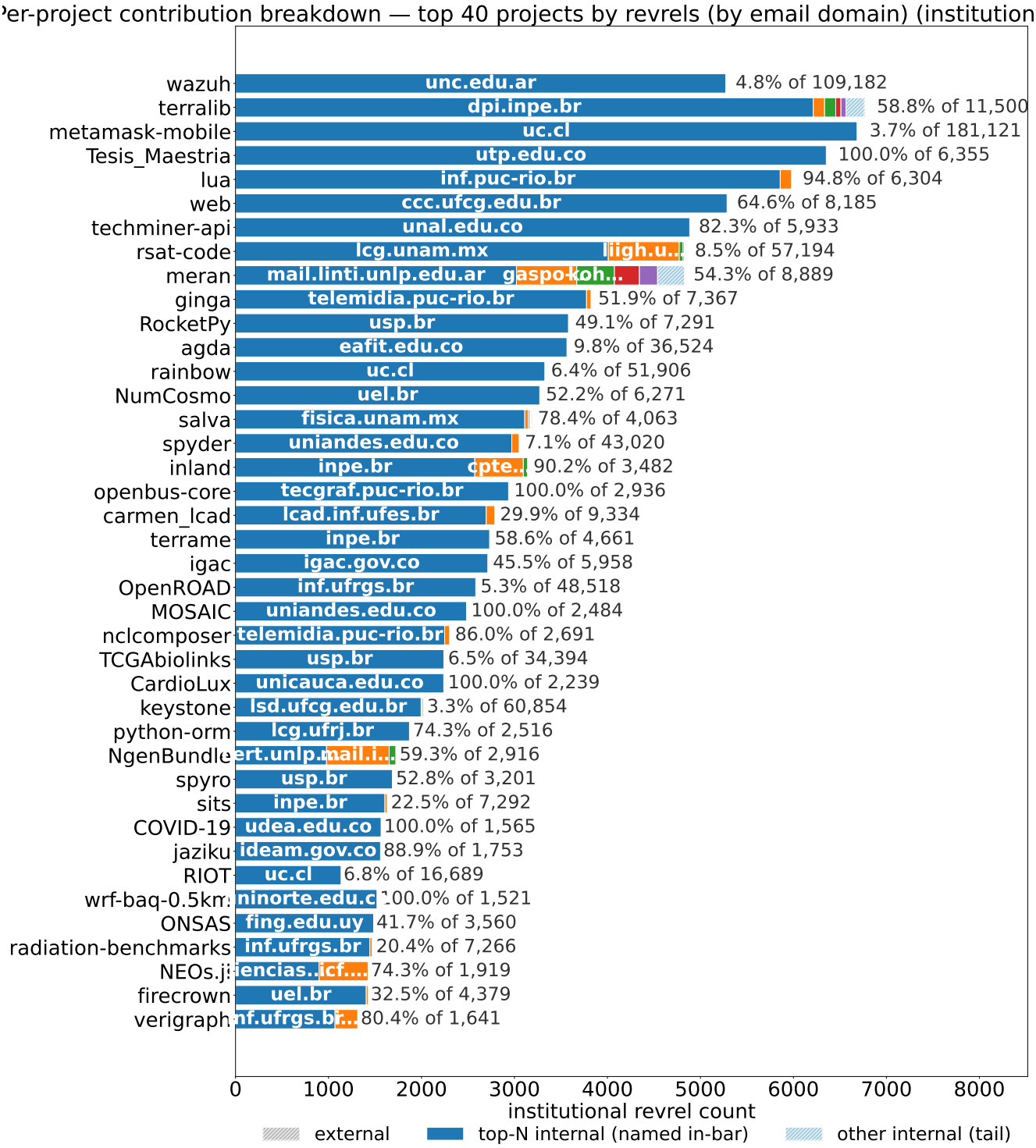



Figure 2: **Institutional contribution inside each of the top 40 projects.** Each bar is one project, and its length is the contribution that came from an institutional address of the perimeter, so the largest contributors can be named inside it. The figure at the end of each bar is that contribution as a share of the whole project. Contributions from outside the perimeter are not drawn, because on a project the perimeter wrote a few percent of they leave no room for a name. The per-segment figures, including the external total, are published beside this report.

# 3 Software Latin American researchers help build

Some of the most heavily worked projects in this perimeter belong to someone else. A Latin American institutional address contributed to them, often substantially, but the institution does not own them. Counting those as institutional output would overstate it; leaving them out would hide the thing an organisation-based measurement can never see, which is staff working in the open on code the world depends on.

| project | institutional contributions | institutional share % |
|---|---|---|
| https://github.com/wazuh/wazuh | 11645 | 10.67 |
| https://github.com/MetaMask/metamask-mobile | 6727 | 3.71 |
| https://github.com/agda/agda | 3700 | 10.13 |
| https://github.com/rainbow-me/rainbow | 3457 | 6.66 |
| https://github.com/spyder-ide/spyder | 3155 | 7.33 |
| https://github.com/openstack/keystone | 2142 | 3.52 |
| https://github.com/Inria-Chile/RIOT | 1535 | 9.2 |
| https://github.com/prefeitura-rio/pipelines.git | 1252 | 11.82 |
| https://github.com/yetanotherco/aligned_layer | 1211 | 9.72 |
| https://github.com/nh-server/Kurisu | 1139 | 21.48 |
| https://github.com/renanrodrigo/ubuntu-advantage-client | 1130 | 17.06 |
| https://bitbucket.org/pmonter/airflow_mlops.git | 960 | 3.49 |
| https://github.com/liferay/liferay-portal | 939 | 0.07 |
| https://github.com/anlix-io/flashman | 868 | 6.83 |
| https://github.com/openstack/ironic | 771 | 1.4 |
| https://gitlab.com/fluidattacks/web.git | 739 | 6.92 |
| https://github.com/folio-org/ui-users | 717 | 5.55 |
| https://github.com/gmh5225/ethrex | 667 | 10.36 |
| https://github.com/aliceoq/devportal | 635 | 34.44 |
| https://github.com/HewlettPackard/c-spiffe | 622 | 27.17 |

The last column is the institution's share of everything the project has ever received. A high share is maintainership; a low one is sustained participation in a large codebase. Both are real, and they are different claims.

# 4 Which countries build software together

The perimeter covers 26 countries, and the projects inside it are not partitioned by border. The matrix below counts, for every pair of countries, how many projects both of them contributed to. It is the one figure in this report that measures the region as a region rather than as a list of institutions.

A dense row is a country whose contributors turn up in other countries' projects. A blank cell means the pair shares fewer than five projects, which is the threshold below which a pair is left out, so a sparse row is a weak signal rather than an absence of collaboration. Where a row is sparser than you expect, the perimeter for that country is the first thing to check: a country whose forges are not yet archived has few projects here to share.

# 5 What was counted, and how we know

**The perimeter.** Every project with at least 1 contribution signed from one of the institutional email domains of Latin American research institutions. The full list is in the appendix, and the machine-readable version is published beside this report, so the perimeter can be checked rather than believed.

**The classification.** Not every repository on an institutional forge is research software: the corpus also holds mailing-list mirrors, distribution trees, dotfiles, coursework and machine-written logs. Every project above the threshold was therefore judged individually by a language model reading its metadata, its top contributor domains and its README as archived. The mapping from a label to a decision was made by code, not by a model, and a human ruling, where one exists, overrides the model.

**How far we can check it.** The classification was tested against a property no model saw: whether a repository declares itself citable with a `CITATION.cff` or a `codemeta.json` file. Among the projects kept, 0.71% declare themselves citable; among those dropped, 0.21%. A ratio of **3.4**. The count of dropped projects declaring citability is 59, and the coursework pass did not change it: not one of the projects it added declares itself citable.

**What that does not tell you, and it matters here.** Citation adoption across this whole corpus is 0.1%, so the test has little power whatever the ratio: there are too few declarations on either side for it to say much. It bounds how wrong the classification can be and it does not show that it is right. 59 dropped projects do declare themselves citable and are worth re-reading by hand. The classification has not been read by a human end to end. Where a judgement was close, the project was kept.

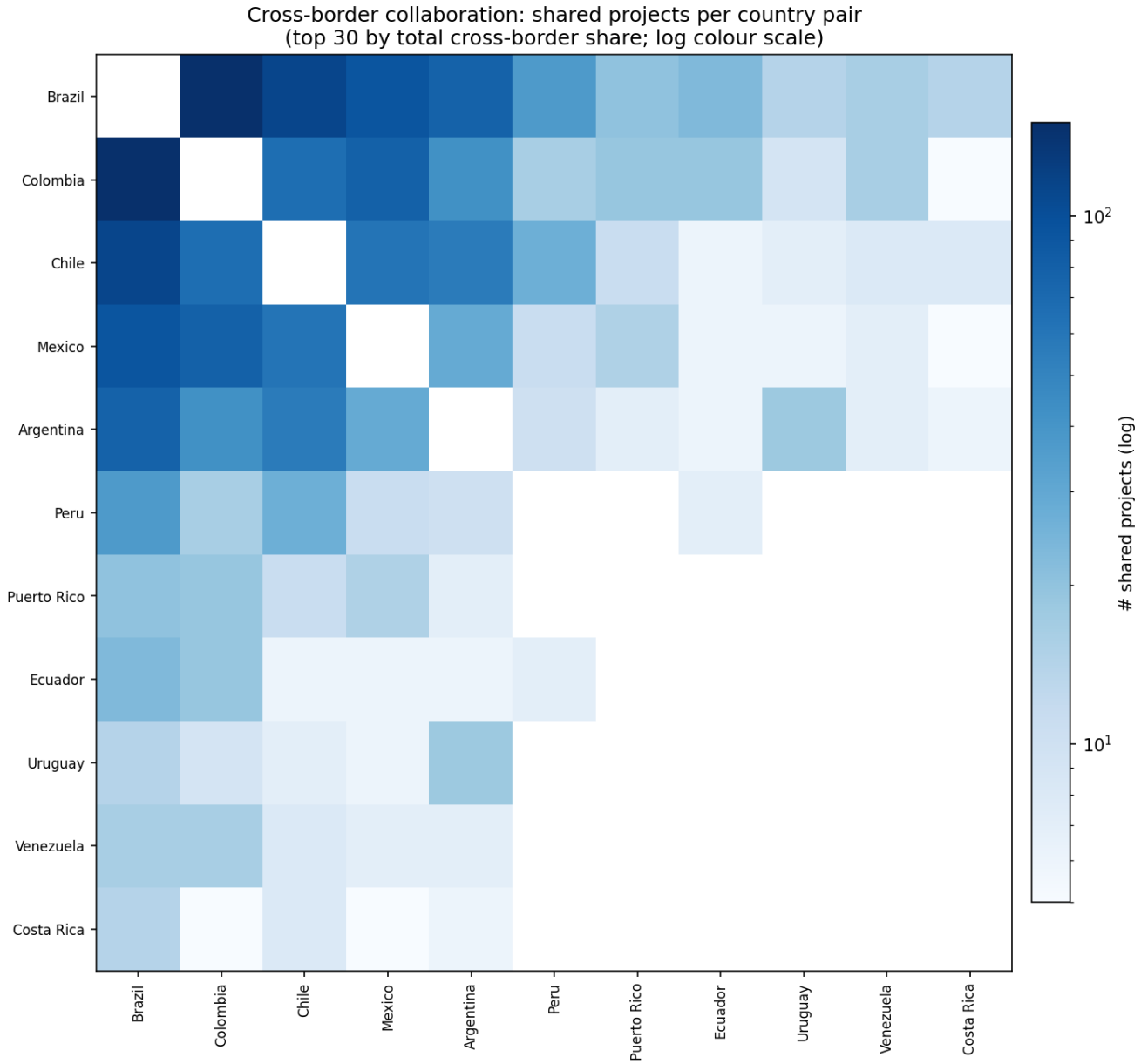



Figure 3: **Projects shared between each pair of countries.** The countries with the most shared projects, up to thirty of them, on a logarithmic colour scale so that small collaborations stay visible beside the largest ones. A pair counts once per project both countries contributed to, and pairs sharing fewer than five projects are left out.

# 6 Where the software lives, and whose forge it is on

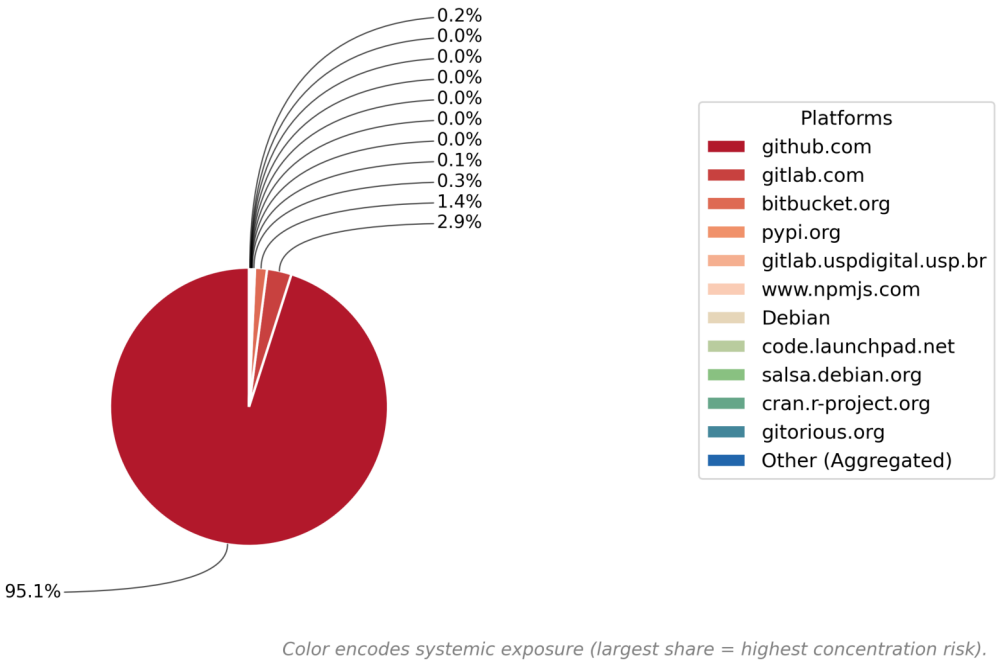



Figure 4: **The platforms this code sits on.** Share of the 620,095 projects of the perimeter, by the platform hosting them.

**The tail is the interesting part.** Behind the general-purpose platforms sit forges run by the institutions themselves, and those are the ones where an institution controls its own preservation. A forge that Software Heritage does not crawl is a single point of failure for everything on it, and a forge that it does crawl is preserved regardless of what happens to the institution that runs it. This is the evidence behind the first ask.

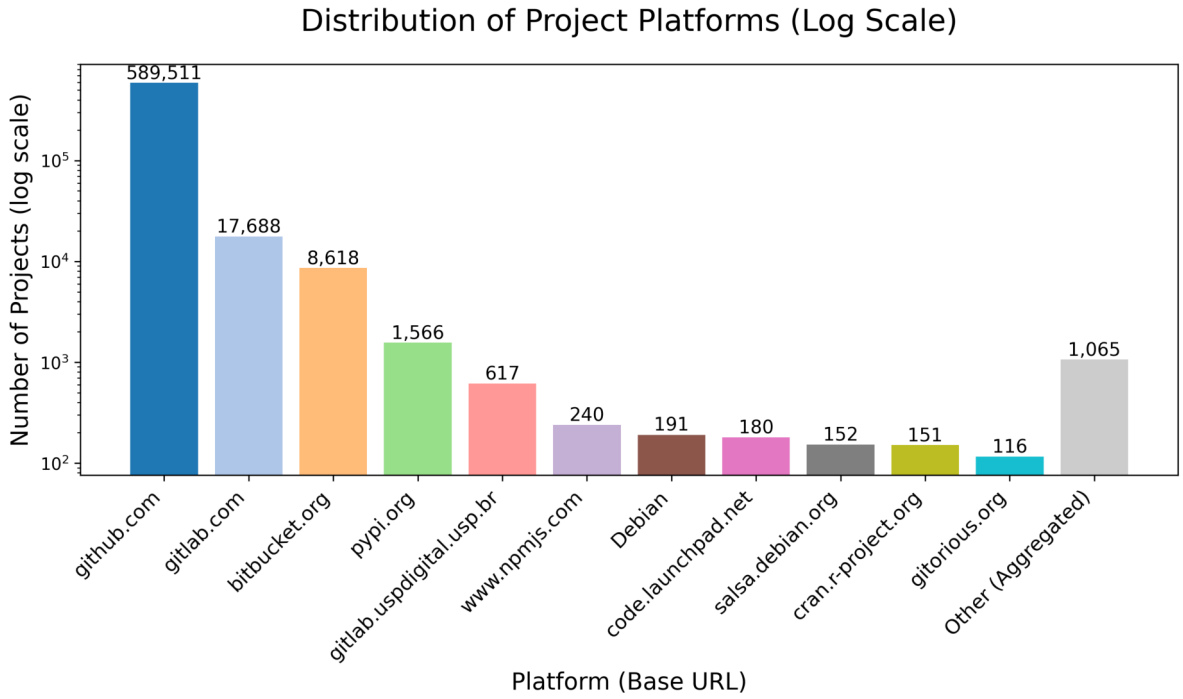



Figure 5: **The same platforms on a logarithmic scale**, so the institutional forges behind the commercial ones are visible. Population: the 620,095 projects of the perimeter.

# 7 Your country, and your institution

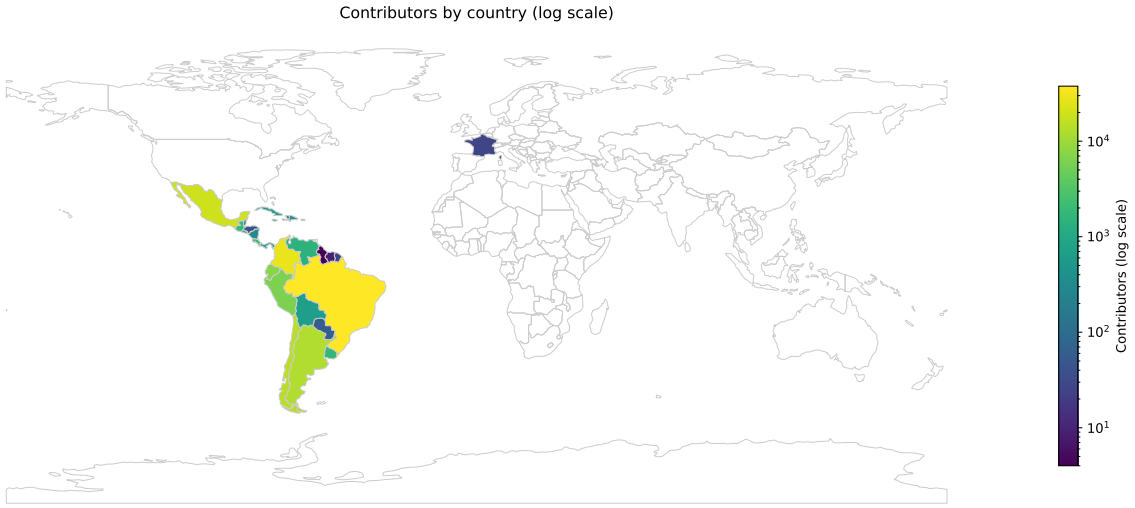



Figure 6: **Contributors by country.** Population: the 139,221 contributors of the perimeter, placed by the country of the institutional domain they signed from.

| group | contributors | share | n domains |
|---|---|---|---|
| Brazil | 37686 | 27.1% | 75 |
| Colombia | 27834 | 20.0% | 31 |
| Mexico | 18718 | 13.4% | 48 |
| Chile | 14735 | 10.6% | 41 |
| Argentina | 12596 | 9.0% | 51 |
| Ecuador | 7507 | 5.4% | 16 |
| Peru | 6192 | 4.4% | 16 |
| Costa Rica | 2218 | 1.6% | 9 |
| Puerto Rico | 2218 | 1.6% | 2 |
| Guatemala | 1931 | 1.4% | 2 |
| Uruguay | 1766 | 1.3% | 15 |
| Venezuela | 1446 | 1.0% | 9 |
| El Salvador | 897 | 0.6% | 1 |
| Cuba | 660 | 0.5% | 8 |
| Bolivia | 659 | 0.5% | 10 |
| Panama | 633 | 0.5% | 2 |
| Jamaica | 527 | 0.4% | 2 |
| Dominican Republic | 407 | 0.3% | 3 |
| Nicaragua | 307 | 0.2% | 3 |
| Belize | 105 | 0.1% | 1 |
| Paraguay | 59 | 0.0% | 4 |
| Honduras | 42 | 0.0% | 1 |
| Haiti | 39 | 0.0% | 2 |
| France | 26 | 0.0% | 2 |
| Suriname | 9 | 0.0% | 1 |
| Guyana | 4 | 0.0% | 1 |

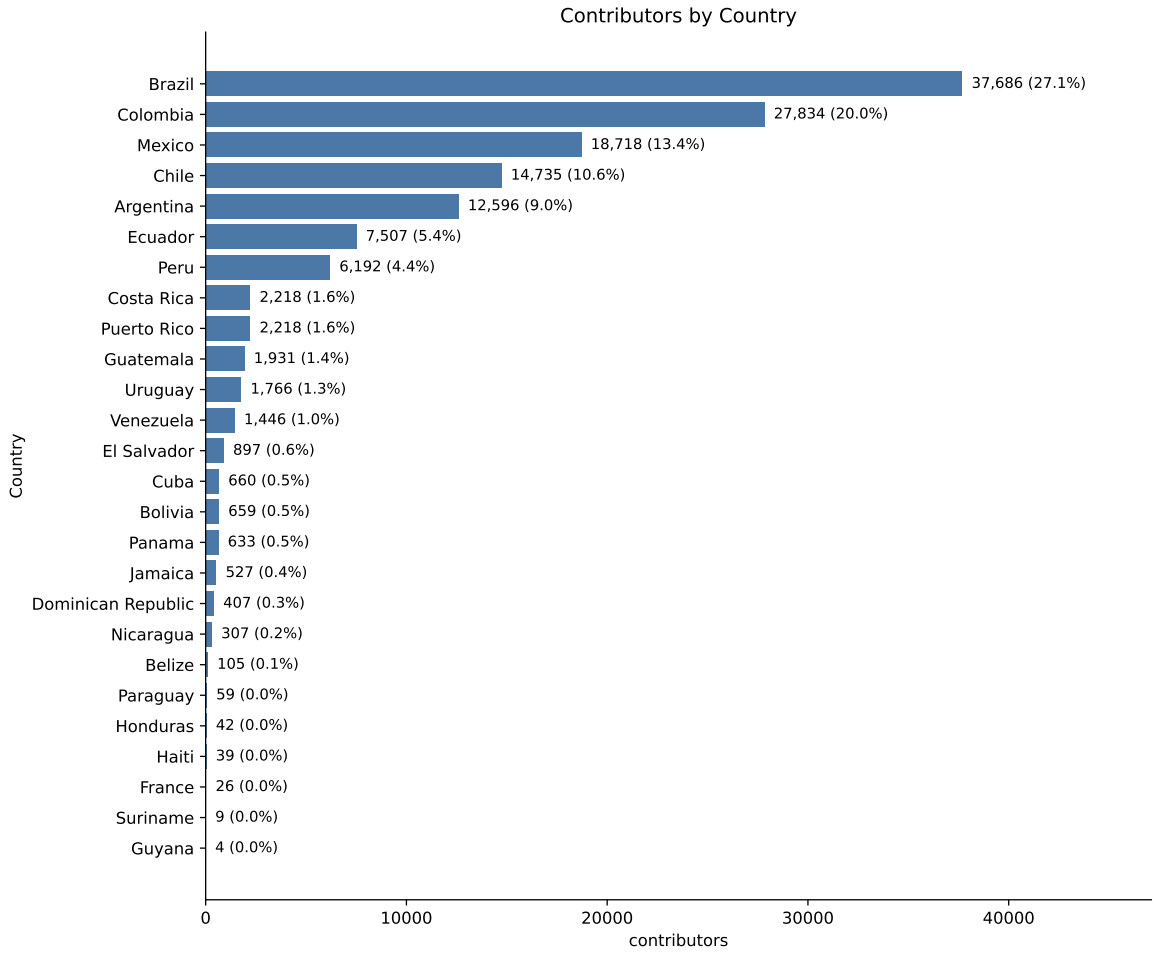



Figure 7: **The same distribution as a ranking.** Population: the 139,221 contributors of the perimeter.

Each row names a country, the contributors who signed from one of its institutional domains, and how many of its domains the perimeter covers. Where a country shows few domains, the perimeter is still thin there, and extending it is a request away.

## 7.1 Which institutions

| domain | count | domain | count |
|---|---|---|---|
| unal.edu.co | 7197 | uninorte.edu.co | 2010 |
| usp.br | 5639 | espol.edu.ec | 1988 |
| unam.mx | 5374 | upr.edu | 1891 |
| uniandes.edu.co | 5209 | udistrital.edu.co | 1716 |
| uc.cl | 4336 | uchile.cl | 1715 |
| utn.edu.ar | 3678 | utfpr.edu.br | 1637 |
| ufcg.edu.br | 2825 | ipn.mx | 1558 |
| ufpe.br | 2740 | pucp.edu.pe | 1544 |
| uba.ar | 2448 | pucminas.br | 1525 |
| udea.edu.co | 2441 | usach.cl | 1448 |
| tec.mx | 2269 | uabc.edu.mx | 1397 |

| domain | count | domain | count |
|---|---|---|---|
| udg.mx | 2257 | itba.edu.ar | 1358 |
| usm.cl | 2045 | epn.edu.ec | 1324 |
| ufrj.br | 2033 | ucr.ac.cr | 1300 |
| unicamp.br | 2017 | ufpb.br | 1278 |

The thirty institutions with the most contributors in the perimeter. The full list, every institutional domain the perimeter covers, ships as a companion table beside this report.

# 8 What would be lost, and is not

**50,605 projects counted here can no longer be reached where they were published**, against 3,580 that received a contribution in the last twelve months. They were not backed up by their authors, and they are not on the platform that hosted them. They exist because they were archived, and they can still be read, cited and built today.

## 8.1 Can this software be cited

743 projects, 0.1% of the perimeter, declare how they wish to be cited. The more interesting number is the complement: everything else in this report cannot be credited automatically in any bibliography, which is a solved problem left unsolved. A `CITATION.cff` file is a few lines at the root of a repository.

## 8.2 Can this software be reused

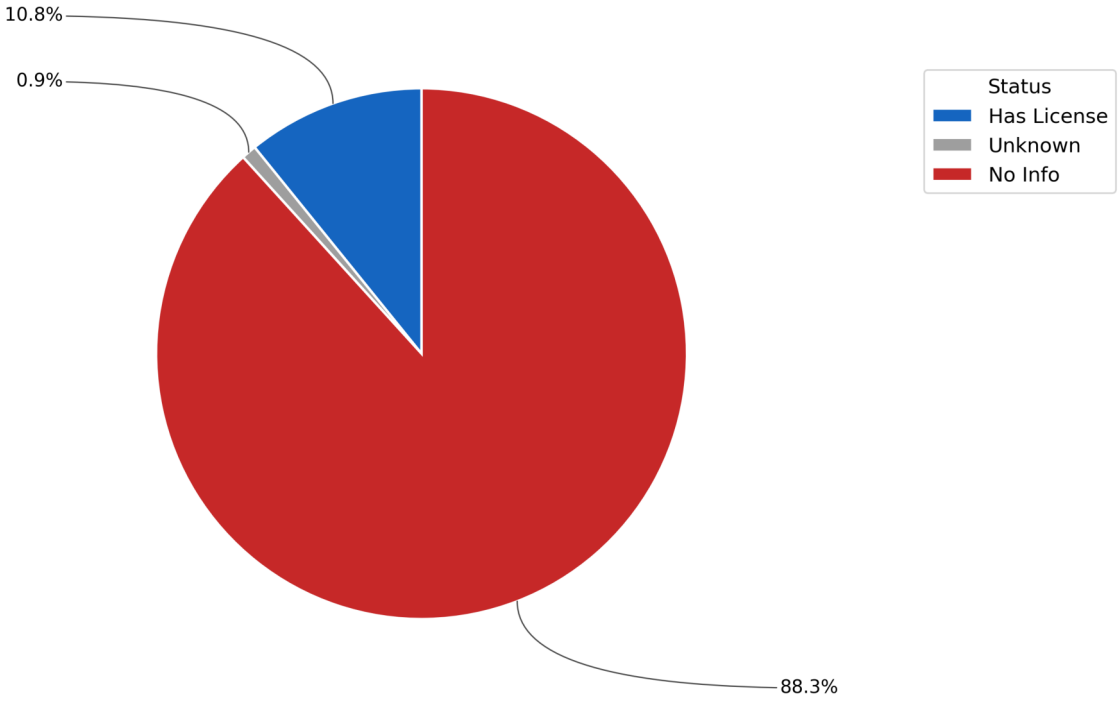



Figure 8: **Whether a licence is declared at all.** Population: the 620,095 projects of the perimeter.

# 9 What a ministry or a research council can do next

**Have the forges archived.** The section on where the software lives shows where this code actually sits. Where it is on a forge a national institution runs, ask for that forge to be archived: the request is a form, the effect is permanent, and it covers everything on the forge rather than one repository at a time.

**Share the cost of the infrastructure.** The archive is a common non-profit infrastructure, paid for by the institutions and public bodies that rely on it, and this measurement exists because the archive does. A contribution from the region mutualises that cost, keeps the archive neutral, and keeps the same measurement available to every institution on the same terms.

**Archive what is at risk now.** Save Code Now takes a repository URL and archives it in seconds. It needs no account and no agreement, and it is the fastest answer to a project whose maintainer is leaving.

## 9.1 What this report does not support

The perimeter binds a contribution to an institutional email address, not to a person, and an address is a weak proxy for affiliation: people move, use personal addresses, and contribute from several institutions at once. The figures are sound at the scale of a country or an institution and unsound at the scale of a researcher, which is not a caveat to be read past.

## 9.2 Where to send someone who asks how



**How to archive and reference your code.** The practical guide: get a repository into the archive, obtain an identifier that will still resolve in twenty years, and cite it in a paper. softwareheritage.org/howto-archive-and-reference-your-code



**Save Code Now.** Paste a repository address and it is archived in seconds. No account, no agreement, nothing to sign. This is the third ask on the previous page, and it can be done from this page. archive.softwareheritage.org/save



**The Source Code of Science.** A doctoral course on archiving, referencing and reproducing research software, CC-BY-4.0: yours to reuse, translate into Spanish or Portuguese, and teach from. github.com/rdicosmo/source-code-of-science

# 10 About Software Heritage

This report is produced from the Software Heritage archive, and is one of the analyses published through the Public Code Observatory. Software Heritage collects, preserves and shares the source code of all publicly available software, with its full development history, and gives every version an identifier that can be cited and resolved. It is a non-profit initiative started by Inria in partnership with UNESCO, and the archive is open to everyone, everywhere, at no cost and with no agreement to sign.

**2016** Public unveiling on 30 June, with the first sponsors.

**2017** UNESCO and Inria sign an agreement on the access to and the preservation of software source code, on 3 April.

**2018** The archive opens to the world on 7 June, at UNESCO headquarters.

**2019** The Paris Call on Software Source Code is published on 26 February, after an expert meeting convened at UNESCO. The archive's graph is released as an open dataset.

**2020** The EOSC SIRS report, *Scholarly Infrastructures for Research Software*, sets out how research software should be archived, referenced, described and credited.

**2026** The Public Code Observatory is unveiled at UN Open Source Week, and this report is one of its analyses.



**The archive.**
softwareheritage.org



**The Public Code Observatory**, where this report and its companions are published.
stateofpubliccode.org

# 11 Appendix: the perimeter

This analysis counts projects with at least 1 contribution from Latin American research institutions, selected by a compiled perimeter of 8,053 characters, sha 353e40ca7926, supplied beside this report as acad-latam_domain_perimeter_regex.txt. Contribution volume is counted as distinct revisions and releases, never as their sum over the origins that reach them.

Reproducibility: graph export 2026-06-04, toolchain 1.1.0, analysis code 8228bc89b.