Software Heritage
THE GREAT LIBRARY OF SOURCE CODE

State of Public Code 2026

Indicators Report

the core per-country indicators and ranking tables

How governments contribute to open source

A Public Code Observatory edition

Preliminary findings, open for comment by all stakeholders, including UN member states.

with support from

unesco

United Nations
Office for Digital and
Emerging Technologies

Part of the State of Public Code 2026 family — Indicators Report · Companion dataset & reproduction guide.

Edition 2026-10

Updates the findings first presented at UN Open Source Week, June 2026.

Contents

Executive summary 2

§1 — Coverage and scope 3
- §1.1 By continent and subcontinent . . . 3
- §1.2 By income group . . . 4
- §1.3 By official language . . . 4
- §1.4 World map . . . 5
- §1.5 By country (top 30) . . . 6
- §1.6 By distinct projects touched (breadth-of-footprint view) . . . 7

§2 — Trends over time 10

§3 — Cross-border collaboration 11

§4 — Sustainability: 1-year activity 12

§5 — What predicts per-country output? 13

§6 — Digital Public Goods 14

§7 — Ministry-type breakdown (with structural caveat) 15

§8 — AI-related projects per government 16

§9 — SDG coverage (indicative lower bound) 17

§10 — Quality and metadata signals 18
- §10.1 Licensing . . . 18
- §10.2 Programming languages . . . 18

§11 — Concentration: Gini-Lorenz curves 20

Annex A — Methodology 21

Annex B — Sources 21

Annex C — Key findings at a glance 22

Annex D — The many paths to contribute 24

Executive summary

Open source is global infrastructure that governments everywhere help build. This report measures the public open-source footprint of the UN member states — every repository in the Software Heritage archive that received at least one contribution from a government-domain email.

Headline findings

- **Coverage.** 240,922 distinct public repositories carry at least one government-email contribution (82,438 distinct contributors). **12,667** had an institutional contribution in the 12 months ending at the graph snapshot.
- **The measured contributions do not yet cover the perimeter.** The published policy table lists 2,785 government domains across 195 countries and territories, all **193 UN member states** plus Palestine and the Holy See. Of those 195, **164** have at least one contribution in this snapshot, so **31 are silent**: their domains are in scope and nothing was measured behind them. A silent country is a gap in what we can see, not evidence that nothing is being built there, and closing it is the most useful thing a reader can help with. If your institution is missing or mis-attributed, propose a correction at world-gov-domain-names or use the participation form in Annex D.
- **The strongest predictor of per-country output is not wealth.** Per-country contributor count correlates with the UN E-Government Development Index at Spearman ρ = **0.53** (n = 160): 3.1× the rank-variance of GDP per capita (0.30); removing the US outlier moves no correlation by more than 0.013.
- **Trends.** Activity grew from the first recorded contribution in 1995 to **24,504 newly-active repositories in 2022** — the most recent year we can confidently quote (see §2 and §4).
- **Cross-border collaboration is rare but real.** 1,997 repositories (0.8%) carry contributors from more than one country; the densest bilateral pair is **France↔United States on 709 shared projects**, and France–Germany now appears on 15.
- **Sustainability is sober.** Across **122,540 repositories in the 2018–2022 cohorts**, roughly 4 out of 5 fall silent within a year.
- **Digital Public Goods: deploy-not-contribute.** Only **11** vetted DPGs have any government-email contribution in our universe, across 8 countries (see §6).
- **AI maturity.** Netherlands leads on AI-share-of-output at 7.8% (3.82× the global mean of 2.0%).

Scope. This report measures government **contributions** to open source — code authored from government email domains and archived by Software Heritage — **not** software usage or adoption. The dataset counts repositories that government contributors *touched*, not repositories that governments *own*: a `.gov.uk` engineer pushing a typo fix to a Linux-kernel mirror puts that mirror in the dataset (surfaced where it matters most — collaboration; see Annex A). Every count is taken after the noise filter described there.

§1 — Coverage and scope

The perimeter is a published policy table, maintained independently of this report at world-gov-domain-names. This run read 2,785 included domains (1,865 subtree, 908 exact, 12 umbrella) against 8,672 exclusions. The pipeline's pattern is compiled from that table; each origin is attributed to a country by longest-suffix match of the author email domain.

- **Origins:** 240,922 public repositories.
- **Contributions (revisions + releases):** hundreds of millions of archived revisions and releases.
- **Single-country resolution:** the overwhelming majority of contributors resolve to a single country; the small remainder are the cross-border collaborators of §3.

§1.1 By continent and subcontinent

Region	Contributors	Share
Americas	48,278	58.6%
Europe	23,500	28.5%
Asia	6,011	7.3%
Oceania	2,759	3.3%
Middle East	924	1.1%
Africa	889	1.1%
Europe (non-EU)	63	0.1%
Unknown	14	0.0%

The Americas dominate because the United States routes nearly all federal contributors through agency subdomains under the bare `.gov` TLD (NIST, NASA, DOE laboratories, USDA; see §7). Europe is second, lifted by the UK (`gov.uk`) and France (`gouv.fr`). Africa, at about 1% of volume, has the most headroom.

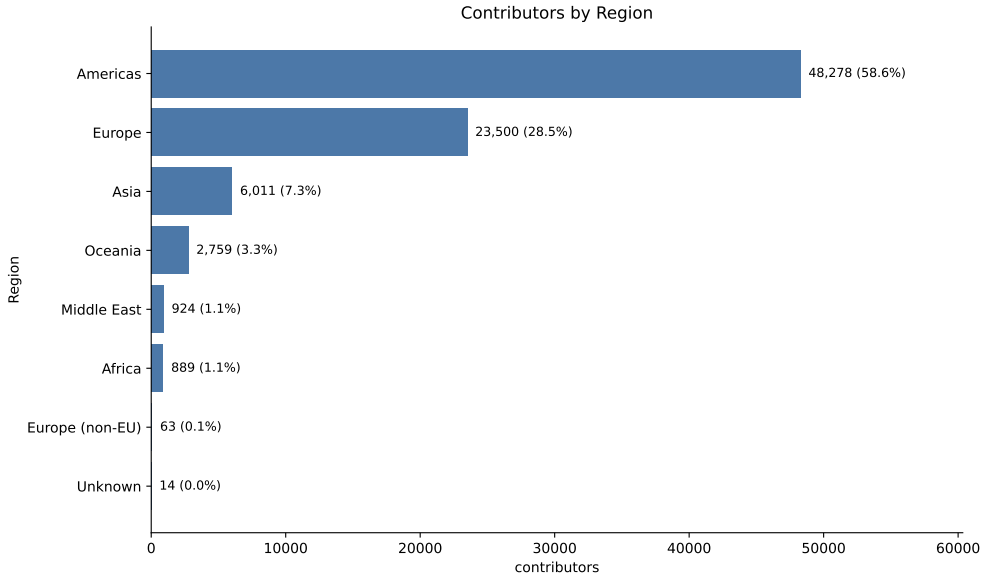


Figure 1: *Contributors by world region (continent-level rollup).*

§1.2 By income group

Income group	Contributors	Share
High income	65,615	79.6%
Upper-middle income	14,469	17.6%
Lower-middle income	2,125	2.6%
Low income	215	0.3%
Unknown	14	0.0%

High-income countries account for roughly three-quarters of measured contribution, but the **upper-middle-income share is meaningful**; that bracket (Brazil, Argentina, Mexico, South Africa, Indonesia, Malaysia, Türkiye, Russia) drives most of the non-US, non-European volume. §5's EGDI correlation predicts this pattern best; raw wealth alone is the weakest of the five tested indicators.

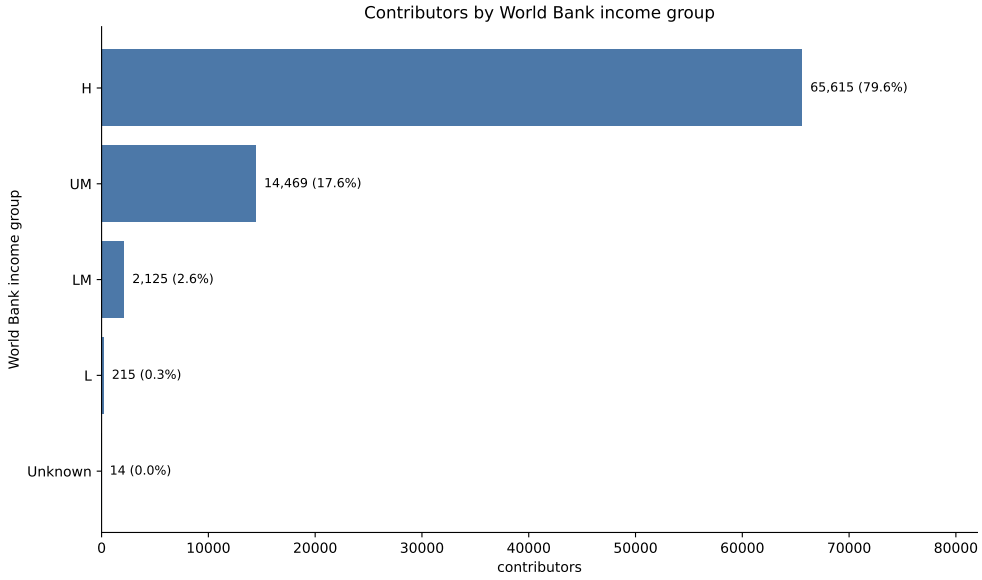


Figure 2: Contributors by World Bank income classification.

§1.3 By official language

Language	Contributors	Share
English	46,728	56.7%
French	10,457	12.7%
Spanish	7,018	8.5%
Portuguese	5,077	6.2%
Norwegian	3,397	4.1%
Unknown	1,038	1.3%
Japanese	905	1.1%
Indonesian	874	1.1%
Arabic	862	1.0%
German	730	0.9%

Portuguese remains prominent largely because of Brazil (the `gov.br` aggregation discussed in §1.1 and §7). Spanish is more diffuse: Mexico, Argentina, Colombia, Spain, Peru, Chile and the smaller

Latin American republics jointly. English's dominance reflects the United States plus the Anglophone Commonwealth (UK, Australia, Canada, New Zealand, Ireland) more than any community-of-practice effect. Language here tracks the largest national contributor blocs, not any intrinsic property of the code.

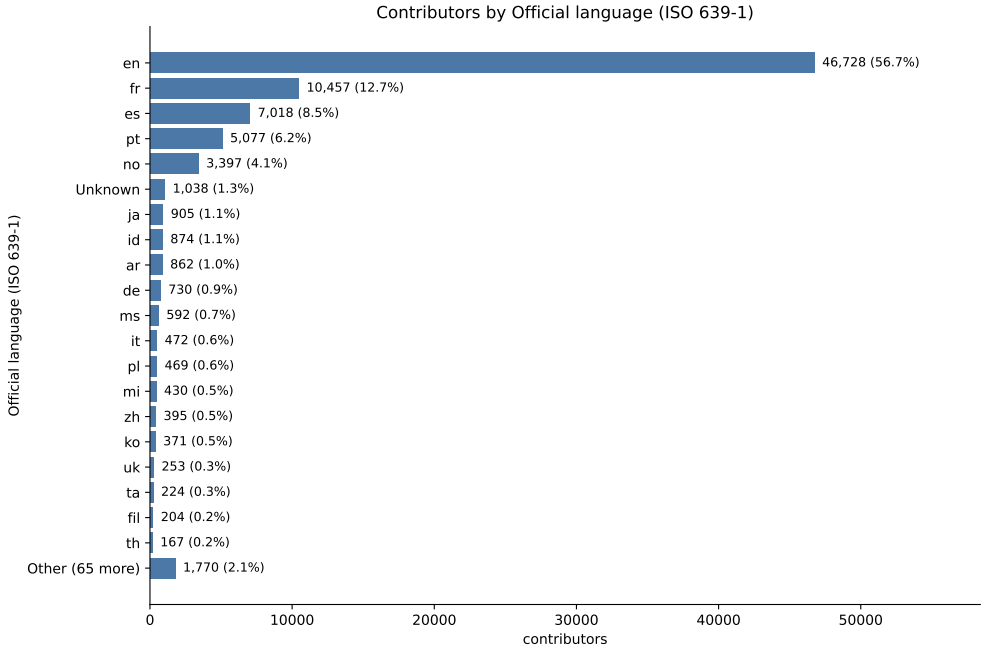


Figure 3: *Contributors by national official language (top 20; the rest grouped as Other).*

§1.4 World map

The US-dominance and Europe-second pattern is visually obvious; the emerging Brazilian, Indian, Indonesian and South African footprints stand out at the second tier, as does the near-absence across the Sahel, the Arabian Peninsula and Central Asia. Countries below the colour-bin floor or with no matched contributor are unshaded.

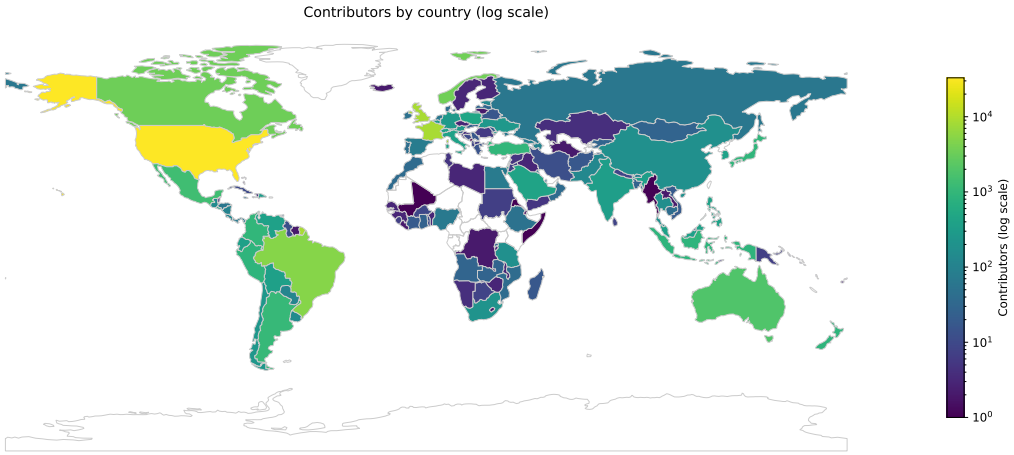


Figure 4: *Contributor count per country in the perimeter (viridis colour map on a log scale).*

§1.5 By country (top 30)

The per-country distribution is right-skewed by orders of magnitude: a dominant United States, then a second tier, then a long tail of governments with modest counts each.

#	Country	Contributors	Share
1	United States	32,933	39.9%
2	United Kingdom	8,603	10.4%
3	France	8,571	10.4%
4	Brazil	4,968	6.0%
5	Norway	3,397	4.1%
6	Canada	3,283	4.0%
7	Australia	1,861	2.3%
8	Mexico	1,373	1.7%
9	Argentina	1,062	1.3%
10	Colombia	948	1.1%
11	Turkey	924	1.1%
12	Japan	905	1.1%
13	Singapore	881	1.1%
14	Indonesia	874	1.1%
15	New Zealand	860	1.0%
16	Peru	857	1.0%
17	Poland	469	0.6%
18	Saudi Arabia	412	0.5%
19	Philippines	408	0.5%
20	Italy	380	0.5%
21	Venezuela	377	0.5%
22	Malaysia	372	0.5%
23	El Salvador	371	0.5%
24	South Korea	371	0.5%
25	Switzerland	370	0.4%
26	Ecuador	357	0.4%
27	Bolivia	354	0.4%
28	India	332	0.4%
29	Dominican Republic	289	0.4%
30	Belgium	278	0.3%

This right-skew is the load-bearing motivation for §11 (Gini-Lorenz concentration): the per-country distribution is as unequal as extreme wealth-concentration measures.

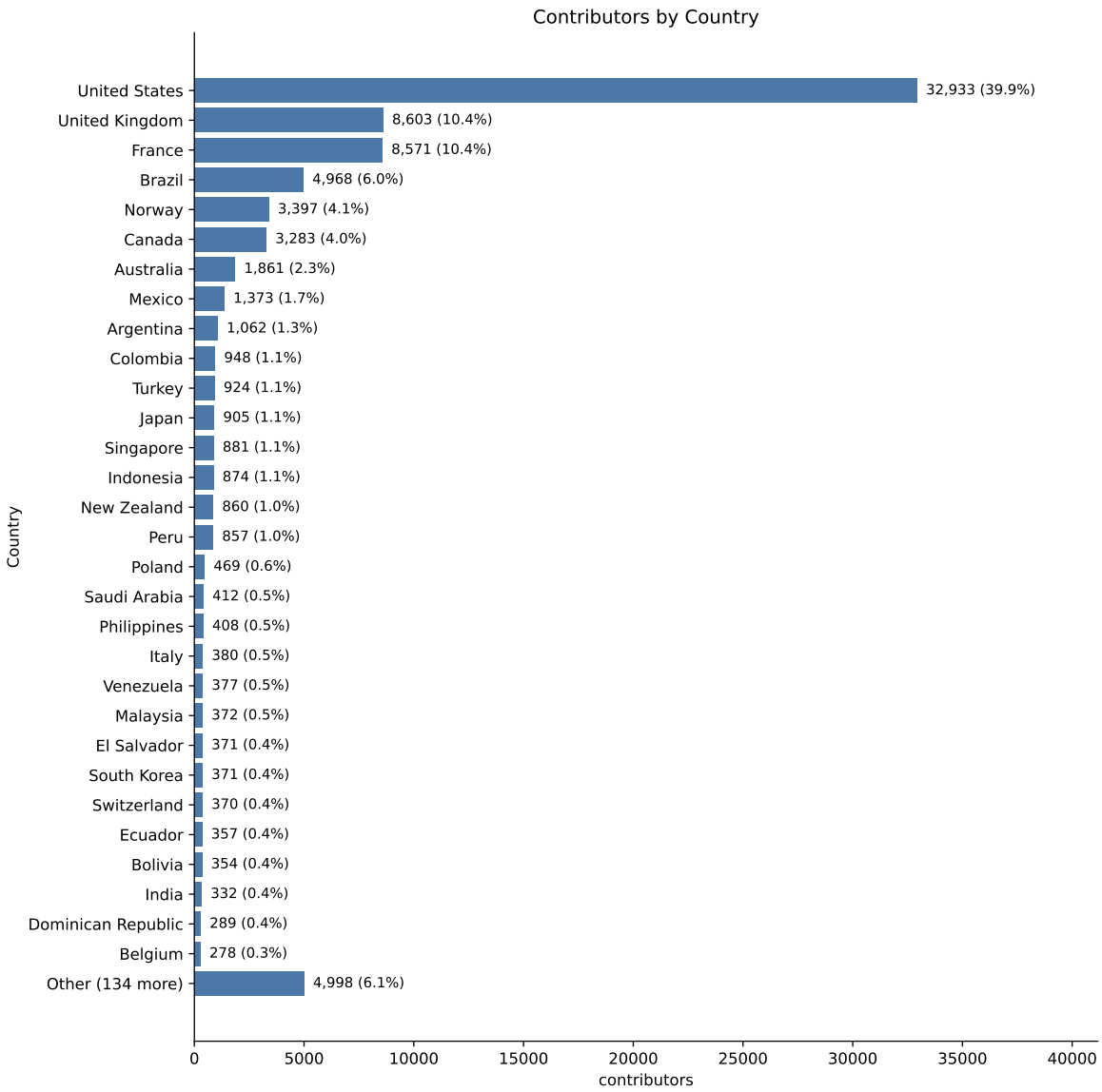


Figure 5: *Top-30 countries by contributor count, the rest grouped as Other (linear scale; the distribution is strongly right-skewed).*

§1.6 By distinct projects touched (breadth-of-footprint view)

The contributor-count view in §1.5 weighs the United States heavily because the federal government splits its volume across many agency subdomains under bare `.gov`. Countries whose entire civil service routes through a single national TLD (`canada.ca`, `gov.uk`, `gouv.fr`) look smaller in that view than their actual programme footprint warrants.

The **distinct-projects-touched** view corrects this: per country, the number of distinct deduplicated repositories that received at least one contribution from a gov-email author resolving to that country. This is **breadth**, not volume.

The two rankings agree at the very top and diverge further down. The movers listed after the table are the governments whose position changes most between the contributor view and the project view.

#	Country	Distinct projects	Share of the corpus
1	United States	98,102	40.7%
2	France	32,525	13.5%
3	United Kingdom	29,009	12.0%
4	Brazil	12,669	5.3%
5	Norway	11,535	4.8%
6	Canada	8,799	3.7%
7	Australia	5,391	2.2%
8	Mexico	3,227	1.3%
9	Japan	3,149	1.3%
10	Argentina	2,890	1.2%
11	Singapore	2,559	1.1%
12	New Zealand	2,343	1.0%
13	Turkey	2,258	0.9%
14	Colombia	1,991	0.8%
15	Indonesia	1,830	0.8%
16	Peru	1,697	0.7%
17	Italy	1,176	0.5%
18	Poland	1,080	0.4%
19	Bolivia	1,027	0.4%
20	Philippines	936	0.4%
21	Ecuador	920	0.4%
22	El Salvador	905	0.4%
23	South Korea	849	0.4%
24	Saudi Arabia	825	0.3%
25	Switzerland	797	0.3%
26	Venezuela	761	0.3%
27	Belgium	745	0.3%
28	Dominican Republic	736	0.3%
29	Austria	713	0.3%
30	Estonia	699	0.3%

Shares are taken against the 240,922 projects of the corpus. A project touched by several governments counts under each of them, so the column adds up to more than 100%. 114 of 164 attributed governments touch ≥ 10 distinct projects (50 below that threshold). Notable breadth-vs-volume movers (contributors-rank minus projects-rank, so a positive value means the government ranks higher by projects touched than by contributors): Japan (+3), New Zealand (+3), Colombia (-4), Italy (+3), Bolivia (+8), Ecuador (+5), Saudi Arabia (-6), Venezuela (-5), Belgium (+3).

For the UN audience the project-count view is arguably the more meaningful indicator of how much of the public-software commons each government engages with; the contributor-count view indicates the size of the gov-OSS engineering workforce that happens to use a gov email.

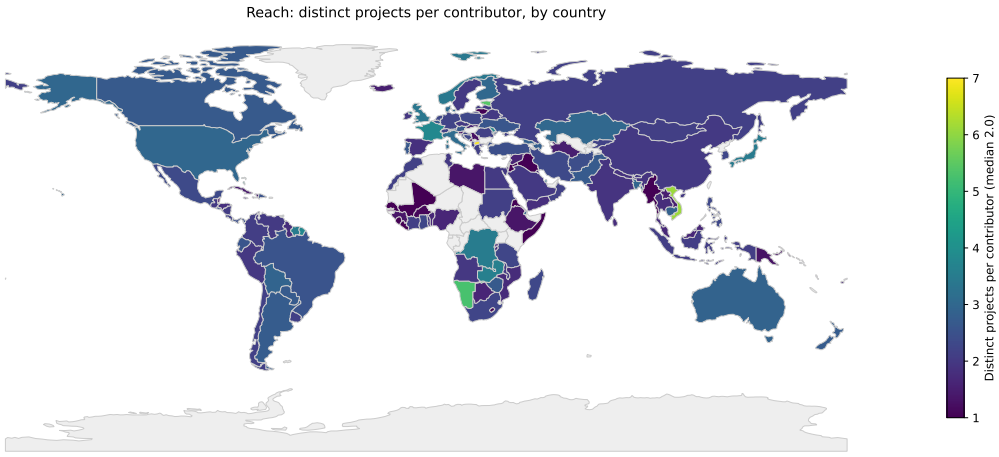


Figure 6: *Reach intensity: distinct projects touched per contributor, by country. Same viridis colour map as the contributor map, but a linear scale on a ratio rather than a log scale on a count, so a given colour does not mean the same thing on the two maps. Because raw contributor and project counts are nearly collinear, this ratio rather than absolute project count is what reveals a distinct geography.*

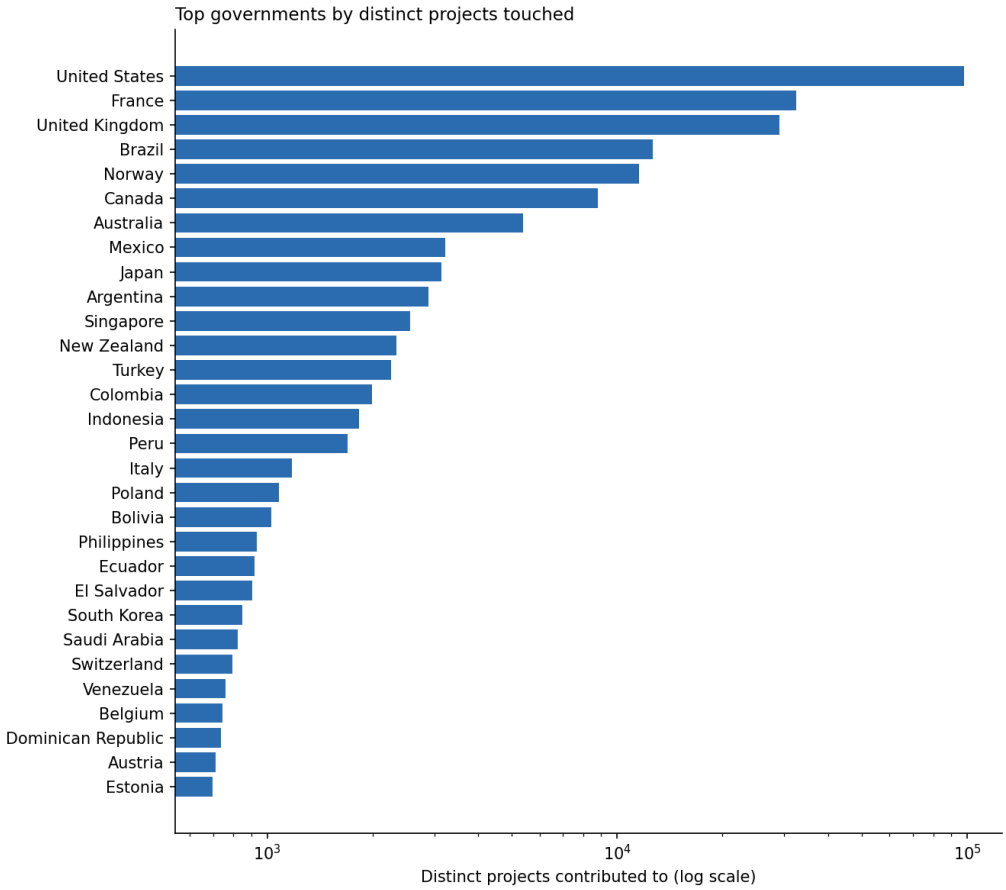


Figure 7: *Top-30 governments by distinct projects touched (log scale).*

§2 — Trends over time

The first government-email contribution observed in our universe is **1995** (10 repositories). Growth is roughly exponential from 2000 to about 2018, then a high plateau of new government-contribution repositories per year.

Archive ingestion horizon. The archive's ingestion of new repositories has lagged since 2022; the snapshot under-counts 2023–2024 activity for crawling-cadence reasons unrelated to government open-source production. Every figure in this section caps the plot at **2022** — the most recent year we can confidently quote — rather than telling a "government open source is collapsing" story the data does not support.

By region: the Americas (largely US `.gov`) dominate; Europe sits second (lifted by `gov.uk` and `gouv.fr`); Asia grew fastest in 2018–2022; Africa visibly emerges from 2015 onward.

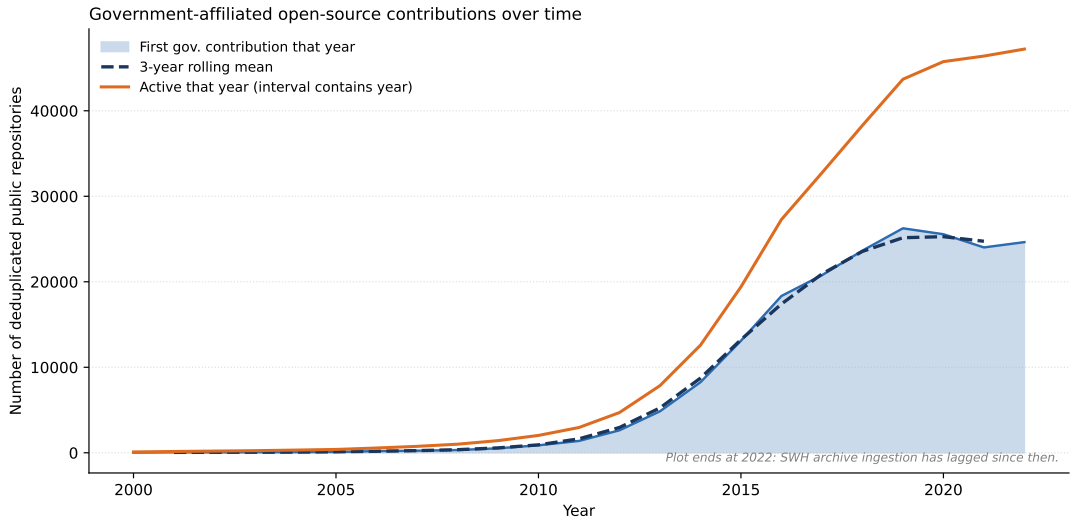


Figure 8: Repositories receiving their first government contribution each year, with a 3-year rolling mean, and repositories active in each year, 2000–2022 (the plot window starts in 2000; the first recorded contribution is earlier).

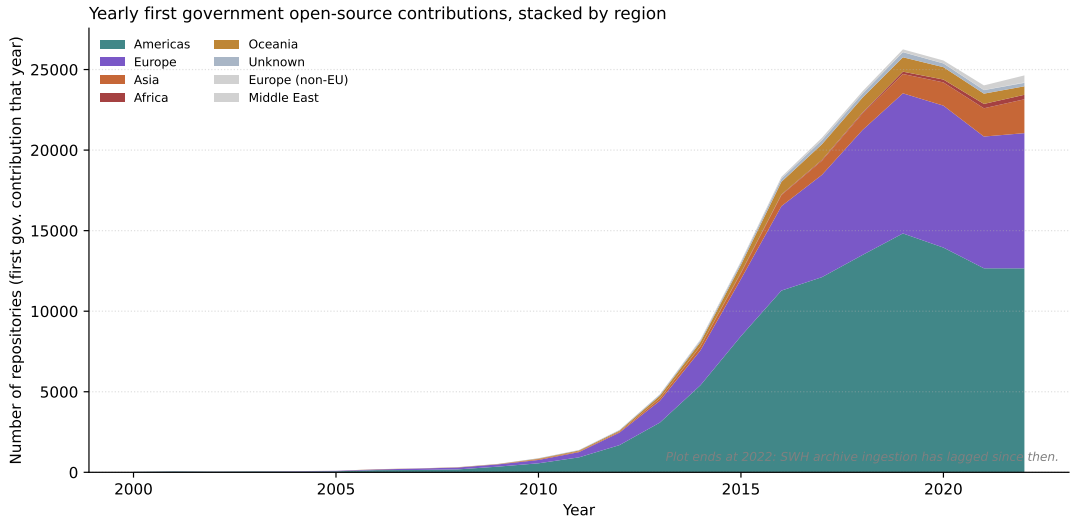


Figure 9: Yearly new contributions stacked by region, 2000–2022.

§3 — Cross-border collaboration

A repository counts as "international" when contributors from more than one UN member state appear on it. Cross-border collaboration on a shared codebase is rare in relative terms but real and concrete — open source lets government technologists co-maintain code across borders with minimal friction, often absent any bilateral agreement.

Metric	Value
Repositories with contributors from > 1 country	**1,997** (0.8%)
France↔United States	709 shared projects
United Kingdom↔United States	347 shared projects
Canada↔United States	206 shared projects
Australia↔United States	194 shared projects
Norway↔United States	107 shared projects
France ↔ Germany	15 shared projects

Interpretation note. The most-international repositories at the very top of the distribution are public mailing-list mirrors (kernel.org netdev / lkml public-inbox archives), not single jointly-developed codebases. This is correct — Linux-kernel governance is genuinely cross-border — but worth surfacing. After filtering the mailing-list shape, the top genuinely cross-border code repositories are shared operational infrastructure (CI recipes, Helm charts, Kubernetes operator catalogues) for public-sector deployments.

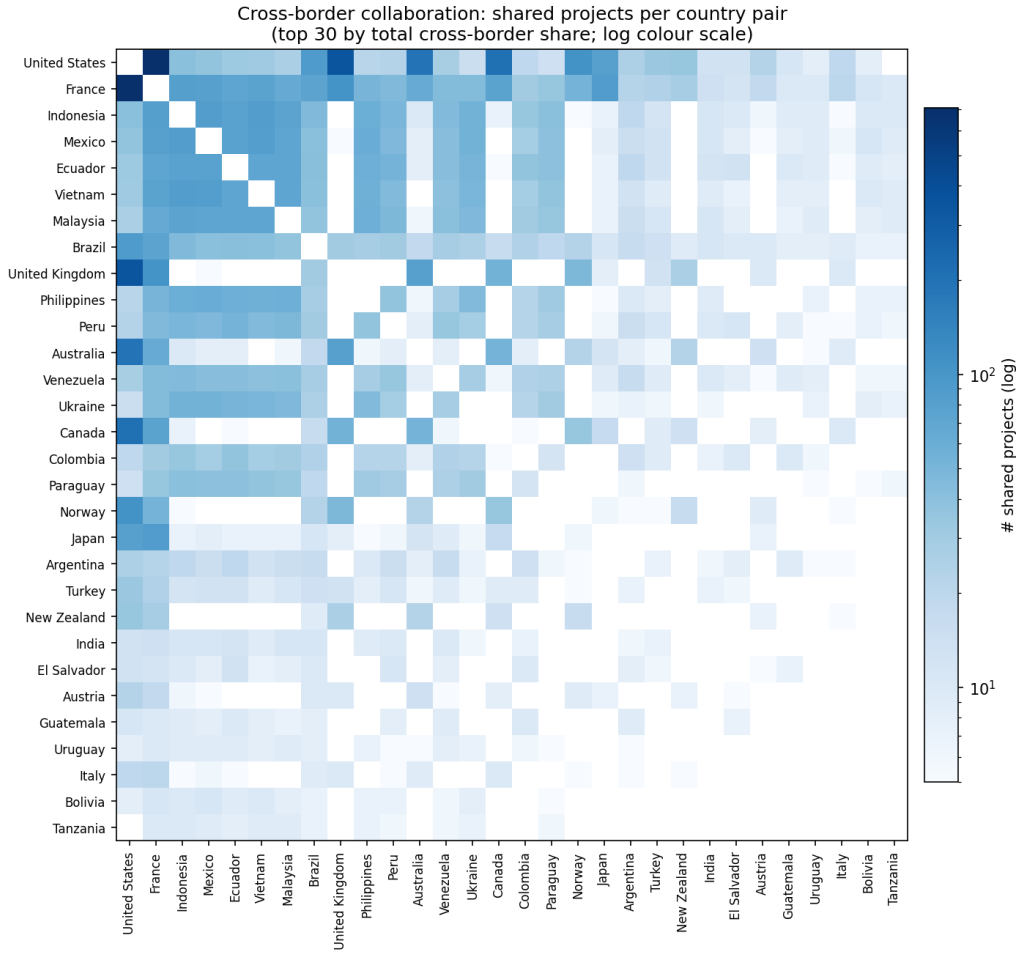


Figure 10: Country pairs by shared-project count, top 30 × top 30 (log scale).

§4 — Sustainability: 1-year activity

We define the cohort as repositories whose *first* government-affiliated contribution falls in calendar year Y; a repository is "active at 1 year" if the *last* government-affiliated contribution arrives ≥ 365 days later. The 2023–2024 cohorts are heavily right-censored against the archive horizon. The sustainability claim therefore pools the complete 2018–2022 cohorts and weights their active share by the measured number of repositories in each cohort.

Cohort year	Cohort size	Share active at 1 year
2018	23,018	21.4%
2019	25,863	21.5%
2020	25,318	22.0%
2021	23,837	19.0%
2022	24,504	16.0%
2023	24,472	11.2%
2024	15,569	6.6%

In plain terms: across 122,540 repositories in the 2018–2022 cohorts, roughly 4 out of 5 fall silent within a year. This is the single most actionable indicator for any framework that wants to measure sustained public-sector engagement with open source.

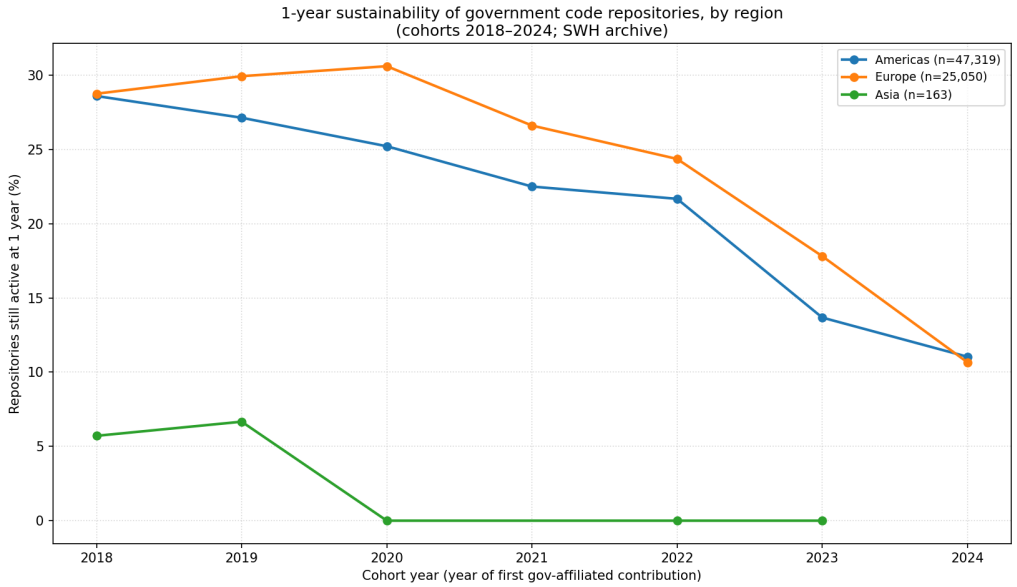


Figure 11: *Share of new-contribution cohorts still active one year later.*

§5 — What predicts per-country output?

Government digital capacity matters more than government budget. Where a government has invested in digital-administration capacity — the substance of EGDI: online-service maturity, telecommunications, human-capital readiness — measurable open-source output follows. Raw wealth, represented by GDP per capita, is the weakest of the tested predictors.

Indicator	n	Spearman ρ (with US)	ρ (without US)	Δ(US)
UN EGDI 2024	160	0.53	0.53	0.007
UN HDI	159	0.37	0.36	0.009
Internet users %	153	0.33	0.32	0.008
ITU IDI 2024	144	0.31	0.30	0.013
GDP per capita	158	0.30	0.28	0.012

EGDI leads GDP per capita by 3.1× in rank-variance (ρ^2). The signal is genuinely cross-country, not an artefact of any single large contributor: every correlation barely moves when the US outlier is removed.

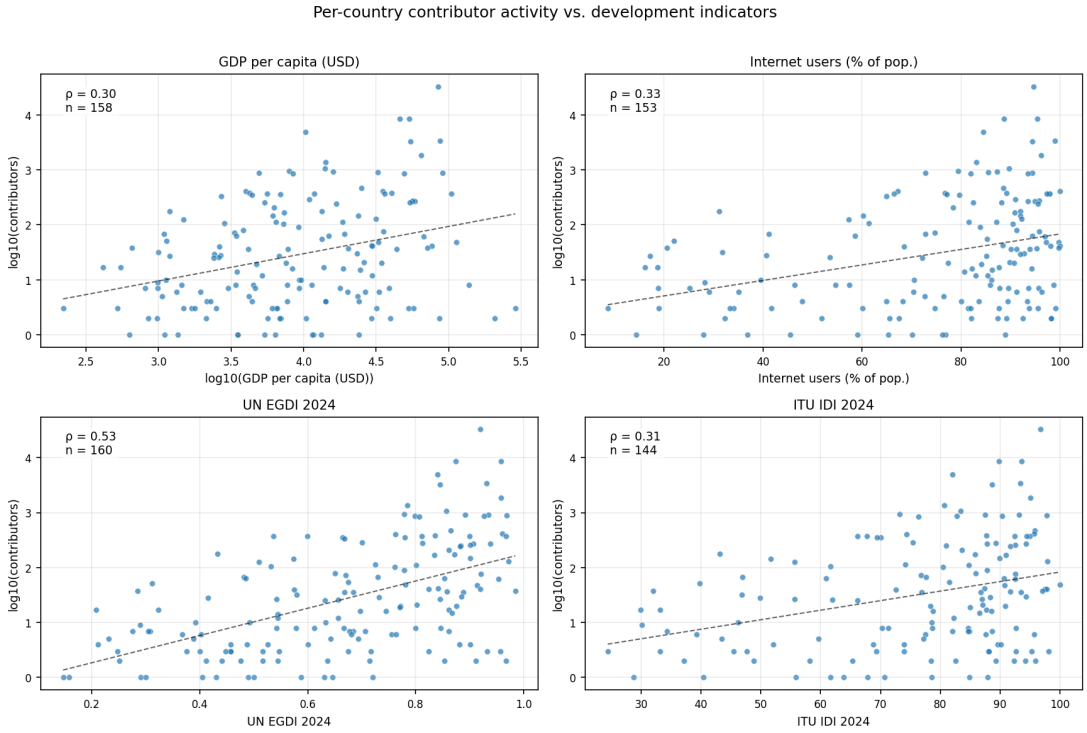


Figure 12: *Per-country contributor count vs four of the five development indicators in the table (HDI is not plotted).*

Sources. UN EGDI 2024 via the World Bank Data360 mirror; UN HDI 2022 from the UNDP HDR composite-indices series; ITU IDI 2024 from the ITU statistics workbook; GDP per capita and internet-users-% from World Bank Open Data.

§6 — Digital Public Goods

The Digital Public Goods Alliance maintains a curated registry of vetted Digital Public Goods: open-source software that advances the Sustainable Development Goals. We cross-referenced the registry against our universe.

- DPGs with a direct government contribution: **11** across **8** countries.
- Registry coverage: government upstream contribution reaches only a minority of the registered Digital Public Goods.

The dominant pattern is: governments deploy DPGs widely; very few contribute upstream. This is potentially actionable — a "contribute back" framing in DPG procurement and digital-cooperation policy could meaningfully grow the upstream contributor pool.

Validation. Re-running the attribution with the full per-revrel join (every gov-email author who ever touched each DPG, against the curated country domain map) held the numbers: the same DPGs, the same contributing countries, no per-DPG growth in country count. The underlying institutional populations are small and concentrated by country, not diffuse — a load-bearing finding, not a methodological floor. Civil servants who contribute to DPGs under personal email addresses are deliberately outside the gov-domain frame.

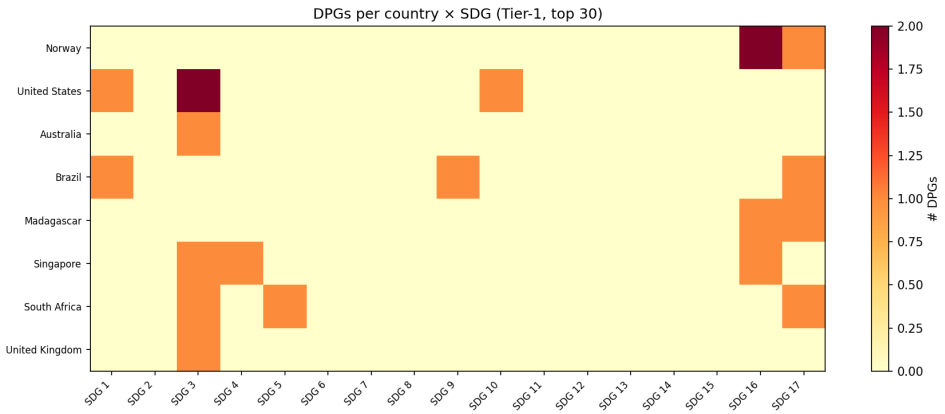


Figure 13: DPG contributions by country and SDG (top countries × 17 SDGs).

§7 — Ministry-type breakdown (with structural caveat)

We classified each government email domain into a 12-bucket ministry taxonomy (Health, Education, Defense, Finance, Justice, Foreign Affairs, Environment, Transport, Science & Research, Digital & Statistics, Interior, Other). Top domains were hand-curated; the long tail uses conservative substring patterns; the rest defaults to "Other government".

#	Ministry bucket	Contributors
1	Other government	50,738
2	Environment	18,574
3	Science & Research	5,656
4	Health	3,543
5	Digital & Statistics	1,744
6	Interior	926

Of the 31,700 contributors attributed to a *specific* ministry (not the catch-all bucket), **98.2% are United States**.

Critical structural finding. The overwhelming majority of ministry-attributed contributors are United States. The US is the only large-volume country that routes ministry-level contributors through agency-specific subdomains (`nasa.gov`, `nih.gov`, `ornl.gov`, `state.gov`, ...). Every other country aggregates ministry contributors at a single national TLD — `gov.br`, `gov.uk`, `gov.au` — all collapsed under "Other government" because the TLD does not expose ministry granularity. This is **a structural feature of government-email-domain hygiene, not a defect of the classification lexicon** — and the strongest argument for a follow-up LLM-based classification pass that reads description/README/metadata to assign ministry where the domain alone cannot.

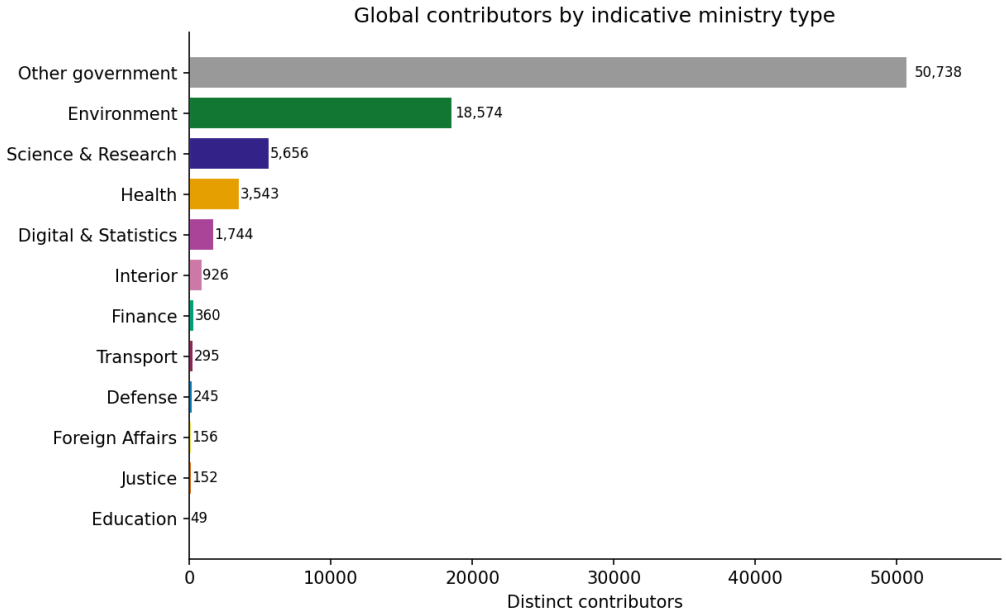


Figure 14: Top ministries globally, ranked by attributed contributor count.

§8 — AI-related projects per government

We classified each repository as AI-related using a rule-based combination of signals on its description, keywords, programming languages, and name. This is *indicative* — roughly a fifth of hits are false positives in a hand sample (acronym collisions: NLP for Non-Linear Programming, ML for Markup Language, ROM for Reduced-Order Models).

Top by absolute AI-repository count: United States (2609), France (1005), United Kingdom (292), Canada (173), Norway (159). **Top by AI-share-of-output** (≥ 100 attributed repos; global mean 2.0%): Netherlands (7.8%, 3.82×); Japan (3.7%, 1.82×); France (3.1%, 1.51×); Egypt (2.7%, 1.34×); United States (2.7%, 1.3×).

By region, among regions with at least 1,000 attributed repositories, the AI share of output is highest in Americas (2.2%) and lowest in Africa (0.8%).

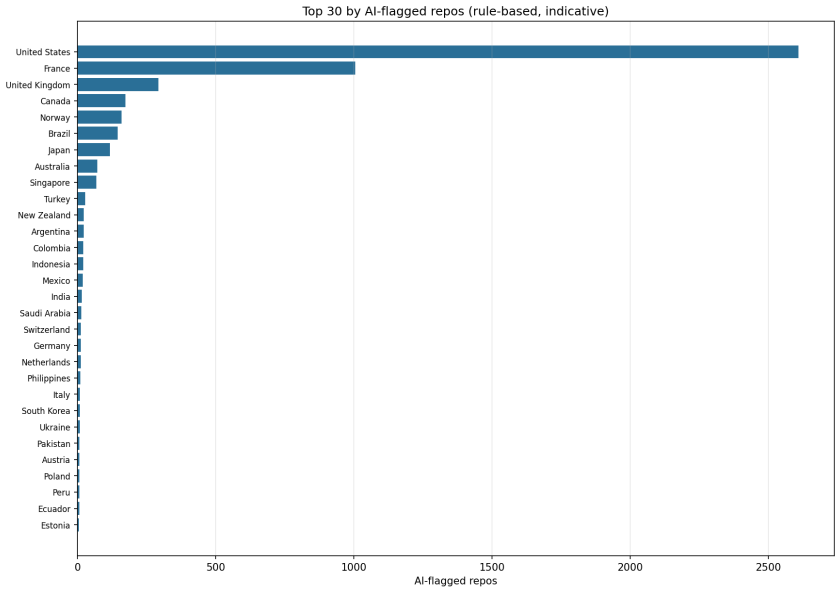


Figure 15: Top-30 countries by absolute count of AI-related repositories.

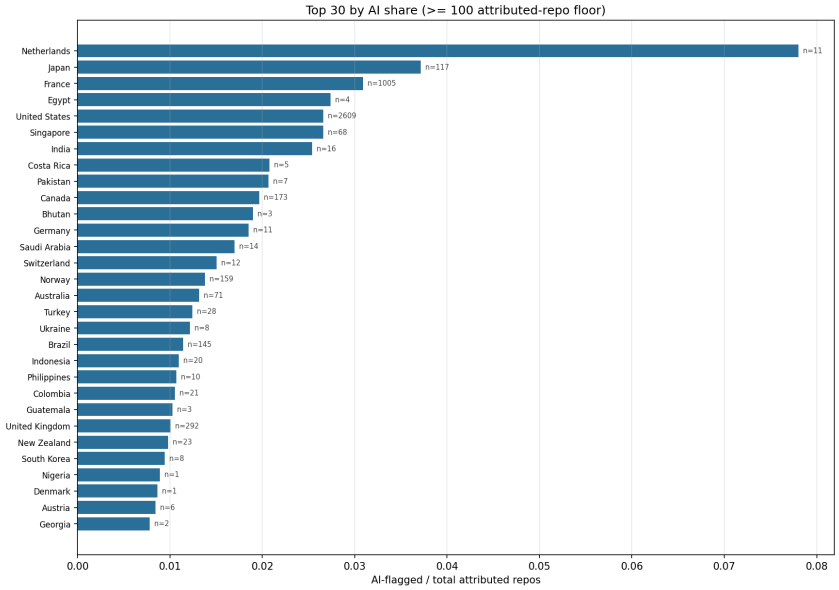


Figure 16: Top-30 countries by AI-related repository share of output.

§9 — SDG coverage (indicative lower bound)

We applied a two-tier classification: **Tier 1** inherits the SDG tags already assigned in the DPG Alliance registry (authoritative); **Tier 2** applies a curated UN-SDG keyword lexicon to repository description, keywords and name (indicative).

- Top SDGs: 3. Health (462); 16. Peace & justice (234); 4. Education (224).
- Bottom SDGs: 10. Reduced inequalities (8); 17. Partnerships (8); 5. Gender equality (3).
- SDG-tagged repositories: **1,765** (union of 1,759 lexicon-matched + 11 DPG-tier − 5 overlap).

The lowest-count SDGs almost certainly reflect lexicon-vocabulary gaps rather than absence of work — policy-shaped goals are rarely tagged with target-specific terms in repository descriptions.

The only Tier-1, DPG-confirmed African contribution is **OpenCRVS — civil registration software contributed to from Madagascar** (SDGs 16 and 17).

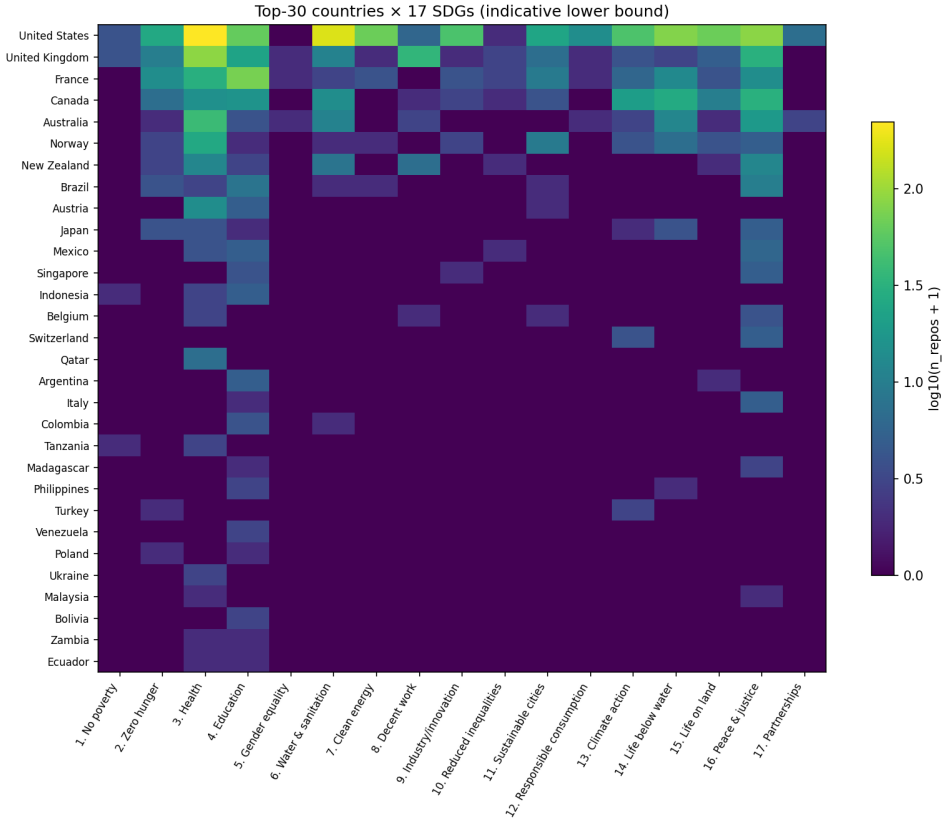


Figure 17: Top-30 countries × 17 Sustainable Development Goals (log scale).

These figures are an *indicative lower bound* on each country's SDG footprint in public software: they count only repositories whose description, keywords or name use language drawn from the UN SDG targets or a small, transparent curated lexicon — a repository using different vocabulary is silently uncounted.

§10 — Quality and metadata signals

§10.1 Licensing

License status	Count	Share
Has a valid license	58,664	24.3%
Unknown license status	16,318	6.8%
No license information at all	165,940	68.9%

Top SPDX identifiers, as a share of all repositories: MIT (11.4%), Apache-2.0 (4.0%), GPL-3.0 (2.8%), BSD-3-Clause (2.0%), GPL-2.0 (0.8%), CC0-1.0 (0.7%). **68.9%** of the universe ships no license information at all, and **75.7%** none that is usable once the unknown statuses are counted. On most platforms (notably GitHub) a missing license defaults the repository to "all rights reserved", which legally precludes reuse. The status is read from the license metadata GitHub exposes, so every origin hosted on another forge counts as "no information" here whatever its actual license, and the share without information is an upper bound. The MIT + Apache + BSD family carries **17.9%** of the universe. Of the licensed quarter, the MIT + Apache-2.0 + BSD family carries the bulk — a strong tilt towards permissive licenses over copyleft. A meaningful CC0 pocket indicates public-domain dedication, mostly from public-sector data/documentation projects.

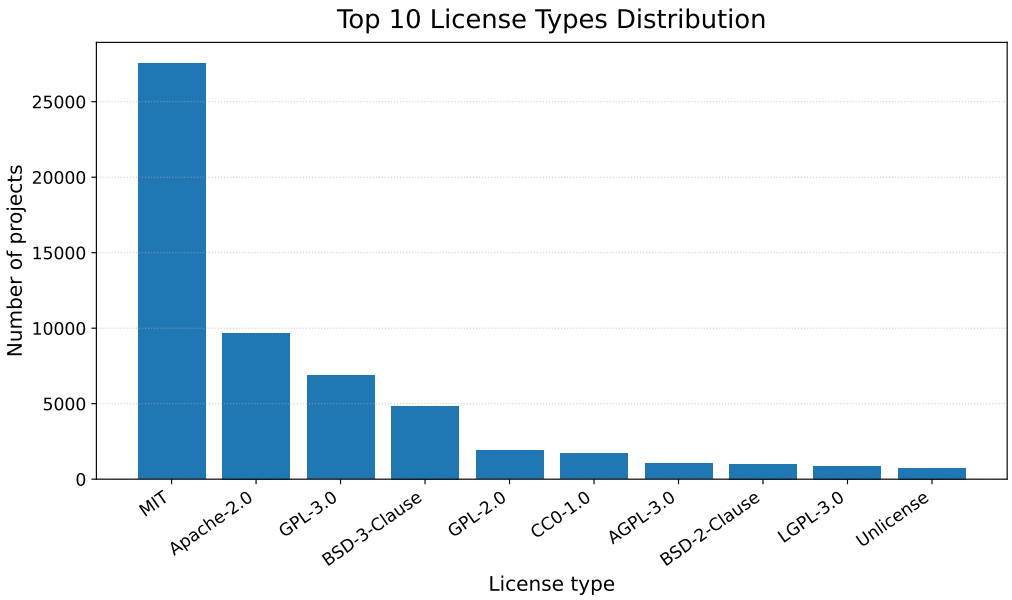


Figure 18: License-type distribution (top SPDX identifiers).

§10.2 Programming languages

The programming-language distribution mirrors the global open-source pattern — Python, JavaScript/TypeScript, Java, Go, C/C++, Ruby, PHP, R, and Jupyter Notebook. Python and JavaScript are typically the top two; Jupyter Notebook's prominence is one of the §8 AI-related signals.

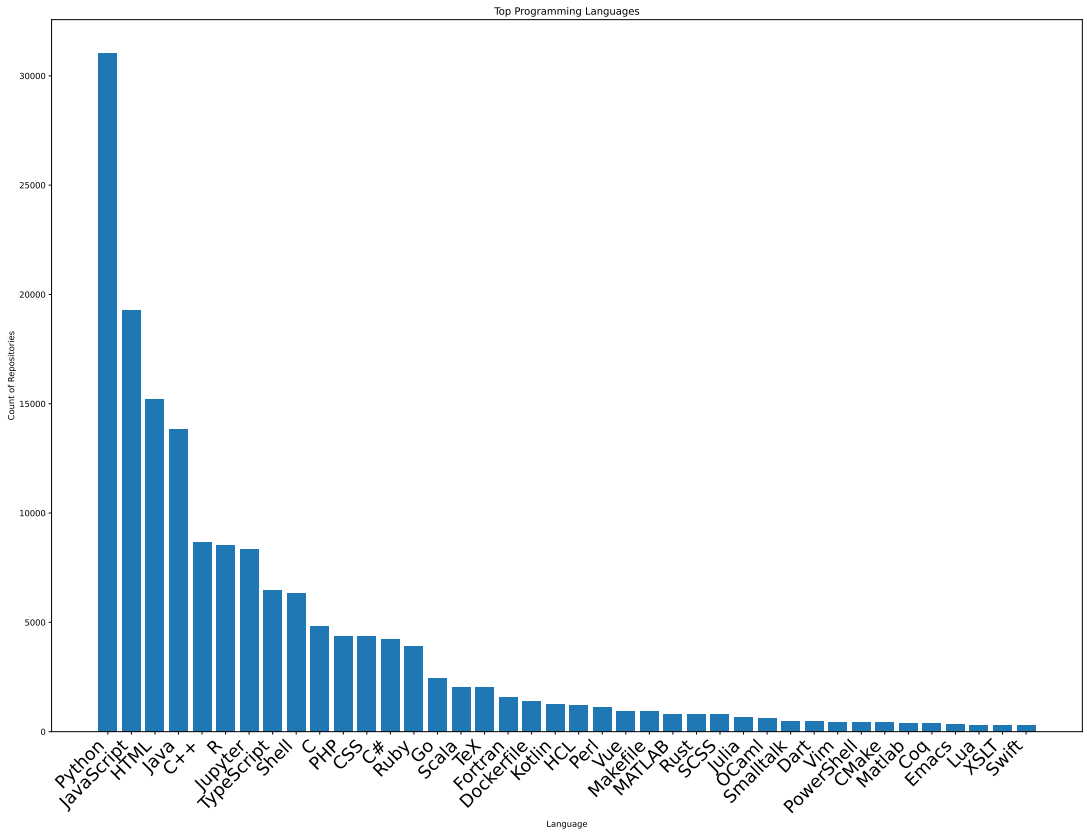


Figure 19: *Top-40 programming languages by repository count.*

§11 — Concentration: Gini-Lorenz curves

The Gini coefficient measures inequality of distribution: 0 = perfectly even, 1 = one entity holds everything. We compute it at three levels of the data.

Level	Gini	n
per_project	0.89	40
per_country	0.91	164
overall	0.90	240,922

The per-country Gini is extreme. For reference, the World Bank's most-recent global income Gini is about 0.62, and Brazil's domestic income Gini — among the highest in the world — about 0.53. Our per-country distribution of measured government open-source output is more unequal than any national income distribution.

Policy implication. Most UN member states are nearly invisible in the measurable commons not because they produce nothing, but because (per §1.1 and §7) their TLD-aggregation hygiene collapses every ministry into one bucket while the US splits into many. A non-aggregated, per-civil-servant view (which a directory-anchored attribution would enable) would soften, but not eliminate, this gap.

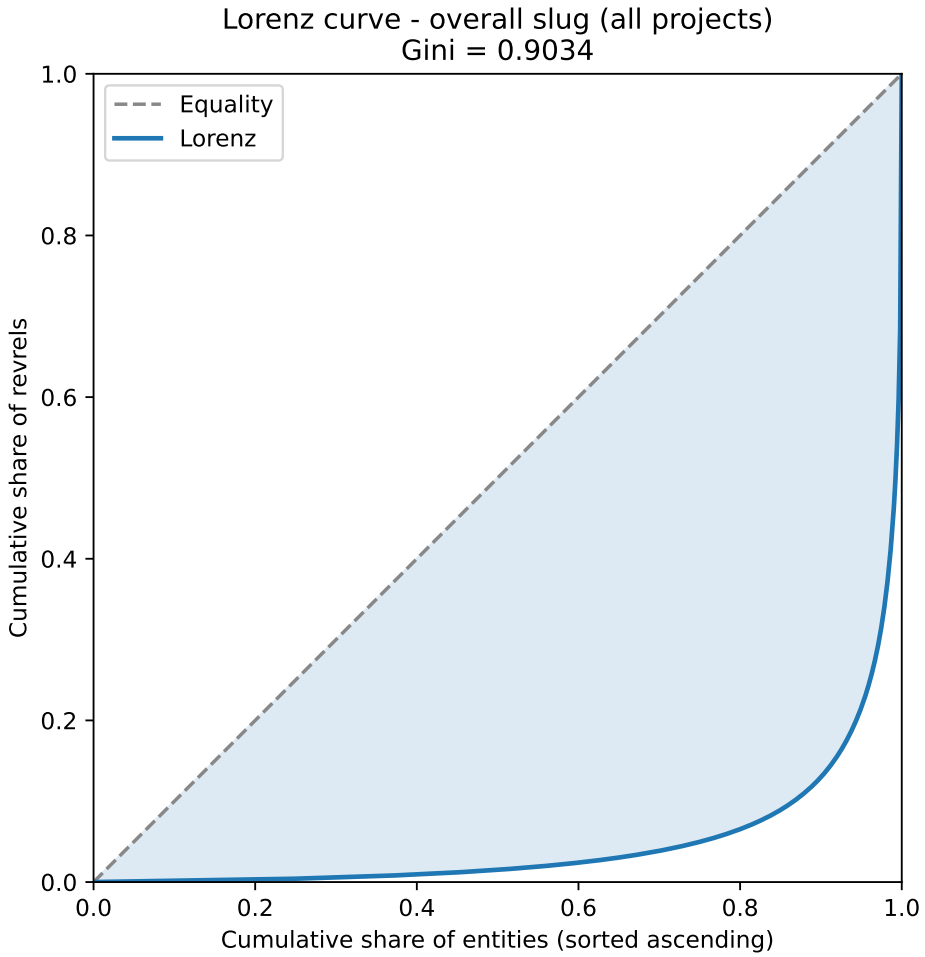


Figure 20: Lorenz curve: cumulative repositories vs cumulative contributors (global).

Annex A — Methodology

Universe. The published policy table at world-gov-domain-names, applied to the Software Heritage archive snapshot. Its counts and the revision this run read are stated below, from the run's own record rather than from this prose. Country/region/income metadata via the World Bank classification and pycountry. **Pipeline.** extract_emails → defork → dedup → analysis → build_report. **Classification cache.** SQLite, content-addressed by canonical origin URL; holds the garbage and mailing-list-mirror verdicts the analysis honours. Garbage verdicts drop an origin from the universe before any metric; mailing-list-mirror verdicts only remove it from the most-international ranking. **Touched ≠ owned.** Country metrics measure repositories where ≥ 1 government-email contributor appeared, not repositories a government owns. **Archive horizon.** Activity after ~mid-2024 is under-represented; time-series figures cap at 2022.

Perimeter of this run. 2,785 included domains against 8,672 exclusions, over 195 country codes, read from the policy table this run archived. The copy it read is fingerprinted `57a993f6bdc1167c`, a SHA-256 prefix over the fetched file rather than a repository revision: it identifies the content, so two runs can be told apart by it, but it cannot be used to fetch that version back. Source: <../data/reference/world-gov-un_domain_policy_source.csv>. The pattern compiled from that table is 200,420 characters, sha `6cc9cc3af5ab`, supplied with the dataset rather than printed here. Engine revision `b6eb57c5`.

Annex B — Sources

- Dataset: Software Heritage archive snapshot.
- Governmental-domain policy table: world-gov-domain-names (Bastien Guerry), at world-gov-domain-names.
- UN EGDI 2024 (World Bank Data360 mirror); UN HDI 2022 (UNDP HDR); ITU IDI 2024.
- World Bank GDP per capita (NY.GDP.PCAP.CD) and internet users % (IT.NET.USER.ZS).
- DPG registry: DPGAlliance/publicgoods-candidates. SDG lexicon: UN SDG goals + targets text.

Data availability. Full per-repository, per-country and per-domain data behind every figure, plus a reproduction guide and a pseudonymised per-project dataset, are published as an open companion dataset (CC BY 4.0): gitlab.softwareheritage.org/swh/products/insights/world-gov-un-insights-report.

Generated from `world-gov-un-insights-report v5.4` on 2026-10-27. Numbers are live from the analysis manifest. Reproducibility: dataset ref v5.4; analysis engine pin b6eb57c5.

Annex C — Key findings at a glance

Four structural signals from the first edition: where government contribution concentrates, the languages public code is written in, its licence status, and its hosting-platform dependency. Each is reproducible from the open companion dataset.

Where government contribution concentrates, it often dominates. Among the projects with the largest government contribution, the public sector frequently accounts for most of all contributions to that project — for example Trilinos — while some larger communities, such as Spack, remain majority-external. These leading projects are predominantly scientific and high-performance-computing libraries (Trilinos, PETSc, Kokkos), earth-system and climate models (NASA's GEOS-ESM, the hydrology codes Amanzi and PFLOTRAN), government digital-service and security platforms (France's etalab, the US design system, CISA tooling), and scientific-imaging tools (SimpleITK).

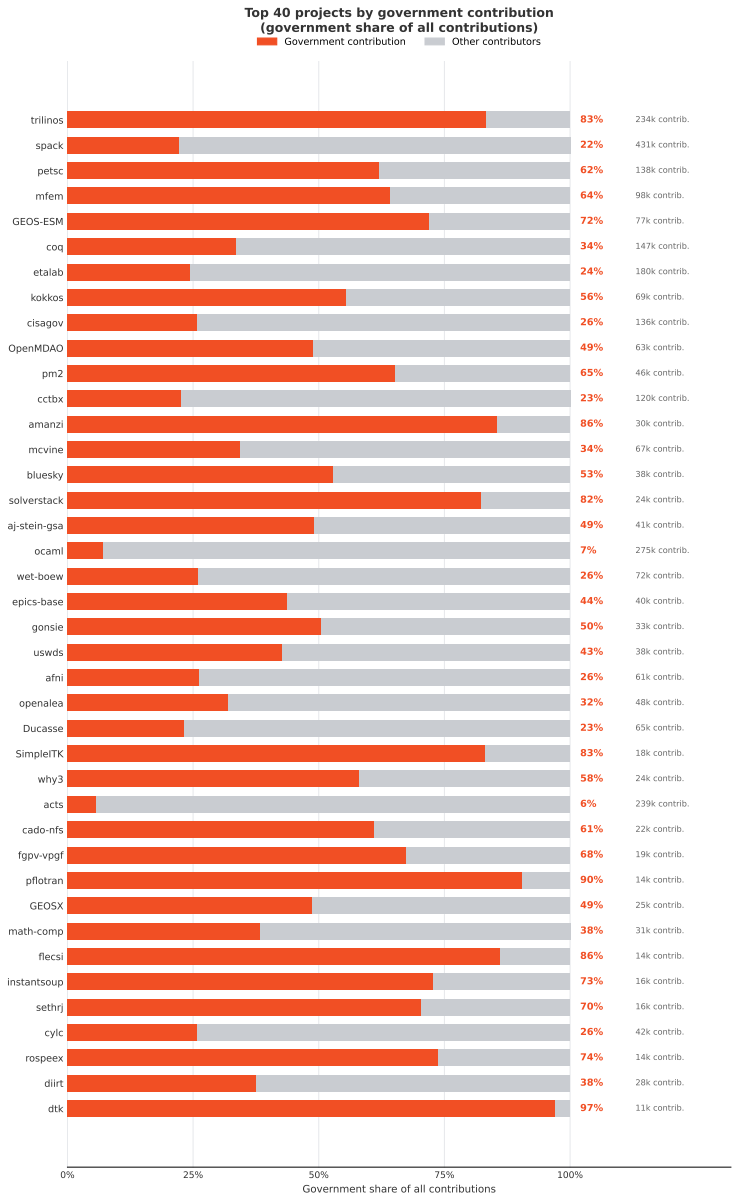


Figure 21: Top 40 projects by government contribution, showing the government share of all contributions (absolute volume annotated).

Public code is written in mainstream languages. Python, C++, C and JavaScript lead, placing public-sector code within the wider open-source ecosystem rather than in a niche of its own.

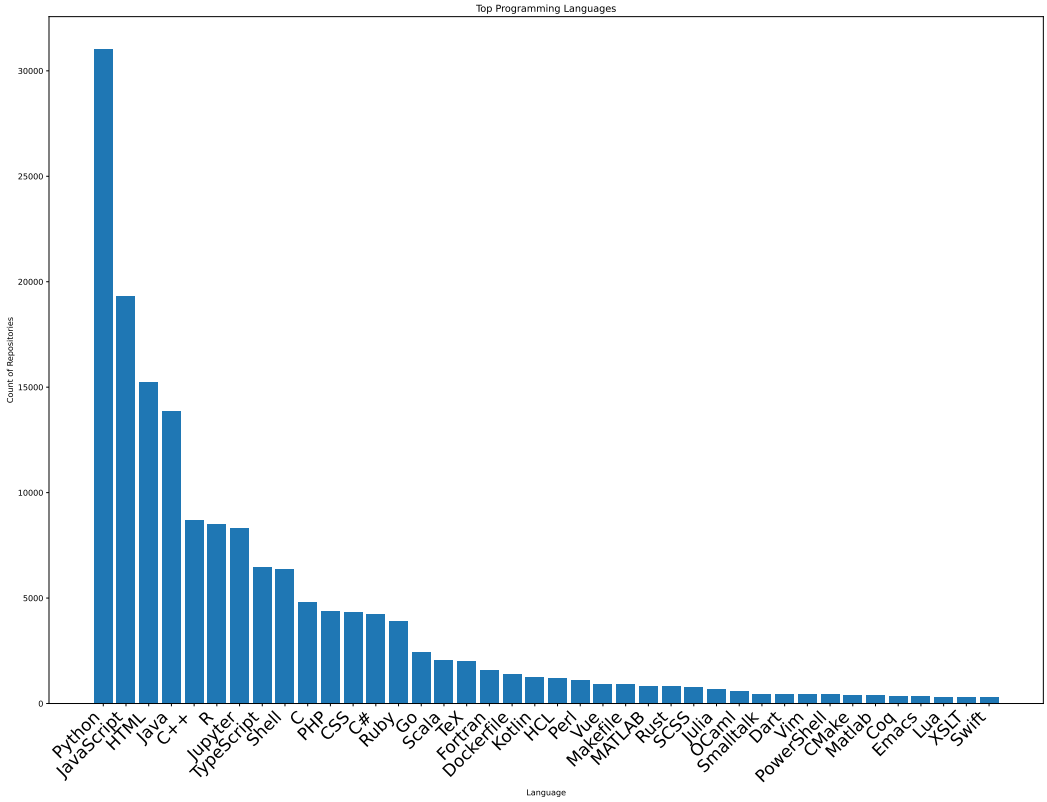


Figure 22: Most-used programming languages across government-associated projects.

Most projects carry no explicit licence. About seven in ten ship without machine-readable licence metadata, and three quarters once unusable entries are counted — a concrete, fixable reuse-readiness gap.

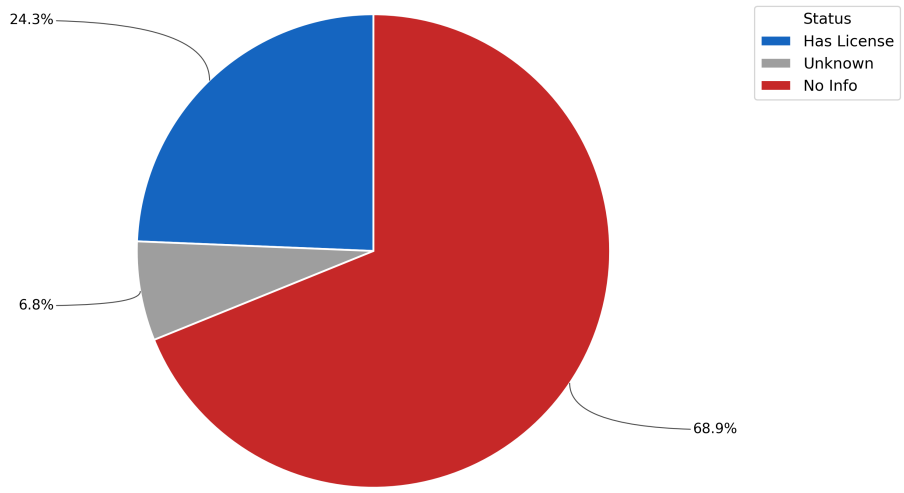


Figure 23: Licence status across government-associated projects.

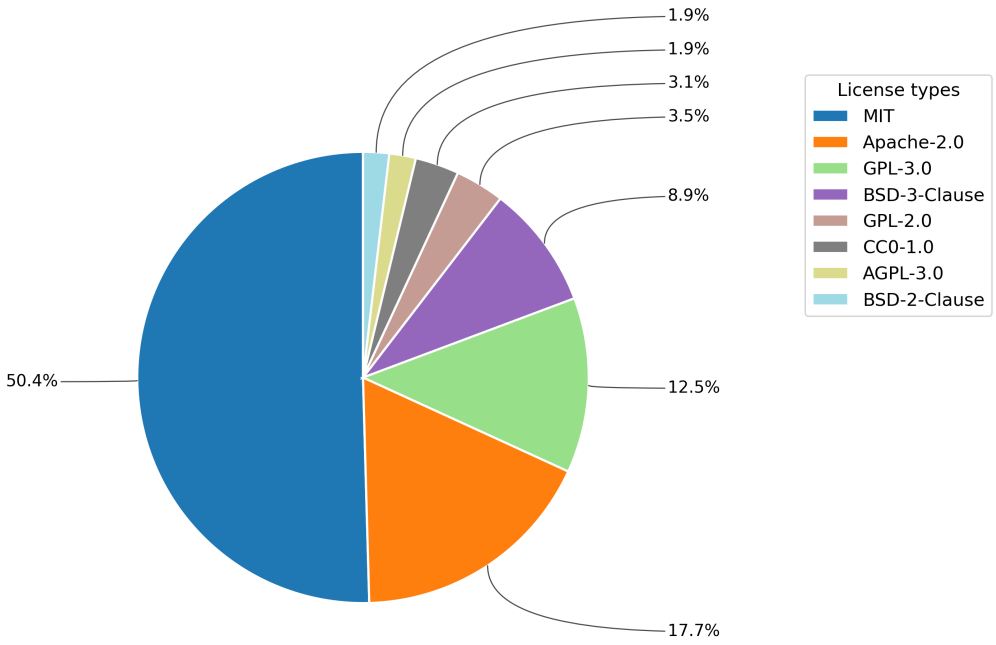


Figure 24: Most common licence types among the licensed projects.

Hosting is highly concentrated. About nine projects in ten sit on a single commercial platform — a dependency that platform-independent preservation directly mitigates.

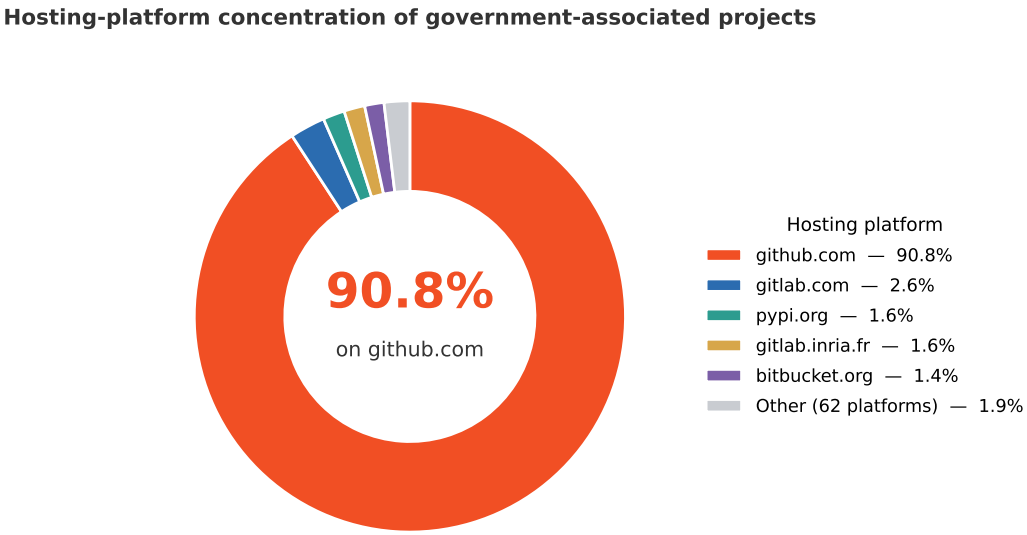


Figure 25: Hosting-platform concentration of government-associated projects.

Annex D — The many paths to contribute

The Public Code Observatory is built to be improved by the very institutions and communities it measures. Its accuracy does not rest on a single authority but on participation: there are several independent, complementary paths to contribute, each suited to a different kind of participant. Every correction, curation and independent re-analysis feeds directly into the next edition.

Where to start. Governments and institutions can reach the Observatory through a single intake form: https://grist.numerique.gouv.fr/o/docs/forms/cD65zwLyCZSqSHxdtFwogz/4. The paths

below explain what each kind of contribution changes.

1. **Tell us about your government's domains.** *(governments, public institutions)* This is the highest-leverage contribution. Government activity is attributed through official email domains, and coverage is still uneven — many countries are currently represented only at the federal level, with regional, agency and sub-national domains still to be added. If your institution is missing, only partially covered or mis-attributed, let us know through the participation form. This directly changes who is counted in the next edition. The underlying coverage is maintained as an open registry, world-gov-domain-names, for those who prefer to propose corrections directly.
2. **Validate your country profile and nominate indicators.** *(governments)* Confirm or contest the profile that emerges for your country, and propose the indicators that matter most for public-code policy, through the participation form.
3. **Flag noise, confirm legitimate projects.** *(open-source communities, maintainers)* Help distinguish genuine government-associated projects from mirrors, automated forks and noise. This improves precision without changing who counts as government. Submit observations through the participation form.
4. **Reproduce, re-aggregate and challenge.** *(researchers, civil society, national statistical offices)* Every figure can be independently verified. The companion dataset is openly licensed (CC-BY) and ships with a reproduction guide and a worked audit trail from any published number back to its source data, so you can build national or thematic profiles and contest the method. The data and guide live in the companion dataset repository.
5. **Preserve your public code.** *(any institution)* Much of what this report measures survives only because it was archived independently of its original host. Any institution can ensure its own source code — and its full development history — is preserved in the Software Heritage archive using Save Code Now. This protects the continuity of the public-code record and, because the Observatory measures what the archive can see, also makes your contribution count in future editions.

Two distinct roles. The government-domain registry decides *who counts as government*; project-level curation decides *what is noise*. Choosing the path that matches your contribution is what makes it effective.

> **An open invitation.** These findings are preliminary and released openly for comment by all stakeholders, including United Nations member states. The Public Code Observatory — developed by Software Heritage with the support of UNESCO and the UN Office for Digital and Emerging Technologies — is a shared and evolving capability that becomes more accurate, more representative and more useful with every participant. We invite governments, open-source communities and researchers to take part.