State of Public Code 2026

Indicators Report

the core per-country indicators and ranking tables
(the complete, fully-reproducible evidence base is the Comprehensive Edition)

How governments contribute to open source

A Public Code Observatory edition
presented at UN Open Source Week

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

with support from

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

DRAFT — pending sign-off

Contents

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.

These are **preliminary findings**, published openly for comment by all stakeholders — including United Nations member states — as part of the Public Code Observatory (Software Heritage, with the support of UNESCO and the UN Office for Digital and Emerging Technologies). They will be refined as the Observatory's collaborative, community-curated effort develops.

Headline findings

- **Coverage.** 288,411 distinct public repositories carry at least one government-email contribution (92,831 distinct contributors). All **193 UN member states** appear, via 199 government domains.
- **The strongest predictor of per-country output is not wealth.** Per-country contributor count correlates with the UN E-Government Development Index at Spearman $\rho = \mathbf{0.52}$ ($n = 158$) — 3.9× the rank-variance of GDP per capita (0.26); the ranking is unchanged with or without the US outlier (max shift 0.014).
- **Trends.** Activity grew from the first recorded contribution in 1995 to **29,813 newly-active repositories in 2022** — the most recent year we can confidently quote (see §2).
- **Cross-border collaboration is rare but real.** 1,409 repositories (0.5%) carry contributors from more than one country; the densest bilateral pair is **United Kingdom↔United States on 404 shared projects**, and France–Germany now appears on 11.
- **Sustainability is sober.** Of newly-active government-contribution repositories, only the 1.7%–17.5% band remains active a year later — roughly 9 out of 10 fall silent within a year.
- **Digital Public Goods: deploy-not-contribute.** Only **10** vetted DPGs have any government-email contribution in our universe, across 8 countries (see §7).
- **AI maturity.** Netherlands leads on AI-share-of-output at 7.2% (5.12× the global mean of 1.4%).

The dataset counts repositories that government contributors *touched*, not repositories that governments *own*. We surface this where it matters most (collaboration, forks); see Annex A for the full caveat. A curated noise-filter (the classification cache) removes known-noise origins — joke repositories, single-author personal projects on work emails, classroom forks; the qualitative tails (most-international, most-forked) are robust to this filtering.

> *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, forks; see Annex A). Coverage depends on a curated, evolving registry of government email domains (bzg/world-gov-domain-names); completeness is an ongoing, community-curated effort.

§1 — Coverage and scope

The pipeline applies the UN-member-state governmental domain set from Bastien Guerry's bzg/world-gov-domain-names registry to the Software Heritage archive snapshot, then attributes each origin to a country by longest-suffix match of the author email domain against the canonical per-country domain list.

- **Origins:** 288,411 public repositories.
- **Contributions (revisions + releases):** ‹REVIEW: doc cited 201,912,422 (rev+rel); manifest TOTAL_REVISIONS has a different/narrower definition — re-derive or drop›
- **Single-country resolution:** ‹REVIEW: doc '99.96%' has no CSV source; collaboration single-country share is a different statistic›

§1.1 By continent and subcontinent

Region	Contributors	Share
Americas	66,982	72.2%
Europe	13,391	14.4%
Asia	5,805	6.2%
Oceania	4,800	5.2%
Middle East	912	1.0%
Africa	880	0.9%

Region	Contributors	Share
Europe (non-EU)	61	0.1%

The Americas dominate because the United States routes essentially all federal-government contributors through agency-specific subdomains under the bare `.gov` TLD (NIST, NASA, the DOE national laboratories, USDA, …) — see §8 for the structural-data caveat. Europe is the second-largest bloc, lifted by the UK (`gov.uk`) and France (`gouv.fr`). Africa, at well under 1 % of global volume, is the region where future capacity-building has the largest relative headroom.

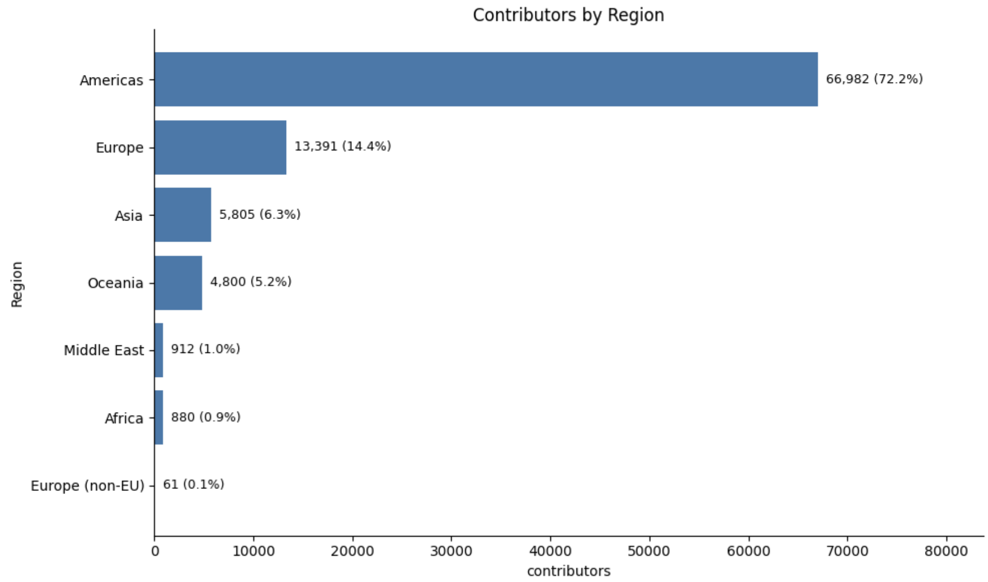

Figure 1: Contributors by world region (continent-level rollup).

§1.2 By income group

Income group	Contributors	Share
High income	63,279	68.2%
Upper-middle income	27,246	29.3%
Lower-middle income	2,095	2.3%
Low income	211	0.2%

High-income countries account for roughly two-thirds 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. §6's EGDI correlation predicts this pattern best; raw wealth alone is the weakest of the five tested indicators.

§1.3 By official language

Language	Contributors	Share
English	56,916	61.3%
Portuguese	17,828	19.2%
Spanish	7,119	7.7%
French	2,068	2.2%
Unknown	1,009	1.1%
Japanese	889	1.0%
Arabic	841	0.9%
Indonesian	813	0.9%
German	705	0.8%
Malay	574	0.6%

Language	Contributors	Share

Portuguese in the second slot is almost entirely Brazil (the `gov.br` aggregation discussed in §1.1 and §8). 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.

§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.

§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	41,253	44.4%
2	Brazil	17,725	19.1%
3	United Kingdom	9,252	10.0%
4	Australia	4,764	5.1%
5	Mexico	1,533	1.7%
6	France	1,427	1.5%
7	Argentina	1,110	1.2%
8	Colombia	922	1.0%
9	Turkey	912	1.0%
10	Japan	889	1.0%
11	Singapore	862	0.9%
12	Peru	833	0.9%
13	Indonesia	813	0.9%
14	Canada	810	0.9%
15	Poland	463	0.5%
16	Saudi Arabia	404	0.4%
17	Philippines	400	0.4%
18	Venezuela	373	0.4%
19	El Salvador	360	0.4%
20	Malaysia	358	0.4%
21	Ecuador	352	0.4%
22	Bolivia	347	0.4%
23	Switzerland	344	0.4%
24	India	338	0.4%
25	South Korea	316	0.3%
26	Dominican Republic	282	0.3%
27	Belgium	271	0.3%
28	Germany	266	0.3%
29	Austria	258	0.3%
30	Italy	252	0.3%

This right-skew is the load-bearing motivation for §12 (Gini-Lorenz concentration): the per-country distribution is as unequal as extreme wealth-concentration measures.

§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.

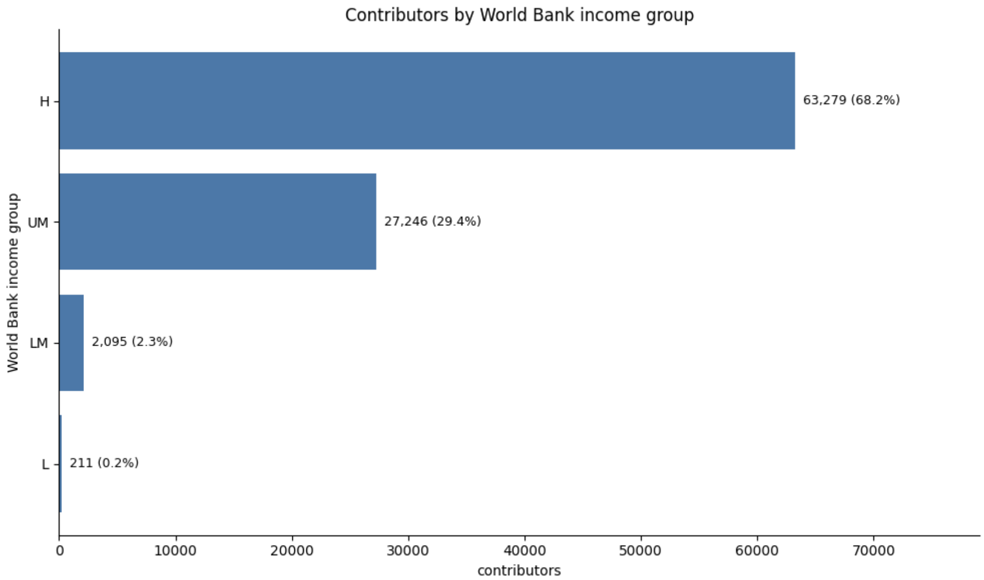

Figure 2: *Contributors by World Bank income classification.*

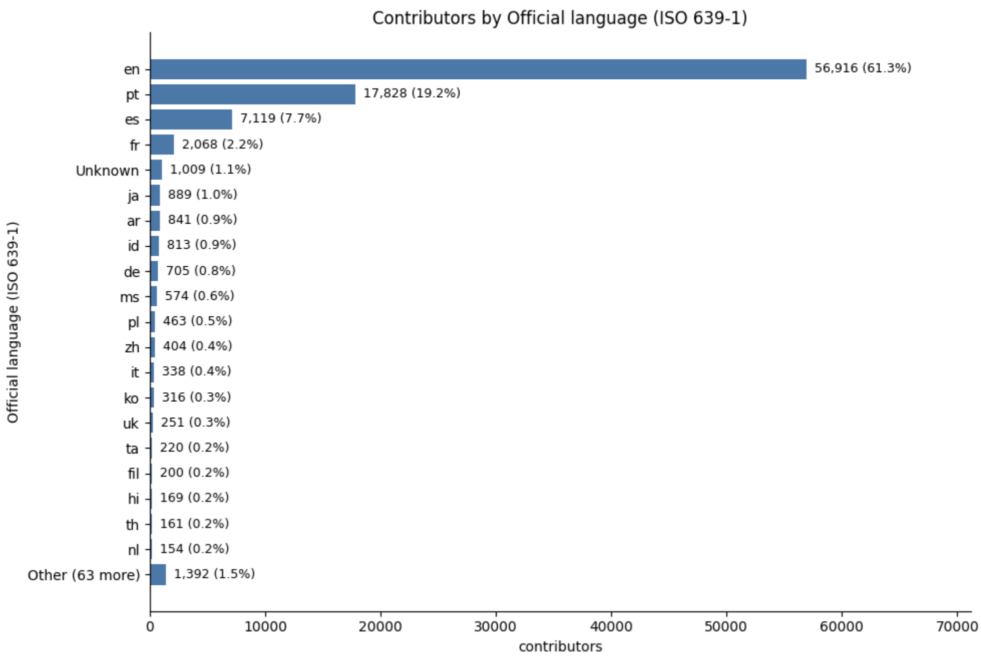

Figure 3: *Contributors by national official language (top 10).*

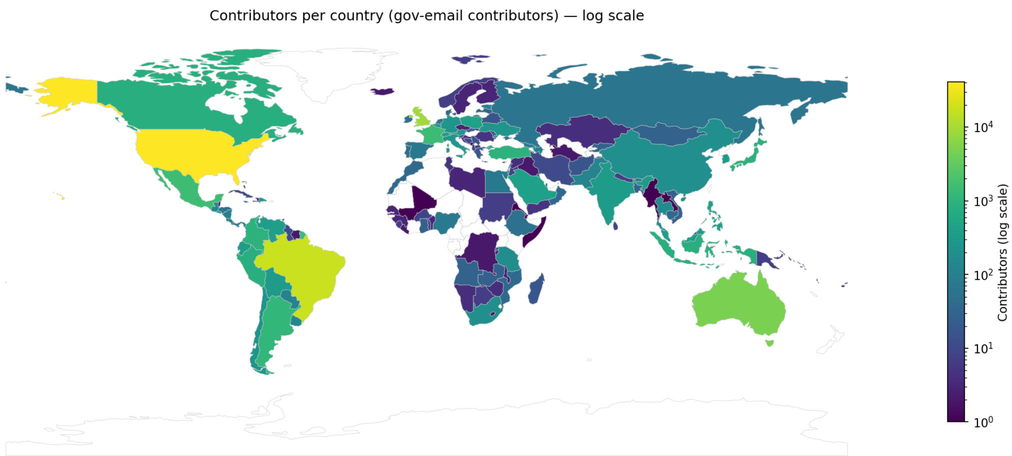

Figure 4: *Contributor count per country (logarithmic colour scale, UN member states). A log scale is used because headcount spans several orders of magnitude (the United States alone is ~44% of all measured contributors).*

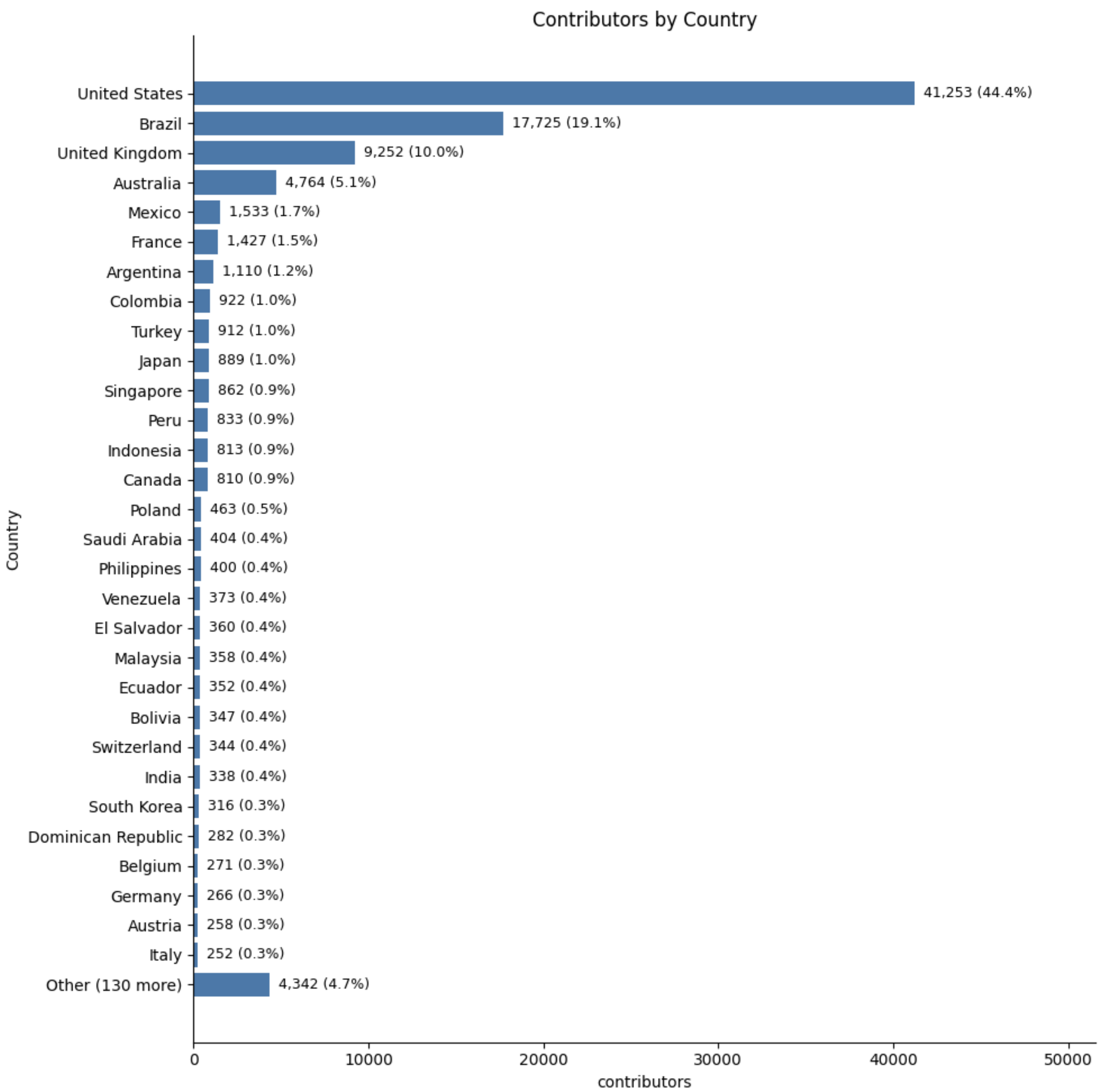

Figure 5: *Top-30 countries by contributor count (right-skewed, log scale).*

The **distinct-projects-touched** count — per country, the number of distinct deduplicated repositories that received at least one contribution from a gov-email author resolving to that country — measures **breadth** rather than workforce volume.

At country level this view is almost perfectly correlated with the §1.5 contributor count (Spearman $\rho = 0.99$): on equal (logarithmic) colour scales the projects choropleth and the contributor choropleth shade nearly identically, so absolute project count does **not** reveal a different geography. What *does* differ between countries is **reach per contributor** (distinct projects ÷ contributors), mapped below; the absolute breadth ranking is given by the table and the top-30 chart.

#	Country	Distinct projects	Share
1	United States	126,323	41.8%
2	Brazil	77,456	25.6%
3	United Kingdom	33,229	11.0%
4	Australia	12,691	4.2%
5	France	5,235	1.7%
6	Mexico	3,581	1.2%
7	Japan	3,332	1.1%
8	Argentina	3,145	1.0%
9	Canada	2,932	1.0%
10	Singapore	2,724	0.9%
11	Turkey	2,319	0.8%
12	Colombia	1,984	0.7%
13	Indonesia	1,848	0.6%
14	Peru	1,695	0.6%
15	Poland	1,119	0.4%
16	Bolivia	1,029	0.3%
17	Philippines	956	0.3%
18	Ecuador	918	0.3%
19	Italy	886	0.3%
20	El Salvador	885	0.3%
21	Saudi Arabia	862	0.3%
22	Switzerland	841	0.3%
23	Venezuela	776	0.3%
24	Austria	772	0.3%
25	South Korea	754	0.2%
26	Belgium	741	0.2%
27	Dominican Republic	725	0.2%
28	Ukraine	725	0.2%
29	India	672	0.2%
30	Estonia	664	0.2%

112 of 160 attributed governments touch ≥ 10 distinct projects (48 below that threshold). Notable breadth-vs-volume movers (projects-rank minus contributors-rank): Japan (+3), Canada (+5), Colombia (-4), Bolivia (+6), Ecuador (+3), Italy (+11), Saudi Arabia (-5), Venezuela (-5), Austria (+5), India (-5).

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.

§2 — Trends over time

The first government-email contribution observed in our universe is **1995** (3 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 undercounts 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.

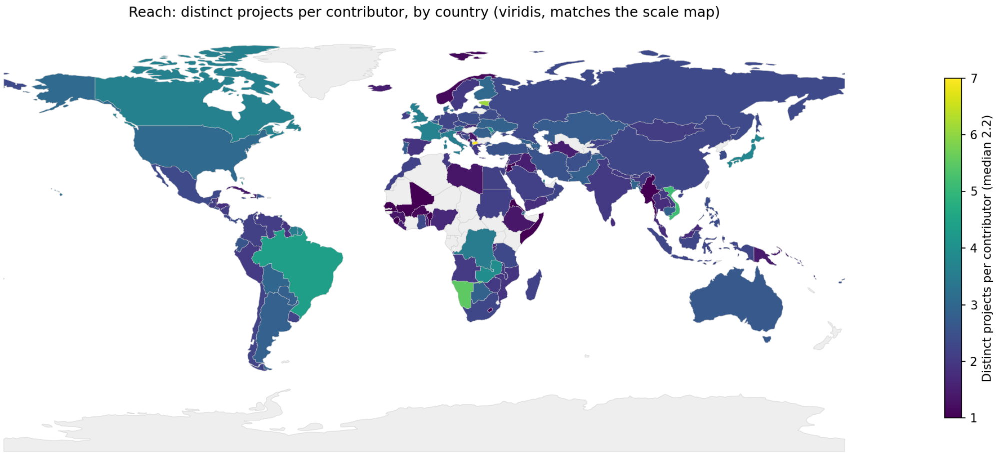

Figure 6: *Reach intensity: distinct projects touched per contributor, by country (same viridis colour scale as the contributor map, so the two are directly comparable; global median ~2.2). Because raw contributor and project counts are $\rho = 0.99$ collinear, this ratio - not absolute project count - is what reveals a distinct geography: Estonia (~6.1), Brazil (~4.4) and several G7 states are high; India, Spain and Ethiopia are low.*

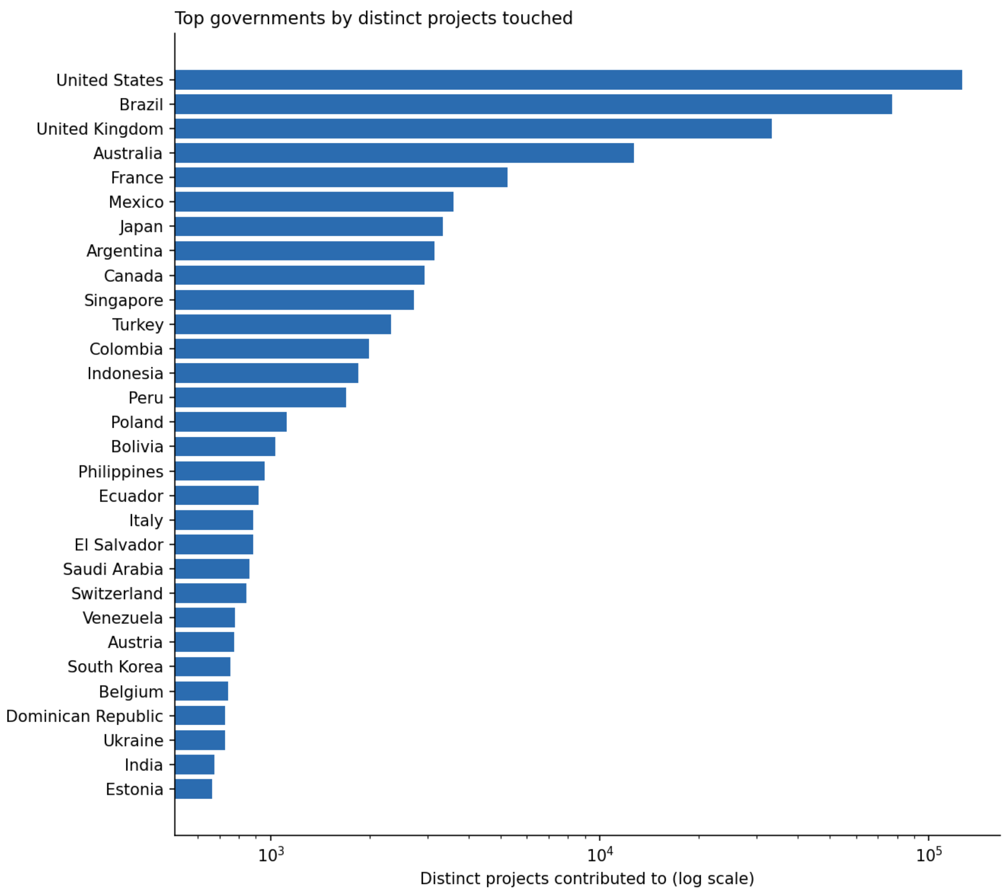

Figure 7: *Top-30 governments by distinct projects touched (log scale).*

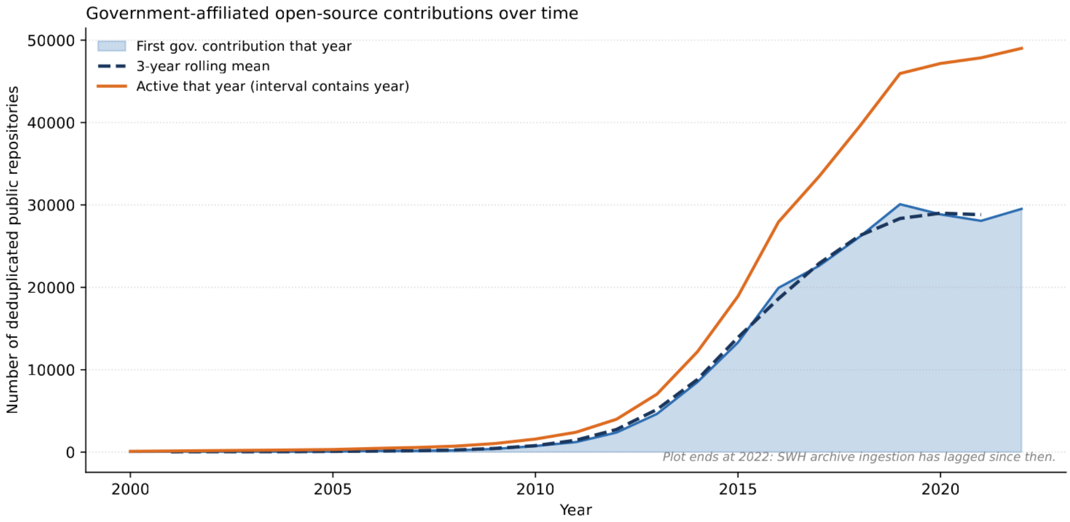

Figure 8: Yearly new-contribution counts, 1995–2022 (3-year rolling mean).

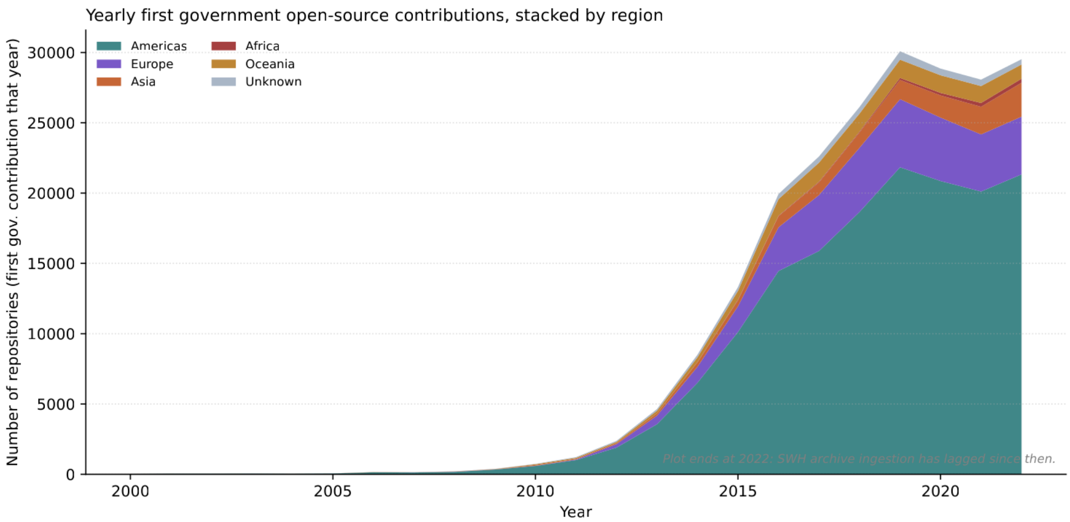

Figure 9: Yearly new contributions stacked by region, 1995–2022.

§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	**1,409** (0.5%)
United Kingdom↔United States	404 shared projects
Australia↔United States	303 shared projects
Brazil↔United States	229 shared projects
Australia↔United Kingdom	134 shared projects
France↔United States	103 shared projects
France ↔ Germany	11 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.

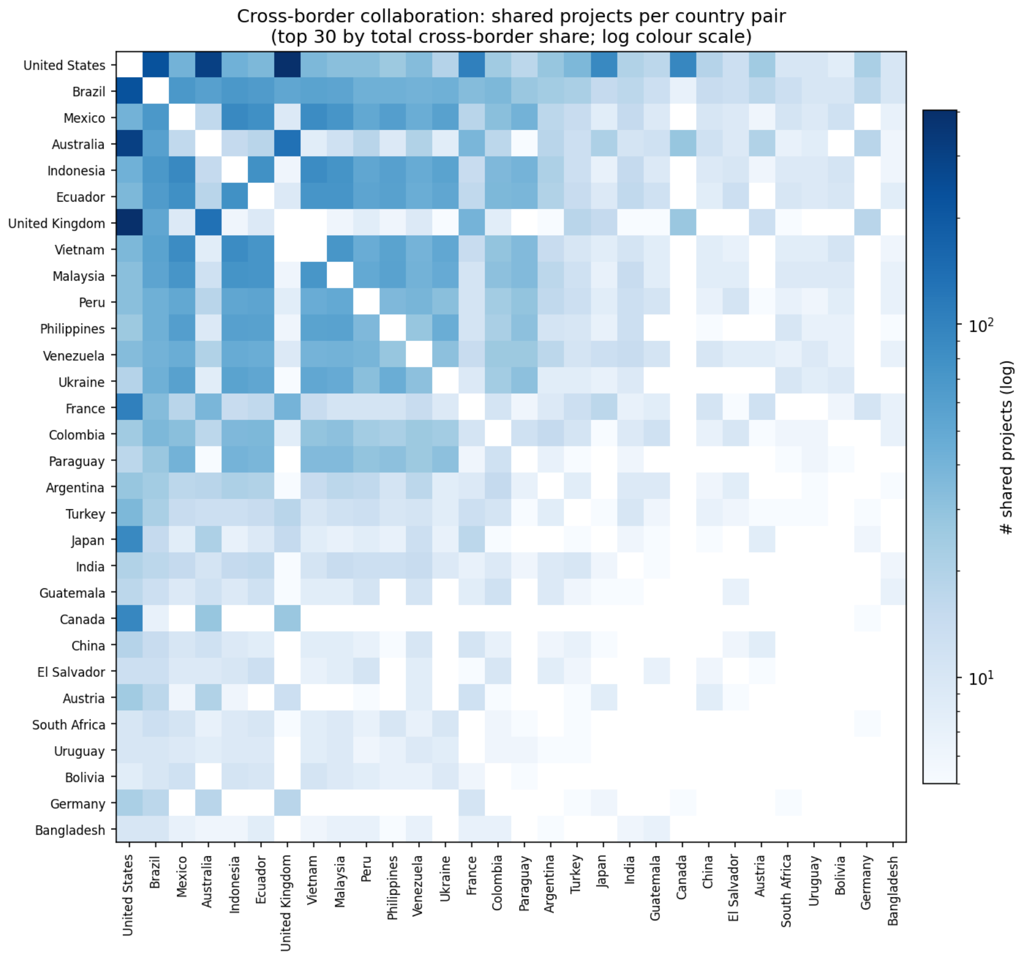

Figure 10: Country pairs by shared-project count, top 30 × top 30 (log scale).

§4 — Repository reuse: forks-received

Forks-received measures the outbound attention to the repositories a country's government has touched: how many third parties chose to work from this code.

#	Country	Total forks received	Mean forks / repo
1	United States	4,566,658	38.5
2	United Kingdom	878,328	28.2
3	Brazil	838,506	11.1
4	Australia	501,497	42.8
5	France	287,721	58.7

Framing caveat. Because the dataset is "repositories touched by gov contributors", the top-forked list is dominated by community OSS (e.g. `tensorflow/tensorflow`, `twbs/bootstrap`, `github/gitignore`) where a gov contributor appeared once and the repository's large fork count attributes to that country's row. The number reads as *fork activity on repositories this government participated in*, not *fork activity on government-owned code.* A strict host-matched view returns almost nothing because government-hosted forges rarely expose public fork pages — the broad view is the only sensible per-country metric.

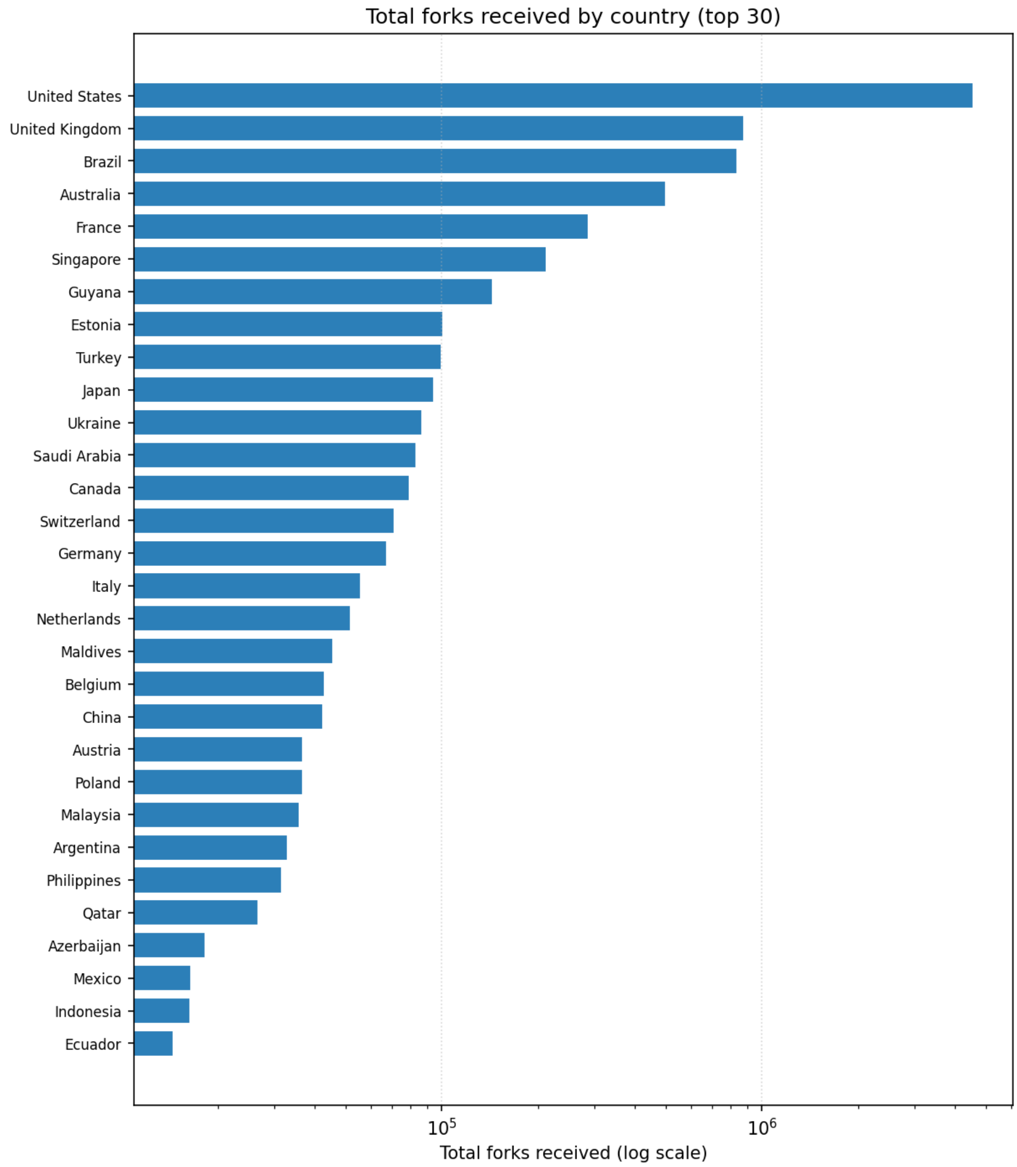

Figure 11: Top-30 countries by total forks received on repositories they touched.

§5 — 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 trustworthy steady state is the earlier band.

Cohort year	Cohort size	Share active at 1 year
2018	25,676	17.5%
2019	29,844	16.9%
2020	28,815	16.4%
2021	28,138	13.5%
2022	29,697	10.2%
2023	31,781	6.2%
2024	35,148	1.7%

In plain terms: about five out of every six government-contribution repositories fall silent within a year of their first government-domain commit. This is the single most actionable indicator for any framework that wants to measure sustained public-sector engagement with open source.

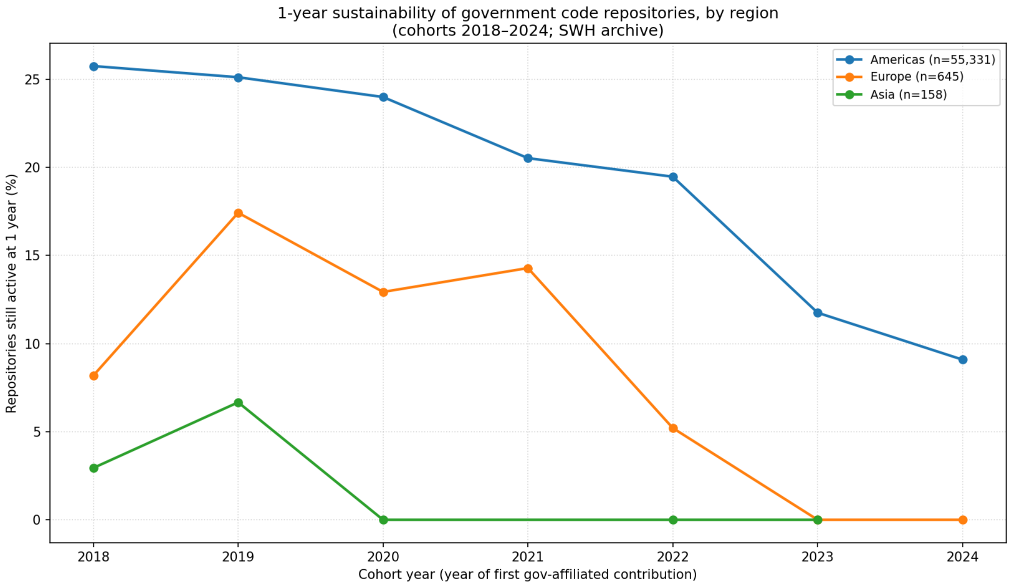

Figure 12: Share of new-contribution cohorts still active one year later.

§6 — 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	158	0.52	0.51	0.007
UN HDI	157	0.34	0.33	0.010
Internet users %	151	0.31	0.30	0.009
ITU IDI 2024	142	0.29	0.27	0.014
GDP per capita	156	0.26	0.25	0.012

EGDI leads GDP per capita by 3.9× 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.

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.

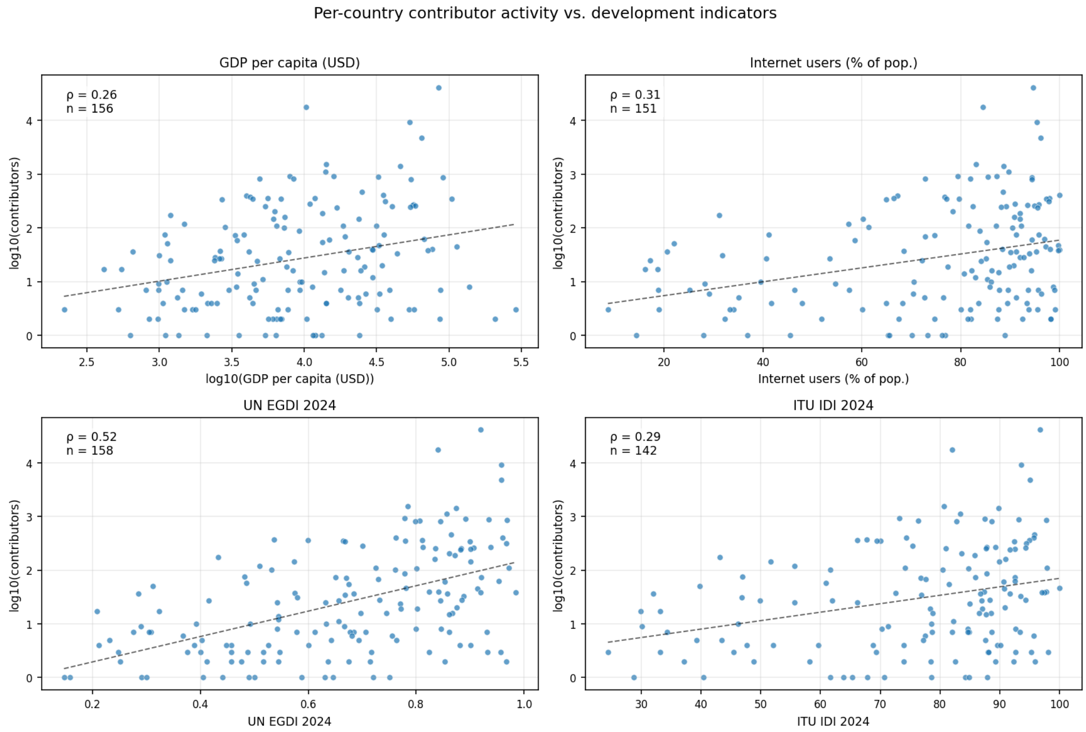

Figure 13: *Per-country contributor count vs four development indicators.*

§7 — 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: **10** across **8** countries.
- Registry size: ‹REVIEW: '165 of 228 (stage==DPG)' not recoverable — shipped dpg_registry.csv has no stage column; re-fetch upstream›

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.

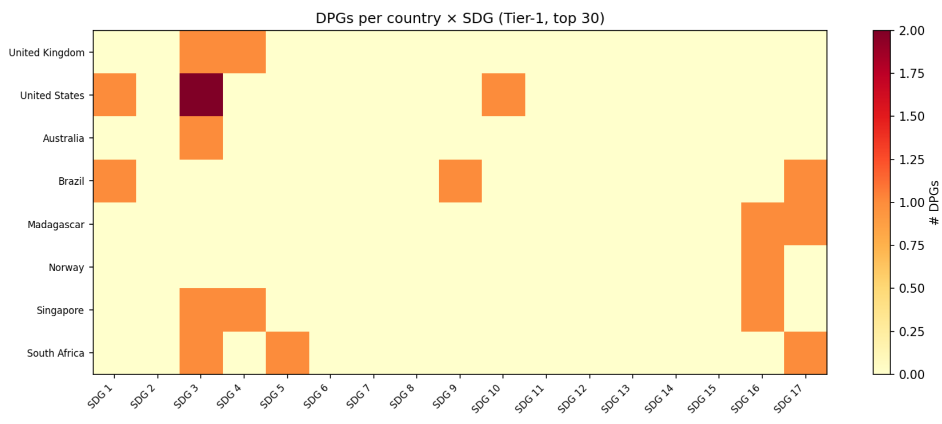

Figure 14: *DPG contributions by country and SDG (top countries × 17 SDGs).*

§8 — 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	60,683
2	Environment	19,034
3	Science & Research	5,368
4	Health	3,575
5	Digital & Statistics	1,683
6	Interior	956

Of the 32,148 contributors attributed to a *specific* ministry (not the catch-all bucket), **99.9% 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.

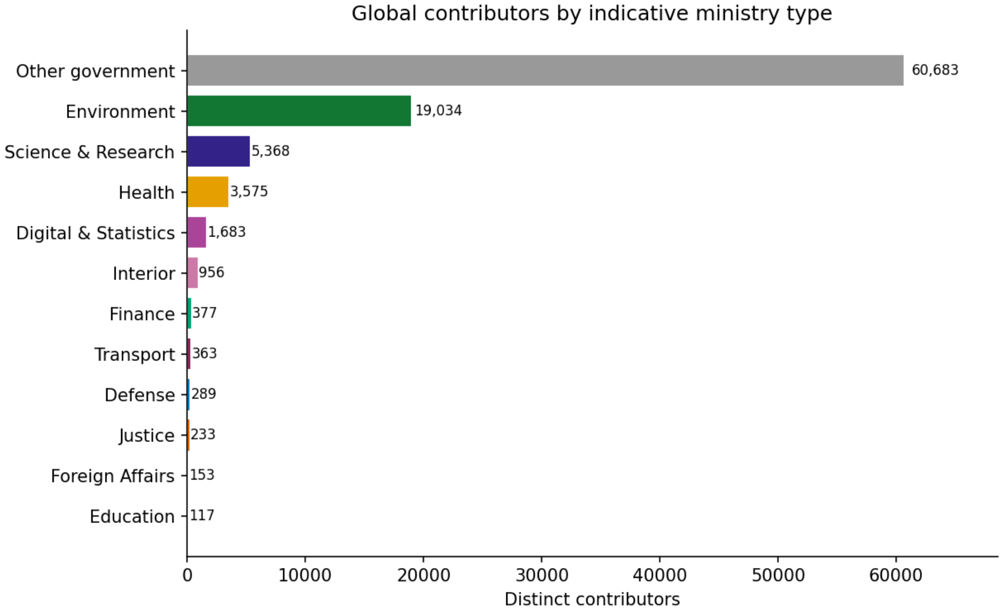

Figure 15: Top ministries globally, ranked by attributed contributor count.

§9 — 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 (2797), Brazil (430), United Kingdom (290), Australia (121), Japan (112). **Top by AI-share-of-output** (≥ 100 attributed repos; global mean 1.4%): Netherlands (7.2%, 5.12×); Japan (3.4%, 2.39×); Egypt (2.6%, 1.84×); Singapore (2.3%, 1.65×); India (2.2%, 1.59×).

Regional pattern: Asia and North America lead on AI-share-of-output; Latin America is lowest — Brazil’s enormous `gov.br` output is infrastructure-and-services dominated, not AI-led.

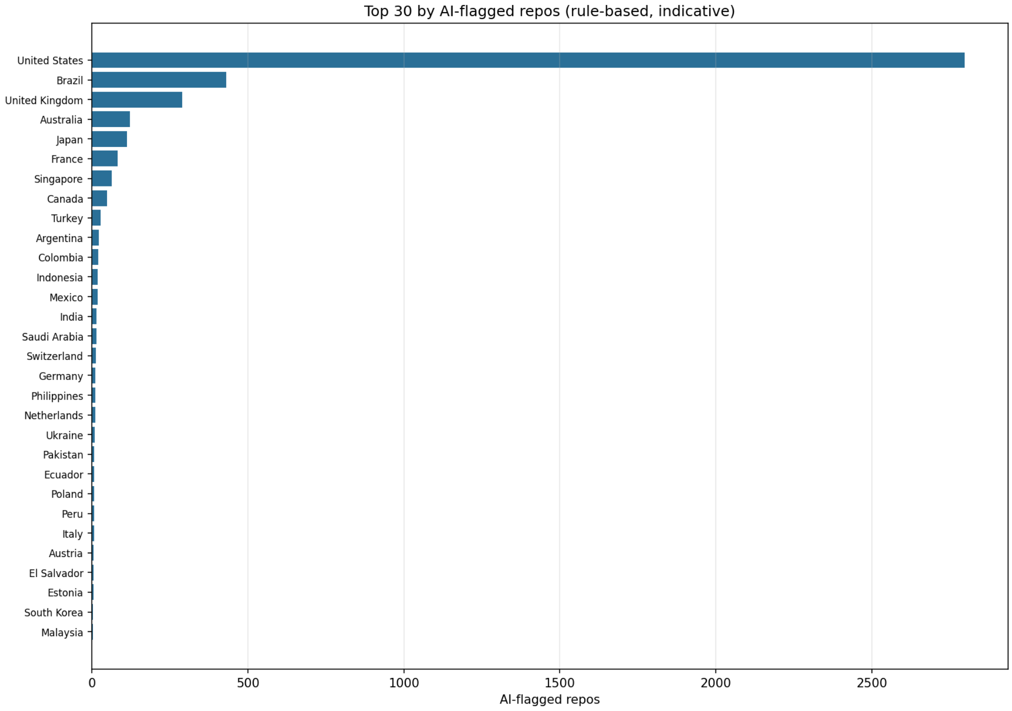

Figure 16: *Top-30 countries by absolute count of AI-related repositories.*

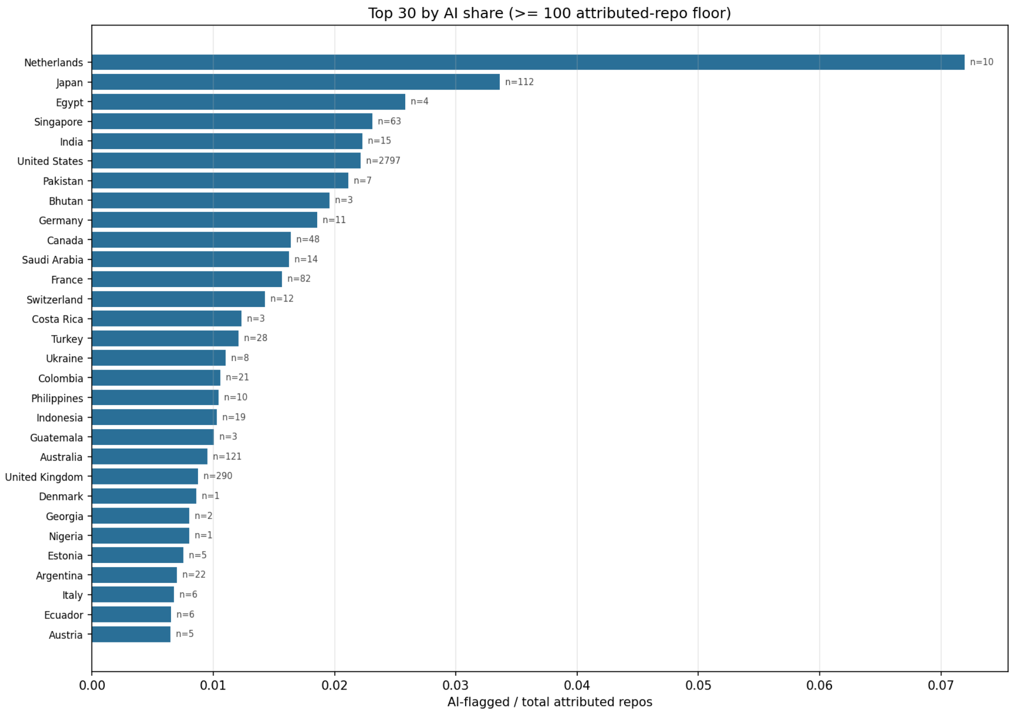

Figure 17: *Top-30 countries by AI-related repository share of output.*

§10 — 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 (460); 16. Peace & justice (316); 6. Water & sanitation (260).
- Bottom SDGs: 17. Partnerships (9); 10. Reduced inequalities (7); 5. Gender equality (3).
- SDG-tagged repositories: **1,949** (union of 1,945 lexicon-matched + 10 DPG-tier − 6 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).

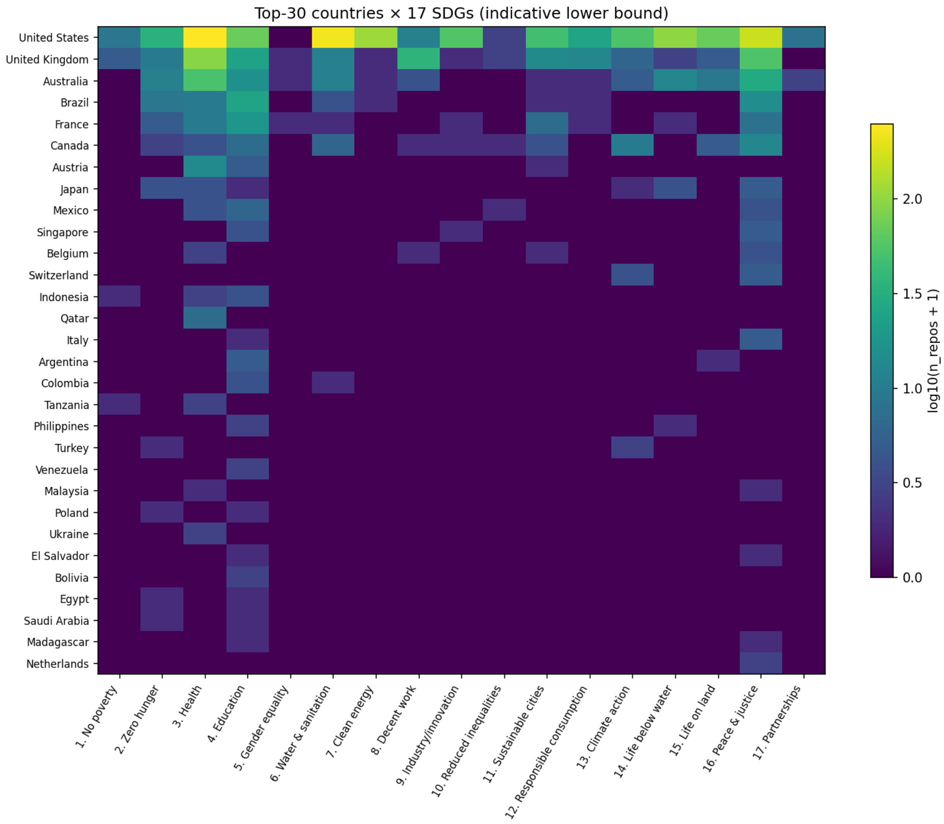

Figure 18: Top-30 countries × 17 Sustainable Development Goals (log scale).

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.

§11 — Quality and metadata signals

§11.1 Citation files

Only **885 of 288,411** repositories (**≈0.3%**) ship a CITATION.cff or codemeta.json — 820 carry a `CITATION.cff` and 87 carry CodeMeta metadata. A repository is "citable" in the academic sense when it ships a `CITATION.cff` file or a `codemeta.json` descriptor (or both). This near-total absence of citation metadata is a material gap for any policy framework wanting to incentivise (or measure) academic reuse of public-sector software; among the citable few, the dominant pattern is a sole `CITATION.cff` — `codemeta.json` adoption is negligible.

§11.2 Licensing

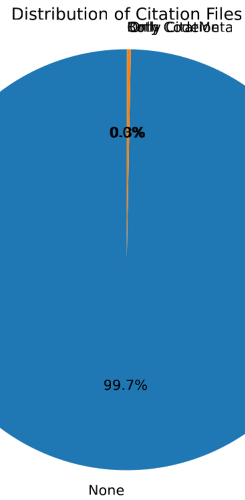

Figure 19: CITATION.cff and CodeMeta presence across the universe.

License status	Count	Share
Has a valid license	57,107	19.8%
Unknown license status	15,174	5.3%
No license information at all	216,130	74.9%

Top SPDX identifiers among the licensed: MIT (10.0%), Apache-2.0 (3.1%), GPL-3.0 (2.1%), BSD-3-Clause (1.6%), GPL-2.0 (0.6%), CC0-1.0 (0.6%). Three quarters of the universe ships no license metadata at all — and on most platforms (notably GitHub) this defaults the repository to "all rights reserved", which legally precludes reuse.

The MIT + Apache + BSD family carries **15.0%** 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.

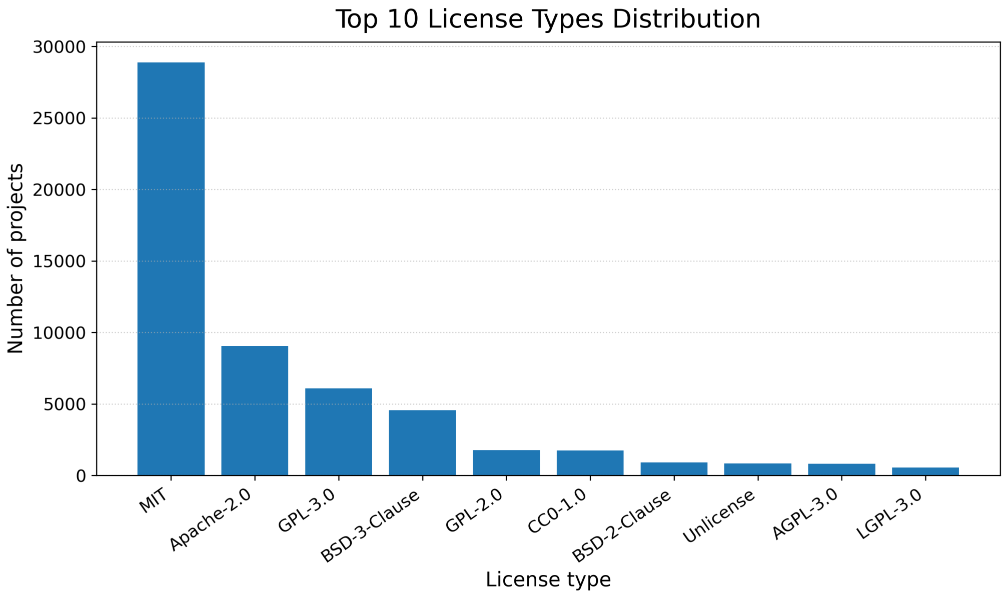

Figure 20: License-type distribution (top SPDX identifiers).

§11.3 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 §9 AI-related signals.

§12 — Concentration: Gini-Lorenz curves

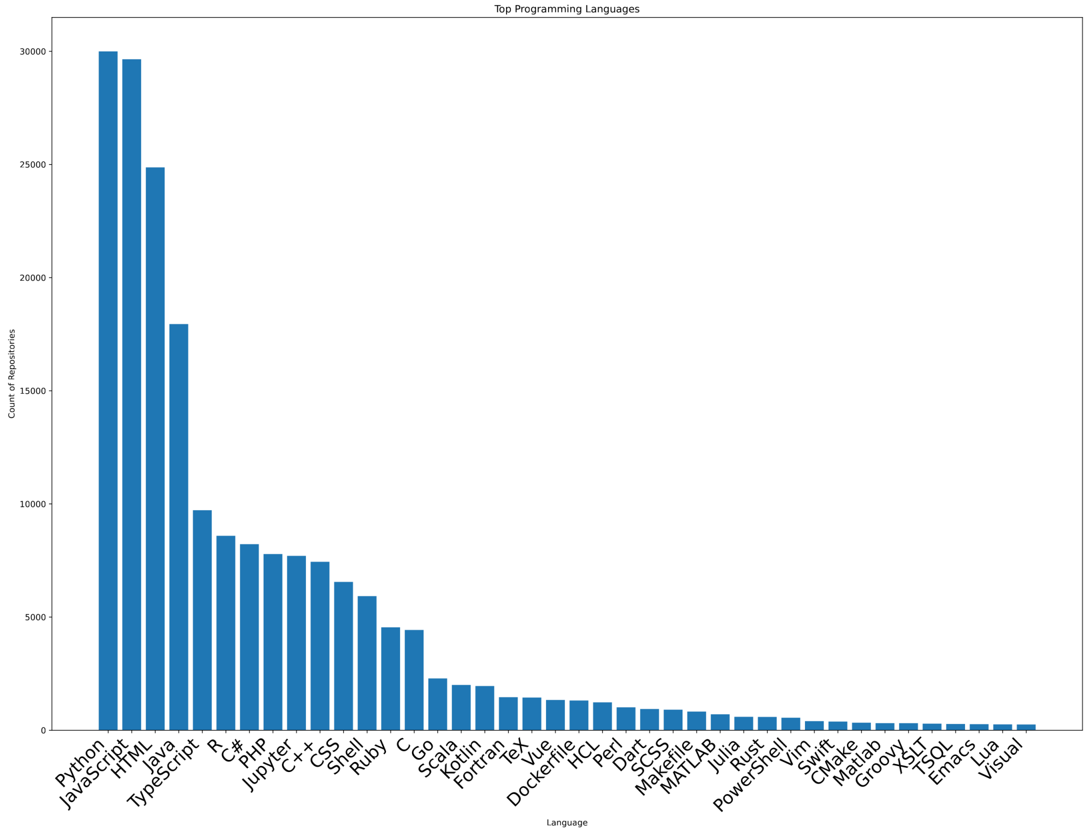

Figure 21: Top-40 programming languages by repository count.

Annex A — Methodology

Universe. The UN-member-state governmental domain set from bzg/world-gov-domain-names, applied to the Software Heritage archive snapshot; 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; seeds garbage and mailing-list-mirror verdicts that the analysis honours. **Touched ≠ owned.** Country metrics measure repositories where ≥ 1 government-email contributor appeared, not repositories a government owns; government-hosted forges rarely expose public fork pages, so a strict ownership view is not recoverable. **Archive horizon.** Activity after ~mid-2024 is under-represented; time-series figures cap at 2022.

Annex B — Sources

- Dataset: Software Heritage archive snapshot.
- Governmental-domain registry: bzg/world-gov-domain-names (Bastien Guerry); UN members only — observer/non-UN entries excluded.
- 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.

Generated from `/tmp` on 2026-06-19. Numbers are live from the analysis manifest; <REVIEW> markers flag figures with no clean machine source. Reproducibility: regex_sha/persons_snapshot/dataset pinned in the run sidecar.

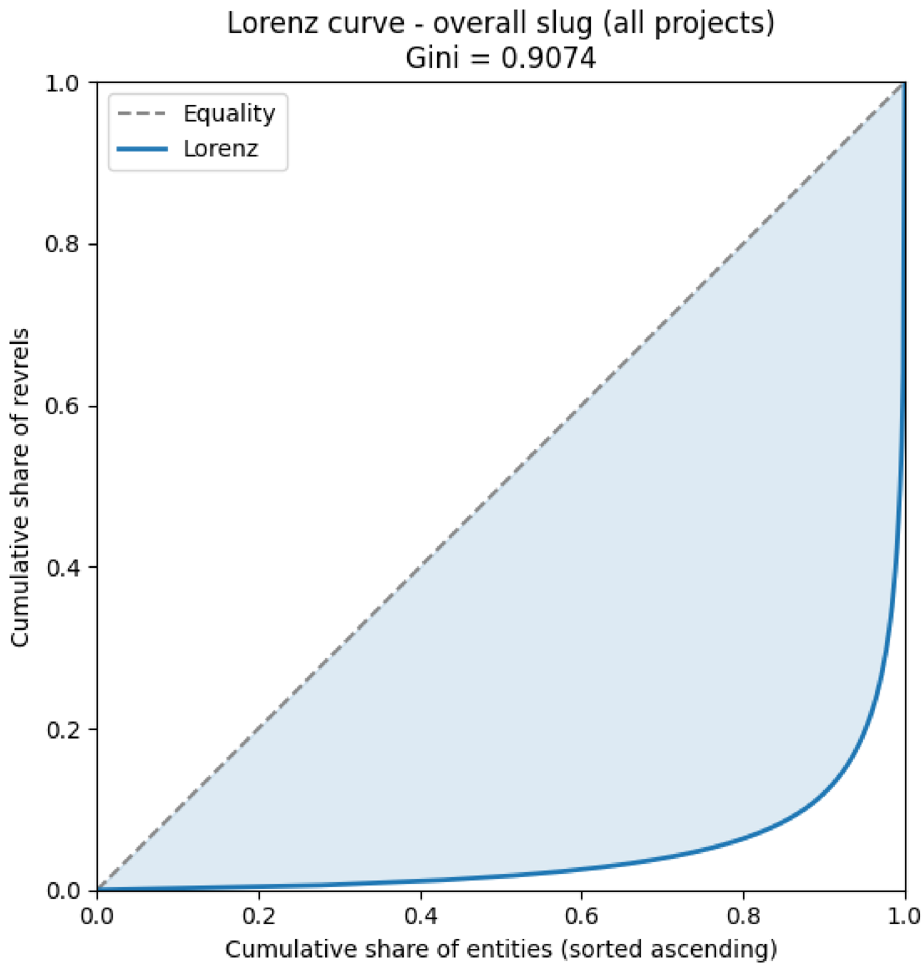

Figure 22: Lorenz curve: cumulative repositories vs cumulative contributors (global).

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 (83%) — 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).

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.

Most projects carry no explicit licence. About three quarters ship without machine-readable licence metadata — a concrete, fixable reuse-readiness gap.

Hosting is highly concentrated. Around 94% of projects sit on a single commercial platform — a dependency that platform-independent preservation directly mitigates.

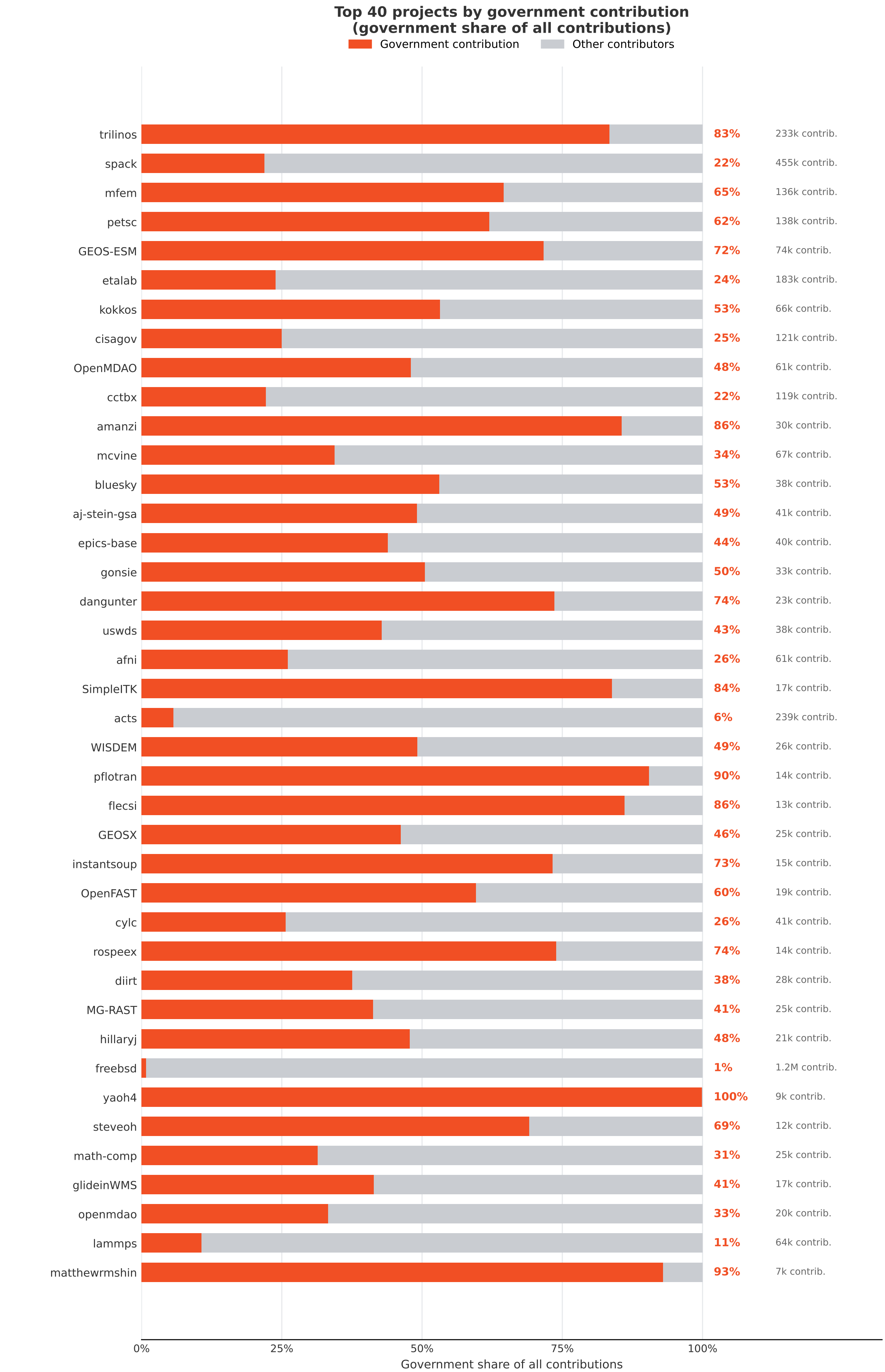

Figure 23: Top 40 projects by government contribution, showing the government share of all contributions (absolute volume annotated).

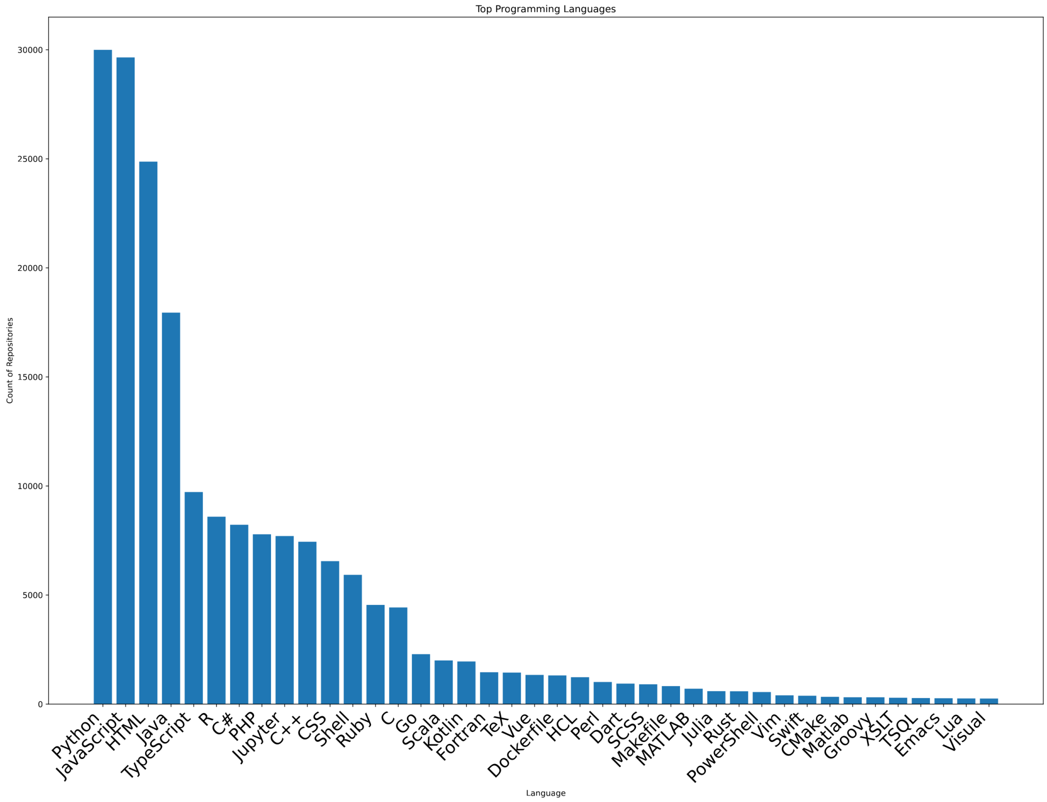

Figure 24: Most-used programming languages across government-associated projects.

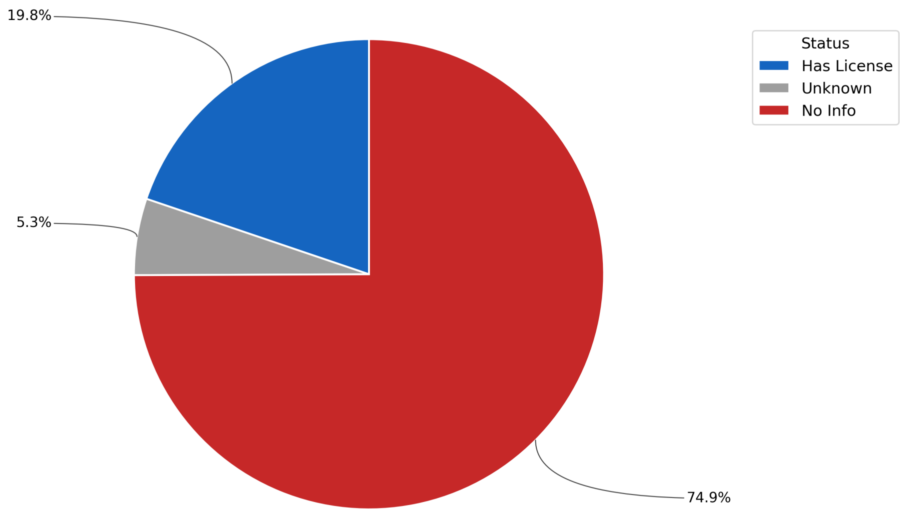

Figure 25: Licence status across government-associated projects.

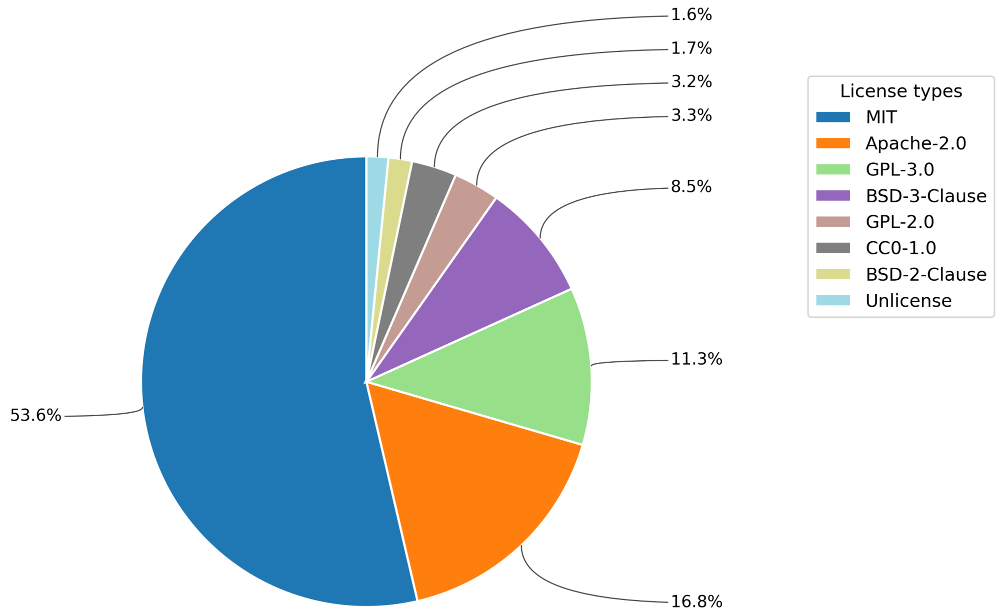

Figure 26: *Most common licence types among the licensed projects.*

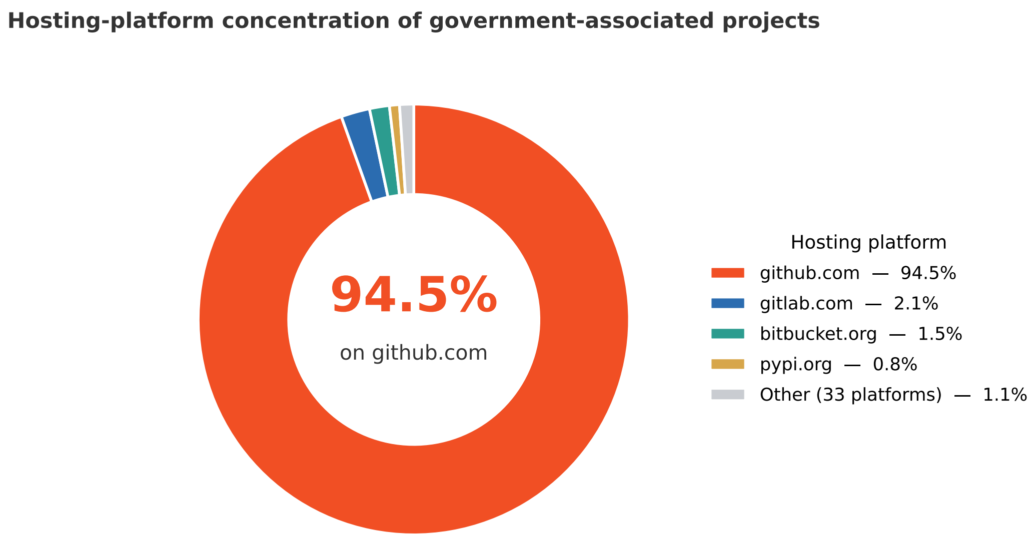

Figure 27: *Hosting-platform concentration of government-associated projects.*

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.