Analytical supplement to the global education atlas

GCyEO v4.0 · 2 November 2026 · accompanies Mapping cybersecurity education worldwide

S1 Evidence layers and counting rules

The descriptive main paper supersedes the measurement-focused framing of v3.4.180. It uses unchanged August inputs and retains later audits as interpretive evidence. The historical full and short manuscripts, source captures and decisions are preserved. This supplement does not reinterpret model agreement as truth or incomplete searches as national zeroes.

Object	Frozen value	Meaning
Evaluated jurisdictions	218	194 candidate-positive, four pilot-only, 20 with no retained rows
Global candidate rows	7167	Heterogeneous evidence units; no verified programme total
Reported delivery instances	8284	Explicit zero preserved; absent or empty count defaults to one
Country institution labels	4646	Distinct stripped nonempty provider strings within country
Pilot records	616	Five countries; Australia overlaps global layer
Credential and scope strings	461 and 319	Exact-label dictionaries; no global semantic recoding

jurisdiction_atlas.csv retains each source-country status, public count, source descriptions, last-check field, inclusion decision and notes alongside separately computed candidate and pilot counts. The regional grouping uses the existing world_bank_region field with whitespace removed. It is not a newly adopted or externally refreshed regional classification. Full labels, including the Middle East, North Africa, Afghanistan & Pakistan category, are preserved.

S2 Probability audit and later identity work

The frozen nine-stratum probability sample contains 400 of 7,167 candidate rows. In the primary audit, 226 distinct URLs were attempted. The exact-title screen supplied model context for 85 cases; 315 without a title lead received no model input and remained unresolved. Two blind model versions and an all-case challenge combined conservatively to produce three provisional inclusions, three provisional exclusions and 394 unresolved decisions. These are source-packet judgements about a stricter eligible-current target, not the descriptive presence of a candidate row.

The design-weighted conditional source-support bounds were 0.005514 to 0.996053. They depend on provisional model judgements and are noninformative for the intended eligible-current fraction; truth-agnostic bounds remain zero to one. A sample drawn from known candidates cannot identify programmes absent from its frame. The atlas therefore uses no acceptance-rate correction, omission estimate or extrapolated active-programme total.

Post-result ledger v98 preserves the same 400 cases and leaves 14 without a same-record identity signal, with unresolved design-weight mass 97.843484. The other 386 have heterogeneous later identity evidence, including structured rows, historical titles, embedded components and training routes. Their identity flags do

not settle strict eligibility, current intake or actual delivery. Primary labels remain 3/3/394; the 14/400 count is a different outcome measured in a later evidence channel.

Primary execution: research/next_study_v3_4/probability_400_execution_report_2026-10-02.md. Identity record: probability_400_unified_source_identity_ledger_v98.csv and probability_400_v98_all14_boundary_report_2026-10-29.md. These files are linked by hashes in artifact_manifest.json; no outcomes were recoded for this revision.

S3 Model coding and historical human review

Earlier four-model coding of 95 enriched blind cases agreed on inclusion in 95/95 and scope in 89/95. Six scope disagreements and nine unanimous controls received a targeted two-person audit in September. That 15-case audit is historical, targeted and nonrepresentative; it is not independent validation of all 95 cases, the 400-case draw or the global atlas. Later model source packets, source recoveries and sensitivity analyses have distinct denominators and cannot validate earlier judgements retrospectively. Subsequent follow-up used no additional human case review, and no external model was called for this revision.

S4 Directed contrasts and incomplete provider searches

The canonical registry retains 24 contrasts: 16 reconstructed historical contrasts and eight prospective contrasts. They vary in source frame, unit, vocabulary and evaluability. Spain's 5 to 57 is jointly a language-and-source contrast; Brazil's 18/23/36 of 40 varies evidence threshold; Ontario's 14 omissions among 52 is sector-specific. The Netherlands 15/17/21 varies the counted object. No pooled international causal effect or national correction factor is computed.

The first prospective round left two predictions unresolved after incomplete institutional recapture. The second selected Luxembourg and Iceland prospectively, but its substantive post-input protocol amendment is a sensitivity specification and 20 declared provider zeroes lack preserved search artifacts. It is not confirmatorily evaluable. These exercises remain technical context rather than headline results of the descriptive paper.

Later completed-provider search gates stand at Estonia 2/18, Ireland 0/18 and Finland 0/35. They measure completeness of specified searches, not the number of cyber providers or a national coverage rate. Partial catalogues, fee lists, blocked routes and absent lexical hits do not validate provider-wide absence. None supplies a national omission denominator.

S5 Temporal follow up and closure of acquisition

The 1 November attempt recorded 80 DNS failures. The 2 November follow-up recovered valid content in five of eleven Estonia/Ireland panel routes. Studyinfo's same ten search pages changed from 1,919 to 1,953 education OIDs while locked eight literal and 16 expanded title-ID sets remained unchanged. HEA still listed the same 18 providers. Uniarts' index changed from 51 to 58 guide URLs, with nine added and two absent; these are source events. Hanken's five-versus-four Vaasa discrepancy remained unresolved. None establishes programme opening, closure or national absence.

The revised descriptive paper closes source acquisition at the already available material. Future December and January observations are outside its completion requirements. A longitudinal claim would require a separately scoped study. Additional source discovery is not a condition for submitting this atlas after manuscript review and venue formatting.

S6 Reproduction and preserved history

Run research/tools/build_global_mapping_v4.py with the bundled Python runtime from the project root. It verifies that the three data-file SHA-256 values match the earlier frozen headline metrics, recomputes the country and regional tables and exact label frequencies, draws the map from the local world-atlas file and writes new v4.0 artifacts. It has no network client or external-model call. Outputs remain separate from app/data and earlier releases.

The map uses a Plate carree view from 60 degrees south to 85 degrees north and local world-atlas 110m polygons. Some small jurisdictions have no visible polygon; the CSV is the authoritative geographic lookup. map_crosswalk.json lists matched and undrawn jurisdictions and the explicit name/numeric-code mapping. Basemap boundaries are cartographic context, not a determination of recognition, sovereignty or the location of individual teaching sites.

The detailed preserved companion is research/outputs/GCyEO_manuscript_v3_4_180_full.docx, with its 93 references and source index. It documents the older analytical trajectory and is not presented as a newly harmonized global dataset. The v3.3 DOI identifies an earlier deposit; the v4.0 files are local working artifacts. External bibliography entries were retained from the existing manuscript without a new literature search or a live-service check.