Biosecurity Export Control Navigator: Cross-Jurisdictional Regulatory Comparison for Dual-Use Items[1]

Yousuf Golding
Independent

Apart Research

[1] *Research conducted at the AIxBio Hackathon, April 2026.*

Abstract

Researchers working with dual-use technologies must consult multiple national export control regulations, each published in a different format with different metadata depth. We present the Biosecurity Export Control Navigator, an open-source web application that ingests export control lists from three jurisdictions (US eCFR, EU Annex I and CELLAR XHTML, Australian DSGL 2024) into a unified Pydantic schema and renders them side by side with regime context, EU General Export Authorisation applicability, and an interactive map. The Navigator covers 3,384 entries across all ten Wassenaar categories, with 73 of 93 EU Category 1 top-level codes enriched from the CELLAR API. The unified data model identifies cross-references not visible in any single source. For example, the EU treats ECCN 1C352 (animal pathogens) inconsistently across its Annex I list and Annex II EU001 exclusions. The codebase and data are released openly. The tool is a navigation aid, not a substitute for legal counsel.

1. Introduction

Researchers and practitioners working with dual-use technologies (biological agents, sensitive chemicals, advanced materials, and restricted software) must navigate a fragmented set of national export control regulations whenever their work crosses borders. The same pathogen, controlled under ECCN 1C351, appears in the US Commerce Control List (eCFR Title 15 Part 774), the EU's Annex I to Regulation 2021/821, and Australia's Defence and Strategic Goods List, but each jurisdiction publishes its rules in a different format, with different metadata depth, and through different licensing mechanisms. Cross-jurisdictional comparison is currently a manual exercise.

This matters because compliance errors carry serious consequences. Improper export of dual-use items can result in legal penalties, license denials, and proliferation risk. Practitioners have no open-source tool to compare regulatory entries side by side. Commercial compliance platforms such as Descartes Visual Compliance and OCR Services exist, but they are proprietary, expensive, and oriented toward export-control professionals rather than researchers, policy analysts, or biosecurity practitioners working with limited budgets.

We built the Biosecurity Export Control Navigator to address this gap. The tool ingests export control lists from three jurisdictions (the United States, the European Union, and Australia) into a unified schema, then renders them side by side in a web interface with regime context, EU General Export Authorisation applicability, and an interactive map. Our contributions are:

1 A unified schema and ingestion pipeline that parses three structurally different regulatory sources (US

eCFR XML, EU Commission Excel and CELLAR XHTML, Australian DSGL EPUB) into 3,384 comparable structured entries.

2 A web-based comparison interface with side-by-side regulatory cards, an interactive Leaflet map with toggleable regime membership overlays, EU General Export Authorisation applicability with three-state coloring, and coverage gap warnings.

3 An open-source codebase and structured data files with primary-source attribution for every regulatory entry.

2. Related Work

Export control regulations are published by national authorities in formats reflecting each jurisdiction's regulatory tradition. The US Bureau of Industry and Security publishes the Commerce Control List as part of the eCFR with an XML API, with each ECCN heading followed by labeled sections (License Requirements, License Exceptions, Items Controlled, Related Controls). The EU Commission publishes Annex I to Regulation 2021/821 as an Excel spreadsheet in a hierarchical dotted notation such as *1C351.a.1.a*, with the consolidated regulation text separately available as XHTML through the EU Publications Office's CELLAR API. Australia's Defence and Strategic Goods List is published by the Federal Register of Legislation as an EPUB, with the HTML rendering returning empty shells when fetched programmatically because the source is a JavaScript-rendered single-page application.

This format fragmentation is the practical obstacle. Each source is open and well-documented, but the metadata depth varies considerably: US entries include reasons for control, country charts, license exceptions, and items controlled paragraphs, while EU entries provide code and label only in the Excel (with full text reachable through CELLAR), and Australian entries include code, label, and items controlled but no licensing metadata. A cross-jurisdictional comparison must therefore handle this asymmetry directly rather than force the sources into a uniform shape. Standards such as the National Information Exchange Model (NIEM) provide one model for how structured regulatory data could be exchanged across agencies, and the Navigator is a small-scale proof-of-concept in the same spirit.

To our knowledge, no open-source tool currently provides cross-jurisdictional comparison of dual-use export controls. The market is dominated by commercial compliance platforms such as Descartes Visual Compliance and OCR Services, which are proprietary and licensed for enterprise use. The AIxBio Hackathon's Track 3 (AI Biosecurity Tools), sponsored by Fourth Eon Bio, included "Prototype a biosecurity policy tracker monitoring regulatory changes across jurisdictions" on its example project list, and the Navigator addresses this prompt directly.

3. Methods

3.1 Data Sources and Ingestion

United States (eCFR API). The US Commerce Control List is published as XML via the eCFR API (Title 15 Part 774, Supplement No. 1). The XML structure is flat: ECCN headings appear as heading tags, and the content for each entry follows in a stream of paragraph, note, and table elements with no explicit section boundaries. To partition this stream into structured fields, we built a four-state parser (*PREAMBLE*, *LICENSE_REQ*, *LICENSE_EXC*, *ITEMS_CONTROLLED*) that detects header phrases such

as "License Requirements", "Reason for Control:", "List Based License Exceptions", and "List of Items Controlled" and routes subsequent content into the corresponding *ControlEntry* fields. The parser yielded 637 entries across all 10 categories, with high field population rates: *items_controlled* 92%, *reasons_for_control* 91%, *country_chart* 89%, *license_exceptions* 91%.

European Union (Commission Excel and CELLAR XHTML). The EU's primary publication is an Excel spreadsheet (September 2024) containing 2,652 entries in a hierarchical dotted notation. The Excel provides only *code* and *label* with no licensing metadata or items controlled paragraphs. To recover the full regulatory text, we integrated the EU Publications Office's CELLAR API. Two non-obvious header and identifier requirements (a 3-letter ISO 639-3 *Accept-Language* code and a date-suffixed CELEX identifier for consolidated text) made initial requests fail. With corrected requests, the API returns a 3.0 MB XHTML document of the consolidated regulation. We parse this XHTML with element-based DOM walking (BeautifulSoup), locating *<p class="norm">* elements matching the ECCN heading pattern and walking subsequent siblings until the next heading or category boundary. Of the 93 Category 1 top-level codes, 72 are enriched from CELLAR and 1 (1C352) is enriched from sub-entry labels in the Excel as a fallback. The remaining 20 unenriched codes include all 1D (Software) and 1E (Technology) entries, plus several nuclear- and missile-related codes in 1A, 1B, and 1C. The 1D and 1E entries are short cross-references to the General Technology Note rather than item lists.

Australia (DSGL 2024 EPUB). The Australian Federal Register of Legislation publishes regulatory documents through a JavaScript-rendered Angular SPA, so direct HTML scraping returns empty shells. After investigating the OData API at *api.prod.legislation.gov.au*, we found that an EPUB version of the DSGL is available at a date-qualified URL. The EPUB contains a single 2.3 MB HTML file with the entire DSGL. ECCN codes appear in a spaced format such as *1. C. 351.* rather than the standard form *1C351*, and a regex normalizes this. Heading paragraphs are distributed across multiple CSS classes because the DSGL assigns different styles to items originating from different multilateral regimes, and the parser scans all of these classes. We scoped ingestion to Category 1 (95 entries) for the hackathon timeframe.

Unified schema. All entries are represented as a Pydantic *ControlEntry* model with nullable jurisdiction-specific fields. The asymmetry is intentional: it reflects what each source actually publishes rather than collapsing differences across sources. Storage is JSON (1.7 MB for US, 2.2 MB for EU, 135 KB for AU), small enough that a relational database is unnecessary, and the JSON format keeps the data human-readable and diffable.

3.2 Search and Query Layer

The query library (*search.py*, 329 lines) is a pure-function module with no Flask dependency, so it can be reused outside the web app. It exposes case-insensitive substring search across *label*, *items_controlled*, and *related_controls*, and structured filters on jurisdiction, category (with human-readable names for all 10 Wassenaar-standardized categories), reason-for-control code, and ECCN prefix. The *get_comparison(code)* function returns a dictionary keyed by jurisdiction containing the matching entries. This is the central operation for the compare page.

3.3 Regime Mapping and EUGEA Integration

Four multilateral export control regimes (Australia Group, Wassenaar Arrangement, Nuclear Suppliers Group, and Missile Technology Control Regime) each define lists of controlled items that member states implement through their national export control regulations. We hardcoded the regime member lists from

official sources (43 AG members, 42 Wassenaar, 48 NSG, 35 MTCR) and a mapping from ECCN prefix to regime. The Australia Group's 43-member count includes 42 states and the EU as an institutional participant. Lookup uses longest-prefix match, so specific codes such as *1C351 → Australia Group* take priority over broad category fallbacks such as *1 → Wassenaar*. Dual-regime codes are supported.

Annex II of Regulation 2021/821 defines eight EU Union General Export Authorisations (EU001 through EU008), each of which permits export of specified items to specified destinations without an individual license. We hardcoded all eight EUGEAs from the BAFA-published PDF (OJ L 206), with their item scopes, destination lists, and explicit exclusions. The lookup function *get_applicable_eugeas(code)* returns the applicability status (applicable, partially excluded, or fully excluded) for each EUGEA against a given code. Three coverage warnings are surfaced on the compare page: one-sided coverage (no equivalent in another jurisdiction), shared regime context (both jurisdictions implement through the same regime), and EU data asymmetry (EU data includes code and label only).

3.4 Web Application

The web app is a Flask application (287 lines in *app.py*) with three routes: index for per-jurisdiction statistics, compare as the centerpiece, and search for keyword and filter queries. Pages are rendered through Jinja2 templates with plain CSS and vanilla JavaScript. The compare page presents up to three side-by-side cards (US, EU, and AU when available), each showing the entry's metadata, items controlled, regime context, EUGEA applicability, and a link to the original source.

The compare page also includes an interactive Leaflet map (Leaflet 1.9.4 via CDN, Natural Earth 110m country boundaries, around 259 KB of GeoJSON). The map supports two-click origin-and-destination selection synchronized with the form dropdowns, a regime overlay toggle for the four multilateral regimes, and per-country tooltips with regime membership data. Figure 2 shows the map with the Australia Group overlay active. The map is progressive enhancement, and the dropdowns work without JavaScript.

4. Results

Data Coverage

The Navigator covers 3,384 entries across three jurisdictions: 637 from the US (all 10 categories), 2,652 from the EU (all 10 categories, with Category 1 enriched), and 95 from Australia (Category 1 only). At the top-level ECCN, 336 codes are present in both the US and EU Commerce Control Lists, 301 are US-only (including US-specific xxx521 entries pending classification and xxx998 catch-all entries), and 43 are EU-only (including nuclear items 0A001 and 0B001 through 0B007). EU enrichment from CELLAR added *items_controlled* text to 73 of 93 Category 1 top-level codes (72 from CELLAR XHTML and 1 from sub-entry fallback).

Jurisdiction Entries Categories Source

US 637 0 to 9 eCFR XML

EU 2,652 0 to 9, plus R Commission Excel and CELLAR XHTML (Cat 1) AU 95 1 only DSGL 2024

EPUB

Total 3,384

Table 1: Per-jurisdiction entry counts, category coverage, and source.

The CELLAR-enriched text includes regulatory specifics not visible in the sub-entry labels: CAS registry numbers (for example, Amiton, CAS 78-53-5), concentration thresholds for mixture exemptions (1% for non-CWC states, 30% for CWC states), vaccine-strain exemptions, and cross-references to military goods controls. These are details a compliance researcher needs to access, and they are not recoverable from a basic scrape of the Excel.

Key Findings

The 1C352 EUGEA finding. Section I of Annex II to Regulation 2021/821 lists the items excluded from EU001, the broadest of the eight General Export Authorisations. The exclusion list explicitly covers 1C351 (human and animal pathogens), 1C353 (genetic elements and GMOs), 1C354 (plant pathogens), and 1C450.a.1 / 1C450.a.2 (specific chemical precursors). It does not include 1C352 (animal pathogens). EU001 therefore covers 1C352 exports to its destination list, even though the surrounding bio-pathogen codes do not. There is a further inconsistency. The EU Commission Excel maintains 1C352 as a distinct entry with its own sub-items, but the consolidated regulation text places animal pathogens under 1C351 rather than under a separate 1C352 code. The same items therefore appear under different codes depending on which authoritative EU source a reader consults.

Structural asymmetry across jurisdictions. US entries are flat, with paragraph-level *items_controlled* on each ECCN. EU entries are hierarchical, with sub-items appearing as separate *ControlEntry* objects in a tree (1C351, 1C351.a, 1C351.a.1, and so on). AU entries follow the US flat pattern but with sparser metadata. The schema preserves these differences rather than collapsing them. The US compare card shows paragraph items, the EU card shows the CELLAR-enriched paragraph items plus a count of sub-entries in the hierarchy, and the AU card shows what the DSGL publishes. This preserves traceability to the original sources.

Coverage gaps. At the top-level ECCN, 301 codes appear in the US but not the EU, and 43 appear in the EU but not the US. The asymmetry reflects scope differences. The US Commerce Control List includes many series not derived from multilateral regimes, such as xxx521 entries pending classification and xxx998 catch-all entries. The EU lists nuclear items, including 0A001 and 0B001 through 0B007, that the US handles through a separate Nuclear Regulatory Commission framework rather than the Commerce Control List.

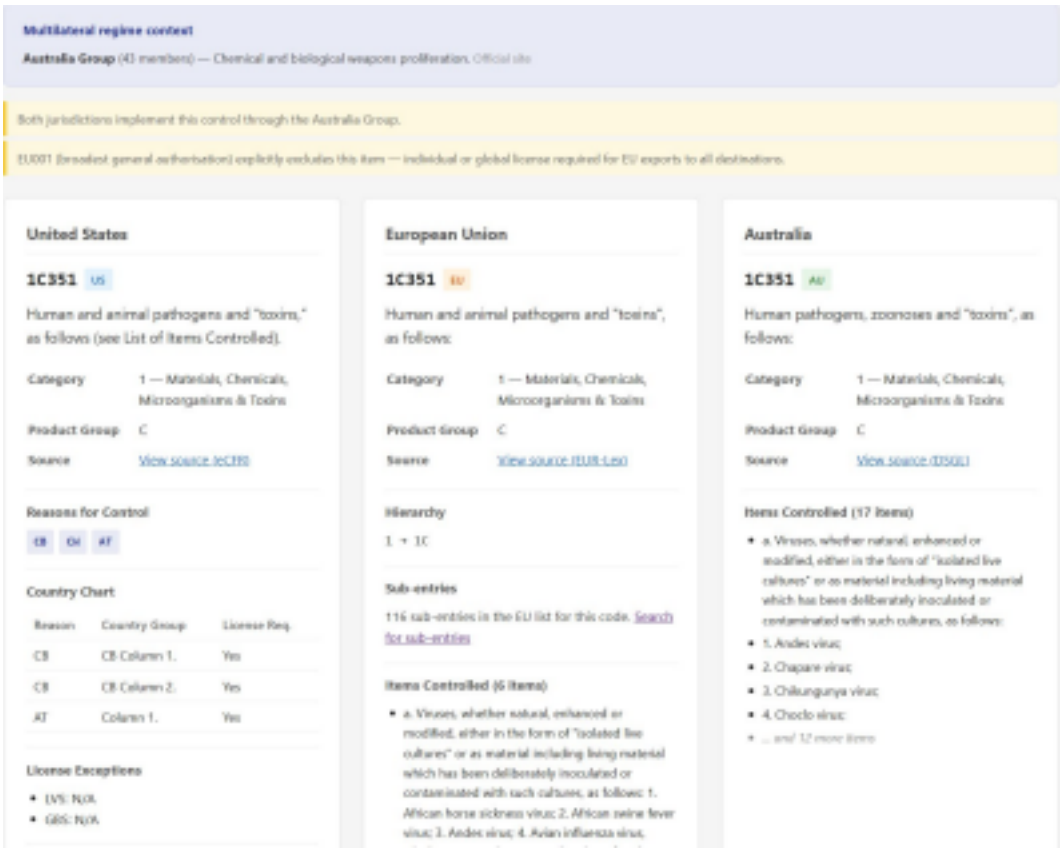

Figure 1: Three-column comparison of ECCN 1C351 (human and animal pathogens) showing US regulatory detail (left), EU entry with EUGEA applicability and CELLAR-enriched items controlled (center), and Australian DSGL entry (right).

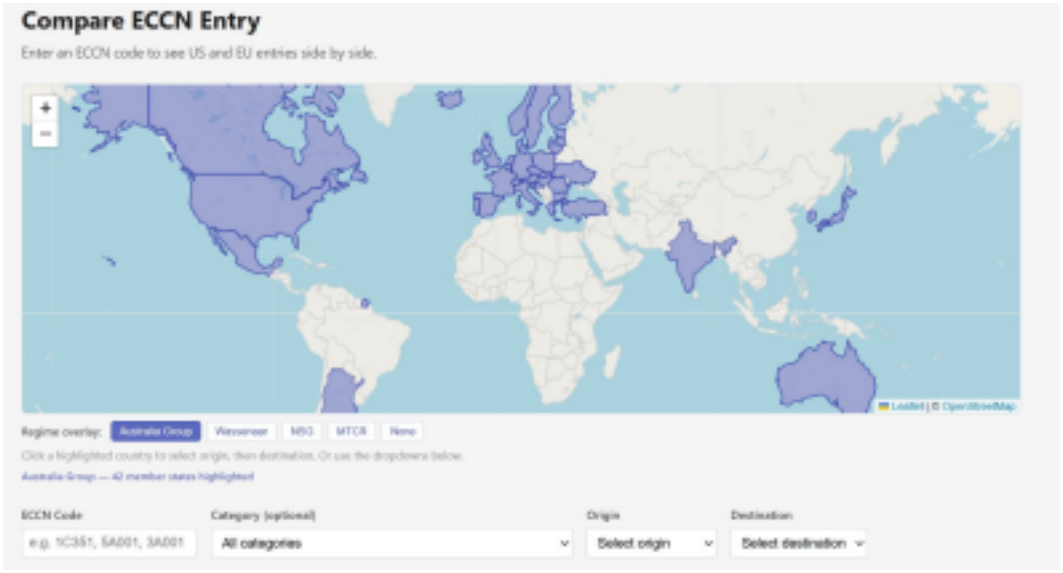

Figure 2: Comparison page map view with the Australia Group regime overlay active. Member states are highlighted in indigo, and non-members appear in pale gray. Origin and destination selections take visual priority over the regime coloring.

5. Discussion and Limitations

The Navigator demonstrates that structured cross-jurisdictional regulatory comparison is feasible using publicly available data sources, even within a hackathon timeframe. The most important architectural decision was the superset schema with nullable jurisdiction-specific fields, which handles regulatory asymmetry directly without forcing false equivalences. Each source is rendered with the metadata it actually publishes, and coverage warnings make the gaps explicit rather than hiding them. The 1C352 EUGEA case shows the value of cross-referencing across sources, since neither source identifies this inconsistency in isolation.

Limitations

EU data outside Category 1 remains code-and-label only. Full CELLAR enrichment was not within scope for the hackathon. The EU Excel dates from September 2024, so the November 2025 Delegated Regulation 2025/2003 (which adds quantum technology and other items) is not reflected. Australian coverage is Category 1 only at 95 entries. Country selection on the map resolves to jurisdiction (US, EU member states, AU) rather than to country-level licensing rules, and EU member states apply national policies that vary, so the tool surfaces EU-level regulation rather than national implementation. The tool is a navigation aid, not a compliance engine. No user testing was conducted with practicing export control specialists. The work was completed by a single developer over a three-day hackathon. A more detailed treatment of edge cases and scalability constraints is provided in the Appendix.

Future Work

Extending CELLAR enrichment to all 10 categories would close the largest data gap. Adding additional jurisdictions is the natural next axis. Canada's Export Control List uses a different code format requiring translation, and the UK Strategic Export Control Lists are PDF-only. Integration with data exchange standards such as NIEM would support interoperability with existing compliance infrastructure. Periodic re-ingestion with diffing would support automated regulatory change detection. User research with practicing export control specialists would validate or revise the design choices made here.

6. Conclusion

We built a tool that ingests export control lists from three jurisdictions (US, EU, Australia) into a unified schema and renders them side by side in a web interface, with regime context, EU General Export Authorisation applicability, and an interactive map. The Navigator covers 3,384 entries with primary-source attribution for each, and it identifies regulatory details that require cross-referencing across sources to discover, including the 1C352 EUGEA exclusion case. The work is a hackathon prototype rather than a research contribution, but the codebase, data, and report are publicly available for extension.

Code and Data

Code repository: *https://github.com/yooleee/dual-use-navigator*

Data: included in the repository under *data/*. US data sourced from the eCFR API, EU data from the Commission Excel and the CELLAR API, AU data from the DSGL 2024 EPUB. All entries include primary-source URLs and last-fetched timestamps.

References

[1] U.S. Bureau of Industry and Security. *Export Administration Regulations*, Title 15 CFR Part 774, Supplement

No. 1, Commerce Control List. https://www.ecfr.gov/current/title-15/subtitle-B/chapter-VII/subchapter-C/part-774/supplement-No.-1-to-part-774.

[2] European Parliament and Council of the European Union. *Regulation (EU) 2021/821 of 20 May 2021 setting up a Union regime for the control of exports, brokering, technical assistance, transit and transfer of dual-use items.* OJ L 206, 11.6.2021. https://eur-lex.europa.eu/eli/reg/2021/821/oj.

[3] European Commission. *Commission Delegated Regulation (EU) 2025/2003 amending Annex I to Regulation (EU) 2021/821.* November 2025.

[4] Commonwealth of Australia. *Defence and Strategic Goods List 2024* (F2024L01024). Australian Federal Register of Legislation. https://www.legislation.gov.au/F2024L01024/asmade/text.

[5] National Information Exchange Model (NIEM). https://www.niem.gov/.

[6] Publications Office of the European Union. *CELLAR, common data repository.* https://op.europa.eu/en/web/cellar.

[7] Apart Research. *AIxBio Hackathon, April 2026.* https://www.apartresearch.com/.

[8] The Australia Group. https://www.dfat.gov.au/publications/minisite/theaustraliagroupnet/site/en/index.html.

[9] The Wassenaar Arrangement on Export Controls for Conventional Arms and Dual-Use Goods and Technologies. https://www.wassenaar.org/.

[10] Nuclear Suppliers Group. https://www.nuclearsuppliersgroup.org/.

[11] Missile Technology Control Regime. https://mtcr.info/.

Appendix: Limitations and Dual-Use Considerations

This appendix addresses the items required for AIxBio Hackathon submissions:

• limitations, including false negatives, edge cases, and scalability constraints, •

dual-use risks,

• responsible disclosure recommendations,

• ethical considerations, and

• suggestions for future improvements.

A.1 Limitations

Beyond the scope limitations noted in Section 5, several technical limitations affect the tool's reliability.

False negatives in keyword search. Search is case-insensitive substring matching across *label*, *items_controlled*, and *related_controls*. The 2,579 EU entries that are not among the 73 enriched Cat 1 top-level codes have *items_controlled* set to None and can be matched only on their short labels. A user searching for a specific reagent or organism will find matches in the US data (full items-controlled coverage) but miss EU entries that would match if their full text were available.

Edge cases in source parsing. The eCFR XML parser depends on heading-string detection ("License Requirements", "Reason for Control:", and others). If a future eCFR revision changes these section labels or their formatting, the parser will silently lose data. The CELLAR XHTML parser depends on specific CSS class names that may change in future regulation versions. The DSGL EPUB parser depends on the

ECCN spaced-format pattern and on a fixed set of heading CSS classes. Style variations within a single document already required scanning multiple distinct classes, and a future DSGL release could introduce additional ones.

Scalability constraints. Reference data for the four multilateral regimes and eight EUGEAs is hardcoded from official PDFs at a fixed point in time. Membership changes (regime additions, EUGEA amendments via Annex II revisions) require manual updates to the source code. The web app uses in-memory JSON stores, which is appropriate for the current data volume but would need to be replaced with a database backend if the corpus grew by an order of magnitude.

A.2 Dual-Use Risks

The Navigator aggregates regulations that are individually public. We do not publish data, methods, or findings that were not already accessible from the source publications. The 1C352 EUGEA inconsistency described in Section 4 is a regulatory artifact of how Annex I and Annex II are constructed. It does not create a usable export pathway, since national licensing requirements still apply to all listed items regardless of whether a General Export Authorisation is available.

A bad-faith user could in principle use the tool to identify jurisdictional differences in coverage, for instance to plan transshipment through a less restrictive jurisdiction. This concern applies to the underlying public regulations rather than to the Navigator specifically. The same comparison can be done by hand from primary sources within an hour for any single ECCN, and the time cost of manual cross-referencing was never the binding constraint on evasion. We judge the marginal uplift to bad-faith users to be low. The Navigator does not include any non-public information. It does not aggregate enforcement-action data, individual licensing decisions, or any structured collection that could function as a target list.

A.3 Responsible Disclosure

We discovered no security vulnerabilities and conducted no adversarial testing. The 1C352 EUGEA inconsistency described in Section 4 is a regulatory inconsistency in publicly available text rather than a vulnerability requiring private disclosure. We note it openly in this report and recommend that interested readers raise it with the European Commission's Directorate-General for Trade if substantive clarification is sought. No private disclosure path is required.

A.4 Ethical Considerations

All data is sourced from official government publications under terms that permit reuse. The codebase and data are released under an open license, with primary-source URLs and last-fetched timestamps recorded for every entry, so any user can verify our parsing against the original sources. The tool collects no user data, requires no authentication, and does not transmit user activity to any third party. The web app is positioned consistently as a navigation aid rather than a compliance engine, both in inline disclaimers within the interface and in the public README. We do not provide a mechanism for users to record their own export decisions or to request licensing assistance.

A.5 Suggestions for Future Improvements

Several improvements would directly address limitations and dual-use considerations identified above. Periodic re-ingestion with diffing would surface regulatory amendments to users and reduce the risk of decisions made on stale data. Engagement with regulatory authorities on inconsistencies identified by the

tool, beginning with the 1C352 EUGEA case, could contribute to upstream resolution rather than only downstream surfacing. Adding national-level licensing data for EU member states would make the country-selection feature more accurate and would reduce the risk of users assuming EU-level uniformity that does not exist in practice. User research with practicing export control specialists would surface unintended use patterns that automated review cannot.

LLM Usage Statement

Claude (Anthropic) was used for project planning, prompt design for coding sessions, research on regulatory data sources, and assistance drafting this report. All regulatory data was sourced from official government publications. All claims and results were verified against primary sources.

Beyond the uses noted above, this report was drafted with assistance from Claude Opus 4.7, with multiple editing rounds covering structure, style, and section-level revisions. I reviewed all the drafts and verified the information against the development log and primary sources. For the next hackathon I plan to allocate more time to writing the report.