The University of Maryland's Cyber Events Database 2.0: A Systematic Framework for Analyzing Global Cyber Threats

Charles Harry, Devin Entrikin and William Lucyshyn
The Center for Governance of Technology and Systems (GoTech) at the University of Maryland's School of Public Policy, College Park, USA

charry@umd.edu
dentrik@umd.edu
lucyshyn@umd.edu

Abstract: Cyberattacks are increasing in scale, scope, and impact, yet systematic, accessible data on these events remain fragmented, narrowly scoped, or methodologically opaque. Existing cyber incident datasets often focus on specific threat types or high-profile cases, are proprietary, or lack transparent coding rules, limiting comparative analysis and cumulative research. This paper introduces the Cyber Events Database 2.0 (CEDB 2.0), a publicly sourced, event-level repository of global cyber incidents from 2014 onward designed to support reproducible research and strategic cybersecurity analysis. The CEDB 2.0 employs a mixed-methods data collection approach that combines automated web scraping with multilingual, near–real-time news monitoring, integrating the Global Database of Events, Language, and Tone (GDELT) Project beginning in 2025. As of February 2026, the database contains 16,382 coded cyber events across 175 countries and 1,431 distinct threat actors. Each event is classified using a theory-informed taxonomy capturing actor type, motive, target sector, affected country, and observable effects. By enabling cross-sectional and longitudinal analysis at scale, the CEDB allows researchers to move beyond anecdotal case studies toward systematic, evidence-based assessment of state and non-state cyber behavior. The paper details the database's methodology, structure, applications, limitations, and future development.

Keywords: Cybersecurity, Cyber operations, Structured event data, Threat landscape, Data-driven analysis, Strategic decision-making

1. Introduction

Cyberattacks are growing in number and impact, carried out by various actors with different motives, targeting multiple industries and causing effects from data breaches to operational disruptions. These incidents, including the 2021 ransomware attack on Colonial Pipeline that interrupted fuel supplies along the eastern U.S. (Easterly, 2023), the 2020 SolarWinds breach attributed to Russian operatives that compromised numerous systems (GAO, 2021), and recent attacks on water utilities linked to pro-Russian and Iranian groups (ODNI, 2024), all represent cyber events significant to academic research. Less severe incidents, such as attacks on schools (Kennedy, 2024), small contractors (Rogers, 2020), and software vendors (JD Supra, 2023), are less frequently discussed, leading much of the analysis to be based on a small data sample subject to sampling bias. Information regarding the individuals or groups responsible, their intentions, the affected sectors, and the actual consequences is typically scattered across news reports, specialized private intelligence agencies, periodic statements by businesses and governments, or subscription-based commercial databases. While some open-source data repositories of cyber events exist, they often contain information that remains fragmented, narrowly scoped, or methodologically opaque.

To address these shortcomings, the Center for the Governance of Technology and Systems (GoTech) has expanded on initial work in conjunction with the Center for International and Security Studies at Maryland (CISSM) to develop and launch a revised Cyber Events Database (CEDB 2.0). CEDB 2.0 integrates publicly accessible news articles by referencing identified cyber-relevant news and trade publications, as well as the Global Database of Events, Language, and Tone (GDELT) Project's Web News NGrams 3.0 (2021) and Article List datasets. The CEDB now offers comprehensive global coverage through real-time monitoring of news sources in over 65 languages, ensuring broader, more timely access to information.

This article outlines the purpose, methodology, structure, and uses of CEDB 2.0. It reviews the relevant literature, explains methods, describes database fields, and covers use cases, limitations, and future prospects. By highlighting cyber events, the CEDB 2.0 helps reassess key issues, including threat locations, motives, patterns, impacts, and responsible actors.

2. Literature Review

Cyber events vary across multiple critical dimensions, including threat actor type and motivation, targeted industry, geographic scope, and the nature and severity of observable effects. Despite this heterogeneity, cyber event reporting remains inconsistent and difficult to standardize. Many existing data collection initiatives fail to

meet foundational criteria of exhaustiveness, exclusivity, and consistency, limiting their suitability for comparative and theory-driven analysis (Harry & Gallagher, 2018). Others focus on specific threat categories or technical artifacts, struggle to capture evolving attack vectors, rely on incomplete or outdated records, or lack adequate coverage of multi-stage and multi-actor campaigns (Cremer et al., 2022; Mvula et al., 2023; Goldschmidt & Chudá, 2024). As a result, much of the empirical work on cyber war and security relies on qualitative methods over quantitative large-N analysis (Kong, Kim, and Lim, 2019) (Buchanan, 2020), with some notable exceptions (Valeriano, Jensen, and Maness, 2018) (Kostyuk & Zhukov, 2017). Collectively, these limitations undermine consistent event classification, obscure meaningful variation across incidents, and introduce selection bias, which complicates inference and policy-relevant analysis.

We reviewed 13 datasets, often cited in academic and professional publications, that are most closely related to the CEDB 2.0. The datasets are broadly reviewed within four distinct groups based on their observed focus, completeness, audience, and description. The first group of cyber incident datasets prioritizes state-linked activity or high-impact events rather than the broader universe of cyber threats. For example, the Center for Strategic and International Studies (CSIS) Significant Cyber Incidents timeline documents notable cyberattacks since 2006, emphasizing state-sponsored operations, espionage, and incidents with estimated losses exceeding $1 million, particularly those affecting government and defense targets (CSIS, n.d.). Similarly, the Council on Foreign Relations (CFR) Cyber Operations Tracker catalogs state-attributed cyber operations with an explicit focus on interstate dynamics and strategic competition (CFR, n.d.).

Both resources provide valuable descriptive context but are not designed as taxonomic or event-level datasets. Events are coded narratively rather than through a standardized, multi-dimensional classification scheme, and inclusion criteria privilege strategic salience over systematic coverage. As a result, these datasets are ill-suited for large-N comparative analysis, replication, or longitudinal assessment of variation in actor behavior, targeting patterns, and effects across different types of cyber activity.

A second group focuses on intrusion detection and specific threats. This includes the Canadian Institute for Cybersecurity (CIC) datasets, such as CIC-IDS2017, which provide network flow data for intrusion detection and machine learning applications that simulate specific attack vectors (e.g., DDoS or malware). However, they lack real-world incident narratives or longitudinal trends (Sharafaldin et al., 2018). While invaluable for detection and model training, these datasets capture technical artifacts rather than socially and politically situated cyber events with identifiable victims and consequences.

A third category of datasets monitors broad cyber incidents rather than specific subtypes. The European Repository of Cyber Incidents (EuRepoC) offers a comprehensive global dataset encompassing over 3,400 politically or security-relevant cyber operations from 2000 to 2024, featuring 60 variables pertaining to initiators, targets, and intensity. Similarly, the Dyadic Cyber Incident and Campaign Dataset (DCID) focuses on campaigns and incidents associated with interstate conflicts. Although these datasets are methodologically robust, their emphasis on actions by major state actors may underrepresent non-state proxies, such as criminal entities and hacktivists. Additionally, their primary focus on interstate conflict constrains their capacity to document cyber operations occurring within intrastate conflicts.

A fourth group comprises specialized, community-driven, or technically focused repositories, each centered on specific attack categories. For example, Ransomware.live automates the scanning of ransomware groups' data leak sites and other sources (Ransomware.live, n.d.; Poireault, 2024). Likewise, Ransomware monitors cryptocurrency payments sent to wallets associated with known ransomware activity (Ransomwhere, n.d.; Cable et al., 2024). The U.S. Department of Health and Human Services Office for Civil Rights (HHS OCR) gathers reports of HIPAA-covered breaches that impact 500 or more people in the healthcare field (HHS, 2023; Alder, 2025). Cloudflare observes major Internet disruptions, including DDoS campaigns or state-mandated shutdowns (Cloudflare Radar, n.d.; Azevedo, 2023). Shadowserver offers global threat telemetry (Shadowserver, n.d.; Kijewski, 2021) while the “World’s Biggest Data Breaches & Hacks” compiles cyber event information into a timeline with source links (Information is Beautiful, n.d.). These tools tend to have a narrow focus, ambiguous event classifications, technical orientations, or limit themselves to certain types of incidents.

Collectively, these four categories of cyber event data repositories demonstrate both the advancements achieved and the ongoing challenges in documenting cyber activity. State-centric timelines prioritize strategic significance over comprehensive coverage and intrusion detection datasets focus on technical indicators but often lack contextual and impact information. Dyadic and interstate repositories limit analysis to specific participants and conflict patterns and specialized or community-driven platforms offer detailed insights into threat vectors. Consequently, no single dataset currently provides thorough, event-level data that is sufficiently

broad, methodologically transparent, and systematically organized for comparative and longitudinal research across actors, sectors, and regions. These gaps highlight the necessity for a unified, publicly accessible repository of cyber events that synthesizes diverse sources, employs a consistent classification scheme, and records the tangible effects of cyber operations, objectives addressed by the CEDB 2.0.

3. Methodology

The CEDB 2.0 is designed to provide a systematic, event-level record of global cyber activity that supports comparative, longitudinal, and policy-relevant analysis. Building on earlier efforts to document cyber incidents using open-source reporting, the CEDB 2.0 applies explicit inclusion criteria, a structured taxonomy, and a hybrid data-collection approach that combines automation with human validation. This section outlines the database's event definition, data collection pipeline, and validation process, highlighting methodological choices intended to mitigate selection bias and improve global coverage. The first source stream relies on automated web scraping of a curated and regularly updated watchlist of open-source outlets known for timely reporting on cyber incidents, including specialized cybersecurity blogs, major news organizations, public feeds, leak sites, and paste sites. Prior to 2025, this Python-based scraping pipeline served as the primary mechanism for identifying candidate events.

The site-specific scraper collects cyber-related articles via RSS feeds (using feedparser, rapidfuzz, dateutil.parser, etc.), then filters entries by date, keyword match, and deduplication. For each retained entry, it records the RSS publication timestamp (reported_date), the article link (source_url), and the RSS summary text (description), as well as the RSS title. Execution timing is captured in log files via a local timestamp. Where possible, preliminary indicators of threat actor type and apparent motivation are inferred heuristically from textual context, though these provisional labels are always subject to subsequent human review.

Beginning in January 2025, the data collection was expanded to include the GDELT Project's Web News NGrams 3.0 (2021) and Article List datasets. GDELT's near-real-time monitoring of broadcast, print, and online media in over 65 languages across more than 100,000 sources. This dramatically increases both temporal recency and geographic breadth of coverage, particularly for incidents reported outside English-language specialist outlets. As an example, the U.S. share of recorded cyber events declined from 48.8% to 36.9% when comparing the 2014–2024 dataset, compiled using the original collection method, to records from 2025 onward, when GDELT-based identification was incorporated. The change likely reflects improved global coverage rather than a substantive shift in cyber activity. Events ingested via GDELT are subjected to the same definitional criteria and manual validation workflow as traditionally scraped incidents; only the upstream acquisition mechanism differs. This methodological shift is explicitly flagged in the dataset schema to enable temporal-comparability analyses.

Harvested records are aggregated daily, deduplicated by URL and semantic similarity, and appended to two intermediary comma-separated values (CSV) files, one containing raw open-web scrapes and a second reserved for post-2025 additions from GDELT-sourced articles. Importantly, the algorithm's sole function is to collect potentially relevant material, which is then reviewed by one of the researchers.

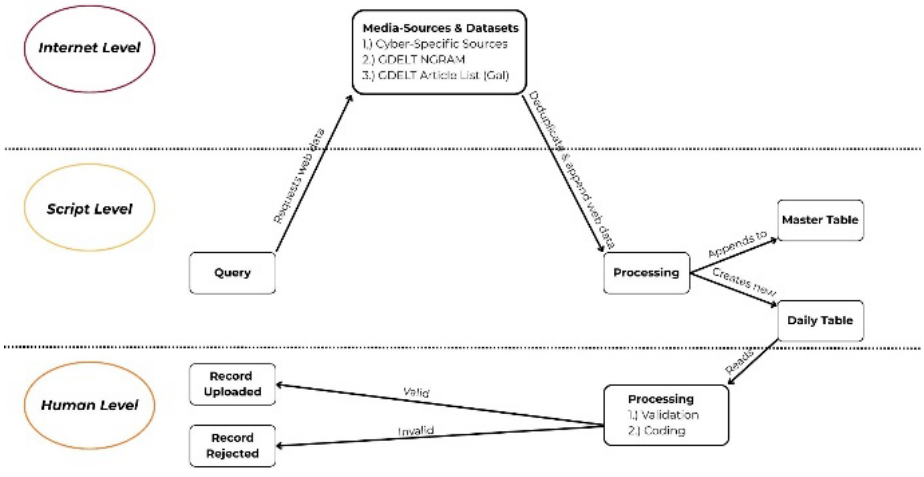

Figure 1: Data Collection and Coding Process

Following daily aggregation, trained researchers restrict inclusion to cyber events that meet two necessary conditions: (1) the operation must have produced a discernible real-world effect, and (2) there must be an identifiable target victim. Events that lack an observable impact against an entity, such as unsuccessful intrusions, or vulnerability disclosures without exploitation are systematically excluded. Each record is linked to at least one publicly accessible or archived primary source containing sufficient narrative detail to support independent verification and consistent coding.

Each cyber event record consists of 45 standardized fields that document what happened, who was affected, and how. The event date reflects the best available information on when the intrusion or effect occurred, while the reported date is the publication date of the cited source URL. Each record includes a specific victim organization (or individual) and its corresponding two-digit NAICS industry code when available. Threat actors fall into one of the following categories: criminal, nation-state, hacktivist, hobbyist, terrorist, or undetermined. Events are categorized by primary effect as exploitive, disruptive, or mixed (both exploitive and disruptive effects). Subtypes specify the nature of an event based on the part of the target organization's IT infrastructure that was most seriously impacted. These include disruptive subtypes such as data attacks, message manipulation, external or internal denial-of-service, physical attacks, or specific forms of exploitation against application servers, end hosts, network infrastructure, sensors, and/or data in transit. Beginning with 2025 data, disruptive and mixed events include severity coding for magnitude (extent of disruption), duration (length of disruption), and scope (systems or assets affected). Exploitive and mixed events record compromised data by category, which includes intellectual property, organizational data, and customer or client data, using quantitative detail whenever possible. Motive is coded as financial, political or industrial espionage, sabotage, protest, reputation, personal attack, or undetermined based on evidential inference. Additional fields capture victim and actor countries using standardized naming and ISO3, binary variables for organization and alliance membership, a concise factual description, and a verifiable source link. Finally, for cyber events targeting the United States, records also capture the state and county where the organization operates, if available.

This process serves three primary objectives:

- Confirmatory validation that the candidate event has both an attributable victim and observable effect;
- Consistent assignment of core taxonomic attributes; and
- Precise classification of affected industry and granular end effects according to the structured impact taxonomy

Processed and validated records are consolidated, deduplicated across both ingestion streams, and released in monthly cumulative updates accompanied by schema documentation and a change log that highlights any coding adjustments. The resulting dataset thus offers scholars a transparent, reproducible, and longitudinally consistent resource for quantitative (and qualitative) analysis of cyber conflict dynamics (Harry et al., 2025).

4. Results

As of January 2026, the database contains 16,382 cyber events across 175 countries involving 1,431 unique threat actors. The average number of events per month is 112, and the yearly average is 1,239 (2014-2025). When we break the data down by types of effects, exploitive events are the most common, driven largely by criminal actors seeking financial gain through data theft and extortion. Mixed-effect attacks, such as ransomware that combines data theft and encryption, saw a spike in 2022 and 2023. However, there has been an observed shift in threat behavior in recent years as more criminal actors use data theft-only extortion methods (Symantec & Carbon Black, 2026). Interestingly, there is a notable spike in disruptive attacks in 2022, which aligns with heightened hacktivist and nation-state activity following Russia's invasion of Ukraine.

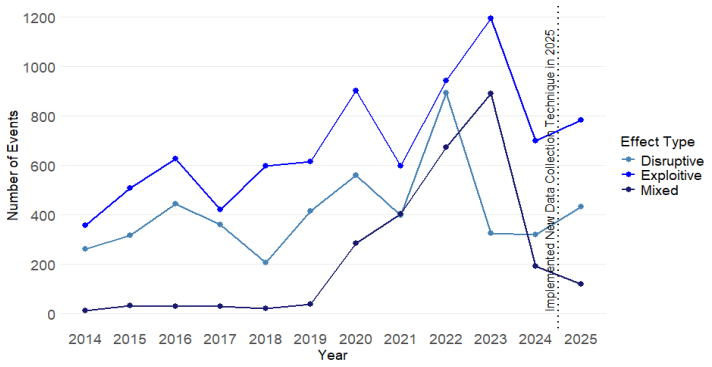

Figure 2: Cyber Events by Effect Type - All Records (2014-2025)

Figure 3 shows that criminals, driven by financial gain, are, by far, the most prolific threat actor, accounting for 74.5% of events tracked in the CEDB 2.0, while hacktivist groups, motivated by ideological or political alignments, are the second most active (with a 13.2% share) for the period 2014-2025. Russia's invasion of Ukraine led to a spike in hacktivist activity and continues to motivate groups such as Russian-aligned NoName057(16) efforts to disrupt NATO-allied countries that support Ukraine's defense. Additionally, Ukraine relied heavily on hacktivist groups such as the IT Army of Ukraine for both offensive and defensive cyber operations in the early days of the war.

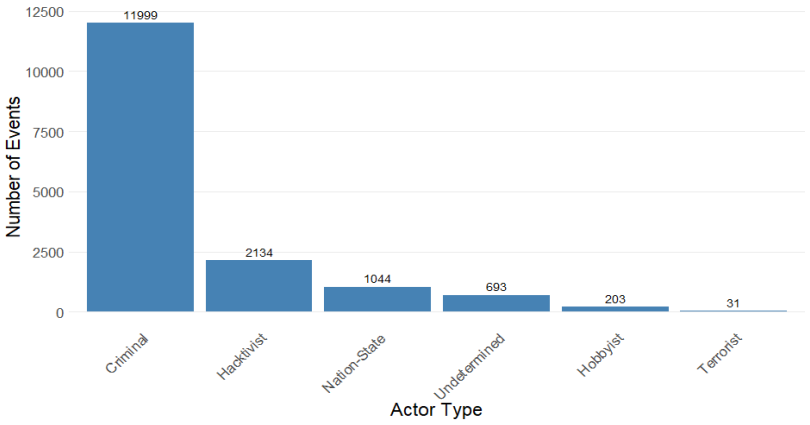

Figure 3: Cyber Events by Actor Type (2014-2025)

Figure 4 spotlights that cyber events are most frequently concentrated in public-facing and data-intensive sectors. Public administration and health care account for the largest shares, reflecting both their high operational dependence on IT systems and relatively strong disclosure requirements. Information, finance, and education also appear prominently, suggesting that sectors managing sensitive data or providing essential services face sustained exposure to cyber activity, while more asset-heavy sectors such as manufacturing and transportation experience comparatively fewer reported events.

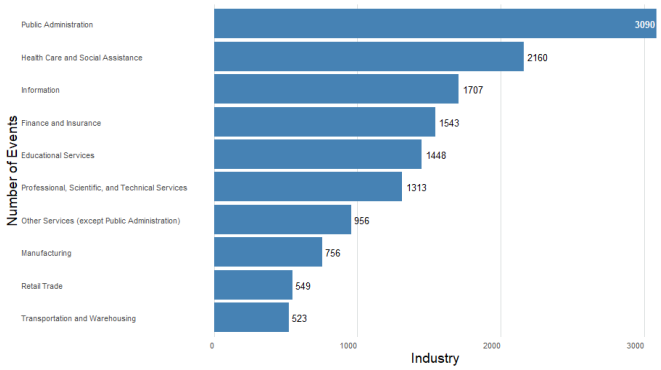

Figure 4: Top 10 Most Targeted Industries (2014-2025)

Figure 5 shows a strong concentration of recorded cyber events in a small number of countries, with the United States accounting for a 47.6% of total events. While reporting volume, disclosure regimes, and media coverage clearly play a significant role in shaping the geographic distribution of events in the dataset, the United States is also a very attractive target considering its robust industrial activity, economy, and geopolitical significance. The remaining nations are all relatively large G20 economies, and Ukraine is included due to its status as a war zone.

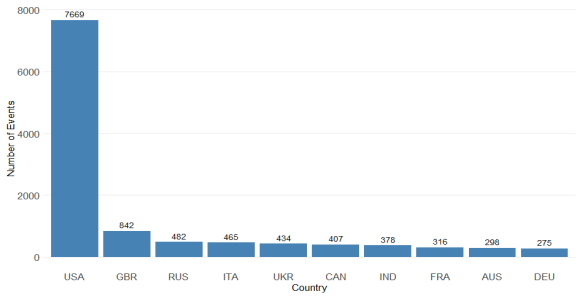

Figure 5: Top 10 Most Targeted Countries (2014-2025)

Finally, the threat actor distribution suggests that *observed* cyber activity is highly concentrated among a small number of persistent and operationally active groups, including criminal, hacktivist, and nation-state actors. cl0p has been the most active criminal group, reflecting both their high operational tempo and relatively robust resourcing, as well as the fact that their activities are more likely to be publicly reported and attributed. The appearance of advanced persistent threat (APT) groups among the top ten actors, such as North Korea's Third Surveillance Bureau and Russia's GRU Main Special Center for Special Technologies (Unit 74455), highlights the sustained visibility of certain state-linked campaigns within the dataset. For instance, the presence of Russia's GRU Unit 74455 reflects its association with highly visible, disruptive campaigns, most notably NotPetya, one of the most damaging cyber operations on record, which generated widespread collateral effects (U.S. Department of Justice). By contrast, the Third Surveillance Bureau-linked activity associated with North Korea blends traditional espionage with overt financial operations such as cryptocurrency theft and revenue-generating intrusions, which are more likely to surface in open-source reporting.

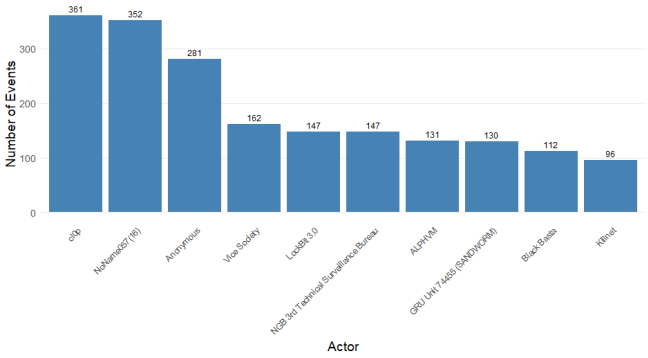

Figure 6: Top 10 Threat Actors (2014-2025)

5. Discussion

The CEDB 2.0 significantly enhances the analytical landscape for cybersecurity research by enabling scholars to address previously challenging questions due to data limitations. By providing structured data on actor types (e.g., nation-state, criminal, hacktivist), motives (e.g., espionage, sabotage, financial gain), and geopolitical affiliations through binary indicators for international organizations like NATO, or the EU, researchers can empirically test hypotheses about the nature, origins, and escalation of threats. Categorizing and contextualizing cyber events positions CEDB 2.0 as a valuable resource for industry and country-level threat landscapes, longitudinal trend analysis, effect and severity analysis, among other uses, all in service of knowledge generation and strategic decision-making.

Moreover, the CEDB 2.0 has also proven to be a vital resource for generating analytical insights and supporting evidence-based cybersecurity policymaking. In a comprehensive study, World Bank researchers integrated the CEDB with advanced artificial intelligence tools and then analyzed millions of cybersecurity-related articles. They

created a robust dataset covering approximately 190 countries and 21 industries, documenting over 30,000 publicly disclosed cyber incidents from 2014 to 2023. This enabled the identification of critical trends, such as a 21% average annual growth rate in global cyber incidents, with upper-middle-income countries experiencing a 37% surge. By linking incident frequency to cybersecurity commitments, the database provided policymakers with actionable insights. These findings highlighted the need for tailored policies, such as enhanced cybersecurity investments and awareness programs, particularly in developing nations, to mitigate cyber risks and foster sustainable digital development (Vergara Cobos & Cakir, 2024).

Another example of its utility is the European Securities and Markets Authority (ESMA) inclusion of CEDB 2.0's data in their Semi-annual Trends, Risk and Vulnerability (TRV) report (ESMA, 2024), incorporating the data into the TRV Statistical Annex (Chart A.131) to illustrate the rising global trend of cyber incidents in finance. The database's standardized information on threat actors, motives, targets, and impacts enabled ESMA to rank Infrastructure and Services at an elevated risk level, projecting ongoing cyber and operational vulnerabilities. The database's insights helped contextualize specific incidents, such as the CrowdStrike software update outage and Russia's cyberattacks against Ukraine, impacting EU entities. (ESMA, 2024).

The CEDB has been widely adopted in both analytical and policy initiatives, demonstrating its value in academia and beyond. It supports projects like CyberIR@MIT and Bamboo Weekly (2024) in compiling and analyzing cyber event data. It informs financial risk assessments by central banks such as the Bank of Japan (2023) and the European Central Bank (2022), among several other organizations.

Future iterations of the dataset will focus on achieving greater fidelity of event-level information. Planned enhancements include the integration of six-digit NAICS industry classifications, expanded organizational attributes, and the systematic backfilling of severity indicators across historical records. Increasing the granularity of these variables will enable more precise sectoral analysis and facilitate deeper empirical research on how cyber operations vary across industries and organizational types. Future development will also explore mechanisms for identifying and tracking cyber events involving artificial intelligence–enabled tools and techniques as these capabilities change the cyber threat landscape.

5.1 Limitations

Like most open-source cyber incident datasets, the Cyber Events Database relies on publicly available reporting, including news coverage, regulatory disclosures, and statements from affected organizations. As a result, events that are undisclosed, classified, intentionally concealed, or that occur in environments with limited media freedom are likely underrepresented. Coverage also varies across regions and sectors due to differences in legal reporting requirements, journalistic capacity, and organizational transparency, resulting in uneven geographic and industry representation. In addition, while the database increasingly incorporates multilingual sources, challenges remain in processing non-English reporting at scale and ensuring real-time accuracy as details about cyber events often evolve after initial disclosure.

6. Conclusion

The CEDB 2.0 provides a publicly accessible repository of structured, event-level data on global cyber incidents, enabling researchers, policymakers, and practitioners to analyze trends, assess risk, and support evidence-based decision-making. The integration of multilingual, near–real-time news monitoring through GDELT in 2025 expanded the database's geographic coverage, enabled timelier updates, and strengthened its utility for cross-sectional and longitudinal analysis across threat actors, motives, industries, and regions. With more than 16,382 recorded events spanning 2014 through 2025, CEDB 2.0 facilitates systematic case selection and comparative analysis, including fine-grained filtering of incident types, motivations, and actor characteristics.

Looking ahead, the CEDB 2.0 will support a comprehensive empirical analysis of cyber activity associated with actors linked to the Russian Federation. Leveraging the database's event-level coding of actor type, motive, sector, and observable effects, this research will build on previous research to examine patterns in targeting behavior, operational methods, and outcomes across time and geographic contexts in relation to the ongoing war in Ukraine (Lilly & Cheravitch, 2020) (Microsoft, 2022). The dataset's structure enables systematic comparison of state-attributed espionage and disruptive operations with hacktivist and proxy-based campaigns, including ransomware and data theft, allowing assessment of convergence, divergence, and potential coordination.

The CEDB 2.0 supports empirical research on cyber conflict, improves situational awareness for organizations assessing sector-specific risks, and informs policy discussions on infrastructure resilience, deterrence, and

international cooperation. In doing so, the dataset helps shift discussions of cyber operations from anecdotal observations toward transparent, evidence-based analysis.

Ethics Statement: This research did not require ethical clearance since we rely exclusively on publicly available media reporting.

AI Declaration: AI tools were used to assist with language editing, translation of non-English source material, and clarity checks. All substantive analysis, interpretation, and final content decisions were made by the authors.

References

Abubakara, A. I., H. Chiromab, S. A. Muazc, & L. B. Ilad. (2015). A Review of the Advances in Cyber Security Benchmark Datasets for Evaluating Data-Driven based Intrusion Detection Systems. The 2015 International Conference on Soft Computing and Software Engineering (SCSE 2015). 2015.

Alder, S. (2025). Healthcare Data Breach Statistics. HIPAA Journal. Oct 26, 2025. https://www.hipaajournal.com/healthcare-data-breach-statistics/

Al-zubidi, A. F., Farhan, A. K., & El-Kenawy, E. M. (2024). Surveying Machine Learning in Cyberattack Datasets: A Comprehensive Analysis. Journal of Soft Computing and Computer Applications. Vol. 1. Iss. 1, Article 1000. DOI: https://doi.org/10.70403/3008-1084.1000

Azevedo, C. (2023). Gone offline: how Cloudflare Radar detects Internet outages. The Cloudfare Blog. September 26, 2023. https://blog.cloudflare.com/detecting-internet-outages/

Bamboo Weekly. (2024). *79 cyber attacks: The problem and the solution*. Bamboo Weekly. https://www.bambooweekly.com/bw-79-cyber-attacks-solution/

Bank of Japan. (2023). *Financial System Report* (April 2023). Bank of Japan. https://www.boj.or.jp/research/brp/fsr/fsr230421.htm

Bertl, M. (2019). News analysis for the detection of cyber security issues in digital healthcare A text mining approach to uncover actors, attack methods and technologies for cyber defense. Young Information Scientist 4 (2019), 1–15.

Buchanan, B. (2020). *The hacker and the state: Cyber attacks and the new normal of geopolitics*. Cambridge, MA: Harvard University Press.

Cable, J., I. W. Gray, & McCoy, D. (2024). Showing the Receipts: Understanding the Modern Ransomware Ecosystem. 2024 APWG Symposium on Electronic Crime Research. August 17, 2024. https://arxiv.org/pdf/2408.15420

CFR. (n.d.). Cyber Operations Tracker. Council on Foreign Relations. n.d. https://www.cfr.org/cyber-operations/

Cloudflare Radar. (n.d.). Cloudflare Radar website. https://radar.cloudflare.com/

CSIS. (n.d.). Significant Cyber Incidents Since 2006. Center for Strategic & International Studies. N.d, https://csis-website-prod.s3.amazonaws.com/s3fs-public/2025-06/250610_Significant_Cyber_Incidents.pdf?VersionId=IAAkHurCCF.s7dd26zpWQUXbumz3JXsq

Easterly, J. (2023). The Attack on Colonial Pipeline: What We've Learned & What We've Done Over the Past Two Years. CISA. May 7, 2023. https://www.cisa.gov/news-events/news/attack-colonial-pipeline-what-weve-learned-what-weve-done-over-past-two-years

Ekstedt, M., Johnson, P., Lagerström, R., Sommestad, T., Ullberg, J., & Buschle, M. (2015). SecuriCAD by Foreseeti: A CAD tool for enterprise cyber security management. In 2015, IEEE 19th International Enterprise Distributed Object Computing Workshop (pp. 152–155). IEEE.

ESMA. (2025). TRV Statistical Annex, ESMA Report on Trends, Risks and Vulnerabilities No.2, 202. European Securities and Market Authority. 50-524821-3445. 29 August 2024. https://www.esma.europa.eu/sites/default/files/2024-08/ESMA50-524821-3445_TRV_2_2024_Statistical_annex.pdfVergara Cobos, E. (2024). Cybersecurity Economics for Emerging Markets. © World Bank. http://hdl.handle.net/10986/42130

EuRepoC. (n.d.). Cyber Incident Dashboard. The European Repository of Cyber Incidents. N.d. https://eurepoc.eu/dashboard/

European Central Bank. (2022). *Cyber risk and financial stability* (Special feature, Financial Stability Review, November 2022). European Central Bank. https://www.ecb.europa.eu/press/financial-stability-publications/fsr/special/html/ecb.fsrart202211_03~9a8452e67a.en.html

GDELT Project Blog. (2020, June 27). IEEE DSC: Longitudinal Analysis of Cyber-Related Articles. https://blog.gdeltproject.org/ieee-dsc-longitudinal-analysis-of-cyber-related-articles/

GDELT Project (2021). *Announcing the new Web News NGrams 3.0 Dataset*, GDELT Project Blog, 15 December. https://blog.gdeltproject.org/announcing-the-new-web-news-ngrams-3-0-dataset/

GDELT Project. (n.d.). Data: Querying, Analyzing and Downloading. https://www.gdeltproject.org/data.html

GDELT Project. (n.d.). The GDELT Project. https://www.gdeltproject.org/

Goldschmidt, P. & D. Chudá. (2024). Network Intrusion Datasets: A Survey, Limitations, and Recommendations. Computers & Security. Volume 156. September 2025.

Götz, A. (2023). "The fair principles: Trusting in fair data repositories", Open Access Government, July 2023, pp.262–263. https://www.openaccessgovernment.org/article/the-fair-principles-trusting-in-fair-data-repositories/162752/.

Harry, C. & N. Gallagher. (2018). Classifying Cyber Events: A Proposed Taxonomy. Journal of Information Warfare. Vol. 17, No. 3 (Summer 2018), pp. 17–31.

Harry, C., N. Gallagher, W. Lucyshyn, & D. Entrikin. (2025). Cyber Events Database Codebook. Center For Governance of Technology and Systems. August 2025. https://gotech.umd.edu/sites/default/files/2025-08/Cyber%20Events%20Database%20Codebook_August%202025.pdf

HHS. (2023). Submitting Notice of a Breach to the Secretary. Department of Health and Human Services, Office of Civil Rights. February 27, 2023. https://www.hhs.gov/hipaa/for-professionals/breach-notification/breach-reporting/index.html

Holm, H., Buschle, M., Lagerström, R., & Ekstedt, M. (2014). Automatic data collection for enterprise architecture models. Software & Systems Modeling, 13, 825–841. https://link.springer.com/article/10.1007/s10270-012-0252-2

Hong, D., Z. Fu, X. Zhang & Y. Pan. (2025). Research on the Development and Application of the GDELT Event Database. Data, 10(10), 158. https://www.mdpi.com/2306-5729/10/10/158

Information is Beautiful. (n.d.). Information is Beautiful website. https://informationisbeautiful.net/visualizations/worlds-biggest-data-breaches-hacks/

JD Supra (2023). *Carvin Software data breach affects 187,360 consumers*. https://www.jdsupra.com/legalnews/carvin-software-data-breach-affects-187-8824546/

Kennedy, P. (2024). *Outwood Academy Acklam cyber attack forces school closure*. Gazette Live, 15 November. https://www.gazettelive.co.uk/news/teesside-news/outwood-academy-acklam-cyber-attack-31606323

Kijewski, P. (2021). Securing your network using Shadowserver reports. Asia Pacific Network Information Centre. June 10. 2021. https://blog.apnic.net/2021/06/10/securing-your-network-using-shadowserver-reports/

Kostyuk, N., & Zhukov, Y. M. (2019). Invisible digital front: Can cyber attacks shape battlefield events? *Journal of Conflict Resolution*, 63(2), 317–347.

Lilly, B., & Cheravitch, J. (2020). The past, present, and future of Russia's cyber strategy and forces. In T. Jančárková, L. Lindström, M. Signoretti, I. Tolga, & G. Visky (Eds.), *2020 12th International Conference on Cyber Conflict (CyCon): 20/20 Vision – The Next Decade*. Tallinn: NATO CCDCOE.

Lin, D., Crabtree, J., Dillo, I. et al. (2020). The TRUST Principles for digital repositories. Sci Data 7, 144 (2020). https://doi.org/10.1038/s41597-020-0486-7

Maness, R. C. (n.d.). The Dyadic Cyber Incident and Campaign Data (DCID), Versions 1, 1.1, 1.5, and 2.0. n.d. https://www.ryanmaness.com/cyber-conflict-dataset

Microsoft. (2022). *Defending Ukraine: Early Lessons from the Cyber War*. Redmond, WA: Microsoft Digital Security Unit.

Moustafa, N. & J. Slay. (2015). UNSW-NB15: A comprehensive data set for network intrusion detection systems (UNSW-NB15 network data set). 2015 Military Communications and Information Systems Conference (MilCIS), Canberra, ACT, Australia, 2015, pp. 1-6.

Mvula, P. K., P. Branco, G. Jourdan, & H. L. Viktor. (2023). A systematic literature review of cyber-security data repositories and performance assessment metrics for semi-supervised learning. Discover Data. April 6, 2023. https://pmc.ncbi.nlm.nih.gov/articles/PMC10079755/pdf/44248_2023_Article_3.pdf

ODNI. (2024). Recent Cyber Attacks on US Infrastructure Underscore Vulnerability of Critical US Systems, November 2023–April 2024. Office of the Director of National Intelligence. June 2014. https://www.dni.gov/files/CTIIC/documents/products/Recent_Cyber_Attacks_on_US_Infrastructure_Underscore_Vulnerability_of_Critical_US_Systems-June2024.pdf.

Poireault, K. (2024). #CyberMonth: Inside Ransomware.live, Free Ransomware Intelligence for Everyone. InfoSecurity Magazine. 3 October 2024. https://www.infosecurity-magazine.com/interviews/ransomwarelive-intelligence/

Ransomeware.live. (n.d.) Ransomeware Live website. https://www.ransomware.live/

Ransomewhere. (n.d.). Ransomewhere website. https://ransomwhe.re/

Rogers, S. (2020). *Breach of city contractor e-mail put some city workers' info at risk*. WTVQ, 9 October. https://www.wtvq.com/breach-of-city-contractor-e-mail-put-some-city-workers-info-at-risk

Rosay, A., E. Cheval, F. Carlier, & P. Leroux. (2017). Network Intrusion Detection: A Comprehensive Analysis of CIC-IDS2017. In Proceedings of the 8th International Conference on Information Systems Security and Privacy (ICISSP 2022), pages 25-36. DOI: 10.5220/0010774000003120

Salem, A.H., Azzam, S.M., Emam, O.E., & A. A. Abohany. (2024). Advancing cybersecurity: a comprehensive review of AI-driven detection techniques. Journal of Big Data **11**, 105 (2024).

Shadowserver. (n.d.). Shadowserver website. https://www.shadowserver.org/what-we-do/

Symantec and Carbon Black Threat Hunter Team (2026) *Ransomware: Tactical Evolution Fuels Extortion Epidemic.* Symantec and Carbon Black. https://www.security.com/threat-intelligence/ransomware-extortion-epidemic

University of Maryland Center for Governance of Technology and Systems. (n.d.). Cyber Events Database. https://gotech.umd.edu/cyber-events-database

University of Maryland School of Public Policy. (2025, September 10). Cyber Events Database Enhanced by GDELT's Global News Monitoring. https://spp.umd.edu/news/cyber-events-database-enhanced-gdelts-global-news-monitoring

U.S. Department of Justice (2020). *Six Russian GRU officers charged in connection with worldwide deployment of destructive malware and other disruptive actions in cyberspace*. Press release, 19 October. https://www.justice.gov/opa/pr/six-russian-gru-officers-charged-connection-worldwide-deployment-destructive

Vergara Cobos, E. & Cakir, S. (2024). A Review of the Economic Costs of Cyber Incidents. Washington, DC: World Bank. 2024. https://documents1.worldbank.org/curated/en/099092324164536687/pdf/P17876919ffee4079180e81701969ad0a18.pdf

VERIS. (n.d.). The Vocabulary for Event Recording and Incident Sharing. The Vocabulary for Event Recording and Incident Sharing (VERIS). n.d. https://verisframework.org/index.html

Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., et al. (2016). The FAIR Guiding Principles for scientific data management and stewardship. Scientific Data, 3(1), 160018. https://doi.org/10.1038/sdata.2016.18

Winter, C. (2024). The TRUST Principles for Digital Repositories, HSSCommons. https://ospolicyobservatory.uvic.ca/the-trust-principles-for-digital-repositories/