START

Significant Multi-Domain Incidents against Critical Infrastructure (SMICI): Codebook v1.3.1

August 2023

UWT

UNCONVENTIONAL **WEAPONS AND TECHNOLOGY**

National Consortium for the Study of Terrorism and Responses to Terrorism
A Department of Homeland Security Emeritus Center of Excellence
Led by the University of Maryland

About START

The National Consortium for the Study of Terrorism and Responses to Terrorism (START) is a university-based research, education and training center comprised of an international network of scholars committed to the scientific study of terrorism, responses to terrorism and related phenomena. Led by the University of Maryland, START is a Department of Homeland Security Emeritus Center of Excellence that is supported by multiple federal agencies and departments. START uses state-of-the-art theories, methods and data from the social and behavioral sciences to improve understanding of the origins, dynamics and effects of terrorism; the effectiveness and impacts of counterterrorism and CVE; and other matters of global and national security. For more information, visit www.start.umd.edu or contact START at infostart@umd.edu.

Contents

Executive Summary

Cyber-physical attacks on critical infrastructure have the potential to damage the physical infrastructure assets and have widespread consequences for countless others. Although cyber-physical attacks on critical infrastructure have been identified as one of the major homeland security challenges for the foreseeable future, there has not been a dataset that aggregates publically available data on cyber-physical attacks against critical infrastructure globally across all of the critical infrastructure sectors. Lack of such a dataset has limited our ability to gain a deeper understanding of the cyber-physical attack phenomenon as well as our ability to hypothesize about the behaviors and motivations of the attackers. In an effort to gain a better understanding of the adversaries' multi-domain behavior and motivations, the Unconventional Weapons & Technology Division (UWT) of the National Consortium for the Study of Terrorism and Responses to Terrorism (START) has developed the Significant Multi-domain Incidents against Critical Infrastructure (SMICI) dataset, a first of its kind, using only publicly available information.

Data and Methodology

The dataset collects on 38 individual variables and currently contains 524 cyber-physical and cyber-operational incidents carried out against critical infrastructure worldwide from January 1, 1992 to the present. We intend to continue this collection effort, improve data granularity, and expand the dataset temporally in the future.

Inclusion Criteria:

For an incident to be considered for inclusion in the SMICI dataset, the incident must meet the following base inclusion criteria:

1) The attack must have originated in the cyber domain.
2) The attack must target a critical infrastructure sector as defined by Presidential Policy Directive 21 (PPD-21), dated February 12, 2013.
3) The attack must be a disruptive cyber-physical incident OR disruptive cyber-operational incident.

Collection of data is by publicly available sources only. Credible social media sources, news reporting, and industry briefs/ reports are the primary sources used in the data collection.

Variables

Note

The dataset is designed under an incident-target per observation format. This means every line is one incident for a target. The observations run from 1992 to the present.

Below are the variables to code for the SMICI dataset and the coding rule for each variable.

Actor_Variables

Attribution

Variable Type: Dichotomous

Description: Is it confirmed publicly who/what threat actor is responsible for an incident?

Coding Rules:

- Cannot code unknown (-99) for this variable.

Attribution	Code
Yes	1
No	2

State_Nonstate

Variable Type: Dichotomous

Description: Is the threat actor a state or non-state actor?

Coding Rules:

- If the threat actor is a state-proxy or hacker-for-hire working for a state, then code as "State".
- Suspected threat actor is coded -99.
- If unknown, encode "-99."

Type	Code
State	1
Non-state	2
Unknown/Suspected	-99

Threat_Actor

Variable Type: Free Text

Description: Who is the threat actor?

Coding Rules:

- If the incident is attributed to a state or is attributed to state affiliated actor (APT29), input the state as the threat actor. All other names for the threat actor go in *Threat_Actor_Notes*.
- If the threat actor is a threat group (e.g. Maze Team, FIN6) that is not aligned with a state or it is not clear they have an affiliation with a state, then input the primary name of the threat actor referenced

across all sources collected for the incident. All other names for the threat actor go in *Threat_Actor_Notes*.

- If threat actor is unknown, input "unknown," not "-99".

Examples:

- **SMICI093:** *Gaza Cybergang*
- **SMICI0148:** *Sven Jaschan*
- **SMICI0217:** Iran

Motive

Variable Type: Categorical

Description: Suspected or confirmed motivation for the threat actor in the incident. If reported motive covers more than one category below or multiple motives have been reported in various reports, please code for the most salient motive here, but list all of them in the *Threat_Actor_Notes.*

- **Financial Gain** – Threat actor(s) primary or sole motivation in executing an attack is for financial gain. Most incidents are the result of the target being an 'easy mark' due to lax security standards. In other cases, the target was third party software/hardware/networking vulnerabilities and once exploited, netted victims exposed to the third party vulnerability.
- **Destruction** – Unlike cybercrime, threats actors pursuing destruction are not holding a victim's data hostage for ransom nor exfiltrating data to sell to elsewhere. They want to cause harm. Disrupt, degrade, and destroy.
- **Espionage** – Whereas destruction is loud, espionage is all about being quiet and exfiltrating as much data as possible.
- **Hacktivism** – Those motivated by hacktivism seek to influence public opinion, make a political statement, have fun, or simply 'do it for the lulz.' These threat actors are the least sophisticated and capable to carry out successful attacks against critical infrastructure beyond web defacements or simple distributed-denial-of-service (DDoS) attacks.
- **Proof-of-Concept** – Incidents coded for this motivation are not 'attacks' in a general sense. Often times this incidents revolves around academic experiments that 'got loose', a grey hat hacker disrupting a target to demonstrate a security vulnerability, or security firms conducting a penetration test on a live target (with target permission).

Coding Rules:

- If motive is unknown, encode "-99."

Motive	Code
Financial Gain	1
Destruction	2
Espionage	3
Hacktivism	4
Proof-of-Concept	5
Unknown	-99

Threat_Actor_Notes

Variable Type: Free Text

Description: In addition to any additional, not-yet-captured, coding from above, please report anything of note about the threat actor involved.

Coding Rules: One paragraph or less depending on available information.

Examples:

- **SMICI0135:** *19-year-old British teenager Aaron Caffrey, 19, was arrested and charged with the DDoS attack against the Port of Houston, however, he was acquitted of all charges in October 2003. Evidence showed that the attack was executed from Aaron's computer, had malicious program signed as "Aaron" and the attack was traced to Aaron's house in the United Kingdom. Additionally, Aaron was a leading member of the hacker group Allied Haxor Elite.*
- **SMICI0141:** *18-year-old Jeffrey Lee Parson of Minnesota author the Blaster worm and executed its propagation on August 13, 2003. He was arrested in September 2003. He was sentenced in January 2005 to 8 months in prison, 3 years of supervised release and 100 hours of community service.*
- **SMICI0164:** *Reporting suggests this to be a non-state cybercriminal. Data encrypted was not accessed nor exfiltrated, possibly meaning actor was not capable of exfiltration or was testing malware.*

Target_Country_Variables

Rivalry

Variable Type: Categorical

Description: Is the target country a rival of the attacking country in the year that the attack occurred?

Coding Rules: Select the numeral that corresponds to the answer for the target and attacker. Coders should conduct brief outside research to determine whether the two state actors were rivals in the year of the attack.

Rivalry	Code
No	0
Yes	1
Attacker is a Non-state Actor	2
Unknown	-99

International Conflict

Variable Type: Dichotomous

Description: Is the target country engaged in an international conflict (such as a war; foreign military deployment; etc.) in the year that the attack occurred?

Coding Rules: Select the numeral that corresponds to the answer for the target. For reference, please see the Uppsala Conflict Data Program UCDP/PRIO Armed Conflict Dataset version 23.1.

International Conflict	Code
No	0
Yes	1
Unknown	-99

Conflict with Attacker

Variable Type: Dichotomous

Description: Is the target country engaged in a conflict against the attacker in the year that the attack occurred?

Coding Rules: Select the numeral that corresponds to the answer for the target. For reference, please see the Uppsala Conflict Data Program Dyadic Dataset version 23.1.

Conflict with Attacker	Code
No	0
Yes	1
Unknown	-99

Domestic Conflict

Variable Type: Dichotomous

Description: Is the target country involved in violent domestic conflict (such as violent protests; civil war; etc.) in the year that the attack occurred?

Coding Rules: Select the numeral that corresponds to the answer for the target. For reference, please see the Uppsala Conflict Data Program Georeferenced Event Dataset Global version 23.1, the UCDP Violent Political Protest Dataset version 20.1, the Armed Conflict Location & Event Data Project, and other outside research to support your coding.

Domestic Conflict	Code
No	0
Yes	1
Unknown	-99

POLITY of Target

Variable Type: Categorical

Description: POLITY 2 score of the **target** for the year observed.

Coding Rules: Use provided Polity IV dataset and enter the corresponding POLITY 2 score for the targeted State.

POLITY of Attacker

Variable Type: Categorical

Description: POLITY 2 score of the **attacker** for the year observed.

Coding Rules: Use Polity IV dataset and enter the corresponding POLITY 2 score for the attacking State.

- Note: For non-state actor, enter -666

GDP PPP of Target

Variable Type: String

Description: Gross Domestic Product (GDP) per Capita Purchasing Power Parity (PPP) for the **target** country for the year of the incident.

Coding Rules: Use provided World Bank data and enter the corresponding GDP per Capita PPP.

Quality of Overall Infrastructure

Variable Type: String

Description: Overall quality of infrastructure of the **target** country.

Coding Rules: Use provided World Economic Forum's Global Competitiveness Index data's Overall Quality of Infrastructure variable and enter the corresponding value for the country for the year of observation.

Online_Connectedness

Variable Type: String

Description: Number of Internet user as a percentage of the total population of the **target** country.

Coding Rules: Use provided World Bank data on "Individuals Using the Internet" and enter the corresponding Internet user percentage for the year of observation.

Adopted Information Security Standards

Variable Type: Dichotomous

Description: Has the **target** country adopted an information security standard (e.g. ISO, NIST, etc.)?

Coding Rules: Select the numeral that corresponds to the answer for the target.

Adopted Information Security Standards	Code
No	0

Yes	1
Unknown	-99

Specific_Target_Variables

Critical Infrastructure Sector

Variable Type: Categorical

Description: The critical infrastructure sector of the target.

Coding Rules: Select the numeral that corresponds to the sector of the target.

- If sector is unknown, encode "-99."

Sectors			
Sector	**Code**	**Sector**	**Code**
Chemical (C)	1	Food and Agriculture (FA)	10
Commercial Facilities (CF)	2	Government (G)	11
Communications (Co)	3	Healthcare and Public Health (HPH)	12
Critical Manufacturing (CM)	4	Information Technology (IT)	13
Dams (D)	5	Nuclear (N)	14
Defense Industrial Base (DIB)	6	Transportation (T)	15
Emergency Services (ES)	7	Water and Wastewater (W)	16
Energy (E)	8	Multiple	17
Financial Services (FS)	9	Unknown	-99

Critical Infrastructure Sector Multiple

Variable Type: Categorical

Description: If Critical Infrastructure Sector is coded as Multiple, please list all sectors here.

Coding Rules: Select all numerals that correspond to the sector of the target.

- If sector is not multiple (17), code this variable "-1"
- If multiple (17) sectors and sectors are unknown, code "-99"

Critical Infrastructure Sub-Sector

Variable Type: Categorical

Description: The subsector of the target.

Coding Rules: Select the numeral that corresponds to the subsector of the target. Please see the **Critical Infrastructure Sub-Sector List** document to find a list and examples of subsectors.

- If sector is multiple (17), then encode to the corresponding subsector multiple code or unknown "-99" depending available information.
- If subsector is unknown, encode "-99."

Critical Infrastructure Sub-Sector Multiple

Variable Type: Categorical

Description: If Critical Infrastructure Sector is coded as Multiple, please list all sectors here.

Coding Rules: Select all numerals that correspond to the subsectors of the target. Please see the **Critical Infrastructure Sub-Sector List** document to find a list and examples of subsectors.

- If sub-sector is not multiple (17), code this variable "-1"
- If multiple (17) sub-sectors and sub-sectors are unknown, code "-99"

Specific_Target

Variable Type: Free Text

Description: What is the target? Please include as much information as is available.

Coding Rules:

- If the target uses a pseudonym, which is standard practice for public reporting from security companies engaged in remediation with the target, then input the pseudonym, but also input it in the *Impact_Notes* column in the dataset for that incident.
- If the target is does not have a pseudonym but referenced by its CI/KR sector and/or its geographic location, then input that in *Specific_Target* column in the dataset for that incident. Furthermore, for these incidents, input "the target identity is not known" in *Impact_Notes*.
- If multiple, write in "multiple" and list all targets.
- If the target is completely unknown, write in "unknown," not "-99."
- If there are varying levels of confidence in targets, please make a note of confidence for each target, such as, suspected, verified, rumored, etc.

Example:

- **SMICI0042:**

Specific_Target	Impact_Notes
Kemuri	OT side compromised, threat actor able to manipulate. IT network relied on a single legacy IBM Application System/400 (AS/400) server, released back in 1988. Kemuri is a pseudonym for the company.

Geographic_Variables

Country/Area and above geocoding

The SMICI dataset utilizes the M49 geocoding standard developed and maintained by the United Nations Statistical Division (UNSD) for country/area level and above coding. There are two minor modifications with regard to intermediate regions and sub-regions. Please see the *Target_Intermediate_Region* and *Target_Sub-region* sections for more detail.

Along with Country-Area/Int. Region/Sub-region/Reg. names and their respective M4 codes, some countries/areas have a note describing what type of entity it is or its political-administrative relationship with another state. The notes are for further reference, not to be coded into the dataset.

Subnational/US specific geocoding

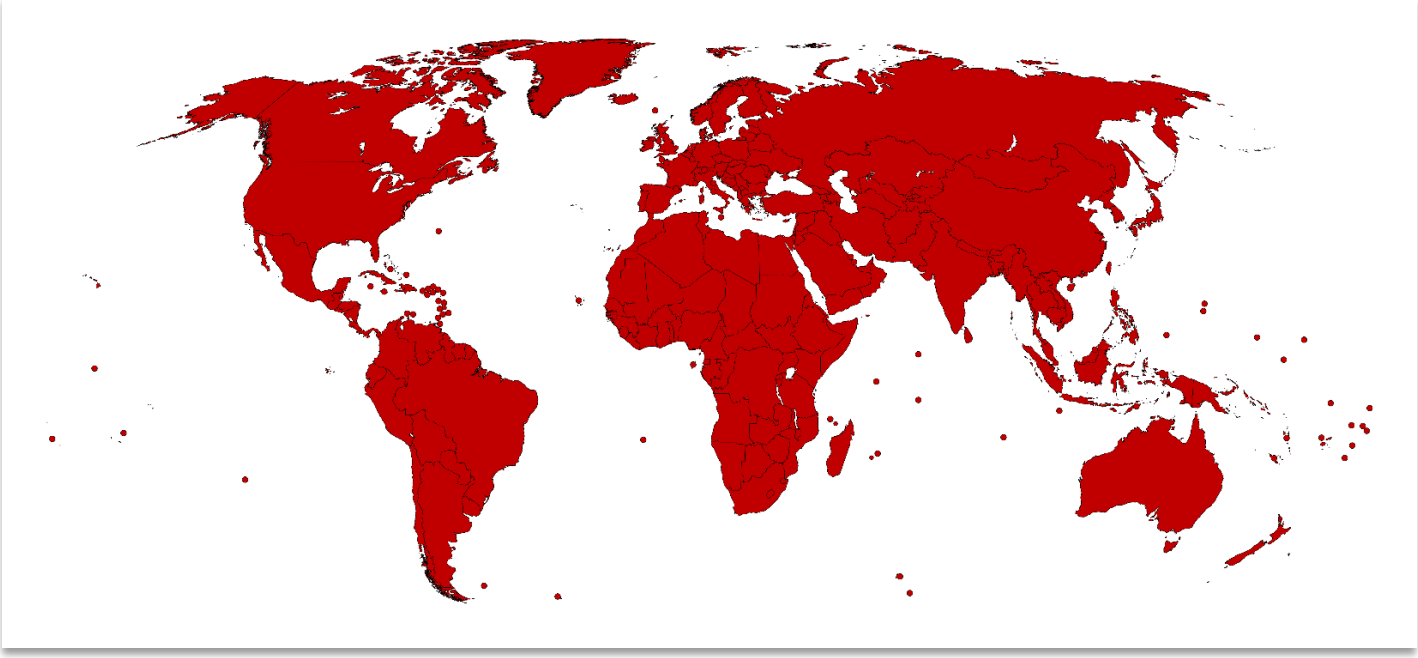
The *Target_State*, *Target_State_Division-region*, and *Target_State_Region* variables are specific to the United States. *Target_State* provides state-level granularity using the American National Standards Institute (ANSI) state-level geocoding standard. *Target_State_Sub-division* and *Target_State_Region* provide further context and use the United States Census Bureau (USCB) coding standard.

Along with State/Sub-division/Region names and respective ANSI or USCB codes, this section also includes the states' corresponding United States Postal Service (USPS) code. The USPS codes are for further reference, not to be coded into the dataset.

As the SMICI dataset is developed further, the subnational geocoding will expand to other countries. Currently, all countries (save for the US) are coded "0" for the subnational coding indicating "Not Applicable".

Note on incidents with multiple targets:

If you are coding for an incident that includes multiple targets, please code for what is determined to be the primary target. If it is unclear what the primary target was, please consult with Rhyner or Megan for clarification.



Target_Region

Variable Type: Numerical

Description: The region of the target.

Coding Rules: enter the Region geocode of the target country's region for this incident. Please see the **Target Country and State Identifiers** file for all region and country codes.

- If the incident impacted a target in multiple regions, encode for multiple as "975." List out the regions impacted in *Impact_Notes* column in the dataset for that incident.
- If target region is unknown, encode "-99."

WORLD	
Name	**Code**
Africa	002
Americas	019
Asia	142
Europe	150
Oceania	009

Target_Sub-Region

Variable Type: Numerical

Description: The sub-region of the target.

Coding Rules: enter the Sub-Region geocode of the target country's sub-region for this incident. Please see the **Target Country and State Identifiers** file for all region and country codes.

- If the incident impacted a target in multiple sub-regions, encode for multiple as "950." List out the sub-regions impacted in *Impact_Notes* column in the dataset for that incident.

- United Nations Statistical Division classifies Iran (364) part of "Southern Asia (034)" sub-region for statistical purposes only. The SMICI dataset codes Iran (364) sub-region as "Western Asia (145)".
- If the target sub-region is unknown, encode "-99."

AFRICA		AMERICAS	
Name	**Code**	**Name**	**Code**
Northern Africa	015	Latin America and the Caribbean	419
Sub-Saharan Africa	202	Northern America	021

ASIA		EUROPE	
Name	**Code**	**Name**	**Code**
Central Asia	143	Northern Europe	154
Eastern Asia	030	Eastern Europe	151
South-eastern Asia	035	Southern Europe	039
Southern Asia	034	Western Europe	155
Western Asia	145		

OCEANIA	
Name	**Code**
Australia and New Zealand	053
Melanesia	054
Micronesia	057
Polynesia	061

Target_Intermediate_Region

Variable Type: Numerical

Description: The intermediate region of the target.

Coding Rules: enter Intermediate-Region geocode of the target country's intermediate region (if applicable) for this incident. Please see the **Target Country and State Identifiers** file for all region and country codes.

- If the incident impacted a target in multiple intermediate regions, encode for multiple as "925." List out the intermediate regions impacted in *Impact_Notes* column in the dataset for that incident.
- A majority of countries are not part of an *Intermediate_Region* in the M49 standard and therefore, are coded "Not Applicable" in the M49 standard. The modification for SMICI is the numerical conversion of "Not Applicable" to "**0**" for this geographic variable.
- If the target's intermediate region is unknown, encode "-99."

AFRICA		AMERICAS	
Name	**Code**	**Name**	**Code**
Eastern Africa	014	Caribbean	029
Southern Africa	018	Central America	013
Middle Africa	017	South America	005
Western Africa	011		

ASIA		EUROPE	
Name	**Code**	**Name**	**Code**
Not Applicable	0	Channel Islands	830

OCEANIA

Name	Code
Not Applicable	0

Target_Country
Variable Type: Numerical

Description: Country/area of the target.

Coding Rules: enter the M49 Country/Area geocode for the target country for this incident. Please see the **Target Country and State Identifiers** file for all region and country codes.

- If the incident impacted a target in multiple countries/areas, encode for multiple as "900." List out the countries impacted in *Impact_Notes* column in the dataset for that incident.
- If the target country/area is unknown, encode "-99."

Target_State_Region (U.S. Only)
Variable Type: Numerical

Description: A state's region. Only applicable to US states.

Coding Rules: enter the geocode for the region of the target state. Please see the **Target Country and State Identifiers** file for all region and state codes.

- If the incident impacted a target in multiple subnational regions, encode for multiple as "975." List out the subnational regions impacted in *Impact_Notes* column in the dataset for that incident.
- If the subnational region is unknown, encode "-99."

REGIONS	
Name	**Code**
Northeast	1
Midwest	2
South	3
West	4

Target_State_Division (U.S. Only)
Variable Type: Numerical

Description: A state's regional division. Only applicable to US states.

Coding Rules: enter the USCB geocode for the regional division of the target state. Please see the **Target Country and State Identifiers** file for all region and state codes.

- If the incident impacted a target in multiple subnational regional divisions, encode for multiple as "950." List out the subnational regional divisions impacted in *Impact_Notes* column in the dataset for that incident.
- If the subnational regional division is unknown, encode "-99."

NORTHEAST(1)		MIDWEST(2)	
Name	**Code**	**Name**	**Code**
New England	1.1	East North Central	2.3
Mid-Atlantic	1.2	West North Central	2.4

SOUTH(3)		WEST(4)	
Name	**Code**	**Name**	**Code**
South Atlantic	3.5	Mountain	4.8
East South Central	3.6	Pacific	4.9
West South Central	3.7		

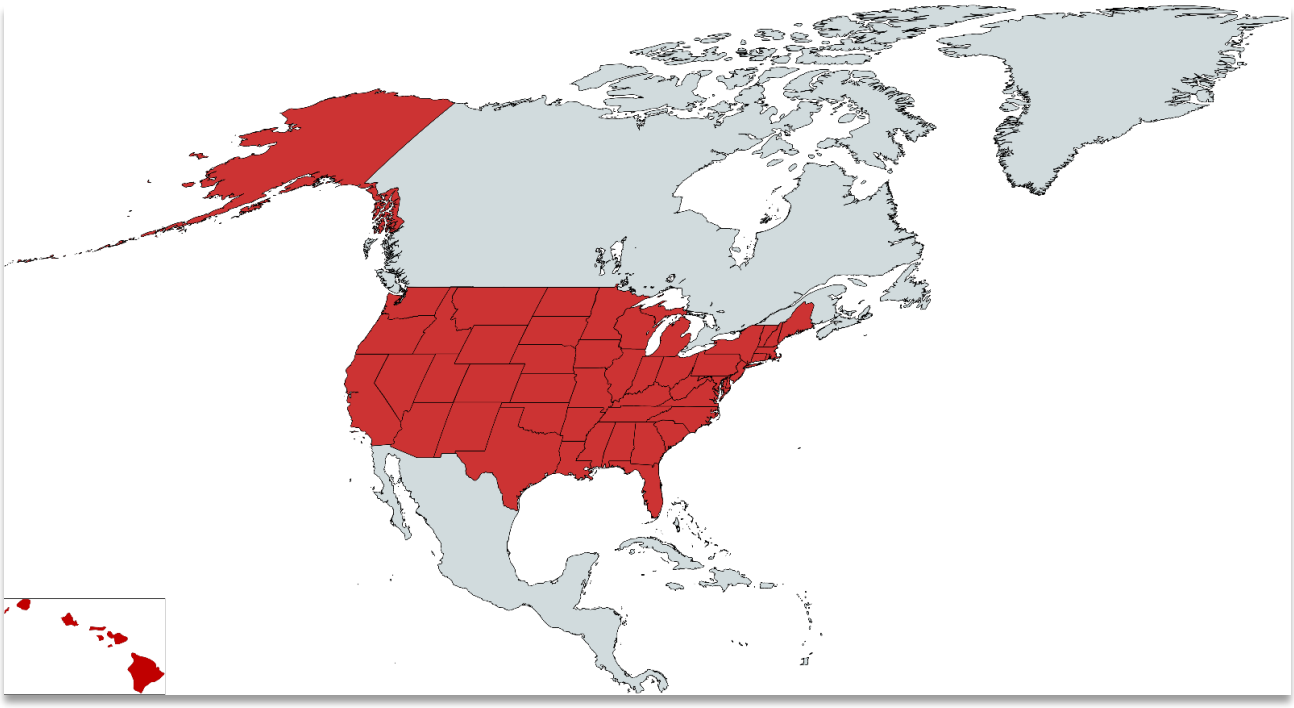
Target_State (U.S. Only)

Variable Type: Numerical

Description: U.S. state of the target. Only applicable to US states.

Coding Rules: enter the ANSI geocode for the target state for the incident. Please see the **Target Country and State Identifiers** file for all region and state codes.

- If the incident impacted a target in multiple states, encode for multiple as "925." List out the states impacted in *Impact_Notes* column in the dataset for that incident.
- 003, 007, 014, 043 do not exist as state codes in the ANSI standard.
- If the target state is unknown, encode "-99."



Malware/Technical_Variables

Malware_Type

Variable Type: Categorical

Description: Malware type used in incident

Coding Rules:

- **Botnet:** Include Permeant Denial-of-Service (PDoS) and specific Distributed Denial–of-Service (DDoS) attacks. The DDoS attack must disrupt functionality and/or services that impact end user/clients/customers immediately.
- **Non-malware:** Select if threat actor gains access and causes cyber-physical or cyber-operational disruption without malware. This will mostly be with insider threat incidents.
- **Other:** Describe the malware type(s) involved in *Malware_Notes* column in the dataset for that incident.
- **Multiple:** If multiple (11), input the types in *Malware_Notes* column in the dataset for that incident.
- If malware type is unknown, encode "-99."

Type	Code		Type	Code
Botnet	1		Trojan	7
Cryptominer	2		Wiper	8
ICS Malware	3		Worm	9
Logic Bomb	4		Other	10
Non-malware	5		Multiple	11
Ransomware	6		Unknown	-99

Malware_Name/Family

Variable Type: Free Text

Description: What is the name of the malware? This can be the family name or variant name.

Coding Rules:

- If identified, input the name from the primary source used for the incident. Any other names for the malware go in *Threat_Actor_Notes* column in the dataset for that incident.
- Hyperlink cell to a credible source with detailed information.
- If more than one type or variant of malware is involved/identified, then input the primary malware referenced across all sources collected for the incident. All other malware named in the incident go in *Malware_Notes* and describe that malware type(s) if possible.
- If malware name or family is unknown, encode "-99."

Malware_Notes

Variable Type: Free Text

Description: Provide information about the malware type(s) involved. Information from CISA, MITRE, or credible industry firms (e.g. Crowdstrike, Dragos, ESET, Mandiant, Kaspersky, etc.) are strongly recommended for use in this section.

Coding Rules:

- List all other names for the primary malware here.
- If other malware is identified in the incident, list their associated names as well.

Temporal_Variables

Execution Dates

Incident_Execution_Year

Variable Type: Numerical

Description: The year the incident is **executed** by the threat actor.

Coding Rules: Select the year the incident is executed.

- If execution year is unknown, encode "-99."

1990	1991	1992	1993	1994	1995	1996	1997	1998	1999
2000	2001	2002	2003	2004	2005	2006	2007	2008	2009
2010	2011	2012	2013	2014	2015	2016	2017	2018	2019
2020	2021	2022	2023	(-99)					

Incident_Execution_Month

Variable Type: Numerical

Description: The month the incident is **executed** by the threat actor.

Coding Rules: Select the numeral that corresponds to the month the incident is executed.

- If execution month is unknown, encode "-99."

Month	Code		Month	Code
January	1		August	8
February	2		September	9
March	3		October	10
April	4		November	11
May	5		December	12
June	6		Unknown	-99
July	7			

Incident_Execution_Day

Variable Type: Numerical

Description: The day the incident is **executed** by the threat actor.

Coding Rules: Select the numeral that corresponds to the day the incident is executed.

- If execution day is unknown, encode "-99."

1, 2, 3, 4, 5, 6, 7,
8, 9, 10, 11, 12, 13, 14,
15, 16, 17, 18, 19, 20, 21,
22, 23, 24, 25, 26, 27, 28,
29, 30, 31, (-99)

Discovery Dates

Incident_Discovery_Year

Variable Type: Numerical

Description: The year the incident is **discovered** by the target, security firm/vendor, law enforcement, independent researcher(s), etc. Sometimes this is revealed when the incident is publicly disclosed and/or executed.

Coding Rules: Select the year the incident is discovered.

- If discovery year is unknown, encode "-99."

1990	1991	1992	1993	1994	1995	1996	1997	1998	1999
2000	2001	2002	2003	2004	2005	2006	2007	2008	2009
2010	2011	2012	2013	2014	2015	2016	2017	2018	2019
2020	2021	2022	2023	(-99)					

Incident_Discovery_Month

Variable Type: Numerical

Description: The month the incident is **discovered** by the target, security firm/vendor, law enforcement, independent researcher(s), etc. Sometimes this is revealed when the incident is publicly disclosed and/or executed.

Coding Rules: Select the numeral that corresponds to the month the incident is discovered.

- If discovery month is unknown, encode "-99."

Month	Code		Month	Code
January	1		August	8
February	2		September	9
March	3		October	10
April	4		November	11
May	5		December	12
June	6		Unknown	-99
July	7			

Incident_Discovery_Day

Variable Type: Numerical

Description: The day the incident is **discovered** by the target, security firm/vendor, law enforcement, independent researcher(s), etc. Sometimes this is revealed when the incident is publicly disclosed and/or executed.

Coding Rules: Select the numeral that corresponds to the day the incident is discovered.

- If discovery day is unknown, encode "-99."

1, 2, 3, 4, 5, 6, 7,
8, 9, 10, 11, 12, 13, 14,
15, 16, 17, 18, 19, 20, 21,
22, 23, 24, 25, 26, 27, 28,
29, 30, 31, (-99)

Disclosure Dates

Incident_Disclosure_Year

Variable Type: Numerical

Description: The year the incident is **publicly disclosed**. Sometimes this is revealed when the incident is executed or discovered.

Coding Rules: Select the year the incident is disclosed.

- If disclosure year is unknown, encode "-99."

1990	1991	1992	1993	1994	1995	1996	1997	1998	1999
2000	2001	2002	2003	2004	2005	2006	2007	2008	2009
2010	2011	2012	2013	2014	2015	2016	2017	2018	2019
2020	2021	2022	2023	(-99)					

Incident_Disclosure_Month

Variable Type: Numerical

Description: The month the incident is **publicly disclosed**. Sometimes this is revealed when the incident is executed or discovered.

Coding Rules: Select the numeral that corresponds to the month the incident is disclosed.

- If disclosure month is unknown, encode "-99."

Month	Code		Month	Code
January	1		August	8
February	2		September	9
March	3		October	10
April	4		November	11
May	5		December	12
June	6		Unknown	-99
July	7			

Incident_Disclosure_Day

Variable Type: Numerical

Description: The day the incident is **publicly disclosed**. Sometimes this is revealed when the incident is executed or discovered.

Coding Rules: Select the numeral that corresponds to the day the incident is disclosed.

- If disclosure day is unknown, encode "-99."

1, 2, 3, 4, 5, 6, 7,
8, 9, 10, 11, 12, 13, 14,
15, 16, 17, 18, 19, 20, 21,
22, 23, 24, 25, 26, 27, 28,
29, 30, 31, (-99)

Impact_Variables

Attack_Type

Variable Type: Dichotomous

Description: What is the attack type?

- **Cyber-Physical:** An incident where a threat actor – state or non-state executes malicious action(s) in the cyber domain that have a disruptive kinetic effect(s) in the physical domain.

  The threat actor causes disruption to operational technology (OT) by bridging the information technology (IT) and OT gap or directly attacking OT. In general, this type of incident occurs when a threat actor targeting critical infrastructure and key resources (CI/KR) compromises Industrial Control Systems (ICS).

- **Cyber-Operational:** An incident where a threat actor executes malicious actions through the cyber domain that have a disruptive kinetic effect in the physical domain.

  However, these incidents do not involve direct action(s) against OT. Rather, these attacks are designed to disrupt the IT systems that are connected to the ICS or Internet-of-Things (IoT) systems and devices in the OT environment.

Coding Rules:

- If attack type is unknown, encode "-99." However, incidents coded -99 for attack type will be reviewed on a case-by-case basis on whether to approve or reject final inclusion into the SMICI dataset.

Type	Code
Cyber Operational	1
Cyber-Physical	2
Unknown	-99

Financial_Impact

Variable Type: Categorical

Description: What is the financial cost of the incident? This can be a culmination of costs or not.

Examples:

- Loss in profit due to incident
- Remediation cost
- Ransom

Coding Rules: Select the monetary range for financial impact of the incident.

- If cost is unknown, encode "-99."

Value Range	Code		Value Range	Code
$1 - $4,999	1		$250,000,000 - $499,999,999	14
$5000 - $9,999	2		$500,000,000 - $999,999,999	15
$10,000 - $49,999	3		$1,000,000,000 - $4,999,999,999	16
$50,000 - $99,999	4		$5,000,000,000 - $9,999,999,999	17
$100,000 - $249,999	5		$10,000,000,000 - $24,999,999,999	18
$250,000 - $499,999	6		$25,000,000,000 - $49,999,999,999	19
$500,000 - $999,999	7		$50,000,000,000 - $99,999,999,999	20
$1,000,000 - $4,999,999	8		$100,000,000,000 - $249,999,999,999	21
$5,000,000 - $9,999,999	9		$250,000,000,000 - $499,999,999,999	22
$10,000,000 - $24,999,999	10		$500,000,000,000 - $999,999,999,999	23
$25,000,000 - $49,999,999	11		>$1,000,000,000,000	24
$50,000,000 - $99,999,999	12		Unknown	-99
$100,000,000 - $249,999,999	13			

Impact_Notes

Variable Type: Free Text

Description: Explain what happened in the incident. If there is an estimated or exact cost to the incident, include in this section, but select the correct cost range in the 'Financial Impact' section.

Coding Rules: One paragraph or less depending on available information.

Examples:

- **SMICI0111:** *Shipping halted. Down market businesses suffered lack of stock causing loss of business and therein revenue. Container World did not pay ransom, but eventually remediated the threat.*
- **SMICI0125:** *Disrupted plant operations in Brazil, Mexico, and the U.S. branch stock suffered when news of the attack went public.*
- **SMICI0151:** *Traffic engineer Kartik Patel and computer expert Gabriel Murillo compromised the L.A. traffic surveillance center system, disrupting traffic signals at major intersections in Los Angeles. Both men were a part of the labor union disrupt with the city and managed to gain access to the traffic system despite the center's efforts to deny union members access. Although no accidents were reported due to the incident, the disruption did cause major backups.*

Miscellaneous Variables and Sources

Lede

Variable Type: Free Text

Description: One sentence summation of the incident.

Coding Rules: Only one sentence with these components.

Required:

- What is the target?
- What type of malware used?

Optional:

- Who is the threat actor?
- What is the malware name/family?
- Unique highlight (e.g. shut down of plant operations, forced to turn patients away, had to use pen and paper, etc.)

Examples:

- **SMICI0181:** *Great Plains Health medical center hit by ransomware and forced to switch to pen and paper.*
- **SMICI0203:** *Travelex, foreign currency exchange firm, hit by REvil ransomware.*
- **SMICI0265:** *Mediterranean Shipping Company (MSC) hit by a cyberattack that disrupts the services of the company across the world.*

Analyst_Notes

Variable Type: Free Text

Description: Any notes or insights about the incident that do not fit into other variables.

Coding Rules: One paragraph or less depending on available information.

Examples:

- **SMICI0185:** *Dentist industry as a whole has poor security practices, as many offices do not want to spend on security. Seen as added cost and not part of core business operations.*
- **SMICI0219:** *BlueScope Steel is a global company; however, their headquarters and areas impacted are in Australia.*
- **SMICI0243:** *Attack was part of a trend of targeted attacks on academic and research commuters. This attack is strongly suspected to be state based, but actors and motives are not known.*

Confidence_Level

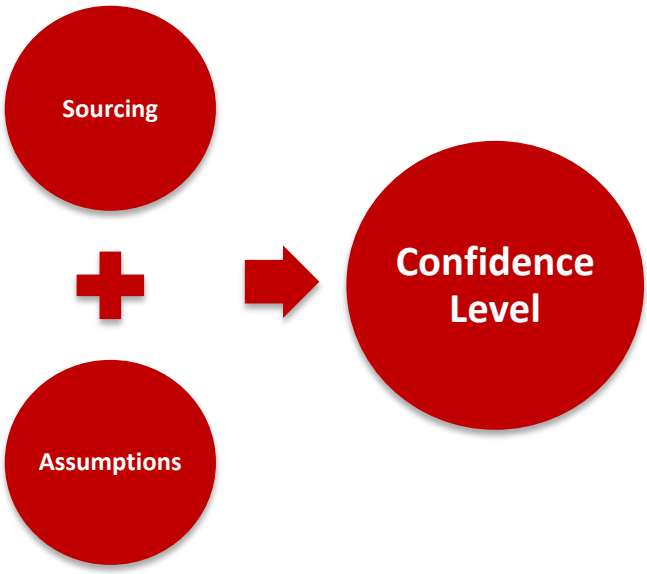
Variable Type: Categorical

Description: The SMICI dataset is built on information disclosed from publicly available sources. Therefore, the vetting of reliable sources and making as few assumptions as possible is needed to affix a high confidence level to the incident. In brief, the confidence variable is a check on the veracity of the data collection for an incident, but also a self-check on the analyst/coder's assumptions.

Confidence Level:
Certainty the incident meets SMICI inclusion criteria. Moreover, the incident actually took place as publicly reported.

Sourcing:
How reliable is the sourcing on the incident?

Assumptions:
Are many assumptions needed to justify "x" level of confidence?



Confidence Level:	Low (11-49%)	Medium (50-79%)	High (80-99%)
Sourcing:	Inaccurate, uncorroborated information from a mix of quasi-proven and unproven sources.	Reasonably accurate and partially corroborated information from a mix of proven and unproven sources.	Very accurate and well-corroborated information from proven, trustworthy sources.
Assumptions:	Many	Several	Minimal to none

Coding Rules: Select one of the confidence levels for the incident.

- Cannot code unknown (-99) for this variable.

Confidence	Code
High	1
Medium	2
Low	3

Sources

Variable Type: Free Text

Description: Sources for each incident.

Coding Rules:

- After each URL link, input three spaces, then semi-colon, followed by three more spaces before inputting the next URL link.
- When inputting the URL, defang the hyperlink function of "http" or "https" by typing in **hxxp** or **hxxps** instead.

Example:

hxxps://www.nbcnews.com/tech/tech-news/computer-worm-forces-hospitals-turn-away-patients-flna118525 ; hxxps://www.csoonline.com/article/2130463/hospital-turns-away-patients-after--virus--disrupts-network.html