MINISTRY OF DIGITAL

CyberSecurity MALAYSIA

University of Kent

Validating a Set of Candidate Criteria for Evaluating Software Tools and Data Sources for National CSIRTs' Cyber Incident Responses

14th International Conference on
IT Security Incident Management & IT Forensics - IMF 2025

Albstadt, Germany

16th September 2025

Sharifah Roziah Binti Mohd Kassim, CyberSecurity Malaysia, Malaysia
Shujun Li, University of Kent, UK
Budi Arief, University of Kent, UK

SECURING OUR CYBERSPACE

Background

- The study focuses on National Computer Security Incident Response Teams (CSIRTs) operational practices concerning tools and data. National CSIRTs are established worldwide to coordinate responses to cyber security incidents at the national level.

- In the present study, we validate the criteria identified in the Focus Groups, using Semi-structured interviews.

- After validating the candidate criteria using semi-structured interviews with these nine interviewees, we applied the criteria to evaluate a selection of software tools and data sources by converting each criterion into one or more relevant metrics.

Sharifah Roziah Binti Mohd Kassim, Shujun Li, and Budi Arief. 2023. Understanding How National CSIRTs Evaluate Cyber Incident Response Tools and Data: Findings from Focus Group Discussions. ACM Digital Threats: Research and Practice 4, 3, Article 45 (2023), 24 pages. https://doi.org/10.1145/3609230

We identified a set of criteria that can be used to evaluate free and open-source tools and public data using focus group discussions with several National CSIRTs worldwide.

Research Questions

- RQ1: How do staff members of national CSIRTs perceive the practical usefulness of the candidate criteria for evaluating tools and data sources?
- RQ2: How do staff members of national CSIRTs perceive the readiness to deploy the candidate criteria for evaluating tools and data sources in national CSIRTs?
- RQ3: How easily can the candidate criteria be applied to evaluate tools and data sources?

The Candidate Criteria -- Tools

Criteria	Definition
Security	
Confidentiality	The degree to which a product or system ensures that data is accessible only to those authorised to have access
Integrity	The degree to which a system, product or component prevents unauthorised access to, or modification of, computer programs or data
Authenticity	The degree to which the identity of a subject or resource can be proved to be the one claimed
Usability	
Learnability	The degree to which specified users can use a product or system to achieve specified goals of learning to use the product or system with effectiveness, efficiency, freedom from risk and satisfaction in a specified context of use
Operability	The degree to which a product or system has attributes that make it easy to operate and control
User interface aesthetics	This refers to properties of the product or system that increase the pleasure and satisfaction of the user, such as the use of colour and the nature of the graphical design
Accessibility	The degree to which a product or system can be used by people with the widest range of characteristics and capabilities to achieve a specified goal in a specified context of use
Maintainability	
Maintainability	The degree of effectiveness and efficiency with which a product or system can be modified by the intended maintainers
Supportability	The degree to which a product or system could provide support and assistance to users when encountering a problem
Analysability	The degree of effectiveness and efficiency with which it is possible to assess the impact on a product or system of an intended change to one or more of its parts, to diagnose a product for deficiencies or causes of failures, or to identify parts to be modified
Modifiability	Modifications can include corrections, improvements or adaptation of the software to changes in the environment and in requirements and functional specifications
Compatibility	
Interoperability	The degree to which two or more systems, products or components can exchange information and use the information that has been exchanged
Functionality	The degree to which the set of functions covers all the specified tasks, appropriateness of the tasks and user objectives
Performance Efficiency	The performance relative to the number of resources used under stated conditions
Time behaviour	The degree to which the response and processing times and throughput rates of a product or system, when performing its functions, meet requirements
Capacity	The degree to which the maximum limits of a product or system parameter meet requirements
Money	The degree of how much money is used in relation to the results achieved
Human effort	The degree of how much human effort is used in relation to the results achieved
Material	The degree of how much material is used in relation to the results achieved
Reliability	
Reliability	The degree to which a system, product or component performs specified functions under specified conditions for a specified period of time
Availability	The degree to which a system, product or component is operational and accessible when required for use
Compliance	The degree to which tools comply with a specific policy, rules and regulations and operations
Certification	The degree to which tools are certified and accredited by reputable accreditation and certification bodies

The Candidate Criteria -- Tools

Criteria	Definition
Context Coverage	
Flexibility	The degree to which a product or system can be used with effectiveness, efficiency, freedom from risk and satisfaction in contexts beyond those initially specified in the requirements
Usability	
Satisfaction	The degree to which user needs are satisfied when a product or system is used in a specified context of use
User experience	The degree of users' perceptions and responses that result from the use and/or anticipated use of a system, product or service
Usefulness	The degree to which user needs are satisfied with their perceived achievement of pragmatic goals, including results and consequences of use
Trust	The degree to which the user has confidence that the product will behave as intended
Comfort	The degree to which user needs are satisfied with physical comfort
Effectiveness	The degree to which accuracy and completeness with which users achieve specified goals
Freedom from Risk	
Sustainability	The degree to which a system, product or component is sustainable with freedom from risk – economic risk mitigation, health and safety risk mitigation and environmental risk mitigation
Harm from use	The degree of negative consequences regarding health, safety, finances or the environment that result from the use of the system
Popularity	The degree to which the security community (large or small) uses the tool

Quality in Use

The Candidate Criteria -Data

Criteria	Definition
Credibility	The degree to which data has attributes regarded as true and believable by users in a specific context of use. Credibility includes the concept of authenticity (the truthfulness of origins, attributions, commitments)
Efficiency	The degree to which data has attributes that can be processed and provide the expected levels of performance by using the appropriate amounts and types of resources in a specific context of use
Confidentiality	The degree to which data has attributes that ensure that it is only accessible and interpretable by authorised users in a specific context of use
Accuracy	The degree to which data has attributes that correctly represent the true value of the intended attribute of a concept or event in a specific context of use
Precision	The degree to which data has attributes that are exact or that provide discrimination in a specific context of use
Understandability	The degree to which data has attributes that enable it to be read and interpreted by users and are expressed in appropriate languages, symbols and units in a specific context of use
Currentness	The degree to which data has attributes that are of the right age in a specific context of use
Completeness	The degree to which subject data associated with an entity has values for all expected attributes and related entity instances in a specific context of us

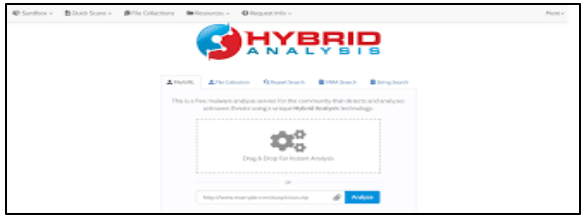
Methodology: Data Collection

National CSIRTs	Website
Uganda CERT	https://www.cert.ug/
Albania CERT	https://cesk.gov.al/
CERT BUND (Germany)	https://www.bsi.bund.de/
NCSC Switzerland	https://www.ncsc.admin.ch/
CERT-MZ (Mozambique)	https://www.cert.mz/
ID-SIRTII/CC (Indonesia)	https://idsirtii.or.id/
NCSC-FI (Finland)	https://www.kyberturvallisuuskeskus.fi/en/our-activities/cert
JpCERT/CC (Japan)	https://www.jpcert.org/
INCIBE-CERT (Spain)	https://www.incibe-cert.es/

1. Online Semi-structured interviews

Methodology: Data Collection

2. Applying the candidate criteria more objectively using several metrics



SHADOWSERVER

Data Analysis – Content Analysis

- Content analysis also allows for categorising, quantifying and describing the data objectively.
- Data analysis method that is flexible, yet systematic and rigorous
- Suitable when an existing theory or the research literature on a particular topic of study is limited [21] – as was the case in our study
- Content analysis is the best fit for exploratory research to gain new insights, opinions and views that could answer research questions and achieve the aim of a study

Codes -- “tags” or “labels”

- In-vivo coding captured exactly what the interviewees had said -- concrete and specific codes.
- We used Erlingsson’s coding model to guide the coding process, which we found easier to follow.
- Codes were developed using a data-driven approach (from the raw interview data) instead of a theory-driven approach since the study was not based on any existing theory.
- Only the primary author did the coding. Hence, no issues in establishing consistency or reliability of the coding process
- We extracted words, phrases and sentences from the interview data as meaningful codes while considering the research questions.
- Focused on extracting “manifest meaning” (what has been said) or surface meaning of the data instead of “latent meaning” (what is intended to be said) or more profound meaning.

Applying the Candidate Criteria

- Evaluate two sample tools and one data source to derive several concrete metrics and values.
- Doing so gives further evidence of the practicality of the criteria in practice.
-
- This supplements the opinions from the semi-structured interviews and makes the study's findings more credible and reliable.

1. All candidate criteria were checked individually.
2. Determine if a criterion is relevant to the evaluated tool.
    *NO -- not relevant, move on to the next criterion.
    *YES -- relevant, identify one or more suitable metrics for the criterion and determine the value for each metric.
3. Value can be derived from:
    *Actual information about the tool's features from its documentation,
    *Actual output and results obtained after inputting an artefact to the tools -- a PDF file.

SHADOWSERVER

VIRUSTOTAL

Choose file

Result -- Interviews

- All nine interviewees generally perceived the candidate criteria **useful** for evaluating tools and data in national CSIRTs.
- The majority (8) of interviewees perceived the candidate criteria very positively **("good", "nice", and "great**").
- Seven interviewees expressed that the candidate criteria could help national CSIRTs **select the right tools and data sources.**
- Three commented that the candidate criteria were **comprehensive and complete.**
- One commented that the candidate criteria were also considered useful for **software tool development** in national CSIRTs.

"We have lots of technical criteria, but not in the end user perspective. So I think this is a good, good starting point." (NCSC-FI)

"Saying that the criteria can be a guideline or best practice." (CERT-BUND)

Yes, I think I'm sure the tool is very important, especially because we are in our early stages, and that can help us use the criteria to actually select the tools that we're going to use." (CERT-Mozambique)

Especially for the criteria, which is already referred to the applicable international standard reports." (IDSIRTII - Indonesia)

Result – Usefulness of the Criteria

How interviewees perceived usefulness of criteria (inductive codes)	Number of interviewees
The criteria provided are good, nice, great	8
Can help national CSIRTs to select tools and data	7
Useful for operations	4
Comprehensive and complete criteria	3
Approach of the criteria is good, helpful and interesting	3
The criteria are important	2
Criteria are valuable	2
A valid research area	1
Criteria have valid points	1
The basic idea around the criteria is interesting	1
The research tackled both sides, tools and data	1
Methodology used and the evaluation is nice	1
Needed by the National CSIRTs	1
The criteria fit	1
Increase quality of incident response reports	1
Positive with the criteria	1
There is no problem with the criteria	1
Would not take out any points from the criteria	1
Easy-to-understand criteria	1
Good point	1
Big help	1

Result – Deployment of Criteria in National CSIRTs

How interviewees perceived deployment readiness of criteria (inductive codes)	Number of interviewees
Can be used in national CSIRTs	7
Don't have criteria like this	2
Can be used in CSIRTs	1
Can be used in all CSIRTs	1
Worth a try	1
Can be a best practice	1
Good if implemented in our organisation	1
Can be a guideline	1
To evaluate new tools, of course	1
They will help us – national CSIRT	1
We want to borrow the criteria	1
It is good for new CSIRTs	1
Beneficial for us	1
Not a must-have	1
Certainly	1

Result – Two Candidate Tools

- Operationalising the Criteria for Evaluating Tools and Data. The evaluation results show that:
  - The criteria can **be operationalised** to evaluate tools and data in national CSIRTs. The evaluation results show how
  - The criteria can be **contextualised and translated into concrete metrics** when applied in real-world operations to evaluate tools and data.

- Demonstrating the Criteria's Usefulness.
  - The **differences** between VirusTotal and Hybrid Analysis were observed in terms of "User interface aesthetics", "Accessibility", "Performance efficiency", "Interoperability", "Learnability" and "Supportability". Such differences show the potential ability of the criteria to **distinguish different tools** from each other.

  - Highlight the **utility of one tool over the other** and potentially help in **decision-making** to identify suitable tools to support incident responses. Moreover, it helps to provide a more systematic way of identifying suitable tools.

Result – Deployment of Criteria in National CSIRTs

Evaluation of VirusTotal and Hybrid Analysis Tools – Product Quality

Criteria	Metrics	Result and Value	
Security		**VirusTotal**	**Hybrid Analysis**
Confidentiality	Not Applicable (NA) because end users of this service (national CSIRTs) are not concerned about the confidentiality of the data to authorised users only.	NA	NA
Integrity	Applicable because end users care about data integrity (such as data is not manipulated). Metric 1: If the data transmission is encrypted end to end using HTTPS (Binary: Yes and No) Metric 2: If the online service is hacked by a malicious party who can manipulate the data before it is sent to the end users (Binary: Yes and No)	Yes No	Yes No
Authenticity	Applicable because end users care about the authenticity of the server. Metric 1: If the HTTPS protocol is used to provide service authentication ((Binary: Yes and No) Metric 2: To what extent can end users be sure about the real identity of the service provider or the developer if there is no service provider (e.g., a company's registration record, paid membership official contract documents, natural person's real-world identity), whose value can be one of the following: • the real identity can be fully obtained • the real identity can be partially obtained (e.g., just an online account, but not real-world identity) • the real identity cannot be obtained	Yes. HTTPS is used 'The real identity can be fully obtained' from a DomainTools whois search, with the below result: IP: 74.125.34.46 IP Location: California – Mountain View – Google ASN: AS15169, Google, US (registered Mar 30, 2000) Company information provided on its website as: Virus Total, USA	Yes. HTTPS is used 'The real identity can be fully obtained' from a DomainTools whois search, with the below result: IP: 104.18.34.183 IP Location: California – San Jose – CloudFlare Inc. ASN: AS13335, Cloudflare-net, US (registered Jul 14, 2010) Company information provided on its website as: Hybrid Analysis, Germany

Evaluation of VirusTotal and Hybrid Analysis Tools – Quality in Use

Criteria	Metrics	Result and Values	
Context Coverage		VirusTotal	Hybrid Analysis
Flexibility	Applicable as users are concerned about the tool's flexibility to specific users (e.g. non-expert users) to achieve intended goals. Metrics: If users perceive the software is flexible for non-expert users. (Binary: Yes and No) These perceptions can be obtained from ``External reports'' and ``Estimations with Co-workers from a national CSIRT''. If the above external reports or estimation with co-workers could not be obtained, the value is ``Not found''. However, if a national CSIRT wants to evaluate a tool based on this criterion, it is best to survey and gather opinions and insights on the software's flexibility to get the value.	Yes	No
Usability			
Satisfaction			
User Experience	Applicable as users are concerned that they perceive and are satisfied with the usefulness of the result. Metric: If users perceive the result from the tool as useful. (Binary: Yes and No) Measurement: Based on how users perceive the usefulness. These perceptions can be obtained from ``External reports'' and ``Estimations with Co-workers from a national CSIRT''. If the above external reports or estimation with co-workers could not be obtained, the value is ``Not found''. However, if a national CSIRT wants to evaluate a tool based on this criterion, it is best to survey to get the value.	Yes	Yes

Result – Evaluation of Data Source

Criteria	Metrics and Values	Result and Value
		Shadowserver Data Feed
Confidentiality	Not applicable (NA) to end users of this service (national CSIRTs) because they don't have concerns about the confidentiality of the data feeds (which are public)	NA
Accuracy	Applicable as users care that the data has accurate information about an incident such as an indicator of compromise (IOC). Metric: If the data accurately indicates or tells about an incident Value: 'Presence of malware URL', 'Presence of phishing URL', 'Hash value of malware'	'Presence of malware URL', 'Presence of phishing URL', 'Hash value of malware'
Precision	Applicable as users care that the data is precise enough for taking further action, such as takedowns. Metric: One metric identified is multi-valued and categorical - 'what specific details the data has about an incident'. Value: 'complete URL', 'source IP address', 'destination IP address', 'timestamp', 'hash value(s) of related software/malware', 'network protocol type', 'port number(s)', 'geo-location(s)'.	'Complete URL', 'source IP address', 'destination IP address', 'timestamp', 'hash value(s) of related software/malware', 'network protocol type', 'port number(s)', 'geo-location(s)'
Understandability	Applicable to ensure the data is in a format that makes data easily understood and facilitate analysis. Metric: The format that can be understood by human users to facilitate analysis Value: 'Can be exported as a CSV file', 'Displayed structured in a table'	'Can be exported as a CSV file', 'Displayed structured in a table'
Currentness	Applicable as users are concerned that the data is current and not outdated. Metric: Shows the current date and timestamp of the data Value: 'Shows current date'	'Shows current date - 12/02/2023 1:28:00 AM' which is the date of evaluation
Completeness	Applicable as users care about the sufficiency of data for further incident response action. Metric: Data is sufficient for a further incident response such as escalation to Service Provider	'Has a complete URL', 'Has IP address', 'Has date', 'Has timestamp', 'Has ASN number', 'Name of malware', 'Has malware hash value'

- The small sample size might make the findings less generalisable
- The candidate criteria are largely very high-level, so translating them into more
- The smooth application of the criteria in national CSIRTs

Despite these limitations, we consider the main findings of our study to be valid and reliable, considering:

- The nine interviewees' opinions are highly consistent.
- The main findings are consistent with the results presented in: https://doi.org/10.1145/3609230
- The main findings also match the first author's experience as an employee of a national CSIRT for over 20 years.

FUTURE WORK

- Suggested on the criterion-to-metric process and creating more detailed guidelines and case studies
- Expand the criteria and metrics from this research to construct an even more comprehensive taxonomy or ontology that will connect the criteria, different types of tools and data sources used by (national and non-national) CSIRT.
- Conducting future work on more in-depth studies of the criteria not currently available in the ISO/IEC 25000 SQuaRE Model, e.g., **Compliance, Popularity, and Certification**.

- The results of both empirical studies confirmed the completeness, comprehensiveness, practical usefulness and deployment readiness of the candidate criteria. Our evaluation

- The results will be released as public resources to help national CSIRTs and other CSIRTs consider adopting the study's criteria in their incident response operational practices.

MINISTRY OF DIGITAL

CyberSecurity MALAYSIA

THANK YOU

CyberSecurity Malaysia
Level 7, Tower 1, Menara Cyber Axis, Jalan Impact, 63000 Cyberjaya
Selangor Darul Ehsan, Malaysia

T +603 8800 7999 | **F** +603 8008 7000 | **H** 1 300 88 2999

www.cybersecurity.my | enquiry@cybersecurity.my

CyberSecurityMalaysia | cybersecuritymy | cybersecuritymy | CyberSecurity Malaysia | cybersecurity_my | cybersecuritymy

DIGITAL MALAYSIA MADANI

ASEAN MALAYSIA 2025 INCLUSIVITY AND SUSTAINABILITY

CERTIFIED TO ISO/IEC 27001:2013 CERT. NO. : ISMS 00114

ACB 021
SAMM 456

Status Company

WINNER

Recognition of Excellence 2023 by OpenGov

CyberSAFE