Understanding How National CSIRTs Evaluate Cyber Incident Response Tools and Data: Findings from Focus Group Discussions

SHARIFAH ROZIAH BINTI MOHD KASSIM, SHUJUN LI, and BUDI ARIEF, Institute of Cyber Security for Society (iCSS) & School of Computing, University of Kent, UK

National Computer Security Incident Response Teams (CSIRTs) have been established worldwide to coordinate responses to computer security incidents at the national level. While it is known that national CSIRTs routinely use different types of tools and data from various sources in their cyber incident investigations, limited studies are available about how national CSIRTs evaluate and choose which tools and data to use for incident response. Such an evaluation is important to ensure that these tools and data are of good quality and, consequently, help to increase the effectiveness of the incident response process and the quality of incident response investigations. Seven online focus group discussions with 20 participants (all staff members) from 15 national CSIRTs across Africa, Asia Pacific, Europe, and North and South America were carried out to address this gap. Results from the focus groups led to four significant findings: (1) there is a confirmed need for a systematic evaluation of tools and data used in national CSIRTs, (2) there is a lack of a generally accepted standard procedure for evaluating tools and data in national CSIRTs, (3) there is a general agreement among all focus group participants regarding the challenges that impinge a systematic evaluation of tools and data by national CSIRTs, and (4) we identified a list of candidate criteria that can help inform the design of a standard procedure for evaluating tools and data by national CSIRTs. Based on our findings, we call on the cyber security community and national CSIRTs to develop standard procedures and criteria for evaluating tools and data that CSIRTs, in general, can use.

CCS Concepts: • **Security and privacy** → **Human and societal aspects of security and privacy**; *Usability in security and privacy*;;

Additional Key Words and Phrases: CSIRT, computer security incident response team, CERT, computer emergency response team, CIRT, cyber incident response team, incident response, national CSIRT, focus group discussion, tools evaluation, data evaluation

ACM Reference format:
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. *Digit. Threat. Res. Pract.* 4, 3, Article 45 (October 2023), 24 pages.
https://doi.org/10.1145/3609230

1 INTRODUCTION

Cyber incidents continue to increase globally, exemplified by the increase in ransomware incidents with an almost 13% increase in 2021 compared to the year 2020 [89], a rise as big as the last five years combined [89]. Increasing threat actors of all motivation and skill levels have increased cyber-attacks in organisations across

Authors' address: S. R. B. Mohd Kassim, S. Li, and B. Arief, Institute of Cyber Security for Society (iCSS) & School of Computing, University of Kent, Giles Lane, Canterbury, Kent, UK, CT2 7NZ; emails: {sm2212, s.j.li, b.arief}@kent.ac.uk.

2576-5337/2023/10-ART45
https://doi.org/10.1145/3609230

geographies and sectors throughout 2021 [73]. One of the contributing factors to the increase in cyber attacks is the delays in applying patches, coupled with the use of outdated systems [7]. Furthermore, with the COVID-19 pandemic [95], people have become much more reliant on digitally connected devices and systems, and this has caused an increase in exposure to cyber threats [6]. For instance, there has been wider adoption and increased usage of working from home, e-learning and online shopping, which may have led to unforeseen new vulnerabilities [54]. This, in turn, has caused a further increase in the number – and severity – of cyber attacks worldwide [44].

Therefore, as prevention, it is essential to ensure a strong and effective defence to protect systems and networks. On top of that, there is a growing importance to being able to respond to a cyber attack, once it occurs, in a timely and efficient manner [21] – quite often, facilitated by a dedicated incident response team [42]. The growth of cyber security science has highlighted the value and contributions that computer security incident response teams can make [29] in defending and mitigating cyber attacks.

Several names and acronyms often refer to an incident response team. An incident response team may be officially known as a ***computer security incident response team (CSIRT)*** [30, 76, 93], a ***computer emergency response (or readiness) team (CERT)*** [4] or a ***cyber (or computer) incident response (or readiness) team (CIRT)*** [47]. For consistency, the term CSIRT is used in this study mainly because it is one of the most widely used terminologies in the research literature and among practitioners.

The primary aim of the computer security incident response team is to mitigate the impact of a potential major incident [2] by minimising the damage, cost and recovery time. Such a team is also instrumental in finding and fixing the root cause(s) of incidents and subsequently preventing future attacks [85]. Therefore, CSIRTs must be prepared and ready to face the growing number of vulnerable systems connected online and the lack of security awareness among Internet users.

National CSIRTs, a special type of CSIRTs, are established at the national level to deal with and coordinate responses to cyber incidents in a specific nation or political region [4, 27, 75, 96]. The CERT Division of **Software Engineering Institute (SEI)** of **Carnegie Mellon University (CMU)** defines a national CSIRT as: "A computer emergency response team (CSIRT) with National Responsibility (or National CSIRT) is a CSIRT that is designated by a country or economy to have specific responsibilities in cyber protection for the country or economy. A National CSIRT can be inside or outside of government but must be specifically recognized by the government as having responsibility in the country or economy" [80]. A list of national CSIRTs from around the world can be found on the websites of the CERT Division, SEI of CMU [80] and the UN agency **International Telecommunication Union (ITU)** [47]. In addition to national CSIRTs, community-based CSIRTs like AfricaCERT, OIC-CERT, and APCERT facilitate discussions on cyber security and incident responses in a larger geographical region [61]. Additionally, Task Force-CSIRT (TF-CSIRT) (https://tf-csirt.org/) promotes collaboration and coordination between CSIRTs in Europe and neighbouring regions, and CERT-EU (https://cert.europa.eu/) is the CERT of all the EU institutions, bodies and agencies to contribute to securing their ICT infrastructure and helping to prevent, detect, mitigate and respond to cyber-attacks.

National CSIRTs have received much attention due to their key role in safeguarding national infrastructures from cyber attacks. To our knowledge, limited research has been done concerning national CSIRTs' operational practices. This is a critical research gap that the present study aims to address. It is known that public data, open data, closed-source data, **open-source intelligence (OSINT)**, and open-source and free tools are frequently used by national CSIRTs to facilitate their incident responses [61, 62]. However, the research literature lacks a systematic discussion on how national CSIRTs would ensure the quality of such tools and data they use in the operations. Some national CSIRTs evaluate the tools and the data they use in an ad-hoc and informal manner to ensure using appropriate tools and data [62]. This raises concerns about how such tools and data are selected systematically and properly. To allow qualified tools and data to be used in operation, it is essential to evaluate such tools and data using systematic procedures or guidelines – valid criteria for evaluating tools and data [38].

Such an evaluation is critical to avoid problematic tools and data, such as "a lack of security" [15] and to ensure only effective tools and data are used [14].

Hence, it would be interesting to examine further the ad-hoc practices in national CSIRTs on tools and data evaluation reported in previous studies [61, 62] and to understand better issues preventing a systematic practice. The first author of this paper (who is a member of staff of a national CSIRT) observed that public data, OSINT, and free and open-source tools are often not evaluated (or not being assessed systematically) due to a lack of standard procedures, resources and expertise [61]. The insights gained from the present study are essential to inform the development of a systematic procedure and guideline – criteria for evaluating tools and data that can be used in national CSIRTs.

It should be noted that, in this study, the tool focused on free and open-source tools only, while data focused on public data only.

The study aims to understand better how national CSIRTs participating in this study evaluate tools and data and to identify a set of candidate criteria that can be used to support the systematic evaluation of tools and data within national CSIRTs. The aim is split into four **research questions (RQs)**:

RQ1: What are the definitions of tools and data used by national CSIRTs participating in this study?
RQ2: What are the current tools and data evaluation practices in the national CSIRTs participating in this study?
RQ3: What challenges and limitations do participating national CSIRTs face in evaluating tools and data?
RQ4: What candidate criteria can be used when evaluating tools and data from the perspective of national CSIRTs participating in this study?

To answer the RQs, we conducted seven online focus group discussions with 20 staff members from 15 national CSIRTs in Asia-Pacific, America, Africa and Europe. The results of the focus group discussions led to five main findings, which are summarised as follows:

(1) All focus groups reported the same definitions of tools and data used by the participating national CSIRTs.
(2) Provides real-world evidence of how tools and data are currently evaluated in the operational practices of the participating national CSIRTs.
(3) The results and efforts from the study are translated into a set of candidate criteria that national CSIRTs, CSIRTs and the broader security operations could use to specify the quality of tools and data.
(4) New candidate criteria identified from the study could help refine requirements defined in the ISO/IEC 25000 standard "Systems and software Quality Requirements and Evaluation (SQuaRE)" [88] and other software evaluation models in the literature.
(5) Key findings from the study lay solid foundations for future research and development activities for different stakeholders: (1) national CSIRTs (towards improving their operational practices); (2) software developers and vendors (towards developing tools and data that are more aligned with users' needs); and (3) researchers (towards conducting more targeted research, e.g., in developing more advanced machine learning methods for tools and data evaluation).

The rest of this paper is organised as follows. Section 2 provides an overview of related work on tools and data evaluation within CSIRTs. Section 3 explains the study's methodology, particularly the data collection and analysis strategies. Section 4 presents the results from the study, while Section 5 discusses the key findings and the implications of the study, along with the study's limitations. Finally, Section 6 concludes the study and provides several suggestions for future research.

2 RELATED WORK

In this section, we review related work in three different areas to clarify the importance of tool and data evaluation for CSIRTs and related work in the context of CSIRTs and other related contexts and to show the research gaps our study will fill.

2.1 Importance of Tool and Data Evaluation for CSIRTs

Tools and data are paramount for CSIRTs and national CSIRTs *to perform successful and effective incident responses* [14, 48]. Having *good and effective mitigation tools for CSIRTs* and companies at large are essential in mitigating cyber attacks [44]. Using appropriate technological tools to facilitate incident response can greatly increase the effectiveness of CSIRTs. The effectiveness of tools may reside in the lead time of solving an incident, on the financial level and in increasing team knowledge and shared situation awareness within a CSIRT [52]. Furthermore, to ensure only qualified tools and data are selected, particularly for incident management and analysis work in national CSIRTs, Bills et al. [14] pointed out the need for evaluation and implementation of tools. On top of what was suggested by Bills et al. (2022) [14], Nowikowska (2022) [69] stressed out the *need for CSIRTs to examine IT devices or software to identify vulnerabilities* – such an examination is crucial to avoid using un-patched tools in the operations. Therefore, evaluating tools and data in a CSIRT is crucial to ensure the team is equipped with quality tools and data to detect, respond, and mitigate security incidents effectively.

2.2 Tool and Data Evaluation Within the Context of CSIRTs

Studies and past research concerning tool and data evaluation within CSIRTs or national CSIRTs could not be identified in the literature. Based on a systematic literature review study we conducted concerning the use of public data and free tools in national CSIRTs, no past studies or research concerning tools and data evaluation in national CSIRTs was observed [63]. Though researchers have investigated the general use of tools in the operational practices of CSIRTs [55, 81], an important area of research – systematic procedures to evaluate tools and data is missing. Such research is vital in ensuring qualified tools and data are used to support incident responses in national CSIRTs and CSIRTs. To enlighten the above gap, Iakovakis et al. [44] developed a list of mitigation tools, selected based on the following criteria – strengths, weaknesses, free trial, cost/price, scalability, technical support, vulnerability assessment, reports and analytics, ease of use, GUI offered, and compatibility. However, methodologically, it remains unclear how such criteria were identified by Iakovakis et al.. Hence, our study extends Iakovakis et al.'s work by using a focus group approach with participating national CSIRTs to establish candidate criteria for evaluating tools and data.

A need-assessment study by Kleij et al. [52] with CSIRTs identified certain features (criteria) that must be present in tools for efficiency. These include that tools must be able to produce reports and output that are more structured and reader-friendly, scalability of tools to handle large-scale incidents, visualisations for a better understanding of the insights and the ability to support different levels of details. Expanding on Kleij et al., the study is interested in identifying a set of candidate criteria (features) that can be used to select effective tools.

Problems concerning data quality have become critical in data management, such as for avoiding inconsistent, duplicate, inaccurate and incomplete data [31]. Vetrò et al. [90] identified common problems with data, such as a lack of metadata, incomplete data, and a lack of a measuring system for geographical information. To address these problems, Vostrovskỳ et al. [91] presented a new design of **data quality management (DQM)** in relation to open data based on SQuaRE standards for data quality management (DQM). The authors proposed a new DQM framework for open data usable in the agricultural sector.

Some work has been done in evaluating data within CSIRTs, more precisely, on threat data feeds. Pawlinski and Kompanek [72] studied evaluating threat intelligence data feeds in helping users to choose qualified threat data feeds available in abundance. The researchers evaluated sample threat feeds using the following factors (criteria): relevance, accuracy, completeness, timeliness and ingestibility with several metrics. Similarly, Kührer et al. [57] evaluated blocklists data feeds based on the following criteria: vantage, volume, timeliness, accuracy, and completeness using several metrics. Expanding on Pawlinski and Kompanek [72] and Kührer et al. [57], this research intended to identify candidate criteria for evaluating data (and tools) using an approach to get collective opinions from the perspective of national CSIRTs. This aligns with de Smale et al. [24], who pointed out that organisations should have ongoing processes to evaluate vulnerability information to address the problem of incomplete and untimely data. The authors concluded that organisations should formalise vulnerability information evaluation with suitable metrics and measurements. While de Smale et al. had a narrow focus on vulnerability data, this research focuses on general data and tools – notably public data, free and open-source tools.

2.3 Tool and Data Evaluation in Other Contexts

In addition to the related work described in the previous subsection, we have also reviewed some studies on evaluating tools and data in a context not directly related to CSIRTs, such as from users' perspectives.

Several studies were conducted to evaluate software from the users' perspective. Ward Jr and Venkataraman [92] stressed the importance of including end users in software evaluation. For example, Boloix and Robillard [16] proposed a comprehensive tool evaluation framework to assess the quality of the tools from the viewpoints of producers, operators, users, managers, and stakeholders. Azizyan et al. [8] surveyed to identify the features indicating the quality of tools from users' perspectives based on the most and least satisfactory attributes of the tools. Gade [34] studied software quality and attributes from the end users' perspective using an online survey with users and discussing the results in a focus group discussion. The factors used for the survey were based on the ISO-9126 model and several other software evaluation models identified from the literature. Similar to Gade (2013), Belinda et al. [11] proposed a software quality model from users' and developers' perspectives using attributes from the ISO 9126 model and other software quality models such as McCall's, Boehm's, FURPS and Dromey's. Meanwhile, Stojkovski et al. [84] used the **User Experience Questionnaire (UEQ)**, a validated instrument for measuring **User Experience (UX)**, to evaluate the quality of open-source software systems.

It should be noted that much of the user-based software evaluation studies from the literature adopted existing criteria or factors from ISO/IEC standards and software evaluation models, e.g., the ISO 9126 model, McCall's, Boehm's, FURPS, Dromey's and also UEQ, to identify the most preferred criteria or factors by users. In essence, the above studies are related to our present study in the context of identifying candidate criteria for evaluating tools and data from a user's perspective. In this study, we attempt to enhance previous studies by exploring what criteria or factors participants perceive for inclusion in a set of candidate criteria for evaluating tools and data. We did not extract the candidate criteria from any standards or software evaluation models but explored candidate criteria from the viewpoints of recruited participants from national CSIRTs.

On data evaluation, Wu et al. [97] studied the quality of governmental open data during the COVID-19 pandemic and proposed an open government health data quality evaluation framework. The study framework was designed to assess the quality of health data in terms of accuracy, completeness, timeliness, accessibility, and interoperability. The article argues that ensuring the quality of health data is crucial for effective policy-making and decision-making during public health emergencies like COVID-19 – relates to this research in the context of assuring the quality of data. Bouwman et al. [18] pointed out the importance of ensuring the quality of threat data in their study on the threat intelligence sharing community, the COVID-19 Cyber Threat Coalition. They highlighted data quality in terms of fewer false positives and data accuracy. It should be noted their study did not delve into the quality aspect of data but focused on the effectiveness of the threat intelligence community in sharing data. This research expands on Bouwman et al. (2022) by identifying criteria that can be used to specify data quality. Several studies showed how open-source tools and public data are used in CSIRTs. One such study was done by Riebe et al. [75] with Germany's state CSIRTs. The authors focused on the technological design of tools and data to support CSIRTs. However, systematically evaluating tools and data for CSIRTs and national CSIRTs is missing from their study – a gap this research intends to address.

2.4 Research Gaps Identified

Related work reviewed in the previous two subsections provides valuable insights into some of the works on tools and data evaluations in the context of CSIRTs, national CSIRTs and from users' perspectives. The need for ongoing research and improvements in this study area, subsequently improving current incident response practices, was highlighted. Four significant gaps in tools and data evaluations from all the reviewed related work were identified; (1) Very few studies focus on establishing criteria for evaluating tools and data in CSIRTs. (2) Novel criteria from users' perspectives are missing; instead, most studies used criteria adopted from existing ISO/IEC standards and a few from the research literature. (3) A lack of a qualitative approach in designing tools and data evaluation criteria, with most studies focusing on quantitative approaches. (4) Studies on establishing criteria specifically for *evaluating tools and data in national CSIRTs* are lacking.

Notably, no single tool and data evaluation model (criteria) suits all organisations; hence, different types of organisations should develop their own models [92] – justifying the present research concentrated on establishing candidate criteria for national CSIRTs.

3 METHODOLOGY

This section explains the methodology we used for the study (data collection and analysis) and how research ethics matters were considered.

3.1 Data Collection – Focus Group Discussion

3.1.1 Methodology Used. The data collection method used in this study is focus group discussions. In this section, we briefly describe the focus group discussion method and reasons for adopting this method for the study. The instrument used for data collection, a focus group schedule and the selection of questions outlined in the schedule are explained. The reliability of the study design with a pilot focus group discussion is briefly described. Recruitment of participants and the process is laid out, with the selection of participants and the number of participants who consented to participate.

Focus Group Discussion, a data collection method widely used for research purposes, was used to collect data for the study. It is a well-known data collection method used in many fields of study, mainly in market research [65], healthcare research [39], social science research [70] and cyber security research [3, 98]. Focus groups consist of participants from similar backgrounds that allow the collection of enriched qualitative data regarding a particular topic of interest [56]. We adopted focus groups to draw out expert knowledge and experiences from staff members of national CSIRTs through collective group discussions [74]. Eventually, this allowed the collection of richer information through active interactions within the focus groups [74]. Notably, all participants have a common background, expertise and experience related to the topic of the study.

Focus groups also allowed the collection of information, as rich as possible, from participants through free-flowing group interactions [39] – aligned with the inductive and exploratory nature of the present study. Notably, focus groups allowed the exploration of the research topic, which is non-sensitive, through collective open opinions, experiences and expertise within a group context, where interviews and surveys may not be suitable [35, 77]. Hence, focus groups fit the purpose of collecting rich data through collective opinions of staff members of national CSIRTs [74] concerning tool and data evaluation practices in the operations. The insights gained from focus groups could inform the development of systematic procedures and guidelines [39] – a systematic procedure for evaluating tools and data for national CSIRTs.

Focus group discussions are typically conducted face-to-face. However, the advancement of telecommunication technologies (such as online meetings and virtual discussion platforms) allowed researchers to conduct focus group discussions in an online environment [32, 82]. Previous studies have successfully demonstrated the feasibility of online focus group discussions, such as the work of Stewart and Williams [83], Fox et al. [33] and Harmsen et al. [40]. This motivated us to consider an online environment for the present study.

Notably, conducting online focus group discussions was vital for our study due to the international nature (where participants are dispersed worldwide) and the global "Covid-19 pandemic" (2020 to mid-2022), which limited travelling and face-to-face interaction with people. Therefore, online focus group discussion was the best option, besides saving cost and time. The Microsoft Teams online platform (https://www.microsoft.com/en-my/microsoft-teams/) was used in this study as the Researcher's Institution has a site license for security and compliance purposes. Microsoft Teams is also considered by us a reliable online communication platform. It allows automatic audio recording of the focus group discussions with reasonable sound quality without an additional external audio recorder. Moreover, all participants preferred Microsoft Teams after we asked them via email for a preferred platform. The interviews were audio recorded, as consented to by participants and transcribed.

The instrument used for the focus group discussions was a *Focus Group Schedule*. It contains the focus group discussion agenda with a list of questions. The questions are arranged in order, from general to specific [56]. The focus group schedule used for the study can be found at https://cyber.kent.ac.uk/research/CSIRTs/Focus-Group/Focus-Group-Schedule.pdf. The questions in the Focus Group Schedule are divided into four sections as below:

(1) Opening questions: to introduce the study and the topic, to get to know the participants,
(2) Warming up questions: to understand some general concepts and terms relevant to the study,
(3) Key questions: to collect the main points of the study, and
(4) Ending questions: collect additional information, de-brief, and summarise the discussion.

Between the 18th of October 2021 and the 16th of December 2021, we conducted seven online focus group discussions with 20 staff members of 15 national CSIRTs with knowledge and experience working in operations. To ensure the credibility and reliability of the study design, the focus group schedule was first reviewed with a domain expert from MyCERT, the national CSIRT of Malaysia, to ensure the questions were correct and appropriate (https://www.mycert.org.my/). Then, a pilot focus group with four senior staff members of MyCERT was conducted to test the Focus Group Schedule for feasibility, appropriateness and time management [50, 78]. The pilot study allowed for refinement and improvement of the Focus Group Schedule. It should be noted the pilot focus group discussion was not used in the data analysis.

3.1.2 Recruitment of Participants. The nature of our study and the definitions of the RQs required participants with specific knowledge and a good understanding of national CSIRTs' operations to provide meaningful and rich insights related to the RQs. Therefore, the selection of participants for the study was "purposive" instead of "random" [25]. More specifically, staff members working at national CSIRTs in different countries. Past studies and experience learned from the research literature showed that the best way to approach potential participants is to rely on the researchers' professional and personal contacts within national CSIRTs and cross-CSIRT organisations [62]. For this study, such contacts were secured via the first author, part of the community, as an employee of **Malaysia's national CSIRT (MyCERT)**.

We recruited a total of 20 participants through five different channels described below. The first author recruited four of her colleagues working at MyCERT as participants. Three participants were recruited using contacts from our previous study that used the same target community (staff members of national CSIRTs) for participant recruitment, for which they consented to be contacted again for future research. Nine participants were recruited through the first author's contact with the CERT Division of the **Software Engineering Institute (SEI)**, the Carnegie Mellon University of the USA (https://www.sei.cmu.edu/about/divisions/cert/). Three participants were recruited through the first author's contact with the **Organisation of Islamic Countries (OIC)** CERT (https://www.oic-cert.org/). A final participant was recruited via the first author's personal contact. Official invitation emails were used to recruit participants. A copy of the invitation email is available at https://cyber.kent.ac.uk/research/CSIRTs/Focus-Group/Email-Invitation.pdf.

All participants willingly consented to participate in the study, and their participation was not part of their official duties within their CSIRT role. Four focus groups consisted of three participants each, two more focus groups consisted of two participants each, and one focus group had four participants. Though four or five

Table 1. List of National CSIRTs in Each of the Seven Focus Groups Discussion

Focus Group (FG)	National CSIRT	#(Participants)
FG 1	NCSC-NL (Netherlands), Sri Lanka CERT/CC, TunCERT (Tunisia)	4
FG 2	BGD eGOV CIRT (Bangladesh), CERT.at (Austria), CERT-SE (Sweden)	3
FG 3	MyCERT (Malaysia), CERT-PH (Philippines)	2
FG 4	MyCERT (Malaysia), SWITCH-CERT (Switzerland), EG-CERT (Egypt)	3
FG 5	MyCERT (Malaysia), CISA-US-CERT (USA), UK-NCSC (UK)	3
FG 6	CTIR-Br (Brazil), CERT-NZ (New Zealand)	3
FG 7	MyCERT (Malaysia), NCSC-Hungary (Hungary)	2
	Total	**20**

participants in each focus group are suggested [39, 51, 56], due to the relatively low number of participants for the study and the difficulties with arranging larger focus groups, we had to keep the focus groups smaller. Nevertheless, on the positive side, smaller groups meant that all participants had a better opportunity and sufficient time to discuss and elaborate their views during the discussion. Since the study is qualitative, we did not consider the low number of participants a real concern.

To ensure the reliability of data collected during the focus group discussion, we considered how to address researcher-specific bias. Notably, the moderator is a staff member of a national CSIRT and has professional relationships with four of the 20 participants. To mitigate potential bias caused by such relationships, debriefing sessions were held at the end of each focus group discussion to collectively summarise and agree on the points discussed among participants in the focus groups [28]. On the positive side, we might get better results in this context because CSIRT personnel tend to trust peers from within CSIRTs, so we could be more open in the focus group discussions. The trust from the professional relationship may help us to get more information than we would otherwise obtain from an unknown moderator. The study's participants list is shown in Table 1. Notably, two participants represented NCSC-NL (the Netherlands), two represented CERT-NZ (New Zealand), and four represented MyCERT (Malaysia). The rest of the national CSIRTs had one participant each.

The 15 national CSIRTs from which we recruited participants can be split into two sub-types. Thirteen of them have national responsibility for the cyber protection of their respective country or economy. In contrast, the other two do not have national responsibility for cyber protection of their respective country or economy but specific roles within specific sectors. They are (1) CTIR-Br, with the role of coordinating and implementing actions for managing computer incidents (monitoring, treatment and response to computer incidents) in governmental bodies and entities in Brazil, and (2) SWITCH-CERT, whose role is to provide incident response services to research and education, the domain registry, and multiple industrial sectors (banks, industry, logistics, and energy) in Switzerland.

3.2 Data Analysis – Content Analysis

3.2.1 Methodology Used. We empirically analysed the focus group data using the content analysis method. In the content analysis, we adopted the coding process to encode the focus group data to highlight information that is of interest to the study. First, the focus group data was overviewed by counting occurrences of particular attributes for the study and translating them into figures, percentages and histograms. Then, a qualitative approach was adopted to generate the main concepts or ideas from the focus group data, from which the major findings were drawn to answer the study's research questions. Notably, for this study, the focus group is the main unit of analysis. Nevertheless, the study also paid attention to individual participants' views during data analysis for inclusion in the findings [9].

Content analysis was the best method to generate findings from the focus group data and put them into the context of the study's research questions [94]. Moreover, content analysis was a flexible approach [20, 87], yet systematic and rigorous [94]. Content analysis is suitable when an existing theory or research literature on a particular topic is limited [41], exemplified by our study topic. The research questions of our study intended to explore staff experiences concerning tools and data evaluation in national CSIRT; hence content analysis was considered the best fit for data analysis [28]. On the other hand, a thematic analysis for the study was ruled out as there was no intention to identify any significant patterns or themes across the data. Our study adopted the data analysis guidelines reported in [12]: familiarising with data, generating codes, generating categories and sub-categories, and identifying concepts or themes.

Within the content analysis, a bottom-up approach was used for this study to analyse the focus group data to capture participants' opinions, essential for developing systematic procedures and guidelines for evaluating tools and data [10]. Similarly, McCall et al. used a bottom-up approach in developing the McCall Software Evaluation Model [59]. A top-down approach was eliminated for the study as the purpose is to gain as many insights from participants and to engage directly with participants to understand the real-world operations at the national CSIRTs participated in the study [10]. Furthermore, a bottom-up approach is more straightforward for the study.

For content analysis, coding [71] was used to guide the identification of significant information in the data [9, 41] that is of interest to the study. After extracting the codes, they were grouped into several code categories [64] based on commonalities among the codes that align with the research questions.

3.2.2 *Coding Process – Inductive Coding.*

Codes and coding are essential in qualitative data analysis to organise and interpret data and capture emerging concepts, ideas and categories underpinning data [36]. Extracting relevant codes from the data involved an iterative process by constantly moving back and forth between the whole data and the code extracts to ensure the data have been fully explored, interpreted and coded [28]. The iteration here meant repeating the process before generating the desired result. The goal is to get closer to the desired results with each repetition. The iteration also helped ensure important information from the data that is of interest to the study was not missed during coding.

It should be noted that only one researcher (the first author of this paper) coded the whole focus group data since the author has rich experience as an employee of MyCERT. Therefore, there were no issues in establishing consistency or reliability of the coding process (in comparison to a situation in which several coders were involved, which would require further checks to ensure consistency and reliability). Although the study has addressed "between-coder" inconsistency by having one coder, the "within-coder" inconsistency may not have been addressed. To overcome this issue, the first author reviewed her codes multiple times and discussed the codes with the other co-authors to get some level of quality control.

During coding, we concentrated 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 deeper meaning [12]. In extracting "manifest meaning," the words in the text were used as they are, rather than finding the underlying meaning of the words or text [12]. Some examples of how we coded the focus group data are shown in Table 2.

The study adopted Erlingsson and Brysiewicz's coding model [28] for content analysis. The steps in the model are easy to follow and understand for coding the focus group data. Before coding, data (transcripts) preparation was done by the following steps:

(1) Transcribed the audio recordings of the focus group discussions manually. Transcription software was not used for privacy and accuracy purposes. The researcher listened to the recordings several times to ensure the accuracy of the transcription. Additionally, "member checking" was used to increase the accuracy of the transcription [22].
(2) Labelled all transcripts with names of the **Focus Groups (FG)** – FG1, FG2, FG3, FG4, FG5, FG6 and FG7. Participants were labelled with the name of the National CSIRTs they represent for privacy purposes. If there is more than one participant from the same national CSIRT, they are labelled as Participant 1 and Participant 2, followed by the national CSIRTs.

Table 2. Coding of the Focus Group Data

Transcripts	**Segments**	**Code**	**Group (Category)**
Focus Group 4 – P1 "We talked to our peers in our vicinity when we do a lot with our European counterparts. We meet about with other teams at FIRST meetings or international conferences and nationally, as well as we talked to fellow teams in the Austria on what they're using, what their experience is and that kind of like gives us the feedback to know, hey, is it worth looking at."	"We talked to fellow teams in the Austria on what they're using, what their experience is."	Talk to fellow teams	Evaluation Practices
Focus Group 1 – P1 "Then for tooling, it's all in usability. For tooling, it's important that the needs of the entire goals we have in processing data, in case of sensors and sensor network it's finding problems."	"It's all in usability."	Usability	Criteria Tool Evaluation
Focus Group 2 – P1 "For tools the main challenge is time if we look through our data, our tracking of what's vulnerable, we get every other day and tool that someone builds somewhere across the world and we just don't have the time to look at all new toys and tools that are popping up globally."	"The main challenge is time."	Time challenge	Evaluation Challenges

(3) Loaded all transcripts into **qualitative data analysis (QDA)** software. We used Atlas.ti version 8.4.5, one of the most widely used QDA softwares [58]. The QDA tool was used to store transcripts and manage the coding process [43].
(4) Data is now ready for the coding process.

After the above steps, transcripts and data are ready for subsequent coding. The coding of data follows the below steps:

(1) Transcripts: Read the transcripts in two steps: (a) read for the first time to get an overall understanding and impression of the data, and (b) read several times (3) to gain a better understanding of the data and grasp the details and gist of the data.
(2) Segments: We highlight and extract the content of the transcripts that are particularly interesting to the study topic and align with the research questions. This is referred to condense the transcripts into segments with significant information only.
(3) Codes: Generated codes (inductive coding) from the segments of the transcripts while keeping the research questions in mind [1, 19], as opposed to deductive coding. The lack of related research on the study's topic justifies inductive coding to extract as much information from the data. Additionally, inductive coding is suitable for developing new procedures and guidelines. Notably, in-vivo coding, instead of descriptive coding, was used to code the actual spoken words or other specialised words uttered by the participants besides better engaging with participants in generating codes [9, 36]. The codes are then scrutinised and refined to ensure consistency.

(4) Code categories or code groups: A category consists of codes dealing with the same issue or similar meanings. This is often short, factual sounding, with manifest meaning visible in the category and with limited interpretation by the researchers [28]. We grouped the codes, using a bottom-up approach, by sorting and appraising the codes to determine which codes belong to a similar group, forming a category aligned with the research questions. Additionally, we also grouped codes based on their semantic meanings. The categories are then inspected for completeness and redundancy.

3.3 Research Ethics Considerations

As the study involved human participants, before recruiting any participants, we submitted a research ethics application to the University of Kent's **Central Research Ethics Advisory Group (CREAG)** under the reference number CREAG087-09-2021 and received a favourable opinion.

An electronic **Participant Information Sheet (PIS)** was provided to participants giving details about the study, explaining the scope of the study, their rights to withdraw without giving a reason, what data will be collected and how it will be processed, stored and used. In addition, an electronic Consent Form was provided to each participant to get their consent. All participants willingly gave their consent to participate in the study and agreed to have their direct quotes included in any research publications resulting from this study, with their personal information anonymised. The **Participant Information Sheet (PIS)** and the Consent Form used in the study can be found at https://cyber.kent.ac.uk/research/CSIRTs/Focus-Group/PIS.pdf and https://cyber.kent.ac.uk/research/CSIRTs/Focus-Group/Consent-Form.pdf, respectively.

4 RESULTS

This section presents the results in the order of the RQs defined in Section 1. First, we present demographic information about the participants, followed by the answers to the RQs. It should be noted the result focuses on qualitative – insights from participants' conversations, though there is a minimal interpretation of the data in numbers and figures. The study used "P" to refer to participants and "FG" to refer to focus groups. The seven transcripts from the focus group discussions were condensed into 266 segments related to topics of interest for the study for subsequent encoding, keeping in mind the research questions. Then from the 266 segments, 289 codes were identified that precisely capture topics of interest for the study, keeping in mind the research questions. The 289 codes were grouped into six code groups or code categories based on semantic similarities. The code groups are (1) tool definition, (2) data definition, (3) evaluation practices, (4) tool evaluation criteria, (5) data evaluation criteria and (6) evaluation challenges. Additionally, we also created 66 memos to facilitate analysis. The code book for the study is available at https://cyber.kent.ac.uk/research/CSIRTs/Focus-Group/Codebook.pdf.

4.1 Demographic Information about Participants

Participants' roles in their respective national CSIRTs are presented in this section. The majority (13) of the participants were involved in the day-to-day technical and operational roles of the participating national CSIRTs. The roles include Team Leader, Analyst, Senior Analyst, Specialist, Security Engineer and Executive. The remaining seven participants were involved in managing and administrating the operations of the participating national CSIRTs. Their roles include Director, Head, Deputy Head, Advisor and Principal Administrator. The more detailed statistics of the roles can be found in Table 3.

4.2 Definitions of Tools and Data – (*RQ1*)

RQ1 aims to understand how tools and data are defined from the viewpoints of the participating national CSIRTs.

We took **Google Hack Database (GHDB)** (https://www.exploit-db.com/google-hacking-database/) as an example and asked the focus groups to define if GHDB is a data or a tool. We referred to GHDB because, in our previous study, it was unclear if GHDB is considered a tool or data. Hence, this study intends to clarify this. All

Table 3. Participants' Roles within the National CSIRTs

Roles	Number of Participants
Director	1
Head	2
Deputy Head	1
Team Leader	2
Principal Administrator	1
Advisor	2
Specialist	3
Senior Analyst	1
Analyst	4
Security Engineer	1
Executive	2

focus groups (seven) agreed and defined GHDB as data consisting of hackers' **tactics, techniques, procedures (TTP)** and **indicators of compromises (IOCs)**, as well as a database or a data resource to search and collect additional information to understand threats and facilitate incident response.

4.2.1 Definition of Tools. All seven FGs defined tools as follows: "*Tools are essential to support specific functions and purposes in the operations of national CSIRTs, e.g., for network analysis, facilitating incident reporting, processing raw data, bringing value to users, improving existing processes and generating more productive outputs, ultimately achieving users' goals and targets.*"

On the definition of tools, one participant pointed out that tools are very loosely defined and also very loosely interpreted in the literature, and it was timely our study raised the question. According to another participant, the literature also tends to be imbalanced in defining tools, covering certain types of tools only, e.g., forensic tools, while under-appreciate other tools, e.g., email and ticketing tools.

4.2.2 Definition of Data. Considering that the research literature supports diversified definitions of data [100], we wanted to understand how data is defined by national CSIRTs.

All seven focus groups defined data as: "*Data can be differentiated based on the access level of the data and how it can be shared with others (e.g., 'open data', 'public data', 'closed-source data' and 'commercial data'), for communication, to facilitate investigations, obtained from the Internet, trusted partners, vendors, shared and disseminated accordingly.*"

The focus groups clarified the difference between open data and public data. All seven focus groups defined "open data" as free for the general public but came with an agreement defined by data owners concerning sharing and circulating open data. While "public data" can be accessed, shared and circulated freely by the public without restrictions or prior agreements with data owners.

When asked about data, all seven focus groups defined data in the context of incident response. The focus groups defined data with examples of data such as IP addresses, log lines, botnets, leaked information, and web services. Furthermore, one focus group explained that data must be massaged to identify actionable information.

One participant informed that without data, national CSIRTs would not know what incident to handle, as stated by a participant:

> "*Without this data we won't be knowing what to do, what to handle.*" (P1, FG3)

The focus group explained that a tool needs data as the input to operate, while data needs a tool that converts the data into more meaningful information.

Table 4. Detail Evaluation Practices of Tools and Data in Participating National CSIRTs

Evaluation Practices	Detail Evaluation Practices (Based on FG Data)	FG
Refer to community reviews	Community and global reviews available on the Internet, interacting with people and peers, getting second opinion from other practitioners, feedback from trusted users within communities	5
Staff self-check on their own	By trial and error – try number of tools, verifying and checking source codes, trust and confidentiality of data sources, own experiences with the tools and data, results from the tool itself, self-weighing data – sensibility, validity of data, trust and confidence with data and tool, check reputation of data – from trusted people, following law and government regulations on data privacy & protection, get impression and rate the data – good or bad	5
Check with other CSIRTs	Communication with other national CSIRTs – through CSIRT networking, interactions at conferences, meetings and talk to fellow teams within a constituency	4
In-house evaluation	Conduct proof of concept (POC), through use cases, stress testing, vulnerability testing, functional testing, conduct trial run	2
Mix of in-house and external party	Have external companies do the testing in addition to in-house evaluation	1

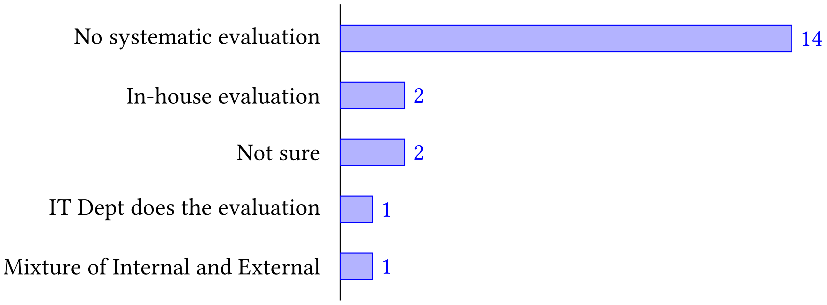

Fig. 1. Current practices in evaluating tools and data in participating national CSIRTs.

4.3 Standards and Practices for Evaluation of Tools and Data – (*RQ2*)

RQ2 aims to identify the current standards and practices in evaluating tools and data in the participating national CSIRTs and the wider community (particularly the relevant industrial sectors) based on participants' knowledge and experience. The study found that all participants (20) from the seven focus groups were unaware of specific standards or best practices in evaluating tools and data. Though one participant mentioned two standards – ISO/IEC 270001 "Information Security Management System" and ISO/IEC:270005 "Information Security Risk Management", based on our checking, these two standards are unrelated to evaluating tools or data. One participant also mentioned "Kitemarking" (https://www.bsigroup.com/en-MY/Kitemark/Kitemark-for-products/). However, based on our checking, "Kitemarking" is frequently used to identify product safety, not to evaluate tools or data.

One participant pointed out that there is not much agreement on standards within the cyber security domain compared to other domains such as the defence sector:

> "*And I guess in terms of standards, I think the problem with a lot of stuff in the threat spaces, there isn't really a huge amount of agreement on standards out there. It's quite limited, whereas I guess in the defence space when they're talking about how do we value by evaluate cyber security or an organisation's cyber security.*" (P3, FG5)

Table 5. Challenges and Limitations in Evaluating Tools and Data in Participating National CSIRTs

Challenges	**Detail Challenges (Based on FG Data)**	**FG**
Lack of awareness & expertise	Lack of knowledge and awareness about tools and data evaluation standards or practices – staff never thought of doing the evaluation, learning new technologies – to facilitate evaluation, dependent on community reviews, knowledge of other evaluation solutions – i.e., riding (piggyback) on the evaluation done by other CSIRTs.	6
Lack of process & procedures	Do not have formal process – formal procedures, guidelines, standards, no guidance, limited standards for tools and data evaluation, no procedure for OSINT and open-source, done on ad hoc.	5
Lack of Resources	People – Lack of staff and manpower to do the evaluation (staff need to be involved in other tasks, i.e., do Vulnerability Assessment and Penetration Testing (VAPT), involved in investigations, balancing between current workloads and other tasks), operating under small team, need resources for post evaluation (i.e., paperwork – documenting the evaluation, ensure system compliance); budget constraint – to setup evaluation environment (i.e., additional servers, network connectivity) and time constraints – more time is dedicated for incident response, don't have time for proper tools and data evaluation, more time needed to handle a high volume of incident reports.	4

All 20 participants disclosed current practices in evaluating tools and data in their respective national CSIRTs, summarised in Figure 1. Most participants (14) mentioned they do not have a systematic, formal and proper process for evaluating tools and data in their respective national CSIRTs. In contrast, very few participants (two) mentioned having in-house evaluation, while one participant said the IT Department does the evaluation, and another participant said of a mixture of internal and external evaluation. This is reflected by the comments below:

> "*Let me confirm this. We also don't have a formal process for evaluating tools or data.*" (P1, FG2)
>
> "*So we don't really have time for proper, detailed process of tool evaluation.*" (P2, FG7)
>
> "*For me, I know there aren't any formal procedures like in my opinion, there aren't any, that's obviously why you were doing your research. There are no standards for using software in national CSIRTs.*" (P1, FG1)

The present study also revealed how tools and data are evaluated in the participating national CSIRTs, summarised in Table 4. Most (five) FGs said they refer to community reviews to identify qualified tools and data. Another five FGs said they self-check the tools and data for quality purposes. None of the participants mentioned having a proper procedure to evaluate tools and data in their national CSIRTs.

4.4 Challenges and Limitations in Tools and Data Evaluation – *(*RQ3)

RQ3 aims at achieving a better understanding of the challenges and limitations that impinge the evaluation of tools and data in national CSIRTs. These challenges are summarised in Table 5.

From the results, most focus groups (6) voiced challenges concerning a lack of knowledge and awareness about tools and data evaluation standards or practices. Furthermore, staff never thought of doing tool and data evaluation. Participants also voiced they need to learn new technologies to facilitate evaluation, which participants found challenging. Therefore, they depend on community reviews and knowledge of other evaluation solutions and even like riding (piggyback) on the evaluation done by other CSIRTs.

Five focus groups revealed their challenge with evaluating tools and data because they did not have a formal process and procedure to conduct the evaluations. The focus groups pointed out that currently, there are no guidelines, frameworks or standards for evaluating tools and data in place in national CSIRTs, as mentioned by a participant:

> "*There's no framework we will be following.*" (P1, FG3)

And another participant added:

> "*We also don't have a formal process for evaluating tools or data.*" (P2, FG2)

The next challenge mentioned by four focus groups is the lack of staff with small teams running the operation. Besides lacking sufficient staff, the focus groups commented on the lack of budget and time, impeding tools and data evaluation in their national CSIRTs. This is mentioned by a participant:

> "*So probably there are some number of tools out there which would be really clever and good and nice to have, but which we just failed to evaluate because of our resource constraints.*" (P2, FG2)

4.5 Candidate Criteria for Evaluating Tools and Data from Staff Members' Perspective–*(*RQ4)

One of the objectives of the current study is to identify candidate criteria for evaluating tools and data based on national CSIRTs' perspectives who participated in the focus groups. In the McCall Model [59], software quality factors were identified from the literature. Our study identified candidate criteria from the opinions of practising staff members of national CSIRTs who participated in the study– improving the previous approach.

Notably, two relevant international standards – ISO/IEC 25010:2011 [46] and ISO/IEC 25012:2008 [45] were reviewed and adopted to leverage the candidate criteria identified from the focus group study. These standards describe principles for assessing software quality [13] and data quality [91] relevant to our current study. These standards were adopted in previous studies for software evaluation [53, 66, 67, 88, 99] and data evaluation [91], perceived as the best model for software quality evaluation compared to other models [60]. Additionally, the above ISO/IEC Standards are the latest. They are not obsolete and were reviewed as valid by the ISO/IEC Standards Committee.

Additionally, we also reviewed and leveraged other Software Quality Models, e.g., McCall Quality Model [59], Boehm Software Quality Model [15], FURPS Model [37, 49] and Dromey Model [26] apart from ISO/IEC 25010:2011 '"**Systems and Software Quality Requirements and Evaluation (SQuaRE)**". We limit to the above models only as these are commonly cited from the literature [5, 23, 68, 79, 86].

We used the term "criteria," which refers to quality characteristics of a software tool and data, as defined in ISO/IEC 25010:2011 "Systems and Software Quality Requirements and Evaluation (SQuaRE)" [46].

4.5.1 Candidate Criteria for Tools. In this study, we identified eight candidate criteria for evaluating tools from participants' opinions, as shown in Table 6. We extracted the candidate criteria from the focus group data using the content analysis method with an inductive coding approach focused on manifest meaning. The candidate criteria were then cross-checked against the ISO/IEC standards to leverage standardised names and definitions for the candidate criteria.

Candidate criteria identified from the focus groups that are not named and defined in the ISO/IEC 25010:2011 were reviewed and decided by the researcher to include as candidate criteria. Those candidate criteria are contributions from the study in improving the current models and standards.

It was found that all seven FGs mentioned that *Usability* and *Maintainability* are criteria that must be included for evaluating tools. According to the focus groups, "Maintainability" concerns up-to-date, good maintenance and support for tools, the ability to fix issues with the tool and the ability to add new features. While "Usability" concerns, for example, learnability, efficiency, trust, effectiveness, accessibility and user interface.

Table 6. Candidate Criteria for the Evaluation of Tools

Corresponding ISO/IEC 25010:2011 Criteria Term	**Candidate Criteria (Based on Focus Group Data)**	**FG**
Usability	Accessibility of tool to CERT, expertise, frees up time, good documentation, good performance tool,how much to train people, interface, limit time of handling incident, solve incident in shortest time, time to produce result, usability, working more efficient result, usability, working more efficient, accurateness of the results, benefit we can get, help our organisation, helps to refine data, make job easier, make our output better, meet our expectation, provide result as close as possible, reputation of tool, tool provide value, trust of tool, trustworthiness, easiest tool, make job more efficient	7
Maintainability	Updated, well maintained, has a license, active development, has support, can audit failures, can add new capabilities, can fix a tool, active developer, active development, feature request, large community maintaining it	7
Security	Confidentiality of tool, data we input into tools, privacy handling, integrity of tool, secure feature, free from bugs, source of tool, secure, security feature, secure interaction with all components of a tool, secure development	6
Functionality	Fundamental capability, fundamental output, functionality of reporting, able to query data, can select type of data, fits for purpose	3
Compatibility	Has API integration, can be connected to other devices, fit with other tools, suitable for a data format, interoperability	4
Reliability	Reliable in producing output, availability of tool, accessibility to CSIRT community, can see the source code, continuous funding, Will be with us over the following years, false-positive rates, some less false-positive	2
Context coverage	Flexible tool, can be customised	2
Others	Compliance, globally accepted tool, certified by regulatory, product certification, used by the community, used by the security community, certified tool, approval from other countries, approval by legal departments, commonly used by the security community, how large is the community using the tool, how many people are using the tool	3

This is followed by *Security* as mentioned by six FGs, concerns with, for example, confidentiality and integrity of the tool. Four FGs perceived *Compatibility* should be a criterion when evaluating tools, for instance, integration and interoperability. *Functionality* was viewed as a candidate criterion by four FGs giving examples of having the fundamental capability and fitting the purposes of the tool. *Reliability* was perceived as a criterion by two focus groups followed by *Context Coverage* by two FGs.

We grouped several criteria identified from the study as "others" – *Compliance to security policies*, *Globally Accepted*, *Used by Community*, *Certified by Regulators and Product certification Services*. We did not find these criteria in the ISO/IEC 25010:2011. Nevertheless, we decided to include them as candidate criteria as perceived by the focus groups. "Used by Community" was expressed by participants as an important criterion for tool evaluation, scarcely mentioned in any standards or models. Notably, the following candidate criteria identified from the focus group study – *Reliability*, *Flexibility* and *Automated Report Generation* are consistent with a previous study about elements of good-quality tools [17].

Table 7. Candidate Criteria for the Evaluation of Data

ISO/IEC 25012:2008 Term	Candidate Criteria (Based on FG Data)	FG
Credibility	Data – legitimate data, reliable data, data integrity, data is secure; source of data – place we download data is trusted, quality of source of data, reliable source, reputable source, from trusted partner, useful data	6
Confidentiality	Less personal identifiable information, data comply to Privacy Act, data is on-premise, policy to protect data in cloud, privacy of data	4
Currentness	Regularly reported, data is updated	2
Understandability	Structured data, has data model	2
Completeness	Backward compatible – There is a history (at least two years) of the same data, contains risk, how important is data, data provides value, data has threat level	2
Precision	Has information for correlation, data that can be acted, data impacting a region or economy	2
Accuracy	Fewer false positives	1
Efficiency	Machine-readable, Can do right thing	2

4.5.2 Candidate Criteria for Data. Eight candidate criteria on data quality characteristics based on participants' opinions were identified and summarised in Table 7. These criteria identified from the focus group discussions are aligned with the criteria defined in the international standard ISO/IEC 25012:2008 "Software product Quality Requirements and Evaluation (SQuaRE) – Data Quality Model" [45]. The same approach we used to identify candidate criteria for tool evaluation was used in identifying candidate criteria for data evaluation.

Most focus groups (6) perceived *Credibility* as a criterion for evaluating data. Participants viewed reliability, integrity, legitimacy and security of data must be followed when evaluating data for incident response. Four focus groups are of the opinion that *Confidentiality* must be a candidate criterion for evaluating data. This includes data complying with privacy acts, and personal information in the data is limited. Other candidate criteria expressed by the focus groups (2 FGs for each criterion) are *Precision* that allows data to be acted upon, *Completeness* related to such as having historical information in the data, *Currentness* related to timeliness of the data and *Understandability* of the data. *Efficiency* and *Accuracy* (one FG for each criterion) were perceived as candidate criteria for evaluating data. The candidate criteria identified from the study are also consistent with previous studies [91].

It should be noted that the focus group study identified candidate criteria for evaluating tools and data based on the opinions of participating national CSIRTs' staff members, who have operational experiences and knowledge about tools and data.

The results of the focus group study are discussed in the next section.

5 DISCUSSIONS

This section discusses our key findings of the study's results, study limitations and future work. The key findings are discussed according to the study's **research questions (RQ)**.

5.1 Key Findings

5.1.1 RQ1: Definitions of Tools and Data. It is clear from the results that tools and data are defined pragmatically by participants based on their knowledge and experience as staff members of national CSIRTs. This gives the impression that the participants have well understanding and knowledge of the tools and data they use daily to facilitate incident responses. The results also show that tools and data depend on and complement each other, playing vital roles in the operations of national CSIRTs. Interestingly, the definition of data identified from the study is consistent with how the literature defines data, though, in principle, the definitions can be quite diverse [100]. Moreover, all national CSIRTs participated in the study had a common understanding of what constitutes tools and data. Such a common understanding is crucial for timely threat information sharing, tools and data exchanges among national CSIRTs, which could help improve overall incident responses.

5.1.2 RQ2: Standards and Practices for Evaluation of Tools and Data. A lack of knowledge about standards and best practices for evaluating tools and data among the focus groups indicates: (1) the ad hoc practices of tools and data evaluations in the participating national CSIRTs, and (2) a lack of systematic procedures for evaluating tools and data in the participating national CSIRTs. These are consistent with findings from previous studies [61, 62].

Findings from the focus group discussions give the impression that participants are positive towards systematically evaluating tools and data. Notably, other national CSIRTs also desire to develop an approach for evaluating tools and data by getting collective opinions from other national CSIRTs, indicating the need and importance of the research undertaken on this topic. A participant revealed this:

> "*When it comes to evaluating, I'm not sure how that is organised in NCSC-XX, but I know that we are still desiring for a more international approach to evaluating the tools that we use in the community and also to get ideas from other countries and other national CSIRTs and to share our best practices.*" (P2, FG1)

Findings from the study are also consistent with what was reported in a recent study [14], which stressed the need to evaluate tools, particularly open-source tools, to ensure the sustainability and success of national CSIRTs' operations. This calls for more research to establish systematic procedures to evaluate tools and data for national CSIRTs, benefiting the broader security community.

5.1.3 RQ3: Challenges and Limitations in Tools and Data Evaluation. There is also a general agreement among all focus groups about the challenges in tools and data evaluation in their national CSIRTs. Among all the challenges, one is particularly important: *lack of formal process and procedures for evaluating tools and data* – which relates to what we are doing now and potential future work. The challenges voiced by the participants are important as they give the impression that establishing a systematic procedure for evaluating tools and data is beyond national CSIRTs' capacity – due to a lack of expertise, staffing, resources and budget issues. This tells why research work by independent researchers like us is necessary to improve the practices in national CSIRTs – one such work is presented in this study. This also indicates an urgent need to develop more cyber security-related procedures by stakeholders and researchers. The focus groups' eagerness to know the outcome of our study also implies how national CSIRTs are receptive to systematic procedures for operations to improve practices, reflected in some candid comments by participants.

5.1.4 RQ4: Candidate Criteria for Evaluating Tools and Data from Staff Members' Perspective. The feedback and discussion among the focus groups on selecting the best candidate criteria for evaluating tools and data give the impression that national CSIRTs are receptive towards working together with researchers to develop systematic approaches and procedures, such as a set of criteria for evaluating tools and data. Such positiveness is important for this study in identifying candidate criteria through interactive discussions among the focus groups. This could help national CSIRTs to make good evaluations of tools and data in the operations. Notably, some candidate criteria from the study could enhance the current tool and data evaluation standards, e.g., ISO/IEC SQuaRe Standards.

Though the candidate criteria reported in this study may not be complete, they can serve as a foundation for further refinements and extensions in future work for more firm and concrete criteria. This is further described in 5.3.

5.2 Limitations

Several limitations were encountered while conducting the study. One limitation is the participants' reluctance to provide details of their operational practices due to privacy and confidentiality concerns. A similar limitation was reported in a previous study as well [62]. This might limit the completeness of the study's information and findings. However, the limitation can be mitigated by considering the fact that the main findings are consistent with the first author's personal experience as an employee of a national CSIRT.

Regarding the limitation of the number of participants, it should be noted that for a very closed community like national CSIRTs, recruiting 20 willing and qualified participants from 15 national CSIRTs was not easy, so this could be considered a success. Nevertheless, we acknowledge the limitation and suggest conducting follow-up work with more participants from a wider range of national CSIRTs.

Another limitation of the study is the quality of the focus group data. Focus group discussions held later might yield better quality data than those held earlier, as the first author improved her skill in moderating the focus groups. However, this limitation can be mitigated by the fact that the quality of the focus group data is incremental. Notably, this study focused only on free tools, open-source tools and public data, not closed-source and commercial tools and data. This is a limitation of the study, but future work could address it by expanding the study to include closed-source and commercial tools and data.

5.3 Future Work

First, future work can be done to address some of the limitations discussed in the previous subsection, e.g., on a follow-up study with more participants from a wider range of national CSIRTs. It is also useful to consider re-running the focus group discussions with different settings. For instance, national CSIRTs may be put into focus groups based on their different types, and the results can be compared to identify any differences.

Findings from this study indicate a critical need to conduct more research to help national CSIRTs (and the wider CSIRT community) develop more standardised procedures for systematically evaluating tools and data. As such, we call on the cyber security community to conduct further research to cross-validate our study findings and apply these findings in a wider cyber security context to improve current practices.

Therefore, one of the future works from this study would be to conduct semi-structured interviews to get opinions from staff members of national CSIRTs on how they perceive the usefulness and deployment of the candidate criteria identified from this study. The opinions and feedback from the interviews could help refine the candidate criteria into more concrete ones. Another future work would be to validate the candidate criteria by applying the criteria to evaluate sample tools and data sources using a number of metrics. Doing so would prove the applicability and feasibility of applying the candidate criteria in practice. This supplements the feedback from the semi-structured interviews and makes the candidate criteria more credible and reliable. This follow-up work can also consolidate the results from the focus groups reported in this paper and verify their generalisability.

Once validated in the above-mentioned future study, we plan to release the candidate criteria as an open resource and recommend them to national CSIRTs as useful references via cross-national-CSIRT channels and platforms such as FIRST and ENISA. We will also release the evaluation results of the sample tool and data source in attachments to the candidate criteria as case studies. Hence, national CSIRTs have more concrete examples or case studies to guide their pilot and deployment of the candidate criteria to evaluate tools and data sources.

Another interesting future research direction is to construct a taxonomy or an ontology that will connect the criteria, different types of tools and data sources used by (national and non-national) CSIRTs, and concrete metrics that map the criteria to characteristics of different tools and data sources. Such a taxonomy or an ontology will help inform the development of more practical operational guidelines and potentially enable partial automation of the tool and data evaluation process.

6 CONCLUSION

This paper presented seven online focus group discussions comprising 20 staff members from 15 national CSIRTs worldwide. The study gained rich insights into the current practices and challenges with tools and data evaluations in the operations. The study also identified candidate criteria for evaluating tools and data through interactive conversations and discussions. It should be noted the topic presented in the study is less studied in the current cyber security research literature. Still, there is a growing need to address the issues and challenges highlighted in this topic.

This study contributes four key findings. First, the study found a lack of staff awareness and fundamental knowledge regarding standards and practices in tools and data evaluation. Second, there is a need to establish criteria for a systematic evaluation of tools and data evaluation criteria for national CSIRTs. Third, the study identified a set of candidate criteria for evaluating tools and data. Fourth, the study revealed several challenges that impinge tools and data evaluation across the participating national CSIRTs, particularly the lack of procedures, guidelines, resources (staff, budget, time), skills and expertise. In summary, it is important for national CSIRTs to understand how to make a good evaluation of tools and data. Hence, the candidate criteria identified from the study can be a starting point to customise and enhance the evaluation of tools and data in national CSIRTs.

REFERENCES

[1] Philip Adu. 2019. *A Step-by-Step Guide to Qualitative Data Coding*. Routledge.

[2] Atif Ahmad, Justin Hadgkiss, and Anthonie B. Ruighaver. 2012. Incident response teams – Challenges in supporting the organisational security function. *Computers & Security* 31, 5 (2012), 643–652. https://doi.org/10.1016/j.cose.2012.04.001

[3] Rabiah Ahmad, Zahri Yunos, Shahrin Sahib, and Mariana Yusoff. 2012. Perception on cyber terrorism: A focus group discussion approach. *Journal of Information Security* 3 (2012), 231–237. https://doi.org/10.4236/jis.2012.33029

[4] Rahayu Azlina Ahmad and Mohd Shamir Hashim. 2011. The organisation of Islamic conference–computer emergency response Team (OIC-CERT): Answering cross border cooperation. In *Proceedings of the 2011 2nd Worldwide Cybersecurity Summit*. IEEE, 5. https://ieeexplore.ieee.org/document/5978783

[5] Rafa E. Al-Qutaish. 2010. Quality models in software engineering literature: An analytical and comparative study. *Journal of American Science* 6, 3 (2010), 166–175. http://www.jofamericanscience.org/journals/am-sci/am0603/22_2208_Qutaish_am0603_166_175.pdf

[6] APCERT. 2021. *APCERT Annual Report 2021*. Annual Report. APCERT. http://www.apcert.org/documents/pdf/APCERT_Annual_Report_2021.pdf

[7] William A. Arbaugh, William L. Fithen, and John McHugh. 2000. Windows of vulnerability: A case study analysis. *Computer* 33, 12 (2000), 52–59. https://doi.org/10.1109/2.889093

[8] Gayane Azizyan, Miganoush Katrin Magarian, and Mira Kajko-Matsson. 2011. Survey of agile tool usage and needs. In *Proceedings of the 2011 Agile Conference*. IEEE, 29–38. https://doi.org/10.1109/AGILE.2011.30

[9] Rosaline Barbour. 2008. *Doing Focus Groups*. SAGE. https://doi.org/10.4135/9781849208956

[10] Nicole Basaraba. 2021. A bottom-up method for remixing narratives for virtual heritage experiences. *Convergence: The International Journal of Research into New Media Technologies* 28 (2021). Issue 6. https://doi.org/10.1177/13548565211048

[11] Ivy Belinda Botchway, Akinwonmi Akintoba Emmanuel, Nunoo Solomon, and Alese Boniface Kayode. 2021. Evaluating software quality attributes using analytic hierarchy process (AHP). *International Journal of Advanced Computer Science and Applications* 12, 3 (2021). DOI: https://doi.org/10.14569/IJACSA.2021.0120321

[12] Mariette Bengtsson. 2016. How to plan and perform a qualitative study using content analysis. *NursingPlus Open* 2 (2016), 8–14. https://doi.org/10.1016/j.npls.2016.01.001

[13] Nigel Bevan. 2001. International standards for HCI and usability. *International Journal of Human-Computer Studies* 55, 4 (2001), 533–552. https://doi.org/10.1006/ijhc.2001.0483

[14] Tracy Bills, Brittany Manley, and James Lord. 2022. *Enabling the Sustainability and Success of a National Computer Security Incident Response Team*. Technical Report. Carnegie Mellon University. https://resources.sei.cmu.edu/asset_files/Handbook/2022_002_001_885865.pdf

[15] Barry W. Boehm, John R. Brown, and Mlity Lipow. 1976. Quantitative evaluation of software quality. In *Proceedings of the 2nd International Conference on Software Engineering*. ACM, 592–605. https://doi.org/10.5555/800253.807736

[16] Germinal Boloix and Pierre N. Robillard. 1995. A software system evaluation framework. *Computer* 28, 12 (1995), 17–26. https://doi.org/10.1109/2.476196

[17] David Botta, Rodrigo Werlinger, André Gagné, Konstantin Beznosov, Lee Iverson, Sidney Fels, and Brian Fisher. 2007. Towards understanding IT security professionals and their tools. In *Proceedings of the 3rd Symposium on Usable Privacy and Security*. ACM, 100–111. https://doi.org/10.1145/1280680.1280693

[18] Xander Bouwman, Victor Le Pochat, Pawel Foremski, Tom Van Goethem, Carlos H. Gañán, Giovane C. M. Moura, Samaneh Tajalizadehkhoob, Wouter Joosen, and Michel Van Eeten. 2022. Helping hands: Measuring the impact of a large threat intelligence sharing community. In *31st USENIX Security Symposium (USENIX Security 22)*. USENIX Association, 1149–1165.

[19] Virginia Braun and Victoria Clarke. 2006. Using thematic analysis in psychology. *Qualitative Research in Psychology* 3, 2 (2006), 77–101. https://doi.org/10.1191/1478088706qp063oa

[20] Stephen Cavanagh. 1997. Content analysis: Concepts, methods and applications. *Nurse Researcher* 4, 3 (1997), 5–16. https://doi.org/10.7748/nr.4.3.5.s2

[21] Felipe Sabino Costa. 2021. *A Practical Approach to Adopting the IEC 62443 Standards*. White Paper. MOXA, Inc. https://www.moxa.com/en/literature-library/moxa-a-practical-approach-to-adopting-the-iec-62443-standards-white-paper

[22] John W. Creswell and Dana L. Miller. 2000. Determining validity in qualitative inquiry. *Theory Into Practice* 39, 3 (2000), 124–130. https://doi.org/10.1207/s15430421tip3903_2

[23] Tharashasank Davuluru, Jayapal Medida, and V. S. K. Reddy. 2014. A study of software quality models. In *Proceedings of the 2014 International Conference on Advances in Engineering & Technology Research*. IEEE, 8. https://doi.org/10.1109/ICAETR.2014.7012958

[24] Stephanie de Smale, Rik van Dijk, Xander Bouwman, Jeroen van der Ham, and Michel van Eeten. 2023. No one drinks from the firehose: How organizations filter and prioritize vulnerability information. In *Proceedings of the 2023 IEEE Symposium on Security and Privacy*. IEEE, 203–219. https://doi.org/10.1109/SP46215.2023.00012

[25] Tom Deliens, Peter Clarys, Ilse De Bourdeaudhuij, and Benedicte Deforche. 2014. Determinants of eating behaviour in university students: A qualitative study using focus group discussions. *BMC Public Health* 14, 1 (2014), 12. https://doi.org/10.1186/1471-2458-14-53

[26] R. Geoff Dromey. 1995. A model for software product quality. *IEEE Transactions on Software Engineering* 21, 2 (1995), 146–162. https://doi.org/10.1109/32.345830

[27] Hanneke Duijnhoven, Tom van Schie, and Don Stikvoort. 2021. *Stimulating the Development and Maturity Enhancement of National CSIRTs*. TNO Publication. https://repository.tno.nl//islandora/object/uuid:e1ba969e-7ab4-4bd5-83fb-7c029db15265

[28] Christen Erlingsson and Petra Brysiewicz. 2017. A hands-on guide to doing content analysis. *African Journal of Emergency Medicine* 7, 3 (2017), 93–99. https://doi.org/10.1016/j.afjem.2017.08.001

[29] Fernando Vela Espín. 2020. Guidelines and their challenges in implementing CSIRT in Ecuador. In *Advances in Emerging Trends and Technologies: Proceedings of ICAETT 2020*. Springer, 239–251. https://doi.org/10.1007/978-3-030-63665-4_19

[30] European Parliament and European Council. 2016. Directive (EU) 2016/1148 of the European Parliament and of the Council of 6 July 2016 Concerning Measures for a High Common Level of Security of Network and Information Systems Across the Union. EU legislation. (2016). https://eur-lex.europa.eu/eli/dir/2016/1148/oj

[31] Wenfei Fan and Floris Geerts. 2012. *Foundations of Data Quality Management*. Springer. https://doi.org/10.1007/978-3-031-01892-3

[32] Rachel Flynn, Lauren Albrecht, and Shannon D. Scott. 2018. Two approaches to focus group data collection for qualitative health research: Maximizing resources and data quality. *International Journal of Qualitative Methods* 17, 1 (2018), 1–9. https://doi.org/10.1177/1609406917750781

[33] Fiona E. Fox, Marianne Morris, and Nichola Rumsey. 2007. Doing synchronous online focus groups with young people: Methodological reflections. *Qualitative Health Research* 17, 4 (2007), 539–547. https://doi.org/10.1177/1049732306298754

[34] Dhananjay Gade. 2013. The Evaluation of Software Quality. (2013). https://digitalcommons.unl.edu/imsediss/38/

[35] Anita Gibbs. 1997. Focus groups. *Social Research Update* 19, 8 (1997), 8. https://sru.soc.surrey.ac.uk/SRU19.html

[36] Lisa M. Given. 2008. *The Sage Encyclopedia of Qualitative Research Methods*. SAGE.

[37] Robert B. Grady. 1992. *Practical Software Metrics for Project Management and Process Improvement*. Prentice-Hall, Inc.

[38] Glenn Gumba, Deborah G. Brosas, and Jessie R. Paragas. 2021. Assessment of SIAS application using software quality model. In *Proceedings of the 2021 3rd International Conference on Research and Academic Community Services*. IEEE, 197–202. https://doi.org/10.1109/ICRACOS53680.2021.9701982

[39] Elizabeth J. Halcomb, Leila Gholizadeh, Michelle DiGiacomo, Jane Phillips, and Patricia M. Davidson. 2007. Literature review: Considerations in undertaking focus group research with culturally and linguistically diverse groups. *Journal of Clinical Nursing* 16, 6 (2007), 1000–1011. https://doi.org/10.1111/j.1365-2702.2006.01760.x

[40] Irene A. Harmsen, Liesbeth Mollema, Robert A. C. Ruiter, Theo G. W. Paulussen, Hester E. de Melker, and Gerjo Kok. 2013. Why parents refuse childhood vaccination: A qualitative study using online focus groups. *BMC Public Health* 13, 1 (2013), 8. https://doi.org/10.1186/1471-2458-13-1183

[41] Hsiu-Fang Hsieh and Sarah E. Shannon. 2005. Three approaches to qualitative content analysis. *Qualitative Health Research* 15, 9 (2005), 1277–1288. https://doi.org/10.1177/1049732305276687

[42] Jan Huck and Frank Breitinger. 2022. Wake up digital forensics' community and help combating ransomware. *IEEE Security & Privacy* 20, 4 (2022), 61–70. https://doi.org/10.1109/MSEC.2021.3137018

[43] Sungsoo Hwang. 2008. Utilizing qualitative data analysis software: A review of Atlas.ti. *Social Science Computer Review* 26, 4 (2008), 519–527. https://doi.org/10.1177/0894439307312485

[44] George Iakovakis, Constantinos Giovanni Xarhoulacos, Konstantinos Giovas, and Dimitris Gritzalis. 2021. Analysis and classification of mitigation tools against cyberattacks in COVID-19 Era. *Security and Communication Networks* 2021, Article 3187205 (2021), 21 pages. https://doi.org/10.1155/2021/3187205

[45] International Organization for Standardization (ISO). 2008. Software Engineering – Software Product Quality Requirements and Evaluation (SQuaRE) – Data Quality Model. web page. (2008). https://www.iso.org/standard/35736.html

[46] International Organization for Standardization (ISO). 2011. Systems and Software Engineering – Systems and Software Quality Requirements and Evaluation (SQuaRE) – System and Software Quality Models. web page. (2011). https://www.iso.org/standard/35733.html

[47] International Telecommunication Union (ITU). National CIRT. Web page. (n.d.). https://www.itu.int/en/ITU-D/Cybersecurity/Pages/national-CIRT.aspx

[48] Internet Governance Forum (IGF). 2014. Best Practice Forum on Establishing and Supporting Computer Security Incident Response Teams (CSIRT) for Internet Security. Online document. Retrieved from https://www.first.org/global/governance/bpf-csirt-2014-outcome.pdf

[49] Ivar Jacobson. 1999. *The Unified Software Development Process*. Pearson Education India.

[50] Yujin Kim. 2011. The pilot study in qualitative inquiry: Identifying issues and learning lessons for culturally competent research. *Qualitative Social Work* 10, 2 (2011), 190–206. https://doi.org/10.1177/1473325010362001

[51] Jenny Kitzinger. 1995. Qualitative research: Introducing focus groups. *BMJ* 311, 7000 (1995), 299–302. https://doi.org/10.1136/bmj.311.7000.299

[52] Rick van der Kleij, Geert Kleinhuis, and Heather Young. 2017. Computer security incident response team effectiveness: A needs assessment. *Frontiers in Psychology* (2017). https://doi.org/10.3389/fpsyg.2017.02179

[53] Toshihiro Komiyama, Shin'ichi Fukuzumi, Motoei Azuma, Hironori Washizaki, and Naohiko Tsuda. 2020. Usability of software–intensive systems from developers' point of view. In *Human-Computer Interaction. Design and User Experience: Thematic Area, HCI 2020, Held as Part of the 22nd International Conference, HCII 2020, Copenhagen, Denmark, July 19–24, 2020, Proceedings, Part I*. Springer, 450–463.

[54] Erik B. Korn, Douglas M. Fletcher, Erica M. Mitchell, Aryn A. Pyke, and Steven M. Whitham. 2021. Jack pandemus – cyber incident and emergency response during a pandemic. *Information Security Journal: A Global Perspective* 30, 5 (2021), 294–307. https://doi.org/10.1080/19393555.2021.1980159

[55] Marko Krstic, Milan Cabarkapa, and Aleksandar Jevremovic. 2019. Machine learning applications in computer emergency response team operations. In *Proceedings of the 27th Telecommunications Forum*. IEEE, 4. https://doi.org/10.1109/TELFOR48224.2019.8971040

[56] Richard A. Krueger. 2014. *Focus Groups: A Practical Guide for Applied Research*. SAGE.

[57] Marc Kührer, Christian Rossow, and Thorsten Holz. 2014. Paint it black: Evaluating the effectiveness of malware blacklists. In *Research in Attacks, Intrusions and Defenses: 17th International Symposium, RAID 2014, Gothenburg, Sweden, September 17–19, 2014, Proceedings*. Springer, 1–21. https://doi.org/10.1007/978-3-319-11379-1_1

[58] R. Barry Lewis. 2004. NVivo 2.0 and ATLAS.ti 5.0: A comparative review of two popular qualitative data-analysis programs. *Field Methods* 16, 4 (2004), 439–464. https://doi.org/10.1177/1525822X04269174

[59] Jim A. McCall, Paul K. Richards, and Gene F. Walters. 1977. *Factors in Software Quality. Volume I. Concepts and Definitions of Software Quality*. Technical Report. General Electric Co. https://apps.dtic.mil/sti/citations/ADA049014

[60] José P. Miguel, David Mauricio, and Glen Rodríguez. 2014. A review of software quality models for the evaluation of software products. *International Journal of Software Engineering & Applications* 5, 6 (2014), 31–54. https://doi.org/10.5121/ijsea.2014.5603

[61] Sharifah Roziah Binti Mohd Kassim, Solahuddin Bin Shamsuddin, Shujun Li, and Budi Arief. 2022. How national CSIRTs operate: Personal observations and opinions from MyCERT. In *Proceedings of the 2022 IEEE Conference on Dependable and Secure Computing*. IEEE, 2. https://doi.org/10.1109/DSC54232.2022.9888803

[62] Sharifah Roziah Binti Mohd Kassim, Shujun Li, and Budi Arief. 2022. How national CSIRTs leverage public data, OSINT and free tools in operational practices: An empirical study. *Cyber Security: A Peer-Reviewed Journal* 5, 3 (2022), 251–276. https://www.ingentaconnect.com/contentone/hsp/jcs/2022/00000005/00000003/art00007

[63] Sharifah Roziah Binti Mohd Kassim, Shujun Li, and Budi Arief. 2023. The use of public data and free tools in national CSIRTs' operational practices: A systematic literature review. *arXiv preprint arXiv:2306.07988v1* (2023).

[64] Francesca Moretti, Liesbeth van Vliet, Jozien Bensing, Giuseppe Deledda, Mariangela Mazzi, Michela Rimondini, Christa Zimmermann, and Ian Fletcher. 2011. A standardized approach to qualitative content analysis of focus group discussions from different countries. *Patient Education and Counseling* 82, 3 (2011), 420–428. https://doi.org/10.1016/j.pec.2011.01.005

[65] David L. Morgan. 1996. *Focus Groups as Qualitative Research*. SAGE. https://doi.org/10.4135/9781412984287

[66] Radka Nacheva and Anita Jansone. 2020. Evaluation of business process modelling tools through software quality metrics. *Baltic Journal of Modern Computing* 8, 4 (2020), 534–542. https://doi.org/10.22364/bjmc.2020.8.4.04

[67] Hidenori Nakai, Naohiko Tsuda, Kiyoshi Honda, Hironori Washizaki, and Yoshiaki Fukazawa. 2016. A SQuaRE-based software quality evaluation framework and its case study. In *Proceedings of the 2016 IEEE Region 10 Conference*. IEEE, 3704–3707. https://doi.org/10.1109/TENCON.2016.7848750

[68] Padmalata Nistala, Kesav Vithal Nori, and Raghu Reddy. 2019. Software quality models: A systematic mapping study. In *Proceedings of the 2019 IEEE/ACM International Conference on Software and System Processes*. IEEE, 125–134. https://doi.org/10.1109/ICSSP.2019.00025

[69] Monika Nowikowska. 2022. The main tasks of the network of computer security incident response teams in the light of the act on the national cybersecurity system in Poland. In *Cybersecurity in Poland*. Springer, 223–241. https://doi.org/10.1007/978-3-030-78551-2_15

[70] Clifford O. Odimegwu. 2000. Methodological issues in the use of focus group discussion as a data collection tool. *Journal of Social Sciences* 4, 2-3 (2000), 207–212. https://doi.org/10.1080/09718923.2000.11892269

[71] Anthony J. Onwuegbuzie, Wendy B. Dickinson, Nancy L. Leech, and Annmarie G. Zoran. 2009. A qualitative framework for collecting and analyzing data in focus group research. *International Journal of Qualitative Methods* 8, 3 (2009), 1–21. https://doi.org/10.1177/160940690900800301

[72] Paweł Pawlinski and Andrew Kompanek. 2016. Evaluating Threat Intelligence Feeds. Slides presented at the 2016 FIRST Technical Colloquium for Threat Intelligence. (2016). https://www.first.org/resources/papers/munich2016/kompanek-pawlinski-evaluating-threat-ntelligence-feeds.pdf

[73] PricewaterhouseCoopers. 2021. *Cyber Threats 2021: A Year in Retrospect*. Technical Report. PricewaterhouseCoopers. https://www.pwc.com/gx/en/issues/cybersecurity/cyber-threat-intelligence/cyber-year-in-retrospect/yir-cyber-threats-report-download.pdf

[74] Roger J. Rezabek. 2000. Online focus groups: Electronic discussions for research. In *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research*, Vol. 1. https://doi.org/10.17169/fqs-1.1.1128

[75] Thea Riebe, Marc-André Kaufhold, and Christian Reuter. 2021. The impact of organizational structure and technology use on collaborative practices in computer emergency response teams: An empirical study. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW2, Article 478 (2021), 30 pages. https://doi.org/10.1145/3479865

[76] Robin Ruefle, Audrey Dorofee, David Mundie, Allen D. Householder, Michael Murray, and Samuel J. Perl. 2014. Computer security incident response team development and evolution. *IEEE Security & Privacy* 12, 5 (2014), 16–26. https://doi.org/10.1109/MSP.2014.89

[77] Jo Samanta and Ash Samanta. 2018. *A Focus Group Content Analysis Study Exploring Cultural and Faith Based Values at End of Life*. SAGE. https://doi.org/10.4135/9781526439239

[78] Helen Sampson. 2004. Navigating the waves: The usefulness of a pilot in qualitative research. *Qualitative Research* 4, 3 (2004), 383–402. https://doi.org/10.1177/1468794104047236

[79] Brijendra Singh and Suresh Prasad Kannojia. 2013. A review on software quality models. In *Proceedings of the 2013 International Conference on Communication Systems and Network Technologies*. IEEE, 801–806. https://doi.org/10.1109/CSNT.2013.171

[80] Software Engineering Institute, Carnegie Mellon University. National Computer Security Incident Response Teams (CSIRTs). web page. (n.d.). https://www.sei.cmu.edu/our-work/cybersecurity-center-development/national-csirts/

[81] Jonathan M. Spring and Phyllis Illari. 2021. Review of human decision-making during computer security incident analysis. *Digital Threats: Research and Practice* 2, 2, Article 11 (2021), 47 pages. https://doi.org/10.1145/3427787

[82] David W. Stewart and Prem Shamdasani. 2017. Online focus groups. *Journal of Advertising* 46, 1 (2017), 48–60. https://doi.org/10.1080/00913367.2016.1252288

[83] Kate Stewart and Matthew Williams. 2005. Researching online populations: The use of online focus groups for social research. *Qualitative Research* 5, 4 (2005), 395–416. https://doi.org/10.1177/1468794105056916

[84] Borce Stojkovski, Gabriele Lenzini, Vincent Koenig, and Salvador Rivas. 2021. What's in a cyber threat intelligence sharing platform? A mixed-methods user experience investigation of MISP. In *Proceedings of the 2021 Annual Computer Security Applications Conference*. ACM, 385–398. https://doi.org/10.1145/3485832.3488030

[85] Sabah Suhail, Sherali Zeadally, Raja Jurdak, Rasheed Hussain, Raimundas Matulevičius, and Davor Svetinovic. 2022. Security Attacks and Solutions for Digital Twins. arXiv:2202.12501 [cs.CR]. (2022), 8 pages. https://doi.org/10.48550/arXiv.2202.12501

[86] Suman and Manoj Wadhwa. 2014. A comparative study of software quality models. *International Journal of Computer Science and Information Technologies* 5, 4 (2014), 5634–5638. https://ijcsit.com/docs/Volume%205/vol5issue04/ijcsit20140504177.pdf

[87] Renata Tesch. 2013. *Qualitative Research: Analysis Types and Software*. Routledge.

[88] Naohiko Tsuda, Hironori Washizaki, Kiyoshi Honda, Hidenori Nakai, Yoshiaki Fukazawa, Motoei Azuma, Toshihiro Komiyama, Tadashi Nakano, Hirotsugu Suzuki, Sumie Morita, Katsue Kojima, and Akiyoshi Hando. 2019. WSQF: Comprehensive software quality evaluation framework and benchmark based on SQuaRE. In *Proceedings of the 2019 IEEE/ACM 41st International Conference on Software Engineering: Software Engineering in Practice*. IEEE, 312–321. https://doi.org/10.1109/ICSE-SEIP.2019.00045

[89] Verizon. 2022. *2022 Data Breach Investigations Report*. Technical Report. Verizon. https://www.verizon.com/business/resources/reports/dbir/

[90] Antonio Vetrò, Lorenzo Canova, Marco Torchiano, Camilo Orozco Minotas, Raimondo Iemma, and Federico Morando. 2016. Open data quality measurement framework: Definition and application to open government data. *Government Information Quarterly* 33, 2 (2016), 325–337. https://doi.org/10.1016/j.giq.2016.02.001

[91] Václav Vostrovskỳ, Jan Tyrychtr, and Roman Kvasnička. 2020. Open data quality management based on ISO/IEC SQuaRE series standards in intelligent systems. In *Applied Informatics and Cybernetics in Intelligent Systems: Proceedings of the 9th Computer Science On-line Conference 2020, Volume 3*. Springer, 625–631. https://doi.org/10.1007/978-3-030-51974-2_58

[92] William A. Ward Jr. and Buvaneswari Venkataraman. 1999. Some observations on software quality. In *Proceedings of the 37th Annual Southeast Regional Conference*. ACM, 2–9. https://doi.org/10.1145/306363.306367

[93] Moira J. West-Brown, Don Stikvoort, Klaus-Peter Kossakowski, Georgia Killcrece, Robin Ruefle, and Mark Zajicek. 2003. *Handbook for Computer Security Incident Response Teams (CSIRTs)* (2nd ed.). Technical Report CMU/SEI-2003-HB-002. Software Engineering Institute, Carnegie Mellon University, Pittsburg, PA, USA. https://resources.sei.cmu.edu/asset_files/Handbook/2003_002_001_14102.pdf

[94] Marilyn Domas White and Emily E. Marsh. 2006. Content analysis: A flexible methodology. *Library Trends* 55, 1 (2006), 22–45. https://doi.org/10.1353/lib.2006.0053

[95] World Health Organisation WHO. 2023. Coronavirus Diseases (COVID-19) Pandemic. Web page. (2023). https://www.who.int/europe/emergencies/situations/covid-19

[96] Johannes Wiik, Jose J. Gonzalez, and Klaus Peter Kossakowski. 2006. Effectiveness of proactive CSIRT services. In *Proceedings of the 18th Annual FIRST Conference*. FIRST, 13. https://www.first.org/conference/2006/papers/kossakowski-klaus-papers.pdf

[97] Dan Wu, Hao Xu, Wang Yongyi, and Huining Zhu. 2022. Quality of government health data in COVID-19: Definition and testing of an open government health data quality evaluation framework. *Library Hi Tech* 40, 2 (2022), 516–534. https://doi.org/10.1108/LHT-04-2021-0126

[98] Zahri Yunos, Ramona Susanty Ab Hamid, and Mustaffa Ahmad. 2016. Development of a cyber security awareness strategy using focus group discussion. In *Proceedings of the 2016 SAI Computing Conference*. IEEE, 1063–1067. https://doi.org/10.1109/SAI.2016.7556109

[99] Mohammad Zarour. 2020. A rigorous user needs experience evaluation method based on software quality standards. *Telkomnika Journal* 18, 5 (2020). https://doi.org/10.12928/TELKOMNIKA.v18i5.16061

[100] Chaim Zins. 2007. Conceptual approaches for defining data, information, and knowledge. *Journal of the American Society for Information Science and Technology* 58, 4 (2007), 479–493. https://doi.org/10.1002/asi.20508

Received 11 November 2022; revised 16 June 2023; accepted 29 June 2023