The "Big Beast to Tackle": Practices in Quality Assurance for Cyber Threat Intelligence

Thomas Geras*
HM Munich University of Applied Sciences
Munich, Germany
thomas.geras@hm.edu

Thomas Schreck*
HM Munich University of Applied Sciences
Munich, Germany
thomas.schreck@hm.edu

ABSTRACT

The quality of Cyber Threat Intelligence (CTI) has a profound impact on the efficacy of an organization's defense against cyber threats, directly influencing its ability to safeguard critical assets and sensitive data. Despite its critical importance, the domain of CTI quality remains a multifaceted and evolving field, often operating at the intersection of theory and practice. Many organizations recognize the need for high-quality intelligence but may struggle to establish systematic processes for assessing and enhancing its quality.

To investigate these issues, our research, encompassing 25 interviews with experts in the field, enriches the understanding of CTI quality in the real world, contributing valuable insights for practitioners and organizations striving to fortify their cybersecurity defenses and information-sharing practices. By bridging the gap between theory and practice, this work aims to inform and inspire advancements in CTI quality measurement.

CCS CONCEPTS

• **Security and privacy → Intrusion/anomaly detection and malware mitigation**.

KEYWORDS

cyber threat intelligence; information sharing; sharing communities; quality assurance

ACM Reference Format:
Thomas Geras and Thomas Schreck. 2024. The "Big Beast to Tackle": Practices in Quality Assurance for Cyber Threat Intelligence. In *The 27th International Symposium on Research in Attacks, Intrusions and Defenses (RAID 2024), September 30–October 02, 2024, Padua, Italy*. ACM, New York, NY, USA, 16 pages. https://doi.org/10.1145/3678890.3678903

1 INTRODUCTION

In our digitally interconnected world, the ever-increasing sophistication of threats presents a significant challenge for organizations of different sizes across the globe. In the ongoing battle against adversaries, CTI has emerged as a vital weapon in the arsenal of security professionals [2]. CTI is defined by NIST as "threat information that has been aggregated, transformed, analyzed, interpreted, or enriched to provide the necessary context for decision-making processes" [12]. CTI equips us with knowledge, insights, and proactive strategies needed to defend against and mitigate cyber threats. With the increasing complexity and sophistication of cyber threats, the need for reliable CTI continues to grow. Therefore, the need for rigorous quality measurements or quality-enhancing procedures are becoming more and more apparent. Examples of threat information that can be collected, analyzed, and processed into CTI include Indicators of Compromise (IoCs) such as malicious IP addresses, URLs, and hashes of malware, which are useful for detecting, analyzing, and responding to potential threats; tactics, techniques, and procedures (TTPs) associated with specific cybercriminal groups; and threat intelligence reports designed for human analysts to aid in threat assessment, incident response, and strategic planning. This information is then contextualized to assess its relevance and potential impact on an organization to develop and implement targeted defensive measures.

As organizations and entities adapt to the evolving threat landscape, with the proliferation of CTI sources and a growing emphasis on information sharing, the question of how to consistently uphold and enhance the quality of CTI remain an open and paramount issue. Organizations use diverse sources to gain access to CTI, such as public sources, information-sharing communities or commercial vendors [30][32][31]. In this context, sharing communities play a crucial role in cyber defence as they provide access to CTI and knowledge about cyber threats. A sharing community comprises individuals or security teams who have cultivated a significant level of trust, either directly or indirectly, enabling them to engage in the exchange of sensitive threat-related information that may not be suitable for public dissemination. Examples of such sharing communities are the Forum of Incident Response and Security Teams (FIRST), Cyber Defence Alliance (CDA), Cyber Security Sharing & Analytics (CSSA), and Financial Services Information Sharing and Analysis Center (FS-ISAC) [7]. A common platform for sharing and managing CTI used by various sharing communities is the Malware Information Sharing Platform (MISP). MISP is an open-source threat intelligence platform used to share, store, and correlate various types of threat intelligence, including IoCs, TTPs, and CTI reports related to cybercriminal groups and their attacks. The overall quality of the mentioned CTI sources may not always be assured, and occasional errors within the dataset are not uncommon. Members of sharing communities, in particular, are confronted with a large amount of data and a high volatility in quality. Numerous studies [36][38][34][44][5][25] highlighted the absence of quality processes in CTI and the subsequent implications of this deficiency as identified by researchers.

RAID 2024, September 30–October 02, 2024, Padua, Italy

ACM ISBN 979-8-4007-0959-3/24/09
https://doi.org/10.1145/3678890.3678903

The lack of rigorous quality assurance processes for CTI not only undermines its reliability and effectiveness but also poses a significant risk to cybersecurity operations and organizations. To uncover innovative strategies that security teams are implementing to maintain high-quality CTI amid the dynamic landscape of technological advancements and increasingly complex cyber threats, we conducted interviews with 25 experts in the field of Cyber Security, with specialization in CTI. During these interviews, we collected valuable data on their operational procedures and their unique perspectives on the definition of quality in regard to CTI. This study aims to bridge between approaches from the research community and practice by uncovering the strategies and criteria employed by professionals in assessing the quality of CTI, thereby contributing to the development of more robust and effective quality assurance practices in the field.

While existing literature, including numerous studies cited earlier, has highlighted the general absence of formal quality assurance processes in CTI, we initially hypothesized that at least the more mature organizations would have well-established procedures in place. However, our findings provided a nuanced perspective on this issue. Contrary to our expectations based on the literature review and preliminary assumptions, our study revealed a more widespread lack of formalized processes across organizations of various maturities, including those perceived as experienced and mature. This discrepancy between our initial assumptions and the study's outcomes highlights the complex and varied nature of quality assurance practices in CTI, underscoring the need for a deeper investigation into the factors contributing to the absence of structured procedures, even among seasoned teams.

Within the scope of this study, we aimed to investigate four pivotal research questions:

RQ1: [Sharing] *What types of CTI are prioritized and deemed most relevant by practitioners?*
RQ2: [Quality relevance] *In what specific ways does the quality of CTI impact security operations, and which key factors determine its level of significance in various contexts?*
RQ3: [Quality aspects] *What do practitioners consider important in regard to the quality of CTI?*
RQ4: [Quality assurance] *What methods are applied by practitioners to assess, quantify, ensure, or enhance the quality of CTI?*

In our first research question, our objective was to gain a comprehensive understanding of the types of information exchanged among the diverse members within sharing communities. Within RQ2, our goal was to delve into the significance of upholding superior quality in CTI and to identify the key factors that underpin this importance. In tandem with our exploration of relevance, we delved into the intricate terrain of quality aspects through the formulation of RQ3. Finally, we sought to gain insights from the practitioners regarding the strategies and methodologies employed in practice by organizations in the realm of quality assurance.

This study results from a broader data collection focused on a better understanding of sharing communities and the quality management of CTI within these groups. This present study delves into the systemic challenges and effective data management strategies employed by professionals in those groups. Another publication resulting from this data collection provides a foundational analysis of the workings, structures, and challenges within CTI sharing communities [7]. Together, these studies present a comprehensive view: The first lays the groundwork for understanding the functioning of sharing communities, and the second addresses the practical aspects of maintaining CTI quality within these communities.

To achieve this, we used an exploratory research design to thoroughly investigate this under-researched topic. The interview questions were formulated in response to the identified research gaps, and as the study progressed, the research questions became clearer, evolved, and were refined. In summary, our contributions can be summarized as follows:

- We analyze the current state of quality assurance procedures utilized by professionals to better understand the methods practitioners employ to determine or enhance the quality of incoming CTI.
- We investigate the diversity of qualitative and quantitative methods applied by CTI professionals, shedding light on the variety of strategies ranging from human judgment and source reputation to sophisticated vetting mechanisms and overlap analytics.
- We reveal the reliance on human competence and intuition in the absence of formal metrics, highlighting the critical role of experience and gut feeling in the quality assessment of CTI among practitioners.
- We delve into the importance of context information, illustrating how practitioners require detailed background to effectively utilize CTI and make informed decisions.
- We discuss the advanced and systematically formalized methods employed by select experts to ensure higher CTI quality, highlighting the need for a balance between formalization and practical applicability.
- We provide actionable recommendations to bridge the gaps identified through our expert interviews in both academic research and practice, aiming to enhance the effectiveness and understanding of CTI quality management.

The rest of the paper is organized this way: Section 2 deals with relevant studies. Our methodological approach is outlined in Section 3, and in Section 4, we describe how practitioners manage quality issues related to CTI. We discuss our findings in Section 5. Furthermore, we provide recommendations for action in Section 6 and conclude our work in Section 7.

2 RELATED WORK

This section provides an overview of the current state of research in the realm of data quality regarding CTI. Our exploration of this field unveils two primary areas of focus: assessment methods for the quality of threat information and overall area of data quality related to CTI.

2.1 Quality Assessment

This area is dedicated to the development and implementation of assessment methodologies and techniques specifically tailored for evaluating the quality of threat information. Researchers in this area endeavor to create comprehensive frameworks and metrics that enable CTI analysts to gauge the reliability and accuracy of

CTI they work with. This extends beyond mere data validation, delving deep into the intricate processes and tools used to ensure the highest standards of data quality.

Numerous studies [41], [29], [13], [3], [6], [22], and [19] have delved into diverse approaches for assessing the quality of CTI and proposed new metrics and methods. However, the majority of this research predominantly concentrates on the data itself, overlooking critical aspects like feedback mechanisms for addressing false positives. Consequently, these valuable methods remain underutilized in practical applications. One prevalent issue we have identified is the dearth of opportunities to assess these proposed methodologies within real-world security operations.

Other notable works, such as [24], [9], [33], [37], [42], [39], and [14], present comprehensive methods and processes, including the development of sophisticated tools and advanced visualization techniques. These innovative approaches are tailored to support analysts in their intricate task of evaluating individual indicators and handling more technical information.

Lastly, certain studies, like [23], [27], [35], [18], and [8], suggest methods for appraising threat information feeds or data shared in sharing communities. These feeds and communities, whether derived from open-source channels or commercial solutions, exhibit varying levels of quality. The evaluation of such information is an ongoing endeavor, necessitating constant adaptation of these methodologies.

2.2 Data Quality

The other area we uncovered in our research concerns a broad perspective on data quality in CTI. This area is not limited to technical aspects alone; it delves into the contextual and operational considerations that shape the quality of data in CTI. Beyond the validation and evaluation of individual data points, it encompasses the broader systemic and procedural factors that influence data quality. Understanding these comprehensive dimensions is vital for developing a robust and adaptable data quality framework in the context of CTI.

In their 2022 study, Zibak and colleagues [43] employed a Delphi study to capture expert opinions to refine and validate their initial literature-based consideration of quality dimensions for threat intelligence. Their research revealed a list of important quality dimensions for threat information and threat intelligence and provided clear definitions for terms like "threat data" and "threat intelligence." However, it is important to note that this study did not investigate actual practices and procedures utilized by organizations to assure or assess the quality of CTI.

Sillaber et al. [34] investigated data quality challenges in threat intelligence sharing by conducting focus group discussions with ten experts from global organizations' security operations centers. The study addressed factors affecting data quality at various stages, including collection, processing, sharing, and storage of threat intelligence. It revealed that the primary challenges stem from integrating diverse information sources to serve a heterogeneous audience. While there are no fundamentally new data quality issues in threat intelligence sharing, the evolving nature of the field and the rapid introduction of sharing tools highlight the need to focus on scalability and data source integration.

3 METHODOLOGY

We conducted 25 semi-structured video interviews with CTI experts from different countries who have been members of various sharing communities for many years to better understand the procedures practitioners employ to determine or enhance the quality of incoming CTI. Many sharing communities are closed networks with no public information. Therefore, we opted for interviews. This approach enabled us to directly access the nuanced insights and experiences of CTI professionals, which are crucial yet often undocumented in publicly available sources. Our methodology can be divided into the phases of preparation, data collection, data handling, and analysis, described below.

3.1 Preparation

In this subsection, we describe the essential aspects of the preparation phase, such as the ethical considerations, what constitutes an expert for this study, the background of our questionnaire, and obtaining consent.

Ethical consideration. At the outset of our research, we referred to a self-assessment form of our institution to monitor the ethical integrity of our research project. This led to the conclusion that we were not required to seek formal approval from our ethics committee, which was also endorsed by our institutional ethics board. To maintain the trust of our participants, we designed our interview questions to revolve around their general experiences and findings.

Experts. An expert in our study is defined as someone with over five years of CTI experience, including roles in incident response, threat hunting, detection, forensics, or management of those activities. Therefore, experts are regularly confronted with CTI quality issues. As mentioned in the introduction, our data collection focused on gaining a better understanding of sharing communities and the quality management of CTI within these groups. Therefore, it was important that our experts were members in various sharing communities. This criterion ensured that our insights are based on those deeply involved in CTI practices and sharing communities, minimizing bias by focusing on experienced practitioners. Our goal was to include experts from a variety of sectors, industries, and expertise areas to enrich the diversity of our study. We believe this mix of experts will paint a holistic and pertinent picture of the topic under study.

Contact. Most of our experts evolved from existing relationships. Those relationships were formed through prior professional collaborations and shared involvement in sharing communities and companies through one of the authors, leveraging existing networks and mutual interests. However, the interviews were conducted by the author, who had no relationship to the interviewees. Interviewees unrelated to the authors were approached on the basis of published works on CTI quality topics at industry conferences. The communication before and after the interviews with the experts took place via email.

Questionnaire. As already mentioned in the introduction, the data collection focused on the dynamics of sharing communities and the CTI quality management within these groups. In this regard, the interview guide was divided into four sections: (1) Demographic information, (2) Workings of sharing communities, (3) CTI and

quality management, and (4) Relevant challenges. The interview guide was semi-structured and based on the SPSS approach (collect, check, sort, and subsume) [11]. The focus of our interview questions was instrumental in filling the research gaps related to sharing communities and quality management of CTI and shedding new light on this critical area. No previous research has examined the workings of sharing communities and how members of such groups, which are confronted with a high volume of incoming CTI, master the challenge with the quality of CTI.

Testing questionnaire. To ascertain that the interview guide adhered to academic norms and maintained the quality requisite of social experiments, we executed two rounds of pre-testing. We initially shared the questionnaire with our first expert, an expert in political science and interview methodology, for analysis. This was followed by a detailed online meeting with this expert to address concerns about each question's wording, sequence, and clarity, ensuring alignment with our research objectives. Subsequently, we conducted pilot interviews with two CTI experts, mirroring the main interview format. After that, drawing upon their input and referencing Kruse and Schmieder [15], we enhanced the questionnaire's utility, language, terminology, and intelligibility. Kruse and Schmieder list key aspects to simplify formulation: Choose unambiguous formulations (express clearly understandable), do not formulate multiple combined questions, use comprehensible wording, and match the choice of words to the interviewee's vocabulary. The pre-testing phase also served to identify any questions that were vague, equivocal, or predisposed to undesired responses. The finalized questionnaire is provided in Appendix A.

Informed consent. Participation in our interviews was voluntary, and the experts signed informed consent before the interviews. In addition, we briefed the experts prior to the interviews about the goals, extent, methods of recording and ensuring anonymity, transcription processes, and data management practices of the study. Finally, a verbal agreement was sought prior to the initiation of audio recordings.

3.2 Data collection

This subsection describes the data collection phase in which we explain our sample size and the generalizability of our results. Moreover, we describe how we conducted the video interviews with the experts.

Sample. The determination of our sample size was chiefly influenced by the concept of data saturation [1] and the suggestions put forth by Marshall et al., advocating for a sample size ranging between 20 to 30 participants [21]. The saturation principle is a recognized qualitative method that mitigates concerns about the diversity and size of our sample by ensuring comprehensive coverage of relevant findings. This approach ensures that data collection continues until no new or relevant information emerges, indicating comprehensive coverage of insights within our sample. In our specific scenario, we observed that at the end of our interviews, even though the experts and their organizations were very mature, the themes and insights we gained were repetitious and consistent. This implied that we reached data saturation and new experts would not significantly change our main findings. Furthermore, this sample size was ample enough to deliver an exhaustive understanding of our research's subject matter while preserving the credibility and validity of our results. In this study, we focus on a deep understanding of the specific phenomena studied rather than aiming for broad generalizability to the population as a whole. This approach is characteristic of qualitative research, which aims to explore in-depth findings rather than to ensure representativeness, as is often the case in quantitative studies.

Interview conduct. We conducted the interview sessions from May 25 through August 30, 2022, by the same interviewer who conducted all video interviews using an institutional version of Zoom [45]. To ensure that both the interviewer and the expert had a common understanding of the critical elements of the research topic, we provided a definition of sharing communities and the NIST definition of CTI at the beginning of the interviews (see Appendix A). While we posed questions based on the interview guide, we also delved into topics that surfaced organically during the conversation. The interview durations ranged from 20 to 105 minutes, with an average span of approximately 53 minutes.

3.3 Data handling

In the following, we describe our data processing phase. In this regard, we detail our transcription process and how we protected the collected interview data.

Transcription. Following the interviews, we used the audio files from the recorded interviews to generate non-verbatim transcripts using Adobe Premiere Pro [20]. This tool does not upload the audio files or the finished transcripts, thus ensuring data protection. We subsequently carried out manual anonymization of any personal data that could potentially reveal the identity of the experts as part of the transcription process, in line with Kuckartz's guidelines [16]. Concurrently, we rectified errors in the transcript by manually cross-checking it against the audio file. Upon completion of the transcripts, we gave the experts an opportunity to verify the transcript for accuracy, as recommended in [4]. Nevertheless, we stressed that they were not permitted to alter the transcript's interpretation. This review aimed to enhance the transcript's trustworthiness, validity, and precision and confirm the adequacy of anonymization. Consequently, we rectified minor grammar issues and incorrect terms in the transcript.

Data protection. The personal data of the interviewees were stored locally in a password-protected archive. The audio files and corresponding transcripts were stored anonymously on an institutional instance of GitLab, which was exclusively accessible to the authors of this work. Keeping personal data segregated from the interview data significantly reduced the likelihood of making inferences about the experts.

3.4 Analysis

This subsection describes the last phase of our methodology, the analysis of the collected data, and shows the limitations of our work.

Content analysis. We subjected the gathered data to a thematic content analysis, following Kuckartz's method [17]. This systematic and structured approach involves a material-guided and

rule-guided analysis of the existing text content. Our analysis procedure incorporated the following steps: (1) Identifying significant sections in the transcribed text and noting distinctive attributes, (2) Inferring primary categories from guiding questions and existing material and developing a coding scheme, (3) Allocating text to the appropriate category by coding, (4) Constructing subcategories, (5) Re-coding the entire material using main categories and subcategories, (6) Accumulating all coded text passages for each main and subcategory, (7) Evaluating and visually representing the results, which can be found in Appendix B. Moreover, the outcomes of the content analysis can be viewed in the findings section.

3.5 Limitations

Given that our research involved expert interviews, it was difficult to completely mitigate the possibility of self-reporting or response bias. Experts could potentially be influenced by the interviewer, questionnaire, or the online milieu during the session. In this context, three experts opted to conduct the interview without enabling their cameras. Also, there was a possibility that experts might provide responses deemed socially acceptable, thereby introducing bias.

In our recruitment, we utilized one author as a gatekeeper, a standard expert-access method. Acknowledging potential bias, interviews were conducted by another author. Furthermore, we want to emphasize the trust established through personal contact with the gatekeeper likely encouraged interviewees to provide more open and in-depth responses, enriching the study's data quality.

4 FINDINGS

In the following, we use the insights of our interviews to answer the research questions outlined in the introduction. We first dive into the various forms of shared CTI and then describe the importance of high-quality CTI. Subsequently, we investigate which aspects of CTI are considered important by practitioners. Finally, we examine the methods practitioners employ to deal with quality issues of CTI.

4.1 Demographics

Professionals from a wide range of backgrounds participated in this study, from specialized security teams to national CERTs to large enterprises. In two cases, two interviewees came from the same organization but from different departments. Some of the experts might know each other since the CTI field is small. However, we can only assume this. Some of the interviewees have been in the field for several decades and were active in the early days of sharing communities. A detailed overview of the demographic details relevant to this study can be found in Table 1, where we have listed our sample in random order, including information on years of work experience, sector, industry, and expertise area. Our sample consisted of experts who were working in the following countries: United States of America (6), Germany (6), Great Britain (3), Netherlands (3), Australia, Poland, Luxembourg, Thailand, Greece, Switzerland, and Brazil. However, this is not inevitably their origin.

4.2 RQ1: Sharing

We now present our findings for **RQ1:** *What types of CTI are prioritized and deemed most relevant by practitioners?* It is fundamental to comprehend what types of shared CTI are relevant to practitioners to better investigate, understand, or develop metrics or procedures to determine the quality of CTI. We grouped the answers in regard to the NIST definition into two groups: Unprocessed Threat Information and Cyber Threat Intelligence:

Unprocessed Threat Information:
- Raw IoCs or vulnerabilities

Cyber Threat Intelligence:

Processed Threat Information
- Interpreted IoCs or vulnerabilities

Threat Actor Insights
- Tactics, techniques, and procedures
- Threat actor profiles

Strategic & Operational
- Reports
- Detection or mitigation techniques
- Courses of action
- Best practices for analysis or forensics
- Prioritization or work focus

Typically, members of sharing communities share and receive a wide variety of information. The highest proportion of what is shared is data or unvetted IoCs or vulnerabilities. In this context, one expert mentioned that primarily "automatically generated indicators" are shared and that he would call most of the shared information "data, not intelligence." Examples of such automatically generated indicators are hash values of malware, IP addresses, or domain names that are created by intrusion detection systems or endpoint detection and response solutions, for example. Another expert added, "when we look at indicators we receive, they may not necessarily meet the criteria for CTI," as the NIST definition states. A predominant part of sharing is the automatic forwarding of unverified and non-interpreted threat information.

Besides unprocessed threat information, CTI is shared to a lesser extent, as the experts stated. We grouped the answers into processed threat information, threat actor insights, and strategic and operational intelligence. A crucial part of sharing is made of processed or contextualized threat information, which is, for example, interpreted IoCs or vulnerabilities linked to specific threats or enriched with context. Furthermore, threat actor insights are shared in the form of threat actor profiles or used TTPs. In this context, one expert mentioned a threat actor spotlight procedure utilized in a smaller trust community, where members introduce relevant and currently active threat actors. In this regard, the same expert mentioned a database of threat actors maintained by this sharing community and its members, where the threat actors are modeled and shared.

Finally, strategic and operational CTI is shared between members of a sharing community, though in a smaller proportion. Sharing CTI reports, detection or mitigation techniques, and prioritization or work priorities of other members is a valuable and critical part of sharing efforts. Besides CTI reports containing contextualized and

No.	Exp.	Role	Sector	Industry	Expertise Area
1	20 +	Vice President of CTI	Private Sector	IT	Incident Response
2	10	Chief Technical Officer	Private Sector	Security and Insurance	Risk Research, Incident Response
3	20 +	Head of CERT	Public Sector	National CERT	Incident Response
4	20 +	Chief Executive Officer	Non-Profit	Cybersecurity	Incident Response, Computer Forensics
5	14	Director Cybersecurity TI	Private Sector	Electronics	Cyber Threat Intelligence
6	7	Head of Defense	Private Sector	Electronics	Incident Response, CTI
7	20	Manager of Outreach	Private Sector	IT	Incident Response, Computer Forensics
8	35	Lead of Incident Responders	Private Sector	IT	Cyber Threat Intelligence
9	20	Cyber Security Consultant	Public Sector	International Relations	Legal Advice
10	15	Chief Regulatory Advisor	Public Sector	Policy and Regulation	Incident Response
11	14	Chief Cybersecurity Advisor	Private Sector	Mobility	Computer Forensics
12	10	Senior Advisor	Public Sector	Cybersecurity	Vulnerability Management
13	18	Expert Advisor	Private Sector	IT	Cybersecurity, CTI Sharing
14	20 +	Principal Analyst	Private Sector	IT	Internet Infrastructure and Security
15	10	CTI Analyst	Private Sector	IT	Threat Landscape and CTI Projects
16	10 +	Security Analyst	Public Sector	IT	Cyber Threat Intelligence
17	7	Security Analyst	Public Sector	National CSIRT	CTI, Incident Response
18	10	Principal Specialist	Public Sector	National CSIRT	Intrusion detection, Information Sharing
19	16	Senior Internet Security Specialist	Non-Profit	IT	Incident Response, Training
20	15 +	CERT Specialist	Public Sector	National CSIRT	Incident Response, Security Management
21	6	CTI Specialist	Private Sector	IT	Cyber Threat Intelligence
22	7	Senior Researcher CSIRT	Public Sector	Higher Education	Incident Response, Vulnerability Mngmt.
23	20 +	Threat Hunter	Non-Profit	IT	Computer Forensics, Threat Hunting
24	6	Key Expert for TI	Private Sector	Electronics	Cyber Threat Intelligence
25	10 +	Chief Architect	Private Sector	IT	Administration of Trust Groups

Table 1: Summary of Experts and Their Expertise Areas

processed information about threats, detection techniques in the form of scripts or YARA rules [40] are shared. In this context, one expert highlighted his interest in how other members do detections and that he is interested in "understanding [members'] methodology, infrastructure, how [members'] systems are set up [...] and the tools" they use. Moreover, intelligence about mitigation, such as solutions to address specific vulnerabilities, incidents, or courses of action, is essential to sharing efforts. In addition, community members share best practices for analysis or forensics. Ultimately, the importance of prioritization and work focus of other community members is a critical knowledge gain, which is shared between members.

4.3 RQ2: Quality Relevance

In this section, we answer **RQ2:** *In what specific ways does the quality of CTI impact security operations, and which key factors determine its level of significance in various contexts?* In the context of this research question, we want to highlight the urgency and relevance of high-quality CTI.

4.3.1 Reasons for high-quality CTI. Our experts emphasized that high-quality CTI is critical for maintaining robust security operations. High-quality CTI is not just a preference but a necessity, as the impact of low-quality information goes far beyond mere operational inconvenience. Experts point out that the proliferation of low-quality CTI jeopardizes the fundamental trust and reliability on which sharing communities rely. One particularly salient point is the direct impact of CTI quality on operational decision-making. Inaccurate or misleading information can lead to poor decisions, misallocation of resources, and potential security failures. This is all the more true in environments where automated systems rely on CTI for real-time decision-making because the margin for error is much smaller. False positives, as highlighted by experts, pose a real challenge, especially on a large scale, as they can distort threat to and response strategies.

In addition, the social dynamics within sharing communities are strongly influenced by the quality of the information shared. Those who contribute low-quality CTI face tangible repercussions, including erosion of their reputation and potential exclusion from critical sharing groups. This loss of trust not only diminishes the individual or organization's standing within the community but also limits their access to valuable information, creating a feedback loop that further isolates them from the important security discourse. Conversely, the expectation of high-quality CTI fosters a culture of care and accountability among contributors and encourages the adoption of rigorous review procedures and quality assurance measures. This not only raises the overall standard of information shared but also strengthens community bonds through a shared commitment to excellence.

In this complex ecosystem, the demand for high-quality CTI is underlined by the dual need to protect operational integrity and

maintain the collaborative framework that underpins effective CTI sharing. The quality of CTI is thus a critical factor that not only influences the tactical and strategic decisions made in security operations but also shapes the fabric of the communities on which these operations rely.

4.3.2 Use case dictates quality requirements. Although we found through the interviews that sharing or receiving high-quality CTI is critical, we learned from four experts that it is vital to reflect that the need for high quality varies depending on the use case of CTI, as different use scenarios have specific quality requirements. One expert described two prominent use cases: automated detections and investigations, where the quality must be higher for automated detections. He explained: "The intelligence you want to use for detection must have extremely high quality." In contrast, CTI for investigation or threat hunting has more tolerance. Here "it is almost the other way around because when you are investigating something, you want to have access to as much information as possible." In such cases, even poor quality can be helpful and give a hint where there was no before.

A second expert compared CTI for automated processes and CTI as input for, e.g., giving pieces of advice, where the latter must not be as accurate. The third expert explained that he takes no responsibility for the quality nor puts "a lot of effort at vetting the information" in the case of automated sharing. He added "in some cases, we do not vet the information at all." Regardless, he highlighted "the assumption on both sides that the quality may vary" is decisive.

The fourth expert sees "the quality as a double-edged sword." Depending on the situation, he outlined that there is an emerging site where sharing has to be as fast as possible, and "you are sharing information as it comes in," "not knowing exact value and veracity of it." In this case, everybody should have the same expectations and know that the quality is not the highest. He added, "it is not exactly CTI at that point, instead incident response sort of data." Conversely, quality becomes essential when "things evolve towards a more actionable state." He added that "at that point, if the information is of questionable veracity, then you diminish their ability to be effective, and you diminish their trust in the data and quality of information."

4.4 RQ3: Quality Aspects

In the following, we present the findings for **RQ3:** *What do practitioners consider important in regard to the quality of CTI?* We identified two main groups of relevant aspects: Context information and essential quality characteristics. Context information helps in understanding the background and the intended use of the incoming CTI and the attack better. Essential characteristics ensure the intrinsic value and reliability of the data itself.

4.4.1 Context. When speaking about the quality of CTI, receiving context information was mentioned as one of the first aspects that came into the mind of the experts. In this regard, the experts mentioned various essential aspects relevant to them in their daily business.

The indispensability of context. For eleven experts, receiving various context information is crucial and indispensable. One of the experts mentioned that anything can be powerful with context. Even a "IP address is really powerful as long as you know what to do with it." Another one emphasized that he "needs to have that context around data to make decisions." "Certain information is not useful without context," another expert added. Moreover, "the finer detail you get, the more you can do with it." Finally, in regard to assessing the quality of CTI, one expert emphasized that "if there is no context information [...] it is very difficult to say what the quality is." Another expert underscored this and encouraged adding extra context, which "is increasing the quality."

Understanding, utilizing, and prioritizing CTI. Context information improves the quality of CTI by increasing the understanding and use of incoming CTI, and decreasing the risk of potential exposure through appropriate prioritization. One expert emphasized that giving context to his customers, such as the ability to understand what they are receiving and how the information can be prioritized and utilized, is essential in using incoming CTI. Furthermore, one expert stated that giving consumers the ability to understand the relevance of information, how the information was generated, or the intended use increases the usability and quality of CTI. Another expert underlined the importance of understanding the process of CTI development: "If you are consuming a feed, you will understand it better [when] you have a bit of a feel for how that information is developed." In addition, one expert emphasized that consumers should know how to use incoming CTI and suggested guidance such as "this data is good for this purpose." Moreover, it is vital to give customers or consumers the ability to understand their potential exposure to the threat related to shared CTI.

Enhancing the understanding of attacks. Further desired context information is related to the attacks, helping better understand them. One expert is interested in the time-related information, such as when the attack was recognized and when was the last activity. In addition, other experts named further desired information such as timestamps of the beginning and end of an attack, frequency of attacks, the potential threat resulting from the attack, who were the threat actors, and what they were trying to do. Furthermore, general network information, like how the attack was routed.

Lack of context in practice. Despite the high relevance of context information for the experts, some experts emphasized the persistent lack of sufficient context. One expert remarked, "people still do not provide enough context around their data and how they developed it." This is also underscored by the results in RQ1, where several experts mentioned that it is mainly data that is shared rather than CTI, highlighting a potential gap between ideal practice and reality.

4.4.2 Vital characteristics. Besides different types of context information, the experts mentioned specific characteristics they consider as critical related to the quality of CTI. In the following, we describe those characteristics briefly.

Accuracy and correctness. Three experts highlighted the importance of the accuracy or correctness of incoming CTI. One expert mentioned that "the data has to be correct, [...] free of error, formatting [must be right]." Another added, "this is where a lot of work and a lot of sharing attempts at sharing organizations fail [...] to effectively use information, it has to be accurate."

Freshness. Moreover, three experts mentioned the freshness of CTI as important and the desire to receive CTI in a timely manner. However, another expert emphasized that receiving timely is very important, "but if you share too quickly, you can end up making mistakes."

Consistency and repeatability. One expert emphasized the inherent value of consistency in CTI. He desires repeatability and wants to receive the same type of CTI in a consistent format at predictable intervals.

Actionability. For three experts, the received CTI must be actionable. One of them mentioned that it is important that the shared CTI is actionable and that he can feed it into detection systems. However, he added that the necessity of actionability depends on the type of CTI. On a tactical or strategic level, actionability becomes less critical. Another expert described actionability differently as a high-level characteristic that can be broken down into sub-characteristics. For him, something is actionable when it comes timely, is complete, and has proven to help against threats. Furthermore, he added, "if it is bad quality, then it is not actionable."

Overlap and sightings. Furthermore, one expert emphasized the importance of overlap or sightings and receiving proof about the reliability. He asked: "How many people have seen this as well? So is it just from one source, or is it from multiple?"

4.4.3 Expert Opinions on Specific Quality Dimensions. After asking the experts about the aspects they consider essential, we showed them the quality dimensions for evaluating CTI derived by the authors Zibak et al. [43]. Their final set of most relevant quality dimensions for threat data and threat intelligence is the following: Relevance, actionability, timeliness, accuracy, source reliability, interoperability, and provenance. The opinions on those quality dimensions were distinct. We learned that each dimension holds importance in its own right. However, their significance is context-dependent, often influenced by the consumer's specific requirements and use cases.

While we identified an agreement about the crucial importance of accurate CTI and receiving it as fast as possible, the experts' opinions on actionability are divided. Some argue that non-actionable information is somehow useless. For others, actionability strongly depends on the type of CTI, the consumer's capabilities, and the use case for CTI. Some information can be used and be actionable for log analysis but not for blocking hosts. Interoperability, particularly in the context of the standardized representation and data models of CTI is "super important" for one expert. However, another expert added that it is not a component of CTI itself. Furthermore, another expert mentioned that interoperability is not similarly important for producers and consumers. Producers sometimes do not want to care about it since it locks consumers to their products. Therefore, it is more relevant from the consumer's perspective.

Provenance can strengthen trust in CTI and is therefore valuable. However, several experts mentioned that it is rarely available for a variety of reasons. Some argue that it is indispensable, while others believe its importance diminishes as long as CTI is accurate and received timely. Finally, relevance is a highly subjective quality dimension and depends on the consumer of CTI. Therefore it is difficult to quantitatively measure it.

4.5 RQ4: Quality Assurance

In this section, we answer **RQ4:** *What methods are applied by practitioners to assess, quantify, ensure, or enhance the quality of CTI?* In this research question, we want to explore the strategies and methodologies employed by practitioners and their organizations in the realm of quality assurance.

4.5.1 Term Clarification. Firstly, we want to differ the terms "assess," "evaluate," and "measure" since the terms are often used interchangeably. However, we found during the interviews that they have different meanings regarding CTI's quality determination.

When the experts talked about assessing CTI quality, they meant getting a sense or overview of quality without metrics or quantification. The results of an assessment are more qualitative or subjective and not systematic or deeply analytical. In contrast, a more systematic or methodical approach was meant when the experts discussed evaluating, measuring, or quantifying CTI quality.

In the context of CTI quality management, "validation" and "verification" are critical initial steps in the broader quality assessment process. They are distinct in focus, while validation ensures the relevance and credibility of CTI, verification confirms its accuracy. Furthermore, they are integral to establishing a foundation of trust in CTI before it undergoes more detailed quality evaluation. These "vetting" procedures, by identifying and filtering out inaccurate or irrelevant CTI, enhance the quality of intelligence that is subject to further, more granular quality assessment metrics and methodologies.

In summary, the methods described by the experts in our interviews can be classified into qualitative and quantitative procedures. In this regard, we noticed that oftentimes, the experts use a combination of qualitative and quantitative approaches. Figure 1 provides a visual representation of the responses of the interviewees. These approaches are described in the following subsections.

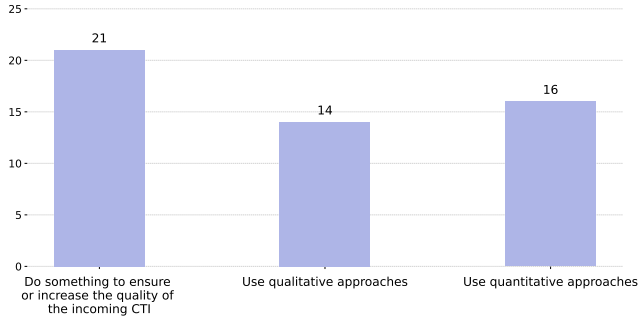

Figure 1: Distribution of Procedures Used by Practitioners

4.5.2 Qualitative Procedures. We noticed that most experts (21) do something to ensure or increase the quality of the incoming CTI. Only four experts mentioned that they have not implemented any procedures to assess, quantify, or increase the quality of CTI they receive. Of those twenty-one experts, fourteen mentioned different qualitative approaches to assess incoming CTI. Figure 2 provides a visual representation of the qualitative approaches used by these fourteen interviewees. In the following, we will describe the characteristics of these approaches.

Human competence. Six experts mentioned various forms of human competence, such as experience, know-how, gut feeling, and personal judgment, as essential for making reasonable quality assessments. One expert highlighted that measuring the quality is "so difficult that most people do not bother and they and they go with gut feeling, which is honestly maybe better than creating a whole bunch of metrics." Another expert added that his "judgment [is], the effort required to do this is not worth the investment" and he can "quickly make assessments if something is valuable" without metrics.

Source trustworthiness and reputation. The source plays an essential role in assessing the quality of CTI for nine experts. For six of them, considering where the information comes from is a decisive factor in whether one trusts the quality and uses incoming CTI. However, those six experts do not use a quantitative approach to track the record with a source, which we describe in Section 4.5.3. Those experts mainly trust the quality based on a source's work history and reputation. One expert underlined the importance of the assessment based on the source's reputation: When assessing the quality of incoming CTI, "a main part is knowing the organization [where CTI] comes from." In addition, another expert is assessing CTI "commonly [based on] trust in the source of the information." He emphasized that "trusting the source is key" and is usually "sufficient to determine whether [the quality] is good to use or not." He added, "we prefer, whenever possible, to spend more time figuring out the best course of action" instead of evaluating or measuring the quality.

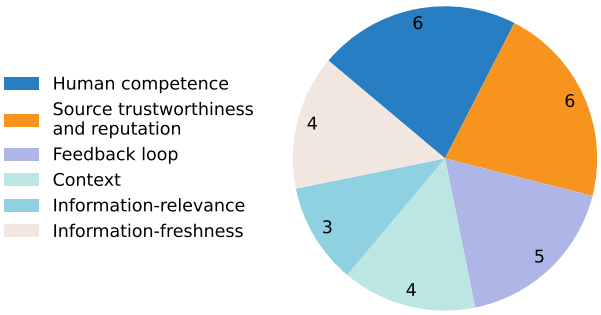

Figure 2: Qualitative Procedures Used by Practitioners

Especially in the case of tactical or strategic CTI, where cross-checking or validating is more complex, the source reputation is critical. As a result, one expert said he thinks "that we rely heavily on the reputation of the source" when assessing the accuracy of this type of CTI. However, unquestioningly trusting the source's reputation is not enough for three of the six experts. They emphasized the need for additional efforts to validate or strengthen trust in CTI received, even for well-trusted sources. In addition, one of those three experts highlighted team discussions on received CTI, which helps assess the quality besides knowing and trusting the source. Furthermore, this expert explained that at least three sources must confirm controversial news to act on it. One expert also considered the service quality of a source in his assessment. For example, how well the communication is between him and the source, which strengthens his trust in CTI received. However, this is more applicable to paid sources. Finally, in the case of a new source, one of the six experts demanded a sample of CTI to make a pre-assessment.

Feedback loop. Feedback is a method that can also be used to help assess the quality of CTI. However, feedback is based on individual interpretations, and results remain subjective. Five experts mentioned the importance of feedback to them. Three of them highlighted the sightings feature of MISP as a feedback mechanism to receive confirmation about the activity of an IoC or event. Another expert explained that he receives feedback on the quality dimension relevance, completeness, accuracy, and timeliness, which helps him share better-aligned CTI with his consumers. In addition, one expert emphasized that he desires anonymous consumer feedback on "how accurate and useful that information was." Moreover, one expert summarized feedback as an excellent way to learn about the quality of CTI.

Finally, one expert described a gamification approach that was based on a feedback mechanism. People could move up a ranking system and receive points for sharing or if others rated their shared CTI well. He said that this "was a sort of a quality metric we used as opposed to the volume metric."

Context. The existing context information related to IoCs, events, or reports is another criterion for four experts in assessing incoming CTI. One expert highlighted that "some kind of transparency and how the information is collected" helps assess the quality. He underlined that he needs a specific level of detail. Otherwise, he "would not act on the information." Another expert added, "Where do you get your data from? What are your sources? How did you figure this out? That is kind of the baseline [context]" that is necessary to make reasonable assessments.

Information-relevance. Three experts further assess CTI based on relevance. One expert emphasized investigating if something is relevant before taking any further steps as necessary. In this regard, the experts assess if sources, threat actors, types of CTI, or vulnerabilities are relevant to their organization or constituency.

Information-freshness. Four experts mentioned the freshness of CTI as something they are focusing on when assessing the quality of CTI. One expert mentioned that when receiving CTI with a "significant delay," he "may choose not to act on," because "it provides too little value."

4.5.3 Quantitative Procedures. In contrast to the approaches described above, where the results are qualitative and subjective, sixteen experts have employed quantitative procedures. Figure 3 provides a visual representation of the quantitative approaches used by these sixteen interviewees.

Vetting mechanisms. Of these sixteen experts, thirteen employ various forms of vetting, such as verifying or validating the authenticity of incoming CTI. Six experts use the MISP warning list to validate incoming CTI, ensure its legitimacy, and filter out known false or misleading indicators. Moreover, occasionally, own or third-party databases (e.g., VirusTotal or ShadowServer), OSINT feeds, other MISP instances, or a top 1000 website list are used to check or validate the received CTI. In addition, manually pinging C2 servers, checking suspected phishing domains, or detonating unknown malware in a sandbox are other types of verification

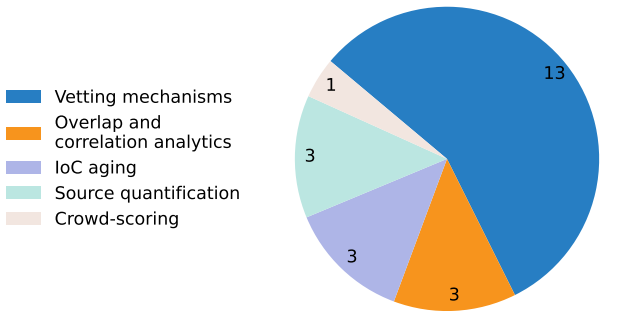

Figure 3: Quantitative Procedures Used by Practitioners

that were mentioned sporadically. Finally, one expert mentioned an automated crosschecking of received hashes with the National Software Reference Library (NSRL) feed.

Overlap and correlation analytics. The overlap matrix in the MISP or self-developed overlap calculations are techniques three experts use to validate and confirm that IoCs are confirmed as true positives by multiple feeds or sources. In this regard, one expert mentioned that he is assessing the quality based on the coverage of CTI received. Moreover, one expert uses the correlation feature of MISP to calculate the correlation between different events or feeds.

IoC aging. Besides the qualitative approach to assessing the freshness of CTI mentioned in Section 4.5.2, three experts consider a quantitative approach, also called "decaying of indicators." Two of them use the decaying model of MISP. The other expert was working with his organization on developing a decaying model. The goal is to find and evaluate the optimal storage time for IoCs. There is always a consideration between keeping an IoC "long enough to catch every sighting" versus "not too long because you have cost of storage," one expert emphasized. Without considering the decaying or aging of IoCs, "the amount of indicator you are ingesting and processing only grows."

Source quantification. In the section above, in the paragraph "source trustworthiness and reputation," we mentioned that for nine experts, the sources are critical in assessing the quality of CTI. Six of them use a qualitative approach, as described above. The three remaining experts mentioned that they use a quantitative approach to evaluate the work history with a particular CTI source. Two of the three experts use the admiralty code or NATO system [10] because of its structured approach to evaluating the reliability of intelligence sources and the credibility of the information, and one expert uses a self-developed ranking system for sources based on experience with a source and the relevance of CTI to his organization, such as specific threat actor groups. The source ratings are maintained in a "wiki" or as labels or tags in a sharing platform.

Crowd-scoring. One expert described a crowd-scoring approach that was utilized in the past at a CERT. Members of a sharing community rated the information provided by contributors regarding its usefulness and accuracy. Over time, this system built a trustworthiness profile for each participant based on the collective feedback. The goal was to increase trustworthiness and accountability within the community. However, such a model requires transparency in the evaluation process and ongoing member participation to be effective. Nevertheless, he added, "I don't know any sharing organizations [that] are actually doing this at this point."

4.5.4 Detailed Procedures of Selected Experts. Across the interviews, we found various approaches to ensure CTI quality. While some use a combination of qualitative and quantitative methods, many of the described procedures by the experts lack formalization, written processes, and automation. Despite this, we pinpointed four instances where the strategies not only exhibit a high degree of formalization but are also systematically designed to enhance the quality of CTI.

Expert #3 explained a process that consists of various features in MISP to crosscheck CTI, such as a filter based on the warning list, the overlap matrix, correlation calculations, and sightings. Based on continuous monitoring of the results of those methods, an analyst manually creates a trust rank for sources. In addition, expert #3 mentioned that a feedback loop, the decaying of an indicator, and timestamps help further ensure high quality.

Expert #13 sought to ensure that CTI he shared adhered to a set of quality dimensions, which he called TRACE. This acronym stands for timely, responsible, accountable, complete, and efficient. (1) Timely: He wanted to share CTI promptly, provide observation provenance, and remove stale data. (2) Responsible: He emphasized the importance of standing behind the data and being willing to assume legal risks associated with CTI he provided, emphasizing the importance of legal accountability associated with the consequences of the data. (3) Accurate: He was keen on the need for CTI to be well-sourced, understood by the producers, and appropriate for its provided purposes. In addition, there should be only a few, if any, false positives. (4) Complete: He emphasized sharing comprehensive information and ensuring no relevant details were omitted. Context is necessary to utilize CTI properly. (5) Efficient: He stressed the importance of efficiently deploying CTI into existing infrastructure and processes, cost and benefit-balancing, and infrastructure that can efficiently utilize CTI in its presented forms.

Expert #15 described an automated process in which CTI is pre-filtered or prioritized based on source ratings. CTI is collected from various sources, such as private security providers, governmental feeds, and open sources. Initially, an experienced person manually assigns each source a score based on the source's reputation, accuracy, or completeness of CTI received. The automated filtering is then based on the source rating and relevance of CTI. For example, paid and trusted sources like Mandiant are rated higher than crowd-sourced platforms like Alien Vault. In addition, CTI from internal teams like incident response is given higher priority due to its immediate relevance. Furthermore, specific types of CTI or threats have different relevance. For example, certain threat actor groups and some malware types or threats (e.g., QakBot, Cobalt Strike) are scored higher based on their relevance to the organization of expert #15. The relevance is based on, for example, previous incidents or the targeted sector of the threat actors.

The automated filtering helps to identify what is relevant, expert #15 underlined. In addition, he uses the MISP warning list. A deep dive into the filtered CTI and a more detailed analysis will be conducted manually in the subsequent steps. The scoring system is simple: low, medium, and high. However, expert #15 emphasized

that "it is not very sophisticated." Nevertheless, "the advantage is simplicity. We try to keep it simple." The ratings for sources and relevance of threat actors, malware, or threats are reviewed and re-assessed at fixed intervals (3-4 times a year). "Critical here is the consumer feedback" by the Security Operation Center (SOC), e.g., on false positives. Expert #15 mentioned that his approach is comparable to [28]. All measurements are tracked in a threat case management system, which also keeps a record of the source, allowing for ongoing evaluation of the source's reliability.

Expert #18 described testing a procedure comparable to [26]. He underlined that this procedure was not implemented "as a continuous process." Instead, he used the procedure twice as a research exercise to "have a better understanding of sources." Nonetheless, he uses "some of [the dimensions] on a pretty regular basis," when he is "adding new data sources with indicators to [a] sharing platform" and needs "to understand the data that is coming in." In addition, this assessment helps him to "get a better understanding of the usefulness" and in the selection of threat intelligence sources." In this regard, a trust metric (three-level scale) is assigned to a source, which is "kind of an expert opinion." He added that this expert opinion is "at least in some cases based on our analysis of the content" with the help of some quality dimensions. However, he emphasized that he is "not computing" the quality dimensions. Instead, he is looking at the received indicators and analyzing, firstly, if it is relevant. Subsequently, he assesses the IoCs based on quality dimensions such as timeliness, accuracy, completeness, or depth of information. It is essential to mention that the source is the central aspect for expert #18 when assessing incoming CTI. Finally, this expert emphasized that he wants to establish quantitative procedures in the future, but currently, the focus is on other, more relevant topics, like automation.

4.5.5 Lack of Evaluation and Measurement. Even though various practitioners use quantitative procedures such as vetting or overlap features, it became apparent during the interviews that most experts and their organizations do not quantify the quality of incoming CTI based on specific quality dimensions and metrics.

Six experts mentioned that the lack of resources and time is one of the main reasons for the lack of measuring CTI quality. In this regard, one expert said: "It is too expensive. It is too time-consuming. Only big organizations can do it." Furthermore, "the added value is not big enough to give this a high priority," another expert explained. He added that it is sufficient when "information is credible enough for us to act, it is kind of good enough for us."

Furthermore, four experts mentioned the difficulty of measuring procedures as another reason for the lack of measuring CTI quality. In this regard, another expert explained that implementing measuring systems requires continuous tuning and maintenance. The introduction of scores also requires additional processes with fail-safe mechanisms for correcting possible scoring errors, which makes the technical implementation particularly complex. Moreover, one expert added, "it is hard to give a high level of confidence at any given time," which makes it hard to quantify the quality of CTI. Another expert highlighted consumers' different requirements, making it challenging to implement measurement systems.

5 DISCUSSION

In this section, we first discuss RQ1 and the importance of understanding what is shared. We then examine RQ2, in which we discuss why high-quality CTI is important and why the importance of high-quality CTI varies. We then address the relevant quality aspects of CTI in RQ3. Finally, we discuss RQ4 and the methods used by practitioners to ensure or enhance the quality of CTI.

RQ1: Sharing. Our findings show that practitioners share and receive a wide variety of CTI. Furthermore, practitioners use different sources besides the sharing group to position themselves more broadly against the dynamic threat landscape. Interestingly, the experts highlighted that the main proportion of shared CTI is made of technical CTI, specifically data. In addition, CTI shared in a sharing community is not tailored to the needs of the consumers, nor is the shared CTI always relevant to them. That is increasing the volume of irrelevant CTI and, therefore, increasing the workload. The lack of quality of CTI and the variety of shared CTI highlights the need for standardized processes, ideally early in the generation of different types of CTI.

RQ2: Quality relevance. The indispensable value of high-quality CTI became evident through the findings of RQ2 and has an impact on the ones who share, consumers, and the community at large. The statement of one expert that "quality is a very, very important criterion for the usability of information" underscores this point. However, the necessary level of high quality is not always equally important. There are scenarios where the quality must be higher and where the quality is a second concern. Therefore, the necessary quality of CTI must be determined regarding a specific use case and type of information. Thus, there is a need for tailored procedures and metrics for specific use cases and types of CTI.

RQ3: Quality aspects. Academic studies like those by Zibak et al. and Schlette et al. outline specific quality dimensions for CTI. However, our RQ3 findings indicate that in practice, these dimensions are not used by practitioners as separate criteria. Instead, context information is critical for practitioners. For them, high-quality CTI means having enough context information to sufficiently understand a threat, understanding its intended use, and being able to properly assess the risk of a potential threat and defend against it. In addition, high-quality means for practitioners to receive not only context-rich CTI but also to receive CTI with high accuracy and as fast as possible. Besides context-richness, accuracy, and freshness, we were not able to find complete agreement on other important characteristics of CTI. This indicates that the different backgrounds of the experts and the use cases dictate what is considered important. One expert also highlighted this, who emphasized that "if I now ask 1000 companies, then I get 2000 requirements" for CTI.

RQ4: Quality assurance. Most experts use different methods to ensure or improve the quality of CTI. This indicates that practitioners are already addressing the problem of low CTI quality. Nevertheless, the quantitative and qualitative procedures varied among practitioners, indicating a lack of formalization and standardization. Furthermore, many of the experts do assessments of the quality of CTI, which are not based on specific quality metrics. Thus, the experience and expertise of the person doing the evaluations is crucial. This leads to the problem that the assessments are inconsistent, which was also pointed out by one expert who

mentioned that he and his colleague should come to the same evaluation, which is not always the case, and discussions are necessary. In addition, practitioners do a lot of vetting to ensure high accuracy and reliability of incoming CTI, which is increasing the quality.

One of our main research goals was to derive from practice advanced and established methods for determining or quantifying the quality of CTI. In this regard, we were only able to find procedures that ensure or increase the quality of CTI, as described above and in Section 4.5. Consequently, this leads to the conclusion that there is a lack of procedures to quantify or measure the quality of CTI. Hence, there is a need for automated procedures for either filtering low-quality CTI or for quantifying the quality, which would help in prioritization and focusing on the relevant CTI.

Bridging the Gap: Extending the Current Understanding of CTI Quality Assurance Practices. While existing research, such as the work by Bouwman et al. on commercial TI feeds and Zibak et al. on quality dimensions, has significantly advanced our understanding of CTI, these studies have often focused on specific subtopics within the field of CTI or did not focus on what actually is utilized by professionals in practice. Our investigation seeks to bridge this gap by providing a comprehensive analysis of the CTI ecosystem, encompassing quality characteristics and procedures to assure the quality of commercial feeds, Open Source Intelligence (OSINT), and community feeds. Recognizing the constraints faced by many security teams, either due to budgetary limitations or restricted access to specialized feeds, our study delves into the real-world practices employed by professionals across diverse organizations to navigate and enhance CTI quality. This exploration of practical applications sets the groundwork for the targeted recommendations, which we will detail in the subsequent section, aimed at both research and practice. By highlighting these practical applications, our work aims to enrich the current discourse on CTI, offering actionable insights that facilitate a deeper, more nuanced understanding of CTI quality assurance in the dynamic landscape of cybersecurity.

6 RECOMMENDATIONS FOR ACTION

Based on the findings from our expert interviews and the identified gaps in both academic research and practice, this section provides specific recommendations for improving the effectiveness and understanding of CTI quality management.

6.1 Research

To elaborate on the academic implications, this subsection offers tailored suggestions for future research efforts to enrich the body of knowledge in areas not yet addressed in the existing literature.

Application areas for CTI and specific quality dimensions. The interviews found that the shared CTI varied widely. In addition, the backgrounds of the experts and the fields of application differed considerably. In this context, we learned that the necessary quality of CTI depends on the type, intended use, and application field. Additionally, in the interviews, it became apparent that different quality aspects of CTI are important for practitioners. For example, timeliness can be important for one person or use case but not as important for another.

Therefore, there is a need to research specific and adjustable metrics for the quality dimensions described by Zibak et al. Furthermore, it is necessary to investigate the optimal match between the type of information, its use case, and the relevant quality criteria. In this regard, specific quality metrics tailored to the type of CTI and its use case need to be developed. It is also vital to delineate which information is to be evaluated. Especially since each type of information, be it IP addresses, hashes, domains, courses of action, or reports, must be considered individually. There is no universal assessment method for these different types of CTI. Moreover, even within each sector, there may be different requirements for the quality of CTI, as one expert pointed out.

Quality assurance processes and quantification of CTI quality. Even though sixteen experts apply quantitative procedures, most vet the incoming CTI and do not quantify the quality of CTI. In addition, only a few experts use procedures like overlap, correlation, or decay of IoCs. Moreover, the described procedures in Section 4.5 varied enormously among the experts and are highly dependent on their knowledge, experience, or organizational maturity. From this follows a lack of reproducibility of assessments and an overflow with low-quality CTI.

Thus, there is a need for standardized procedures in the CTI domain. Such standards would ensure consistency, improve the reliability of threat assessments across organizations, and increase the overall quality of shared CTI. By adopting a consistent approach to quantifying CTI quality, organizations can reduce the influx of low-quality CTI and improve the actionable insights they gain from their CTI.

The experts highlighted the importance of transparent calculations and simple, comparable, and understandable results. Since the intended use of CTI differs a lot, future solutions need to be tailored to specific areas of applications and types of CTI, as described above. Future research should research solutions that utilize already existing features and metrics, such as overlap, correlation, and decaying calculations. This allows immature practitioners to be reached and introduce quality assurance processes quickly and easily.

Crowd-based accuracy measurement. In the cybersecurity landscape, there is a recurring problem with validating the authenticity and accuracy of CTI, confirmed by our interviews, as verifying incoming CTI is the most commonly used process. For this purpose, future research should research how crowd-based automated verification systems could help to address this challenge. Frequently, the ground truth or the completeness is unknown. Furthermore, revealing sensitive information is another aspect that makes the validation of CTI difficult. In this context, automatically initiating an anonymized query to verify incoming CTI with other already verified sources could help solve this challenge. Future research should research matching procedures to match true and false positives across multiple data sources without revealing sensitive company information. This could help to improve and increase the visibility of the threat landscape and close the gap in CTI reports.

Research of feedback mechanisms for quality determination. The critical significance of feedback became clear in our interviews. There are existing solutions like the sightings feature in MISP or reaching out to the source. However, there is a lack

of feedback as we learned from the experts. Therefore, future research must investigate the most effective feedback methods and how they affect CTI quality in the long term. This could be facilitated through integration with existing communication or sharing platforms. Feedback could be anonymized and based on CTI quality-improving characteristics. The lack of feedback could be overcome with the research on incentives, which encourage members of sharing communities or practitioners, in general, to give feedback to the source of CTI. In this context, motivation-enhancing approaches such as gamification could help increase the motivation for giving feedback. If feedback is deployed ideally, it can be "the quality management in the hands of consumers." However, feedback must be accessible to both the consumers and the sharing entity. This can improve the tailored creation of CTI and increase confidence in unknown sources.

Research on context. As we described in Section 4.4, the experts desire different context information. However, context is still a missing property of CTI. Even though it is crucial, IoCs without context remain data, and utilizing their full potential and value is difficult. Therefore, future research should investigate the most critical context information and whether it varies by CTI type or application area. Moreover, it is important to determine what minimum standards of context information are required to make CTI data valuable, applicable, and most effective. For this purpose, creating a checklist based on a comprehensive survey would be a target-oriented method. Additionally, gamification approaches could be innovative to increase motivation for creating and sharing context.

6.2 Practice

We now want to bridge the gap between research and practice. In this subsection, we outline recommendations for practitioners based on the findings of our experts and the shortcomings identified in the field.

Quality management and measurements. To increase the accuracy and reliability of CTI, practitioners could consider adopting a two-step review approach. In this regard, there is a publisher who provides CTI and a reviewer who verifies it. Alternatively, practitioners could consider an approach used by the FS-ISAC, where a dedicated team reviews the information provided, adds context, and assesses trustworthiness. However, it should be noted that this approach incurs additional costs and may not be feasible to implement everywhere. A third approach could be to introduce crowd-based scoring. Here, members of a sharing community can provide ratings and feedback on shared CTI, which could help to increase the quality and trustworthiness.

Operationalization of feedback loops. Feedback is a great technique to ensure continuous improvement and effectiveness of CTI. Feedback on sightings, techniques observed in the wild, and their relevance to decision-making is critical. Practitioners should encourage other practitioners or members of sharing communities to share their experiences with the quality and usefulness of CTI sources. Practitioners could gain insights into the reliability of various (unknown) sources if they would share their experiences and ratings with others. The operationalization of feedback could be carried out with a simple thumbs-up/thumbs-down mechanism. In addition, structured routines could be performed with CTI sources and consumers to discuss the feedback. This promotes continuous improvement and adaptation to changing needs.

Standardization and processes. A significant gap exists in the practice regarding standardization and process management. The absence of uniform standards for generating CTI means that a consistent minimum quality is lacking. Furthermore, there is a lack of consistent standardization, especially for CTI reports. In addition, we could not find any generally adopted standardized procedures for quality assurance or quantification of incoming CTI. Against this background, we encourage addressing these existing deficits and implementing standards and processes for creating, assuring, or quantifying CTI quality. A systematic approach to standardization and assessment processes will improve the quality and consistency of CTI and strengthen confidence in incoming CTI.

Context and clarity. As already mentioned, context is a critical quality aspect of CTI that is lacking in practice, which makes it difficult to understand and utilize incoming CTI. In this regard, practitioners should implement the following recommendations to improve the quality of CTI: (1) It is crucial to provide the desired and necessary context information in coordination with the sharing partner. (2) It is indispensable to receive clear instructions on using, understanding, and prioritizing to improve the effectiveness of incoming CTI. (3) Create incentives for providing additional context.

7 CONCLUSION

We conducted 25 expert interviews to explore how practitioners deal with the issue of CTI quality. Our findings highlight that practitioners receive a large amount of CTI of various kinds. In this context, we showed the critical relevance of high-quality since bad-quality CTI leads to bad decisions. Nevertheless, the level of quality required can vary and depends on the type of CTI and its application area. The amount of context information strongly influences the quality of CTI and its effective use. Practitioners want as much context as necessary to understand the intended use of incoming CTI but also to understand attacks, properly assess the risk of a threat, and be able to defend against it. In addition, it is critical to receive accurate CTI in a timely manner.

Moreover, we derived valuable insights into various procedures of practitioners and their quality management, which can be used to develop processes or help other security teams to ensure or increase the quality of CTI. In this regard, we noticed a lack of formalized processes and procedures to quantify or measure the quality of CTI. Finally, we provide useful recommendations for action based on the interviews and our perceptions of gaps in research and practice. An undeniable finding from our research is that determining CTI quality remains a daunting challenge, aptly summarized by one expert's comment, "it is a big beast to tackle."

ACKNOWLEDGMENTS

We thank Alexandra Paulus for her help with the questionnaire, and the experts for their invaluable insights. We also appreciate the sharing communities for their altruistic contributions. This research was funded by the German Federal Ministry of Education and Research (BMBF) under the DEVISE Project.

REFERENCES

[1] G. A. Bowen. 2008. Naturalistic inquiry and the saturation concept: a research note. *Qualitative Research* 8, 1 (2008), 137–152. https://doi.org/10.1177/1468794107085301

[2] Sarah Brown, Joep Gommers, and Oscar Serrano. 2015. From Cyber Security Information Sharing to Threat Management. In *Proceedings of the 2nd ACM Workshop on Information Sharing and Collaborative Security* (Denver, Colorado, USA) *(WISCS '15)*. Association for Computing Machinery, New York, NY, USA, 43–49. https://doi.org/10.1145/2808128.2808133

[3] Hyeisun Cho, Seulgi Lee, Nakhyun Kim, Byungik Kim, and Junhyung Park. 2018. Method of Quantification of Cyber Threat Based on Indicator of Compromise. *2018 International Conference on Platform Technology and Service (PlatCon)* (2018), 1–6. https://doi.org/10.1109/platcon.2018.8472733

[4] Central University Research Ethics Committee (CUREC). 2020. *Elite and Expert Interviewing: Best Practice Guidance 03_Version 4.0.* University of Oxford.

[5] ENISA. 2013. Detect, share, protect—Solutions for improving threat data exchange among CERTs. (2013).

[6] Jelle Ermerins and Niek van Noort. 2020. *Scoring model for IoCs by combining open intelligence feeds to reduce false positives.* PhD Thesis.

[7] Thomas Geras and Thomas Schreck. 2023. Sharing Communities: The Good, the Bad, and the Ugly. In *Proceedings of the 2023 ACM SIGSAC Conference on Computer and Communications Security* (Copenhagen, Denmark) *(CCS '23)*. Association for Computing Machinery, New York, NY, USA, 2755–2769. https://doi.org/10.1145/3576915.3623144

[8] Seonghyeon Gong, Jaeik Cho, and Changhoon Lee. 2018. A Reliability Comparison Method for OSINT Validity Analysis. *IEEE Transactions on Industrial Informatics* 14, 12 (2018), 5428–5435. https://doi.org/10.1109/tii.2018.2857213

[9] George Grispos, William Glisson, and Tim Storer. 2019. How Good is Your Data? Investigating the Quality of Data Generated During Security Incident Response Investigations. *Proceedings of the 52nd Hawaii International Conference on System Sciences* (2019). https://doi.org/10.24251/hicss.2019.859

[10] James M. Hanson. 2015. The Admiralty Code: A cognitive tool for self-directed learning. *International Journal of Learning, Teaching and Educational Research* 14, 1 (2015).

[11] Cornelia Helfferich. 2011. *Die Qualität qualitativer Daten (The quality of qualitative data).* Vol. 4. Springer, Germany.

[12] Chris Johnson, Lee Badger, David Waltermire, Julie Snyder, Clem Skorupka, et al. 2016. Guide to cyber threat information sharing. *NIST special publication* 800, 150 (2016).

[13] Vadim Kartak and Nail Bashmakov. 2022. Method for Selecting Indicators of Data Compromise. *2022 International Siberian Conference on Control and Communications (SIBCON)* 00 (2022), 1–5. https://doi.org/10.1109/sibcon56144.2022.10002962

[14] 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. *Digital Threats: Research and Practice* (2023). https://doi.org/10.1145/3609230

[15] Jan Kruse and Christian Schmieder. 2014. *Qualitative interviewforschung (Qualitative interview research).* Beltz Juventa, Germany.

[16] Udo Kuckartz. 2007. Einführung in die computergestützte Analyse qualitativer Daten (Introduction to computer-assisted analysis of qualitative data).

[17] Udo Kuckartz. 2018. Qualitative Inhaltsanalyse. Methoden, Praxis, Computerunterstützung, 4. Aufl. (Qualitative content analysis. Methods, Practice, Computer Support, 4th ed.) Beltz Juventa.

[18] Marc Kührer, Christian Rossow, and Thorsten Holz. 2014. Paint It Black: Evaluating the Effectiveness of Malware Blacklists. In *Research in Attacks, Intrusions and Defenses*, Angelos Stavrou, Herbert Bos, and Georgios Portokalidis (Eds.). Springer International Publishing, Cham, 1–21.

[19] Vector Guo Li, Matthew Dunn, Paul Pearce, Damon McCoy, Geoffrey M. Voelker, and Stefan Savage. 2019. Reading the Tea leaves: A Comparative Analysis of Threat Intelligence. In *28th USENIX Security Symposium (USENIX Security 19)*. USENIX Association, Santa Clara, CA, 851–867. https://www.usenix.org/conference/usenixsecurity19/presentation/li

[20] Adobe Systems Software Ireland Limited. 2022. Audio and video editing software, Version 22.5 (Build 62). https://www.adobe.com/

[21] Bryan Marshall, Peter Cardon, Amit Poddar, and Renee Fontenot. 2013. Does Sample Size Matter in Qualitative Research?: A Review of Qualitative Interviews in is Research. *Journal of Computer Information Systems* 54, 1 (2013), 11–22. https://doi.org/10.1080/08874417.2013.11645667

[22] Kadir Burak Mavzer, Ewa Konieczna, Henrique Alves, Cagatay Yucel, Ioannis Chalkias, Dimitrios Mallis, Deniz Cetinkaya, and Luis Angel Galindo Sanchez. 2021. Trust and Quality Computation for Cyber Threat Intelligence Sharing Platforms. *2021 IEEE International Conference on Cyber Security and Resilience (CSR)* (2021), 360–365. https://doi.org/10.1109/csr51186.2021.9527975

[23] Aziz Mohaisen, Omar Al-Ibrahim, Charles Kamhoua, Kevin Kwiat, and Laurent Njilla. 2017. Assessing Quality of Contribution in Information Sharing for Threat Intelligence. *2017 IEEE Symposium on Privacy-Aware Computing (PAC)* (2017), 182–183. https://doi.org/10.1109/pac.2017.39

[24] Sami Mokaddem, Gerard Wagener, Alexandre Dulaunoy, and Andras Iklody. 2019. Taxonomy driven indicator scoring in MISP threat intelligence platforms. *arXiv* (2019).

[25] Kris Oosthoek and Christian Doerr. 2021. Cyber threat intelligence: A product without a process? *International Journal of Intelligence and CounterIntelligence* 34, 2 (2021), 300–315.

[26] Pawlinski and Kompanek. 2016. Evaluating Threat Intelligence Feeds. https://www.first.org/resources/papers/munich2016/kompanek-pawlinski-evaluating-threat-ntelligence-feeds.pdf

[27] Alex Pinto and Kyle Maxwell. 2015. Defcon22 - Measuring the IQ of your Threat Intelligence feeds. https://www.youtube.com/watch?v=uMJSOYA9xoM

[28] MISP project. 2019. Decaying of indicators. https://www.misp-project.org/2019/09/12/Decaying-Of-Indicators.html/

[29] Li Qiang, Jiang Zhengwei, Yang Zeming, Liu Baoxu, Wang Xin, and Zhang Yunan. 2018. A Quality Evaluation Method of Cyber Threat Intelligence in User Perspective. *2018 17th IEEE International Conference On Trust, Security And Privacy In Computing And Communications/ 12th IEEE International Conference On Big Data Science And Engineering (TrustCom/BigDataSE)* (2018), 269–276. https://doi.org/10.1109/trustcom/bigdatase.2018.00049

[30] Clemens Sauerwein, Irdin Pekaric, Michael Felderer, and Ruth Breu. 2019. An Analysis and Classification of Public Information Security Data Sources Used in Research and Practice. *Comput. Secur.* 82, C (may 2019), 140–155. https://doi.org/10.1016/j.cose.2018.12.011

[31] Clemens Sauerwein, Christian Sillaber, and Ruth Breu. 2018. Shadow cyber threat intelligence and its use in information security and risk management processes. *Multikonferenz Wirtschaftsinformatik (MKWI 2018)* (2018), 1333–1344.

[32] Clemens Sauerwein, Christian Sillaber, Andrea Mussmann, and Ruth Breu. 2017. Threat intelligence sharing platforms: An exploratory study of software vendors and research perspectives. (2017).

[33] Daniel Schlette, Fabian Böhm, Marco Caselli, and Günther Pernul. 2021. Measuring and visualizing cyber threat intelligence quality. *International Journal of Information Security* 20, 1 (2021), 21–38. https://doi.org/10.1007/s10207-020-00490-y

[34] Christian Sillaber, Clemens Sauerwein, Andrea Mussmann, and Ruth Breu. 2016. Data Quality Challenges and Future Research Directions in Threat Intelligence Sharing Practice. In *Proceedings of the 2016 ACM on Workshop on Information Sharing and Collaborative Security* (Vienna, Austria) *(WISCS '16)*. Association for Computing Machinery, New York, NY, USA, 65–70. https://doi.org/10.1145/2994539.2994546

[35] Rogerio Machado da Silva, João José Costa Gondim, and Robson de Oliveira Albuquerque. 2023. Methodology to Improve the Quality of Cyber Threat Intelligence Production Through Open Source Platforms. *Lecture Notes in Networks and Systems* (2023), 86–98. https://doi.org/10.1007/978-3-031-30592-4_7

[36] Florian Skopik, Giuseppe Settanni, and Roman Fiedler. 2016. A Problem Shared is a Problem Halved. *Comput. Secur.* 60, C (jul 2016), 154–176. https://doi.org/10.1016/j.cose.2016.04.003

[37] Breno Tostes, Leonardo Ventura, Enrico Lovat, Matheus Martins, and Daniel Sadoc Menasché. 2023. Learning When to Say Goodbye: What Should be the Shelf Life of an Indicator of Compromise? (2023), 8. https://doi.org/10.48550/ARXIV.2307.16852

[38] Wiem Tounsi and Helmi Rais. 2018. A Survey on Technical Threat Intelligence in the Age of Sophisticated Cyber Attacks. *Comput. Secur.* 72, C (jan 2018), 212–233. https://doi.org/10.1016/j.cose.2017.09.001

[39] Manfred Vielberth, Ludwig Englbrecht, and Günther Pernul. 2021. Improving data quality for human-as-a-security-sensor. A process driven quality improvement approach for user-provided incident information. *Information & Computer Security* 29, 2 (2021), 332–349. https://doi.org/10.1108/ics-06-2020-0100

[40] VirusTotal. 2024. YARA. virustotal.github.io/yara/

[41] Menghan Wang, Libin Yang, and Wei Lou. 2022. A Comprehensive Dynamic Quality Assessment Method for Cyber Threat Intelligence. *2022 52nd Annual IEEE/IFIP International Conference on Dependable Systems and Networks Workshops (DSN-W)* 00 (2022), 178–181. https://doi.org/10.1109/dsn-w54100.2022.00037

[42] Cagatay Yucel, Ioannis Chalkias, Dimitrios Mallis, Evangelos Karagiannis, Deniz Cetinkaya, and Vasilios Katos. 2020. On the Assessment of Completeness and Timeliness of Actionable Cyber Threat Intelligence Artefacts - Multimedia Communications, Services and Security, 10th International Conference, MCSS 2020, Kraków, Poland, October 8-9, 2020, Proceedings. *Communications in Computer and Information Science* (2020), 51–66. https://doi.org/10.1007/978-3-030-59000-0_5

[43] Adam Zibak, Clemens Sauerwein, and Andrew C. Simpson. 2022. Threat Intelligence Quality Dimensions for Research and Practice. *Digital Threats* 3, 4, Article 44 (mar 2022), 22 pages. https://doi.org/10.1145/3484202

[44] Adam Zibak and Andrew Simpson. 2019. Cyber Threat Information Sharing: Perceived Benefits and Barriers. In *Proceedings of the 14th International Conference on Availability, Reliability and Security* (Canterbury, CA, United Kingdom) *(ARES '19)*. Association for Computing Machinery, New York, NY, USA, Article 85, 9 pages. https://doi.org/10.1145/3339252.3340528

[45] Inc. Zoom Video Communications. 2022. Video conferencing software, Version 5.10.6 (5889). http://zoom.us/

A INTERVIEW MATERIAL

Showed Definitions:

Cyber Threat Intelligence: Threat information that has been aggregated, transformed, analyzed, interpreted, or enriched to provide the necessary context for decision-making processes.

Sharing Community: A sharing community is an association of two or more participants sharing CTI with other participants.

Questions related to the interviewee:

(1) In which field(s) do you work?
(2) How is your position called?
(3) How long have you been working in this position?
(4) In which context do you work with CTI?
(5) How long have you been working with CTI?
(6) Which goals do you have through CTI-sharing?

Questions related to sharing communities:

- These questions are not applicable to this work.

Questions related to CTI & quality management:

(1) What kind of CTI, shared in a sharing community, is relevant for you?
(2) How important is the quality of CTI for you?
(3) Why is the quality of CTI important for you?
(4) How do you assess the quality of CTI?
(5) Which quality metrics do you know for the assessment of CTI?
(6) Which quality metrics do you use to assess your own and external CTI?
(7) Do you use quality metrics from sharing platforms or other external service providers?
(8) Which quality metrics we have not discussed yet, do you think are relevant, and would you like to implement? Which ones would be helpful for you? (What are the reasons you can't implement them?)
(9) What do you think about those quality dimensions? (Accuracy, actionability, interoperability, provenance, relevance, reliability, timeliness)

Questions related to innovations and challenges:

(1) What innovations or actions are required to better assess the quality of CTI?
(2) Where do you see general challenges in the CTI assessment?

Closing questions:

(1) Can you think of anything else you would like to tell us?
(2) Can you recommend some interview partners?

B CATEGORIES ACCORDING TO KUCKARTZ