Threat Intelligence Quality Dimensions for Research and Practice

ADAM ZIBAK, Department of Computer Science, University of Oxford, United Kingdom
CLEMENS SAUERWEIN, Department of Computer Science, University of Innsbruck, Austria
ANDREW C. SIMPSON, Department of Computer Science, University of Oxford, United Kingdom

As the adoption and diversity of threat intelligence solutions continue to grow, questions about their effectiveness, particularly in regards to the quality of the data they provide, remain unanswered. Several studies have highlighted data quality issues as one of the most common barriers to effective threat intelligence sharing. Furthermore, research and practice lack a common understanding of the expected quality of threat intelligence. To investigate these issues, our research utilised a systematic literature review followed by a modified Delphi study that involved 30 threat intelligence experts in Europe. We identified a set of threat intelligence quality dimensions along with revised definitions for *threat data*, *information* and *intelligence*.

CCS Concepts: • **Information systems** → *Collaborative and social computing systems and tools*; *Enterprise information systems*; • **Social and professional topics** → *Management of computing and information systems*; *Computer supported cooperative work*.

Additional Key Words and Phrases: Threat intelligence, Data quality, Information quality, Intelligence quality, Delphi study, Systematic literature review

ACM Reference Format:
Adam Zibak, Clemens Sauerwein, and Andrew C. Simpson. 2021. Threat Intelligence Quality Dimensions for Research and Practice. *Digit. Threat. Res. Pract.* 1, 1, Article 1 (January 2021), 22 pages. https://doi.org/10.1145/3484202

1 INTRODUCTION

A cast of increasingly persistent and sophisticated threat actors, along with the sheer speed at which cyber attacks unfold, have made timely decision-making imperative for an organisation's security [14]. High profile incidents such as WannaCry [39] have shown the extent of the damage an attack could cause and the shortened time-window available for an organisation to put appropriate countermeasures in place.

Threat intelligence, or cyber threat intelligence, has become of increasing importance as organisations continue to generate, process and share forensic data and analytical reports around cyber threats and vulnerabilities [8]. It is also seen as an opportunity for small organisations to benefit from mature organisations' experience, as many of the former do not typically have the resources to develop an independent threat intelligence programme [53, 68]. This has prompted organisations to establish or expand threat intelligence programmes [19] by deploying sharing ontologies and intelligence management solutions [62].

A wide array of sources providing information on emerging threats, attackers' tactics and indicators of compromise (IoCs) have become available for security teams [51]. These sources range from open and commercial data feeds [51, 52] to threat intelligence service providers [4, 53, 66]. But, as the adoption and diversity of threat

Authors' addresses: Adam Zibak, Department of Computer Science, University of Oxford, United Kingdom; Clemens Sauerwein, Department of Computer Science, University of Innsbruck, Austria; Andrew C. Simpson, Department of Computer Science, University of Oxford, United Kingdom.

2576-5337/2021/1-ART1 $15.00
https://doi.org/10.1145/3484202

intelligence solutions and sources continue to grow, questions about their effectiveness, particularly in regards to the quality of the data they provide, remain unanswered.

Moreover, along with the increasing amount of intelligence available, there is a need to support human analysts with automation where possible. Appropriate structured data will be necessary for this, as will a machine-readable representations of quality.

Several studies have highlighted data quality issues as one of the most common barriers to effective threat intelligence exchange [18, 43, 61, 62, 66, 75] and information security processes are facing problems with the quality of the used input data [24]. Nevertheless, academic research has yet to systematically investigate the issues of data quality in threat intelligence. Little is known about the quality requirements and dimensions for threat intelligence data artifacts [38, 53].

In this paper we attempt to empirically investigate these issues by: (1) providing a clear and comprehensive definition of threat data, information and intelligence and (2) deriving quality dimension associated with these definitions.

The methodological approach employed in the work described in this paper consisted of a systematic literature review (SLR) followed by a modified two-round Delphi study in an attempt to bridge the gap between theory and practice. Through the SLR, we identified widely-used definitions of threat data, information and intelligence, and derived a set of literature-based quality dimensions. The preliminary definitions and dimensions served as an input to the Delphi study in which a panel of 30 experts refined and validated our results.

The remainder of this paper is structured as follows. Section 2 contextualises the research by highlighting the key theoretical concepts and related work in the fields of threat intelligence sharing and data quality. Section 3 describes the methodology employed for this study. The analysis and the results are presented in three sections: Section 4 discusses the participants' understanding of the terms threat data, information and intelligence, and introduces their revised definitions of the terms; Sections 5 and 6 present and discuss the quality dimensions. Section 7 outlines the key findings of the study. Finally, Section 8 concludes the paper and identifies areas for further research.

2 BACKGROUND AND RELATED WORK

This section aims to situate our research within the context of the *threat intelligence* and *data quality* fields (Sections 2.1 and 2.2). It also explores the attempts to address quality issues in threat intelligence (Section 2.3).

2.1 Threat intelligence sharing

Despite still being an emerging practice, threat intelligence sharing has become an essential part of organisations' information security and incident response programmes. The growing volumes of data being generated and shared, as well as the diversity of sources, prompted Dandurand and Serrano [14] to outline high-level requirements for a knowledge management infrastructure specifically designed for threat intelligence sharing. A threat intelligence management platform (TIMP) [8, 50], according to Dandurand and Serrano, aims to: accelerate information sharing; enable automation; and facilitate the generation, refinement and vetting of data through collaboration or outsourcing [14].

Proponents of these platforms argue that they help improve organisations' overall cyber security posture by allowing the streamlining of the core processes of collecting, classifying, enriching, correlating, visualising, analysing and sharing threat data from a variety of sources [19]. This enables the user to produce reports, formulate assessments and execute some actions through processes integrated into the platform.

The increasing interest in these solutions is fostering the growth of the global threat intelligence market. Many companies are currently offering a range of solutions (with varying levels of ambition) in this area. These include traditional endpoint security vendors, security service providers, and a new category of specialist threat

intelligence vendors. While some solutions serve mainly as data aggregators, others focus on providing a holistic threat management service [8, 53].

2.2 Data quality research

In its basic form, quality is defined as "fitness for use" [28] or "conformance to requirements" [13]. These definitions imply that quality is not absolute and cannot be determined independently of the consumers [63, 65]. Accordingly, *data quality* is defined as "data that are fit for use by data consumers" [63, 73].

In practice, defining dimensions or attributes that describe data quality levels has long been considered a cornerstone for most quality-related activities [3, 65]. The literature explores three approaches for identifying data quality dimension: *intuitive*, *theoretical* and *empirical* [73].

The majority of studies fall under the intuitive approach in which dimensions are defined based on common sense or practical experience [49]. The theoretical approach utilises formal models and focuses on the inconsistencies observed with the real world to define the dimensions [72]. The drawback of both of these approaches is that they fail to accurately represent the consumer's voice [73]. The empirical approach, on the other hand, utilises experiments, interviews and surveys in order to capture the data quality attributes that matter to data consumers [3, 73].

Examining the data quality literature reveals that there is no consensus on which set of dimensions defines data or information quality, nor on the precise meaning of each dimension [3]. Nevertheless, identifying data quality dimensions and metrics provide organisations with a reference framework which facilitates comparison with certain benchmark or target values [3].

2.3 Quality in threat intelligence

While the literature on intelligence quality goes back decades, data quality issues in threat intelligence in the literature is understandably a more recent and fragmented area of study.

Through a series of focus group discussions with security practitioners, Sillaber et al. [61] were the first to shed light on data quality issues facing intelligence practitioners and end users of threat intelligence management platforms. The study highlighted the need for more empirical research into the quality of the information being shared through these platforms. This study was followed by a number of attempts to investigate various aspects of threat intelligence quality. We present a list of these studies as part of our systematic literature review results in Section 6.

As discussed in Section 2.2, defining relevant quality dimensions is widely considered as the starting point for data quality research. However, in the context of cyber security, the available literature shows a lack of agreement on which data quality dimensions matter the most. This is partly due to the different understandings of what constitutes threat intelligence.

In [37] and [44], the authors claim that measuring the quality of shared threat indicators allows the assessing of the members' individual contributions to the sharing community, which could help identify free-riders. The model proposed by Chandel and colleagues in [10] factors in the size of the sharing community in assigning a quality score to the community itself. In [36] Meier and colleagues focus on ranking collections of IP addresses associated with malicious activities to determine each feed's relative quality. Other studies [7, 21, 67] deal with the comparison of different threat intelligence feeds based on selected quality criteria. Formal definitions of data quality dimensions are used by Li and colleagues in [32] to assess and compare data feeds providing IP addresses and malware hashes. They are also used by Schaberreiter and colleagues in [54] to evaluate trust in a particular source. In [55] the authors focus on how to measure and visualize cyber threat intelligence quality. A small number of authors propose criteria to evaluate and compare existing threat intelligence management platforms and feeds. In this context, they mention quality as a criterion without defining it further [4, 15].

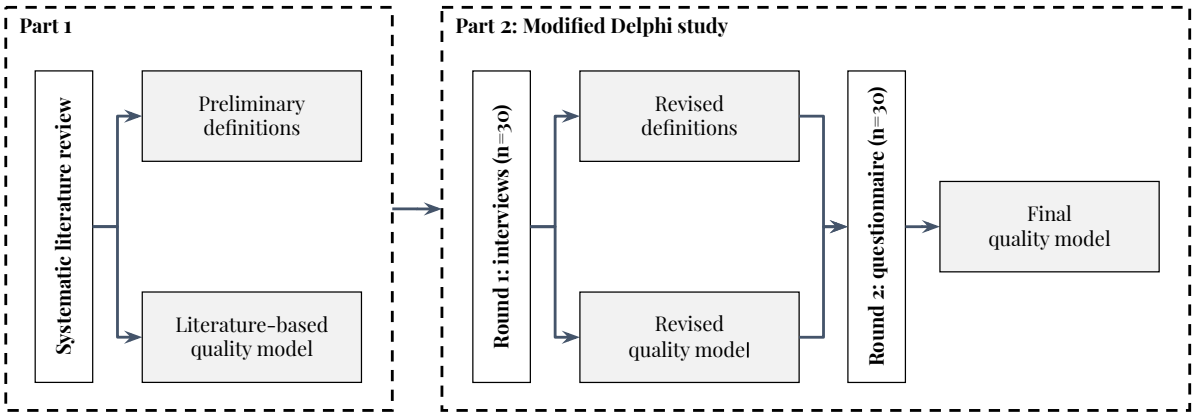

Fig. 1. Overview of the research methodology.

The majority of studies fail to involve stakeholders in determining and defining data quality dimensions. Researchers have adopted a theoretical or intuitive approach in deriving those dimensions based on their experience, previous studies or what they have deemed to be common sense. However, as discussed in Section 2.2, the empirical validity of these approaches is debatable. In this study, we adopt an empirical approach, which allows us to capture stakeholders' perspectives and define the attributes they rely on in determining whether the data is fit for use [73].

3 APPLIED RESEARCH METHODOLOGY

As depicted in Figure 1, our research methodology consisted of two parts. First, we conducted a systematic literature review to identify preliminary definitions of the terms *threat data*, *information* and *intelligence*, as well as to extract a set of literature-based quality dimensions. Secondly, we used a modified two-round Delphi study with a panel of 30 experts to refine the definitions of the terms and to revise and validate the final set of quality dimensions.

3.1 Part 1: Systematic literature review

A systematic literature review (SLR) is an explicit and reproducible research method that is typically used to systematically identify, evaluate and interpret the existing body of research pertinent to a study's area of interest [20]. We performed an SLR to extract quality dimensions mentioned in the literature. We designed and conducted the review in accordance with the guidelines proposed by Kitchenham and Charters [29], which included the following steps: creating search and selection strategy; conducting the search, selecting the papers and extracting the relevant data; and reporting the results. In this section we describe the first two steps, while the results of the final step are presented in Sections 4 and 6.

3.1.1 Search and selection strategy. A systematic search of the literature was carried out between August and September 2019 in order to identify the review's primary sources.

The following search string was derived from our research objectives and previous work, and was adapted to comply with each engine's settings: *(security) AND (("information quality") OR ("data quality") OR ("intelligence quality"))*. We decided to use this broad search term as a search string because our initial search term including the term "threat" did not yield any useful results. This might be due to the fact that there were only a few papers focusing on the quality of threat intelligence at the time of search. The search string was applied to the title,

abstract and keywords of the articles included in eight repositories: IEEE Xplore, ScienceDirect, Wiley, ACM Digital Library, AISEL, Taylor & Francis, MISQ and Springer. Given the relatively new area of research, we decided to include both journal papers and conference proceedings.

The search was limited to studies published between 2014 (the year the term *threat intelligence management platform* was coined by Dandurand and Serrano [14]) and 2019 (the year that the review took place). This yielded a total of 523 papers.

After discarding duplicates, the full list of papers identified by the searches was evaluated against our initial exclusion criteria. They were as follows.

- Papers that are written in languages other than English.
- Papers that are not available in full-text.
- Papers that are not relevant to information security.
- Papers that are not published in a peer-reviewed conference proceeding or journal.
- Papers that do not focus on quality criteria, dimensions, characteristics, attributes, models, metrics or assessment methods for data, information or intelligence.

As a result, a total of 24 papers were selected. A preliminary review of the papers' titles, followed by a complete or partial reading of the articles that had not been excluded in the previous phase, led to the identification of 11 irrelevant articles. Using the remaining 13 studies as a foundation, a snowballing search [74] was conducted. The reference list of each of the 13 articles, as well as their citations, were reviewed for additional potentially relevant articles. A further nine relevant articles met our inclusion criteria and were added to our sources list.

Finally, 22 papers were selected for the next phase of data extraction. Of these studies, 12 focused mainly on threat intelligence while the rest focused on quality dimensions in other aspects of information security but were considered useful to our research.

3.1.2 Data extraction. Each of the 22 selected papers was read thoroughly by two of the co-authors to extract relevant quality dimensions. Where available, we also recorded the corresponding definition or context for each of the dimensions as stated in the studies to better understand what they meant. This allowed us to merge some dimensions with identical or similar meanings in order to simplify the resulting dimensions. For example the *Volume* dimension in [33] was merged with *Amount of data* dimension mentioned in [57] into a single dimension.

According to each paper's stated focus, we categorised the identified dimensions into three groups: data quality dimensions, information quality dimensions and intelligence quality dimensions.

3.2 Part 2: Modified Delphi study

We chose the Delphi method [34] to systematically capture expert opinions with the aim to refine and validate our initial literature-based consideration of quality. The Delphi method is an iterative process through which a group of subject experts is consulted anonymously over several rounds [34]. In each round, they receive feedback and have the opportunity to revise their assessments. The process usually concludes when a defined consensus is reached [31]. In this study, consensus was defined a priori as agreement between the experts on rating each quality dimension within a specific round. We set 66.7% as a minimum level of agreement on any particular dimension in order for it to be included. The consensus rate is in line with similar Delphi studies in which the researchers viewed the two-thirds cut-off as a statistically significant threshold for consensus [11, 30, 69].

Although the number of rounds varies depending on the purpose of research, two or three iterations are usually sufficient [16]. Our study is a modified two-round Delphi study since the first round is structured around the results of the literature review [1].

Potential expert panel participants were selected from the Open Web Application Security Project (OWASP) and several European security interest groups, as well as through LinkedIn and personal contacts. Participant selection was based on the following criteria: hands-on experience in threat intelligence; availability and willingness to

participate; and working for an organisation that operated globally. In total, around 78 invitations were sent, resulting in 30 experts agreeing to participate.

Table 1 provides an overview of the panel experts and their employers. At the time of the study, most of the participants were employed at multinational companies operating in the information, communication, financial or insurance sectors. Over half of the experts (53.3%) were located in the UK, with the rest spread across six other European countries. About two-thirds of the participants (68%) held managerial positions. The majority (73.3%) had more than five years of experience in cyber security and almost all of them (93.3%) had over two years experience in threat intelligence. Two-thirds of the organisations (67%) had a security operations centre (SOC) in place and a similar share (73%) were using a threat intelligence sharing platform.

3.2.1 Round 1. The first round took place between October 2019 and March 2020, during which we interviewed each expert individually. A detailed protocol was developed to guide the interviewer and to guarantee the reproducibility, objectivity and comparability of the interviews. In addition to background questions related to the participants and their organisations, each interview was split into two main parts.

The first part focused on the expert's understanding of the terms *threat data*, *information* and *intelligence*. To achieve this, we presented the participants with a widely-used definition from the literature for each of the terms, and asked them to indicate their level of agreement with each one of them on a five-point Likert-type scale with 1 = 'strongly disagree' and 5 = 'strongly agree'. This was followed by an open-ended discussion of the meaning and use of each of the terms. The three definitions were chosen based on the work of [46] where the researchers synthesise the literature and discuss the strengths and shortcomings of multiple definitions. They are also in line with the commonly accepted depiction of traditional intelligence activity as a progressive refinement of data and information [27].

Interview transcripts were analysed to produce qualitative summaries and extract information relevant to our research [9]. This resulted in an updated version of our definitions of threat data, information and intelligence, and a revised set of quality dimensions.

3.2.2 Round 2. The second round was designed to be more specific. All 30 experts were invited to participate in the second round. We presented the panelists with the revised set of quality dimensions that resulted from the previous round. Using a questionnaire, we asked each of them to indicate the extent to which they agreed with each quality dimension. They were also asked to rank these dimensions in order of importance. Their agreement was measured on a five-point Likert-type scale with 1 = 'strongly disagree' and 5 = 'strongly agree'. The percentages of agreement and disagreement were calculated, as were the median and interquartile range of all ratings. As mentioned in Section 3.2, consensus was defined a priori as agreement between the experts on rating each quality dimension within a specific round. We set 66.7% as a minimum level of agreement (LA), which meant that consensus was achieved when at least two-thirds of the experts agreed or strongly agreed to include the dimension ('4' or '5' on a five-point Likert-type scale). Experts reached agreement on the refined set of dimensions in the second round with all dimensions achieving consensus and, consequently, it was not necessary to conduct a third round [69]. Based on the respective ranking of the dimensions by each participant, the mean value of the ranking per each dimension was calculated. The results were used to derive an overall ranking of the quality dimensions for threat data and information.

4 DEFINING THREAT DATA, INFORMATION AND INTELLIGENCE

We noticed that the terms threat data *threat data*, *threat information* and *threat intelligence* are being used inconsistently in research and practice (e.g. [46] and [35]). Furthermore, the generation of threat intelligence can be seen as an iterative process in which threat data is transformed into information and subsequent intelligence [17]. Accordingly, there needs to be a clear distinction between these concepts in order to facilitate the discussion

Variable		Frequency (n=30)	Percentage
Sector	Information or communication	13	43.3%
	Finance or insurance	9	30.0%
	Manufacturing	5	16.7%
	Food or hospitality	2	6.7%
	Utilities	1	3.3%
Location	United Kingdom	16	53.3%
	Austria	4	13.3%
	Switzerland	4	13.3%
	Netherlands	2	6.7%
	Germany	2	6.7%
	Ireland	1	3.3%
	Italy	1	3.3%
Number of employees	>1,000	23	76.7%
	250–999	2	6.7%
	50–249	1	3.3%
	<50	4	13.3%
Current position	Security manager	12	40.0%
	Security analyst	5	16.7%
	Chief information security officer (CISO)	5	16.7%
	Security engineer	4	13.3%
	Chief information officer (CIO)	2	6.7%
	Chief client officer (COO)	1	3.3%
	SOC team leader	1	3.3%
Experience in cyber security	<2 years	0	0.0%
	2–5 years	8	26.7%
	>5 years	22	73.3%
Experience in threat intelligence	<2 years	2	6.7%
	2–5 years	18	60.0%
	>5 years	10	33.3%
Organisation's threat intelligence activity	Threat intelligence producer	0	0.0%
	Threat intelligence consumer	8	26.7%
	Threat intelligence producer & consumer	22	73.3%
Does your organisation operate a SOC?	Yes	20	66.7%
	No	10	33.3%
Does your organisation use a TIMP?	Yes	22	73.3%
	No	8	26.7%

Table 1. Overview of the Delphi panel experts and their organisations.

Definition	Strongly disagree	Disagree	Neutral	Agree	Strongly agree
Data	1 (3.30%)	1 (3.30%)	6 (20.0%)	19 (63.3%)	3 (10.0%)
Information	1 (3.30%)	5 (16.7%)	4 (13.3%)	18 (60.0%)	2 (6.70%)
Intelligence	0 (0.00%)	3 (10.0%)	5 (16.7%)	15 (50.0%)	7 (23.3%)

Table 2. Experts' agreement with each of the literature-derived definitions.

around the quality criteria associated with each of them. Therefore, the first goal of our study was to explore the experts' understanding of each of the three terms and create a common understanding.

As shown in Table 2, the majority of the experts either agreed or strongly agreed with the literature-based definitions of threat data, information and intelligence (73.3%, 66.7% and 73.3% respectively). However, during the discussion a number of experts highlighted aspects of the definitions that they did not agree with and suggested some changes.

4.1 Threat data

The authors of [46] define threat data as: "lower-level raw logs that have been produced by sensors such as payloads hash values, network artifacts, internet protocol (IP) addresses and uniform resource locators (URLs)."

Participants generally agreed with this definition albeit with three caveats. First, although threat data is primarily raw, lower-level data points such as indicators of compromise (IoCs), it does not always have to be produced by sensors. Secondly, in order for the data to be considered as threat data it has to be related to some sort of malicious activity. Finally, according to the participants, threat data is predominantly machine-readable. Examples of threat data include SHA1 and MD5 hashes that correspond to specific suspicious files or samples of malware; IP addresses of suspected command and control servers; and network artifacts that identify malicious activity from that of legitimate users [6].

Based on the participants' comments and suggestions, we propose the following revised definition of threat data:

> **Revised definition:** Threat data is a machine-readable set of raw recorded facts that can help in identifying or mitigating malicious activity in a system or network. It may include indicators of compromise such as malware signatures, IP and URL addresses, file and domain names, or registry keys.

4.2 Threat information

In the same report [46], threat information is defined as: "data that have undergone additional processing to provide enhanced high-level insight that may help decision makers in reaching a well informed decision."

Overall, participants agreed with the notion that threat information is data enriched with context. However, they stated that it does not necessarily provide high-level insights or allow for well-informed decisions. They also challenged the idea that threat information is strictly the result of extra processing of the data stating that threat data grouped together or with some context is usually enough to make the information relevant to the organisation. For example, a series of raw logs (threat data) collated together indicates a spike in suspicious activity (threat information) [12].

A couple of experts stated that it is difficult to differentiate between data and information and therefore did not see value in distinguishing between the two in practice. Instead they offered a simpler model that only differentiates between intelligence inputs and outputs. In light of the discussion above, we adopt the following definition for threat information:

> **Revised definition:** Threat information is information pertaining to a threat to, or vulnerability of, a system or network. It is the result of contextualising and interpreting threat data.

4.3 Threat intelligence

A widely-cited definition of threat intelligence in the literature is the one presented in [35], where the term is defined as: "evidence-based knowledge, including context, mechanisms, indicators, implications, and action-oriented advice about an existing or emerging menace or hazard to assets. This intelligence can be used to inform decisions regarding the subject's response to that menace or hazard."

The majority of the experts agreed with this definition noting that it adequately captures the meaning and function of intelligence in practice. Nevertheless, they stressed the importance of manual analysis and therefore the analyst's role in vetting, analysing, interpreting and applying hypotheses to the information. A number of participants referred to the forward-looking nature of threat intelligence, emphasising that threat intelligence involves not only impact assessments and recommendations but also requires deduction and prediction. One participant raised doubt about using the word "knowledge" in the definition, arguing that intelligence is more about reaching a hypothesis that one could agree or disagree with and that it does not have to be 100% knowledge or truth per se. For example, indication of a suspicious activity, when contextualised with information on prior incidents involving similar activity, could allow for the development and deployment of a mitigation strategy to stop the incident [12].

In reflecting on the experts' feedback we build on the definition introduced in [35] to formulate the following definition:

> **Revised definition:** Threat intelligence is evidence-based forward-looking assessment including context, implications, and action-oriented advice about a threat to, or vulnerability of, a system or network. This intelligence is produced through the application of individual or collective cognitive methods, and can be used to inform decisions regarding the subject's response to that threat or vulnerability.

5 IDENTIFYING THREAT INTELLIGENCE QUALITY DIMENSIONS

In this section we first report the results of the systematic literature review in the form of a literature-based quality dimensions in Section 5.1. This set of dimensions was fed into the modified two-round Delphi study where the experts panel reviewed and refined the quality dimensions. We discuss their adjustments in Sections 5.2 and 5.3. The panel reached consensus in the second round resulting in a final set of quality dimensions presented in Section 5.3.

5.1 Results of the systematic literature review

The systematic literature review, described in Section 3.1, delivered a final set of 22 papers discussing quality dimensions in the context of cyber security. We extracted 32 quality dimensions in total as listed in Table 3. The most common dimensions were: timeliness (n=14), completeness (n=13), accuracy (n=12), consistency (n=8), relevance (n=8) and reliability (n=7). Both uniqueness and understandability were mentioned in four different studies, whereas the remaining quality dimensions were mentioned once or twice. Table 3 also records whether the dimensions were mentioned in the context of data, information, intelligence, or any combination of the three.

In order to simplify the preliminary set of dimensions, we decided to only include the dimensions that were mentioned in all three contexts: data, information and intelligence. Figure 2 shows the intersections of the three groups of dimensions, with the grey box showing those mentioned in all of them. With the exception of uniqueness and interoperability, the other five dimensions were among the ones mentioned most frequently.

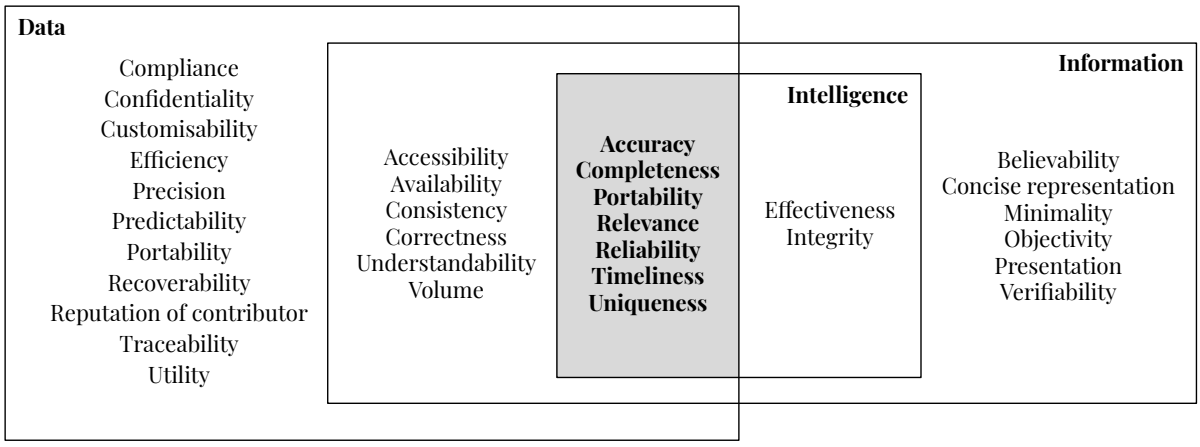

Fig. 2. Literature-based quality dimensions grouped by scope.

The preliminary set of dimensions along with their definitions as stated in the ISO/IEC 25012 standard [26] are as follows:

- **Accuracy:** The degree to which threat intelligence has attributes that correctly represent the true value of the intended attribute of a concept or event in a specific context of use.
- **Completeness:** The degree to which threat intelligence contains all expected information and ensures that no information is missing.
- **Interoperability:** The degree to which threat intelligence is in a form that allows it to be imported, replaced and moved into an organisation's information management systems while preserving its existing quality.
- **Relevance:** The degree to which threat intelligence is useful for the purposes for which they were collected.
- **Reliability:** The degree to which threat intelligence comes from legitimate sources and has not been tampered with.
- **Timeliness:** The degree to which threat intelligence has attributes that are of the right age in a specific context of use.
- **Uniqueness:** The degree to which threat intelligence is similar to previously seen threat intelligence from other sources.

5.2 Results of Delphi round 1

The aforementioned set of literature-based quality dimensions served as a starting point for the first round of the modified Delphi study. The results of this round are manifested in the following three modifications to the dimensions.

5.2.1 Dropping the uniqueness and completeness dimensions: After careful examination of the proposed literature-based quality dimensions, the majority of experts saw little importance in including the uniqueness and completeness dimensions.

Threat intelligence received from one source only often calls for further investigation to ensure that it is not the product of faulty analysis. Despite being promoted as a selling point by some vendors, for consumers, unique data or intelligence plays little, if any, role in determining the overall quality of the product. Truly unique and reliable intelligence according to the panel can only be delivered if a provider has access to unique sources or capabilities, which is applicable mostly to government intelligence agencies. The experts also stressed that unique

	Mohaisen et al. [37]^{x}	Meier et al. [36]^{x}	Park et al. [44]^{x}	Qiang et al. [47]^{x}	Rashid et al. [48]^{y}	Chandel et al. [10]^{xy}	Li et al. [33]^{x}	Wagner et al. [70]^{xyz}	Sillaber et al. [61]^{x}	Li et al. [32]^{xyz}	Schaberreiter et al. [54]^{xyz}	Wagner et al. [71]^{x}	Montesdioca et al. [40]^{y*}	Talha et al. [64]^{x*}	Shamala et al. [57]^{y*}	Sicari et al. [58]^{x*}	Sillaber et al. [60]^{xy*}	Sillaber and Breu [59]^{y*}	Bertino et al. [5]^{y*}	Phua et al. [45]^{x*}	Mu et al. [41]^{x*}	Grispos et al. [22]^{xy*}	**Frequency**
Timeliness		■		■	■	■		■	■		■		■	■	■	■		■			■	■	14
Completeness		■		■	■		■	■			■		■	■	■	■	■		■				13
Accuracy		■		■			■	■	■		■			■	■	■		■			■	■	12
Consistency					■									■	■		■	■	■		■	■	8
Relevance	■		■	■	■			■				■			■				■				8
Reliability					■	■		■	■	■			■		■								7
Understandability									■					■	■		■						4
Uniqueness	■		■			■					■												4
Correctness	■		■						■														3
Accessibility					■									■									2
Availability														■				■					2
Precision																	■		■				2
Utility	■		■																				2
Compliance											■			■									2
Portability											■					■							2
Verifiability											■				■								2
Believeability															■								1
Concise representation															■								1
Confidentiality														■									1
Customisability				■																			1
Effectiveness						■																	1
Efficiency				■																			1
Ingestibility								■															1
Integrity					■																		1
Minimality																			■				1
Objectivity															■								1
Predictability				■																			1
Presentation															■								1
Recoverability																■							1
Reputation									■														1
Traceability																■							1
Volume																■							1

Table 3. Quality dimensions extracted from the literature.

Scope: ^{x} data; ^{y} information; ^{z} intelligence; * non-TI.

intelligence should not be confused with bespoke intelligence products resulting from a specific request by the client.

Similarly, completeness in an absolute sense is difficult to determine, attain or measure. Alternatively, intelligence completeness according to the experts is mainly about the totality of the assessment, in terms of how methodical and rigorous the analysis is. This is usually accomplished by following a certain model like the kill chain [25] to provide a reasonable amount of coverage. Pursuing intelligence completeness according to the panel could undermine its timeliness and actionability. By the time more information is collected and a comprehensive analysis is conducted, the intelligence might become less actionable. As for lower-level data like IoCs, all else being equal, a feed with more data is potentially better than others, but no one expects one source to provide a 100% complete picture. Organisations with more resources minimise these limitations by purchasing or subscribing to more than one service.

Given their minimal impact on determining intelligence quality and following the experts' recommendations, we decided to exclude the two dimensions.

5.2.2 Adding the actionability and provenance dimensions. When asked about what quality dimensions are missing, the majority of experts referred to two main dimensions: actionability and provenance. Further discussion with the panelists highlighted the importance of these two dimensions in evaluating the quality of threat intelligence and therefore were added to the modified set of dimensions in the second round. We discuss the actionability and provenance dimensions in Section 5.3.

5.2.3 Distinguishing between threat data and threat intelligence. In commenting on the validity of the literature-based quality dimensions, the experts highlighted the need to differentiate between threat data and threat intelligence when it comes to determining quality dimensions. This is in line with our discussion in Section 4 where we distinguished between the two terms and the different purposes they serve. Threat information, on the other hand, was excluded as the experts did not see a practical need to evaluate its quality separately. Therefore, the next round sought to capture the participants' agreement with the modified set of quality dimensions in the context of threat data and threat intelligence independently of each other.

5.3 Results of Delphi round 2

All 30 experts from the previous round participated in the second round. The experts were asked whether they agreed with the modified dimensions of the previous round. They were also asked to rank the dimensions in order of importance for each of threat data and threat intelligence. The responses to each of the dimensions were tested against our predefined definition of consensus: LA $\geq$ 66.7%. As shown in Table 4, a third Delphi round was not required because consensus was achieved on all dimensions. Consequently, the final set of threat data and intelligence quality dimensions are: *accuracy*, *actionability*, *interoperability*, *provenance*, *relevance*, *reliability* and *timeliness*. In the following section we discuss each of these dimensions and their significance to threat intelligence practice.

6 THREAT INTELLIGENCE QUALITY DIMENSIONS

In this section we discuss our threat intelligence quality model. The final set of quality dimensions and their definitions are presented in Table 5.

6.1 Accuracy

In an absolute sense, accuracy is the degree to which information has attributes that correctly represent the true value of the intended attribute of a concept or event in a specific context of use. In the context of threat intelligence, inaccurate and false-positives might result in undesired effects and wasted resources [56]. However,

Dimension	Threat Data					Threat Intelligence				
	N_d	N_a	LA	M	IQR	N_d	N_a	LA	M	IQR
Accuracy	0	26	86.7%	5	1	0	30	100%	5	1
Actionability	1	23	76.7%	4	1	1	29	96.7%	5	1
Interoperability	3	25	83.3%	4	1	6	20	66.7%	4	2
Provenance	0	23	76.7%	4	1	0	21	70.0%	4	1
Relevance	2	24	80.0%	5	1	0	29	96.7%	5	1
Reliability	0	26	86.7%	4	1	0	25	83.3%	4	1
Timeliness	0	29	96.7%	5	1	1	26	86.7%	5	1

Table 4. Experts' agreement with the dimensions in Delphi round 2.

N_d : Number of experts who disagreed or strongly disagreed.
N_a : Number of experts who agreed or strongly agreed.
LA : Level of agreement ($LA = N_a/n * 100, n = 30$).
M : Median, IQR : Interquartile range.

Dimension	Definition
Accuracy	The degree to which threat data or intelligence is correct, objective or without false-positives.
Actionability	The degree to which threat data or intelligence allows a decision to be made or action to be taken without the need for further analyses.
Interoperability	The degree to which the formats of threat data or intelligence is compatible with consumers' internal systems allowing it to be accessed and integrated seamlessly.
Provenance	The degree to which a threat intelligence consumer is able to track the evolution of a piece of threat data or intelligence including its origins and the process through which it was produced.
Relevance	The degree to which threat data or intelligence meets the consumer's specific requirements and is useful for the purpose for which it was produced.
Reliability	The degree to which the source of threat data or intelligence is trustworthy, authentic and competent.
Timeliness	The degree to which threat data or intelligence is available without delay in a specific context of use.

Table 5. Final set of quality dimensions for threat data and intelligence.

the experts stated that, in practice, determining the absolute truth is difficult if not impossible. Accordingly, alternative interpretations of accuracy were elicited.

For threat intelligence producers, accuracy means ensuring the objectivity of the analysis and that the way the product was communicated enables the message to arrive and be understood the way that it was meant to. Although a low number of false-positives in IoCs is one indication of the accuracy of threat data, the panel agreed that consumers do not expect a vendor to provide 100% accurate information in an absolute sense. However, in order to determine the relative accuracy of the intelligence product, they ask how the data and intelligence were

collected, whether the analyst come up with the right hypothesis, and what evidence supports the conclusions. Analysts have some latitude based on their own experiences but it is crucial to ensure that their analysis is logical and they do not leap to conclusions.

The need to distinguish between assessments and facts is crucial prompting practitioners to adopt traditional intelligence language that conveys confidence levels such as the NATO or Admiralty coding [23] that is used by some vendors to indicate what they perceive as accurate and reliable information.

6.2 Actionability

According to the panel of experts, actionability is the extent to which threat data or intelligence allows a decision to be made or action to be taken without the need for further analysis.

The production of actionable intelligence in practice usually requires an analyst to provide some recommendations or courses of action. And there is always one course of action, which is to do nothing. So, in reality, what one is asking intelligence analysts to do is to make calls, regardless of whether they are right or wrong. This allows the consumer to use or deploy the intelligence directly before it loses its value. In practice, what makes intelligence actionable can take a range of forms. For high-level analytical reports it could mean that the consumer is going to use that intelligence the next time they make an important decision about security architecture. For lower-level threat data, it could mean using these indicators to block a threat, prioritise patching or develop mechanisms to detect specific adversary tactics, techniques and procedures (TTP).

However, in reality (according to some experts), if the intelligence producer understands their client's requirements, then everything they produce should be actionable in one way or another.

6.3 Interoperability

The increasing volumes of threat data and diversity of sources have elevated the importance of interoperability in threat intelligence. For organisations using more than one service or vendor, it is vital to be able to access and process all of the data in one place and in a timely manner. Similarly, vendors realise that offering products that are aligned to particular standards is important to allow consumption across their clients' different platforms.

The importance of interoperability varies depending on the exact nature of the intelligence. For machine-readable threat data like IoCs, it is crucial to be able to ingest them smoothly and with minimal analyst effort. However, for higher-level analytical reports, recommendations or assessments, the importance of being able to ingest that into a platform is less critical.

The panel also reinforced the importance of standardised formats in ensuring data is exchanged in consistent and machine-readable manner. The Structured Threat Information Expression (STIX) language [2] continues to attract traction and support.

Threat intelligence platforms and APIs continue to evolve and provide new capabilities including auto-scrubbing and auto-validation of indicators in order to prevent false positives or neutral indicators from being added automatically. However, despite the increasing reliance on platforms and APIs, some of the experts on the panel stated that some of the APIs of commercial vendors are not well documented (or not as well as the consumers would like them to be), which in some cases makes it very complex to pull files and generate reports.

6.4 Provenance

Provenance (or traceability) according to the experts panel is the extent to which a threat intelligence consumer is able to track the evolution of a piece of threat data or intelligence including its origins and the process through which it was produced. Moreover, it ensures the integrity of the intelligence during the iterative revisions in collaborative networks. For example, it makes it possible to trace which participant has made changes at which point in the production of the intelligence. Good provenance allows the consumer to trace the intelligence in the

same way the author did based on the evidence provided. This includes the source of information at a granular level and detailed documentation of the processes through which it was produced. A provenance chain would also show what analytical models were used and what hypotheses were tested in formulating the conclusion or assessment.

Provenance is also seen as an important factor in establishing trust and determining usefulness of a vendor. It allows the consumer to identify vendors that merely serve as information aggregators giving rise to issues such as circular reporting. However, establishing provenance is a complex problem, and the community is yet to establish a standardised provenance process, which could be the reason why this dimension is often overlooked.

6.5 Relevance

Overall, the experts agreed that relevance is an important dimension in evaluating the quality of threat intelligence. Increased noisy and irrelevant data result in wasted resources and time. However determining what makes a particular piece of data relevant or not cuts across the organisation's industry, sector, geography, technologies in use, its assets, etc.

In addition to the organisation's business activity, the amount of allocated resources for its threat intelligence programme plays a vital role in determining how tailored the received intelligence needs to be. In big organisations including those operating in multiple sectors, a dedicated internal team collects as much data, information and intelligence as possible before translating it into intelligence that is relevant for the organisation's different business entities.

Our experts agreed that irrelevant intelligence can often be attributed to a failure to understand the consumer's requirements. However, intelligence producers sometimes face cases where their clients have an immature threat intelligence programme and the clients themselves are unable to identify their intelligence needs, which further complicates the producers.

6.6 Reliability

According to the Delphi experts, determining the reliability of a source is critical in deciding whether to rely on the information received or not. The inferential value of the assessments or conclusions is constantly considered with the reliability of their source in mind. The experts stated that the reliability of a source encompasses other factors including its trustworthiness, authenticity competence and objectivity.

Assessing a source's reliability in the context of threat intelligence is currently a multi-layered process and involves subjective elements. The process, according to the experts, should take into account the following three considerations: the historical reliability of the source for similar incidents or topics including its own confidence in the intelligence; the intelligence's consistency with known facts or confirmed post-mortem findings; the intelligence's consistency with information gleaned from other sources. Therefore, perceived reliability in a source is established over time, throughout the progression of the analyst's experience and encounters with the source, as well as the organisation's overall experience. A source with a demonstrated track record of accurate and credible reporting is perceived as more reliable than an unknown source. For example, experts reported high levels of perceived reliability of official sources such as national intelligence agencies and national CERTs, as opposed to a new or untested vendor.

It should be noted that, despite the importance of source reliability, the experts pointed to the challenge of ensuring source anonymity in some cases, as highlighted by Murdoch and Leaver [42]. They also noted that the reliability of a source is not constant and therefore continuous re-evaluation might be required. This is in line with the findings of Schaberreiter and colleagues who present a method that allows an organisation to re-evaluate trust in one or more sources every time it receives new threat intelligence [54].

6.7 Timeliness

Timely decision-making is imperative for an organisation's security [14]. Accordingly, the experts overwhelmingly agreed that timeliness is one of the most important dimensions in evaluating the quality of threat data and intelligence. Organisations are constantly looking for new ways to reduce the delay in receiving important information and in acting upon it.

Although the value of intelligence does not drop to zero if it is late, it could diminish significantly. The shelf life of the intelligence also depends on the exact type of the information being shared. For example, threat data like malware signatures have a relatively short shelf life as they are constantly developing. An organisation responsible for providing those to a consumer needs to ensure they are delivered in a timely manner. However, a more strategic intelligence assessment that looks at 3–6 months of the threat landscape is still timely but could be delivered later. The experts also pointed to the inconsistency in the metadata surrounding the intelligence received by some sources. While some sources state both when the data was observed and when it was reported, others might fail to do so.

7 DISCUSSION

The analysis of the results of the systematic literature review and the modified Delphi study provided insights on certain dimensions and metrics used to assess the quality of threat data, information and intelligence. In this section we discuss the key findings and reflect on the limitations of our research. In doing so we: (1) provide an experts' ranking of the quality dimensions; (2) discuss why measuring intelligence quality remains a challenging issue; and (3) argue that the nature of quality is inextricably linked with the task of defining requirements.

7.1 Determining dimensions priorities

For each of threat data and threat intelligence, we asked the panel's experts to rank (top-to-bottom) the final set of quality dimensions in order of importance. The results as depicted in Figure 3 show that the importance of a dimension varies depending on whether it is being used to assess the quality of threat data or threat intelligence.

Threat data as discussed in Section 4 is used to monitor and detect threats and vulnerabilities. The shelf life of threat data is often short and its value diminishes considerably with time. Data feeds are predominantly machine-readable and are expected to be easily processed and integrated into different platforms. Therefore, it is not surprising that the panel considered timeliness, relevance and interoperability as the three most important quality dimensions.

On the other hand, the three most important threat intelligence quality dimensions are considered to be relevance, actionability and timeliness. As examined in Section 4, many experts stressed that threat intelligence by definition is forward-looking and involves applying hypotheses and making recommendations. Actionable, relevant and timely intelligence therefore is understandably key to avoid waste of resources and overburdening security operation centres.

7.2 Measuring intelligence quality remains a challenging issue

As part of the cost-benefit analysis of their threat intelligence programmes, organisations would ultimately like to know how intelligence products are being used and which paid source is providing good quality intelligence products. Discussions with the experts revealed that, despite growing interest in the idea, stakeholders are yet to identify and develop suitable metrics to measure any of the quality dimensions and most of them do not have a specific process to filter and evaluate threat intelligence based on these dimensions. Moreover recent research introduced the first set of formal metrics to assess threat intelligence quality, but pointed out that these metrics might be subjected to changes as more knowledge is gained about threat intelligence processes, platforms and stakeholders [55].

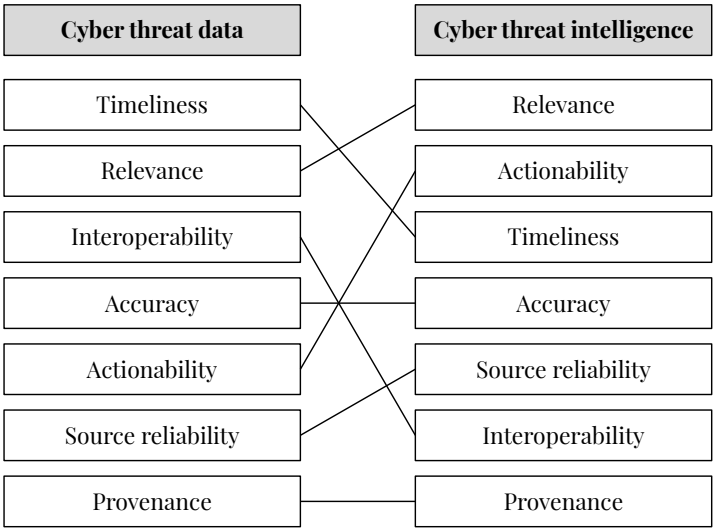

Fig. 3. Experts' ranking (top-to-bottom) of quality dimensions in order of importance.

Simple quantitative metrics like the number of intelligence reports received or number of IoCs fed into a platform are considered unhelpful in measuring intelligence quality. They are however useful for the consumers to know how much data they will receive and the quantity of resources required to process it. Traditional intelligence evaluation frameworks like the Admiralty System are used in some cases to indicate the reliability of the source and the credibility of the intelligence. Whether the intelligence has been validated by other sources, or the threat has materialised, inaccurate elements as well as the source reputation and history are factors that can negatively or positively affect the evaluation of the source.

The experts also stated that consumer feedback is essential for vendors in order to identify areas for further improvement. Feedback collection ranges from informal conversations to more rigorous and regular processes such as questionnaires attached to intelligence reports. Through these processes, vendors solicit feedback on the quality of their products, systems and services. Metrics integrated into the platforms (e.g. number of times a report was read or downloaded) would also be considered. However, despite being an essential part of the traditional intelligence cycle [27], the feedback loop according to some experts does not exist in commercial intelligence since only few vendors actually implement it, and, where it does, it is slow and inefficient.

From a consumer point of view, measuring the quality of a threat intelligence product is intertwined with determining its effectiveness. This is typically achieved by examining the decisions that were made based on the received intelligence and the impact of these decisions on preventing exploitation or reducing vulnerability. The percentage of intelligence that has led to control changes (e.g. deciding to monitor or stop monitoring a specific part of the network), or the number of incidents detected or foiled because of intelligence received from a specific source, are examples of metrics that could be employed to determine the effectiveness of a threat intelligence service. One way to achieve this is by using a tagging system that enables an organisation to track the number and nature of actions and provide it with a direct link back to the source of the intelligence. This would allow it to identify the most used sources and compare their impact to their cost.

The study's panelists stated that threat intelligence vendors and their clients should work together to develop and implement some of these metrics. However they warned that issues such as privacy, confidentiality and

willingness can hinder the progress of the collaboration. Metrics development and tracking also requires dedicated efforts and resources adding more work to already stretched and understaffed cyber security teams.

7.3 Intelligence quality is influenced by its iterative production process

According to the authors of [17] and [15] the production of threat intelligence can be defined as an iterative process called the intelligence cycle. This process includes: planning, collection, processing, analysis, production, dissemination and feedback. The last of these occurs continuously throughout the intelligence process and steers it. This iterative process can be crucial for dimensions like accuracy or completeness as these quality dimensions might be improved through several iterations. Moreover, as the cycle transforms threat data into intelligence it clearly shows why a differentiation of quality criteria for threat data and intelligence is necessary.

However, recent research and our investigations showed that current production of threat intelligence lacks a common understanding and implementation of a formal process like the intelligence cycle. In this context Oosthoek and Doerr showed that describe current threat intelligence as a product without a process [43].

7.4 Determining quality is intertwined with defining requirements

A common view amongst the panel experts is that a high-quality intelligence product is one that aligns closely with the consumer's requirements. An organisation does not necessarily need to know which of the vendors provide the most accurate intelligence, but rather which one of them is the most useful, that is to say, which vendor's services satisfy the organisations' requirements and business area. The experts also argue that in reality if the provider knows their client's requirements well then most intelligence products should be of suitable quality. Consequently, producing an irrelevant or unactionable intelligence product is mostly attributed to a failure at the requirements level.

Requirements are ideally captured at the beginning of the interaction and reviewed regularly through ongoing conversations and requirement exercises. Setting the requirements entails understanding what the business wants from the threat intelligence products. This requires vendors to understand the risks and the worries the client has, the industry's threat landscape, underlying infrastructure, and what they can do to try alleviate some of these concerns or allow the client to make more informed decisions.

However, given that it is a relatively new area, most organisations still do not fully understand how to consume threat intelligence and are therefore unable to accurately identify and convey their needs. Organisations with no existing threat intelligence programme expect the provider to fully understand the organisation's environment including its critical assets, personnel, systems, third party suppliers, and so on. A number of experts also pointed out that expectations from threat intelligence vendors are high and somehow unrealistic, and cautioned that threat intelligence services are not a panacea.

7.5 Limitations

Although the study has several limitations, the decisions taken during the planning and execution attempted to mitigate or minimise their impact.

Similar to any other literature review, the review described in this paper is limited by the search terms used and the selected scientific databases. Given that the terms data and information are sometimes used interchangeably, we have included both of them in the search string along with their disjunctions across eight academic databases. Nevertheless, there still can be some relevant publications that have not matched our search criteria.

The process of literature screening and data extraction is inherently subjective and may suffer from authors' bias or different interpretations. To minimise the bias, both of the first two authors independently conducted the screening before the results were compared and adjusted. We should note that, while the articles were thoroughly screened, it could be the case that the criteria were too strict or that a paper's abstract did not reflect its relevant

contribution. Using the identified papers as a foundation, we conducted a snowball search in order to identify additional relevant publications.

As for the Delphi technique, an inherent limitation of its approach is the generalisability of the findings as it relies on a limited number of experts and the possibility of a selection bias. Although we tried to minimise the impact by recruiting a large panel of 30 experts from around Europe, future research might replicate this study with a different panel or use our results as the basis for a different data collection approach designed to test for generalisability. To minimise the burden on respondents and ensure high response rate, we kept the questionnaire in the second round short by including a brief description of the quality dimensions. However, this is in contrast with the complex nature of the research area. In some cases, this might cause ambiguous misinterpretation of the quality dimensions. For that reason, a comment box was included to allow respondents to report any vagueness.

8 CONCLUSION AND OUTLOOK

As the interest in threat intelligence sharing continues to grow, questions about its quality and effectiveness have surfaced. The current academic literature highlights data quality issues as a concern for threat intelligence producers and consumers. However, it falls short of empirically investigating the issue further. This study set out to develop and validate a set of threat intelligence quality dimensions. A systematic review of the threat intelligence literature followed by a modified Delphi study resulted in identifying seven quality dimensions: accuracy, actionability, interoperability, provenance, relevance, reliability and timeliness. It also examined the literature definitions of threat data, information and intelligence, and suggested areas for improvement. The study has found that practitioners' quality priorities vary depending on the nature of the intelligence. It has also shown that, despite increasing awareness of their potential value, organisations are yet to develop concrete metrics to measure any of the quality dimensions, and that they largely rely on consumer feedback and anecdotal evidence. The generalisability of the findings is limited by the inherent limitations of the Delphi technique. Nevertheless, the study provides empirical insights into the state of threat intelligence quality evaluation and extends our knowledge of what quality attributes are most relevant to practitioners. We lay the groundwork for further research that might explore the relationships between threat data, information and intelligence, and the identified quality dimensions, and offer a framework for the development of associated threat intelligence quality metrics. Additionally, future work might explore the extent to which the quality and value of threat intelligence intersect to maximise benefits and how quality affects user satisfaction across threat intelligence sharing platforms.

ACKNOWLEDGMENTS

The authors thank the anonymous reviewers for their helpful and constructive comments. Adam Zibak's research is funded by EPSRC via the Centre for Doctoral Training in Cyber Security at the University of Oxford.

REFERENCES

[1] J. R. Avella. 2016. Delphi panels: Research design, procedures, advantages, and challenges. *International Journal of Doctoral Studies* 11, 1 (2016), 305–321.

[2] S. Barnum. 2014. Standardizing Cyber Threat Intelligence Information with the Structured Threat Information eXpression (STIX). https://www.mitre.org/publications/technical-papers/standardizing-cyber-threat-intelligence-information-with-thee.

[3] C. Batini and M. Scannapieco. 2016. *Data and Information Quality: Dimensions, Principles and Techniques.* Springer International Publishing. https://books.google.co.uk/books?id=kJ_WCwAAQBAJ

[4] S. Bauer, D. Fischer, C. Sauerwein, S. Latzel, D. Stelzer, and R. Breu. 2020. Towards an Evaluation Framework for Threat Intelligence Sharing Platforms. In *Proceedings of the 53rd Hawaii International Conference on System Sciences*. 1947–1956.

[5] E. Bertino, A. Jabal, S. Calo, D. Verma, and C. Williams. 2018. The Challenge of Access Control Policies Quality. *Journal of Data and Information Quality* 10, 2, Article 6 (2018), 6 pages. https://doi.org/10.1145/3209668

[6] D. J. Bianco. 2013. The Pyramid of Pain. http://detect-respond.blogspot.com/2013/03/the-pyramid-of-pain.html. Accessed: 2020-06-12.

[7] Xander Bouwman, Harm Griffioen, Jelle Egbers, Christian Doerr, Bram Klievink, and Michel van Eeten. 2020. A different cup of {TI}? The added value of commercial threat intelligence. In *29th {USENIX} Security Symposium ({USENIX} Security 20)*. 433–450.

[8] S. Brown, J. Gommers, and O. 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)*. ACM, 43–49. https://doi.org/10.1145/2808128.2808133

[9] J. L. Campbell, C. Quincy, J. Osserman, and O. K. Pedersen. 2013. Coding in-depth semistructured interviews: Problems of unitization and intercoder reliability and agreement. *Sociological Methods & Research* 42, 3 (2013), 294–320.

[10] S. Chandel, M. Yan, S. Chen, H. Jiang, and T. Ni. 2019. Threat Intelligence Sharing Community: A Countermeasure Against Advanced Persistent Threat. In *Proceedings of the 2019 Conference on Multimedia Information Processing and Retrieval (MIPR)*. IEEE, 353–359. https://doi.org/10.1109/MIPR.2019.00070

[11] W. Chang, Y. Lo, and Y. Hong. 2009. A Heuristic Model of Network-Based Group Decision Making for E-Services. In *Proceedings of the 3rd International Conference on Information Technology: New Generations*. IEEE, 326–331. https://doi.org/10.1109/ITNG.2009.140

[12] CREST. 2019. What is Cyber Threat Intelligence and how is it used? https://www.crest-approved.org/wp-content/uploads/CREST-Cyber-Threat-Intelligence.pdf. Accessed: 2020-06-12.

[13] P. B. Crosby. 1979. *Quality is free: the art of making quality certain*. McGraw-Hill.

[14] L. Dandurand and O. Serrano. 2013. Towards improved cyber security information sharing. In *Proceedings of the 5th International Conference on Cyber Conflict* (Tallinn, Estonia) *(CyCon 2013)*. IEEE, 1–16. https://doi.org/10.1109/HICSS.2014.252

[15] Alessandra de Melo e Silva, João José Costa Gondim, Robson de Oliveira Albuquerque, and Luis Javier García Villalba. 2020. A Methodology to Evaluate Standards and Platforms within Cyber Threat Intelligence. *Future Internet* 12, 6 (2020), 108.

[16] A. L. Delbecq, A. H. Van de Ven, and D. H. Gustafson. 1975. *Group techniques for program planning: a guide to nominal group and Delphi processes*. Scott, Foresman.

[17] M Dempsey. 2013. Joint intelligence. *Joint Publication* (2013), 2–0.

[18] ENISA. 2013. Detect, SHARE, Protect - Solutions for Improving Threat Data Exchange among CERTs. https://www.enisa.europa.eu/publications/detect-share-protect-solutions-for-improving-threat-data-exchange-among-certs.

[19] ENISA. 2017. Exploring the opportunities and limitations of current Threat Intelligence Platforms. https://www.enisa.europa.eu/publications/exploring-the-opportunities-and-limitations-of-current-threat-intelligence-platforms.

[20] A. Fink. 2019. *Conducting research literature reviews: From the internet to paper*. Sage publications.

[21] Harm Griffioen, Tim Booij, and Christian Doerr. 2020. Quality Evaluation of Cyber Threat Intelligence Feeds. In *International Conference on Applied Cryptography and Network Security*. Springer, 277–296.

[22] G. Grispos, W. B. Glisson, and T. Storer. 2019. How Good is Your Data? Investigating the Quality of Data Generated During Security Incident Response Investigations. *CoRR* abs/1901.03723 (2019). arXiv:1901.03723 http://arxiv.org/abs/1901.03723

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

[24] Martin Husák, Václav Bartoš, Pavol Sokol, and Andrej Gajdoš. 2021. Predictive methods in cyber defense: Current experience and research challenges. *Future Generation Computer Systems* 115 (2021), 517–530.

[25] E. M. Hutchins, M. J. Cloppert, and R. M. Amin. 2011. Intelligence-driven computer network defense informed by analysis of adversary campaigns and intrusion kill chains. *Leading Issues in Information Warfare & Security Research* 1, 1 (2011), 80–106.

[26] International Organization for Standardization. 2008. ISO/IEC 25012:2008: Software engineering — Software product Quality Requirements and Evaluation (SQuaRE) — Data quality model.

[27] Joint Chiefs of Staff, United States Department of Defense. 2013. Joint Intelligence, Joint Publication 2-0. https://www.jcs.mil/Portals/36/Documents/Doctrine/pubs/jp2_0.pdf. Accessed: 2020-06-10.

[28] J. M. Juran and A. B. Godfrey. 1998. *Juran's quality handbook* (5th ed. ed.). McGraw Hill.

[29] B. Kitchenham and S. Charters. 2007. *Guidelines for performing systematic literature reviews in software engineering*. Technical Report. Technical report, Ver. 2.3 EBSE Technical Report. EBSE.

[30] H. Lehmann, J. Kuhn, and F. Lehner. 2004. The future of mobile technology: findings from a European Delphi study. In *Proceedings of the 37th Annual Hawaii International Conference on System Sciences*. IEEE, 1–10. https://doi.org/10.1109/HICSS.2004.1265225

[31] M. Lewis-Beck, A. E. Bryman, and T. F. Liao. 2003. *The SAGE Encyclopedia of Social Science Research Methods*. SAGE Publications.

[32] L. Li, X. Li, and Y. Gao. 2017. MTIV: A Trustworthiness Determination Approach for Threat Intelligence. In *Security, Privacy, and Anonymity in Computation, Communication, and Storage (SpaCCS 2017)*, G. Wang, M. Atiquzzaman, Z. Yan, and K. R. Choo (Eds.). Springer, 5–14. https://doi.org/10.1007/978-3-319-72395-2_1

[33] V. G. Li, M. Dunn, P. Pearce, D. McCoy, G. M. Voelker, and S. Savage. 2019. Reading the Tea leaves: A Comparative Analysis of Threat Intelligence. In *Proceedings of the 28th USENIX Security Symposium (USENIX Security 19)*. USENIX Association, 851–867. https://www.usenix.org/conference/usenixsecurity19/presentation/li

[34] H. A. Linstone and M. Turoff. 1975. *The Delphi Method: Techniques and Applications*. Addison-Wesley Publishing Company, Advanced Book Program.

[35] R. McMillan. 2013. Definition: Threat Intelligence. https://www.gartner.com/en/documents/2487216. Accessed: 2020-04-07.
[36] R. Meier, C. Scherrer, D. Gugelmann, V. Lenders, and L. Vanbever. 2018. FeedRank: A tamper- resistant method for the ranking of cyber threat intelligence feeds. In *Proceedings of the 2018 10th International Conference on Cyber Conflict (CyCon 2018)*. IEEE, 321–344. https://doi.org/10.23919/CYCON.2018.8405024
[37] A. Mohaisen, O. Al-Ibrahim, C. Kamhoua, K. Kwiat, and L. Njilla. 2017. Assessing Quality of Contribution in Information Sharing for Threat Intelligence. In *Proceedings of the 2017 IEEE Symposium on Privacy-Aware Computing (PAC)*. IEEE, 182–183. https://doi.org/10.1109/PAC.2017.39
[38] A. Mohaisen, O. Al-Ibrahim, C. Kamhoua, K. Kwiat, and L. Njilla. 2017. Rethinking Information Sharing for Threat Intelligence. In *Proceedings of the 5th ACM/IEEE Workshop on Hot Topics in Web Systems and Technologies* (San Jose, California) *(HotWeb '17)*. ACM, 6:1–6:7. https://doi.org/10.1145/3132465.3132468
[39] S. Mohurle and M. Patil. 2017. A brief study of wannacry threat: Ransomware attack 2017. *International Journal of Advanced Research in Computer Science* 8, 5 (2017).
[40] G. P. Z. Montesdioca and A. C. G. Maçada. 2015. Quality Dimensions of the DeLone-McLean Model to Measure User Satisfaction: An Empirical Test on the Information Security Context. In *Proceedings of the 2015 48th Hawaii International Conference on System Sciences*. IEEE, 5010–5019. https://doi.org/10.1109/HICSS.2015.593
[41] C. Mu, M. Yu, Y. Li, and W. Zang. 2014. Risk balance defense approach against intrusions for network server. *International Journal of Information Security* 13, 3 (2014), 255–269. https://doi.org/10.1007/s10207-013-0214-9
[42] S. Murdoch and N. Leaver. 2015. Anonymity vs. Trust in Cyber-Security Collaboration. In *Proceedings of the 2nd ACM Workshop on Information Sharing and Collaborative Security (WISCS '15)*. ACM, 27–29. https://doi.org/10.1145/2808128.2808134
[43] 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.
[44] J. Park, H. Alasmary, O. Al-Ibrahim, C. Kamhoua, K. Kwiat, L. Njilla, and A. Mohaisen. 2018. QOI: Assessing Participation in Threat Information Sharing. In *Proceedings of the 2018 International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 6951–6955. https://doi.org/10.1109/ICASSP.2018.8462036
[45] T. W. Phua and R. K. L. Ko. 2018. Data Provenance for Big Data Security and Accountability. In *Encyclopedia of Big Data Technologies*, S. Sakr and A. Zomaya (Eds.). Springer, 1–6. https://doi.org/10.1007/978-3-319-63962-8_237-1
[46] PROTECTIVE, Horizon 2020, European Commission. 2016. Threat Intelligence Sharing: State of the Art and Requirements. https://ec.europa.eu/research/participants/documents/downloadPublic?documentIds=080166e5b2a13ee7&appId=PPGMS. Accessed: 2021-06-29.
[47] L. Qiang, J. Zhengwei, Y. Zeming, L. Baoxu, W. Xin, and Z. Yunan. 2018. A Quality Evaluation Method of Cyber Threat Intelligence in User Perspective. In *Proceedings of the 2018 17th IEEE International Conference On Trust, Security And Privacy In Computing And Communications/ 12th IEEE International Conference On Big Data Science And Engineering*. IEEE, 269–276. https://doi.org/10.1109/TrustCom/BigDataSE.2018.00049
[48] Z. Rashid, U. Noor, and J. Altmann. 2019. Network Externalities in Cybersecurity Information Sharing Ecosystems. In *Economics of Grids, Clouds, Systems, and Services (GECON 2018)*, M. Coppola, E. Carlini, D. D'Agostino, J. Altmann, and J. Á. Bañares (Eds.). Springer, 116–125. https://doi.org/10.1007/978-3-030-13342-9_10
[49] T. C. Redman. 1996. *Data Quality for the Information Age*. Artech House. https://books.google.co.uk/books?id=UExPAAAAMAAJ
[50] T. Sander and J. Hailpern. 2015. UX Aspects of Threat Information Sharing Platforms: An Examination & Lessons Learned Using Personas. In *Proceedings of the 2nd ACM Workshop on Information Sharing and Collaborative Security* (Denver, Colorado, USA) *(WISCS '15)*. ACM, 51–59. https://doi.org/10.1145/2808128.2808136
[51] 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. *Computers & security* 82 (2019), 140–155.
[52] C. Sauerwein, C. Sillaber, and R. Breu. 2018. Shadow Cyber Threat Intelligence and Its Use in Information Security and Risk Management Processes. In *Proceedings of Multikonferenz Wirtschaftsinformatik 2018* (Lüneburg, Germany) *(MKWI '18)*. 1333–1344.
[53] C. Sauerwein, C. Sillaber, A. Mussmann, and R. Breu. 2017. Threat Intelligence Sharing Platforms : An Exploratory Study of Software Vendors and Research Perspectives. In *Proceedings of 13th International Conference on Wirtschaftsinformatik* (St. Gallen, Switzerland) *(WI 2017)*. 837–851. https://www.wi2017.ch/images/wi2017-0188.pdf
[54] T. Schaberreiter, V. Kupfersberger, K. Rantos, A. Spyros, A. Papanikolaou, C. Ilioudis, and G. Quirchmayr. 2019. A Quantitative Evaluation of Trust in the Quality of Cyber Threat Intelligence Sources. In *Proceedings of the 14th International Conference on Availability, Reliability and Security (ARES '19)*. ACM, Article 83, 83:1–83:10 pages. https://doi.org/10.1145/3339252.3342112
[55] 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.
[56] O. Serrano, L. Dandurand, and S. Brown. 2014. On the Design of a Cyber Security Data Sharing System. In *Proceedings of the 2014 ACM Workshop on Information Sharing and Collaborative Security* (Scottsdale, AZ, USA) *(WISCS '14)*. ACM, 61–69. https://doi.org/10.1145/2663876.2663882

[57] P. Shamala, R. Ahmad, A. Zolait, and M. Sedek. 2017. Integrating information quality dimensions into information security risk management (ISRM). *Journal of Information Security and Applications* 36 (2017), 1 – 10. https://doi.org/10.1016/j.jisa.2017.07.004

[58] S. Sicari, A. Rizzardi, D. Miorandi, C. Cappiello, and Coen-Porisini. A. 2016. Security policy enforcement for networked smart objects. *Computer Networks* 108 (2016), 133 – 147. https://doi.org/10.1016/j.comnet.2016.08.014

[59] C. Sillaber and R. Breu. 2015. Using Stakeholder Knowledge for Data Quality Assessment in IS Security Risk Management Processes. In *Proceedings of the 2015 ACM SIGMIS Conference on Computers and People Research* (Newport Beach, California, USA) *(SIGMIS-CPR '15)*. ACM, 153–159. https://doi.org/10.1145/2751957.2751960

[60] C. Sillaber, A. Mussmann, and R. Breu. 2019. Experience: Data and Information Quality Challenges in Governance, Risk, and Compliance Management. *Journal of Data and Information Quality* 11, 2 (2019), 1–14. https://doi.org/10.1145/3297721

[61] C. Sillaber, C. Sauerwein, A. Mussmann, and R. 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)*. ACM, 65–70. https://doi.org/10.1145/2994539.2994546

[62] F. Skopik, G. Settanni, and R. Fiedler. 2016. A problem shared is a problem halved: A survey on the dimensions of collective cyber defense through security information sharing. *Computers & Security* 60 (2016), 154–176. https://doi.org/10.1016/j.cose.2016.04.03

[63] D. M. Strong, Y. W. Lee, and R. Y. Wang. 1997. Data Quality in Context. *Commun. ACM* 40, 5 (1997), 103–110. https://doi.org/10.1145/253769.253804

[64] M. Talha, A. Abou El Kalam, and N. Elmarzouqi. 2019. Big Data: Trade-off between Data Quality and Data Security. *Procedia Computer Science* 151 (2019), 916 – 922. https://doi.org/10.1016/j.procs.2019.04.127 The 10th International Conference on Ambient Systems, Networks and Technologies (ANT 2019) / The 2nd International Conference on Emerging Data and Industry 4.0 (EDI40 2019).

[65] G. K. Tayi and D. P. Ballou. 1998. Examining Data Quality. *Commun. ACM* 41, 2 (1998), 54–57. https://doi.org/10.1145/269012.269021

[66] W. Tounsi and H. Rais. 2018. A survey on technical threat intelligence in the age of sophisticated cyber attacks. *Computers and Security* 72 (2018), 212–233. https://doi.org/10.1016/j.cose.2017.09.001

[67] Andrea Tundis, Samuel Ruppert, and Max Mühlhäuser. 2020. On the Automated Assessment of Open-Source Cyber Threat Intelligence Sources. In *International Conference on Computational Science*. Springer, 453–467.

[68] D. F. Vazquez, O. P. Acosta, C. Spirito, S. Brown, and E. Reid. 2012. Conceptual framework for cyber defense information sharing within trust relationships. In *Proceedings of the 4th International Conference on Cyber Conflict* (Tallinn, Estonia) *(CYCON 2012)*. IEEE, 1–17. https://ieeexplore.ieee.org/document/6243990

[69] H. A. von der Gracht. 2012. Consensus measurement in Delphi studies: Review and implications for future quality assurance. *Technological Forecasting and Social Change* 79, 8 (2012), 1525–1536. https://doi.org/10.1016/j.techfore.2012.04.013

[70] T. D. Wagner, K. Mahbub, E. Palomar, and A. E. Abdallah. 2019. Cyber threat intelligence sharing: Survey and research directions. *Computers & Security* 87 (2019), 101589. https://doi.org/10.1016/j.cose.2019.101589

[71] T. D. Wagner, E. Palomar, K. Mahbub, and A. E. Abdallah. 2017. Relevance Filtering for Shared Cyber Threat Intelligence (Short Paper). In *Information Security Practice and Experience*, J. K. Liu and P. Samarati (Eds.). Springer, 576–586.

[72] Y. Wand and R. Y. Wang. 1996. Anchoring Data Quality Dimensions in Ontological Foundations. *Commun. ACM* 39, 11 (1996), 86–95. https://doi.org/10.1145/240455.240479

[73] R. Y. Wang and D. M. Strong. 1996. Beyond Accuracy: What Data Quality Means to Data Consumers. *Journal of Management Information Systems* 12, 4 (1996), 5–33. https://doi.org/10.1080/07421222.1996.11518099

[74] C. Wohlin. 2014. Guidelines for snowballing in systematic literature studies and a replication in software engineering. In *Proceedings of the 18th international conference on evaluation and assessment in software engineering*. 1–10.

[75] A. Zibak and A. Simpson. 2019. Cyber Threat Information Sharing: Perceived Benefits and Barriers. In *Proceedings of the 14th International Conference on Availability, Reliability and Security* (Canterbury, United Kingdom) *(ARES '19)*. ACM, Article 85, 9 pages. https://doi.org/10.1145/3339252.3340528