COMPUTERS & SECURITY 82 (2019) 140–155

Available online at www.sciencedirect.com

ScienceDirect

journal homepage: www.elsevier.com/locate/cose

Computers & Security

An analysis and classification of public information security data sources used in research and practice

Clemens Sauerwein[a,*], *Irdin Pekaric*[a], *Michael Felderer*[a,b], *Ruth Breu*[a]

[a] *University of Innsbruck Department of Computer Science Technikerstraße 21a, Innsbruck A-6020 Austria*
[b] *Blekinge Institute of Technology Valhallavägen 1, Karlskrona 371 41, Sweden*

ARTICLE INFO

Article history:
Received 8 May 2018
Revised 18 December 2018
Accepted 19 December 2018
Available online 25 December 2018

Keywords:
Cyber threat intelligence sharing
Cyber security information source
Taxonomy
Classification
Characteristic
Information security and risk management
Data format
Research
Practice

ABSTRACT

In order to counteract today's sophisticated and increasing number of cyber threats the timely acquisition of information regarding vulnerabilities, attacks, threats, countermeasures and risks is crucial. Therefore, employees tasked with information security risk management processes rely on a variety of information security data sources, ranging from inter-organizational threat intelligence sharing platforms to public information security data sources, such as mailing lists or expert blogs. However, research and practice lack a comprehensive overview about these public information security data sources, their characteristics and dependencies. Moreover, comprehensive knowledge about these sources would be beneficial to systematically use and integrate them to information security processes. In this paper, a triangulation study is conducted to identify and analyze public information security data sources. Furthermore, a taxonomy is introduced to classify and compare these data sources based on the following six dimensions: (1) *Type of information*, (2) *Integrability*, (3) *Timeliness*, (4) *Originality*, (5) *Type of Source*,and (6) *Trustworthiness*. In total, 68 public information security data sources were identified and classified. The investigations showed that research and practice rely on a large variety of heterogeneous information security data sources, which makes it more difficult to integrate and use them for information security and risk management processes.

1. Introduction

In the last couple of years the number of cyber attacks and their complexity increased significantly, while the time frame for organizations to react shrinks constantly (Jang-Jaccard and Nepal, 2014). To counteract these threats, organizations are implementing vulnerability management processes and conduct awareness trainings as part of their information security program (Puhakainen and Siponen, 2010; Soomro et al., 2016).

In order to improve these countermeasures, a trend to automation, collaboration and consultation of shared information security data sources can be observed in research and practice (Fenz et al., 2014; Harel et al., 2017; Sauerwein et al., 2016). We define public information security data sources as *information sources that provide information regarding vulnerabilities, threats, attacks, risks, affected assets or available countermeasures (based on* ISO and Std, 2009).

Research and practice introduced several data formats, messaging protocols, technologies and frameworks that enable the automated and standardized exchange of

* Corresponding author.
E-mail address: clemens.sauerwein@uibk.ac.at (C. Sauerwein).
https://doi.org/10.1016/j.cose.2018.12.011
0167-4048/

information security data (Johnson et al., 2016; Kampanakis, 2014; Menges and Pernul, 2018; Skopik et al., 2016; Steinberger et al., 2015). As a result to this trend, several platforms under the umbrella term *threat intelligence sharing platforms* entered the market (Sauerwein et al., 2017). Apart from these threat intelligence sharing platforms, information security data can be found on public information security data sources as well (Sauerwein et al., 2018). According to Mittal et al. (2016), these information sources can be divided into formal or rather official sources, such as the National Vulnerability Database (NVD), and informal sources, such as social media platforms, like Twitter.

Up to date, empirical research in the field of threat intelligence sharing and public information security data sources is rare. Therefore, research and practice lack a systematic and comprehensive overview of the availability, characteristics and dependencies of these public information security data sources and how they are used in practice. Moreover, a taxonomy to classify these information security data sources and compare them against each other is missing (Mavroeidis and Bromander, 2017).

The research objective of this paper is to address theses gaps through introducing the needed taxonomy and providing a qualitative and quantitative analysis of public information security data sources used in research and practice. In doing so, we address the following three research questions:

- (RQ1) *How can public information security data sources be classified?*
- (RQ2) *What are the characteristics of currently available information security data sources?*
- (RQ3) *What are the dependencies between the identified information security data sources?*

In order to answer these research questions, we conducted a triangulation study (Jick, 1979) consisting of a literature review (Webster and Watson, 2002; Wohlin, 2014), an analysis of security expert discussions on Twitter and an exploratory survey (Fowler Jr, 2013) on information security data sources used by medium to large-sized organizations in Europe. Finally, the classification and analysis of the results was conducted based on the proposed taxonomy. Our research methodology delivered a final set of 68 information security data sources for classification and analysis.

This paper provides a taxonomy to classify information security data sources and make them comparable against each other. The taxonomy can be used to systematically evaluate heterogeneous information security data sources and form the basis of a knowledge base on how to integrate these sources into information security risk management processes. Secondly, such a knowledge base was developed by identifying and classifying 68 public information security data sources used in research and practice. For example, the results can be used for attack graph model generation. Current attack modeling approaches mostly use data from various information security data sources. However, it is very difficult to determine which data sources are the most appropriate ones. In order to address this issue, the taxonomy can be used to choose the right data sources and the developed knowledge base provides an overview of highly relevant sources used in research and practice.

The remainder of this paper is structured as follows. Section 2 discusses related work regarding studies and the usage of information security data sources in research and practice. Section 3 presents the applied research methodology. Section 4 answers our research questions by introducing the taxonomy and classification of the identified information security data sources. Section 5 discusses the key findings of our analysis and limitations of the applied research methodology. Finally, Section 6 concludes the paper and provides outlook on future work.

2. Related work

Several contributions in the field of information security data sharing, which are related to the research at hand, can be identified. These contributions focus on the analysis of different information security sharing standards or models, information security data sources, threat intelligence sharing and corresponding platforms.

Steinberger et al. (2015) provided a structured overview of existing information security data exchange formats and protocols by analyzing their use cases, interoperability and scalability. Moreover, Hernandez-Ardieta et al. (2013) propose models for real-time information security data sharing based on information security data exchange formats adopted from the *Making Security Measurable (MSM)* (Making Security Measurable, MSM) initiative.

Rader and Wash (2015) conducted a comparison of three different types of information security data sources, namely news articles, web pages and personal experiences. Their results showed that most of the analyzed sources focus on attacks and their consequences. Massacci and Nguyen (2010) investigated the quality of 14 different vulnerability databases by comparing various types of information security metrics. Tripathi and Singh (2011) performed an analysis of classification schemes for vulnerability databases. The goal was to develop an improved scheme in order to accelerate research regarding information security risk levels. In doing so they considered seven vulnerability databases for their analysis. A further group of publications classifies and analyzes threat intelligence. Tounsi and Rais (2018) classified different threat intelligence types. In doing so they focus on new standards, trends and technical issues. Similarly, Mavroeidis and Bromander (2017) developed a taxonomy to classify sharing standards and ontologies. Brown et al. (2015) presented multiple challenges related to threat intelligence sharing platforms, community requirements and expectations. Furthermore, Zhao and White (2012) outlined the importance of information security data sharing and provided a list of types of information security data types which are important to share. Qamar et al. (2017) developed a threat analysis framework based on the Web Ontology Language. The framework gets network associated threats based on shared data feeds. Burger et al. (2014) proposed a taxonomy for classifying threat intelligence sharing platforms. Menges and Pernul (2018) introduced an incident reporting process model used for comparing various incident

reporting formats. Abu et al. (2018) conducted a literature review on existing definitions of threat intelligence. Their results showed that vendors and organizations lack an understanding what information should be considered as threat intelligence. Johnson et al. (2016) provided guidelines how to participate in threat intelligence sharing. Sauerwein et al. (2017) conducted a systematic study of multiple threat intelligence sharing platforms and compared them. Moreover, Sillaber et al. (2016) identified data quality challenges in threat intelligence sharing practice. Finally, Sauerwein et al. (2018) the phenomenon of shadow threat intelligence sources is discussed by outlining several information security data sources.

Most of the presented research focus on information security data exchange or threat intelligence sharing while only a handful of studies analyze information security data sources, like vulnerability databases. However, they don't identify information security data sources in a systematic way. The sources are mostly identified based on author's knowledge or intention. Furthermore, most of them include well known vulnerability databases, such as National Vulnerability Database (2018), by missing other information security data sources. Moreover, a taxonomy to classify these sources is still missing. To the best of our knowledge no prior empirical research has been conducted that introduced a classification taxonomy for information security data sources and a systematic analysis of information security data sources.

In general, differing from software engineering, where many taxonomies have been developed to understand, compare or tailor different approaches (Usman et al., 2017), are taxonomies in information security still rare. Taxonomies related to threat intelligence sharing mentioned before in this section were proposed by Burger et al. (2014) on threat intelligence sharing platforms as well as by Mavroeidis and Bromander (2017) on sharing standards and ontologies. But there is no taxonomy or at least a mapping study on public information security data sources, as provided in this paper, available. Mapping studies Felderer and Carver (2017) provide a 'map' of a specific area by systematically searching, selecting and classifying sources based on relevant categories. A mapping study can be applied to systematically develop and evaluate a taxonomy, which has for instance been performed for model-based security testing (Felderer et al., 2016) and security regression testing (Felderer and Fourneret, 2015). Furthermore, there is even an evidence-based taxonomy development method available (Usman et al., 2017), which is applied in this paper to develop the taxonomy for public information security data source classification and analysis (see Section 3.5).

3. Research methodology

As stated in Section 1 our contribution focuses on the following three research questions:

- (RQ1) *How can public available information security data sources be classified?*
- (RQ2) *What are the characteristics of currently available information security data sources?*
- (RQ3) *What are the dependencies between identified information security data sources?*

In this context, research question (RQ2) can be divided into the following five sub research questions:

- (RQ2.1) *What types of public available information security data sources can be identified?*
- (RQ2.2) *How is the information structured?*
- (RQ2.3) *What interfaces are provided to access the information?*
- (RQ2.4) *Who provides the information?*
- (RQ2.5) *When is the information shared?*
- (RQ2.6) *Do the sources provide original contents?*

Research question (RQ3) can be divided into the following two sub research questions

- (RQ3.1) *What are the relationships between the different types of provided information that are found within the sources?*
- (RQ3.2) *How do interfaces relate to different types of provided information?*

In order to address these research questions, we conducted a triangulation study combining qualitative and quantitative methods to answer the research questions and limit the threats to validity of the respective methods. As depicted in Fig. 1 our triangulation study consisted of a systematic literature review (Wohlin, 2014) on information security data sources and classification criteria (see Section 3.1), a data collection of information security related Tweets on Twitter (see Section 3.2), and an exploratory survey on public information security data sources used in practice (see Section 3.3). Based on the results of these three studies we compiled a list of information security data sources used in research and practice (see Section 3.4). In addition, based on the results of the systematic literature review we developed a taxonomy to classify the identified information security data sources and addressed RQ1 (see Section 3.5). Finally, we classified and analyzed the identified information security data sources and addressed RQ2, RQ3 and respective sub research questions (see Section 3.6). In this section we explain the applied research methodology in detail.

3.1. *Literature review on information security data sources and classification criteria*

The systematic literature review, carried out between December 2017 and January 2018, is based on the snowballing methodology (Wohlin, 2014), which builds on the ideas of Webster and Watson (2002). It is comparable to the prevalent method of database searches since a comparison of their results did not show any serious differences (Jalali and Wohlin, 2012). The steps associated with the snowballing procedure are: (a) *definition of a start set of papers*, and (b) *execution of snowballing iterations* (Wohlin, 2014). The latter involves forward snowballing, such as identifying new papers citing the one being examined, and backward snowballing, such as looking at the references of the considered paper. In order to guarantee the replication of the applied methodology a review protocol was developed including search strategy and selection criteria.

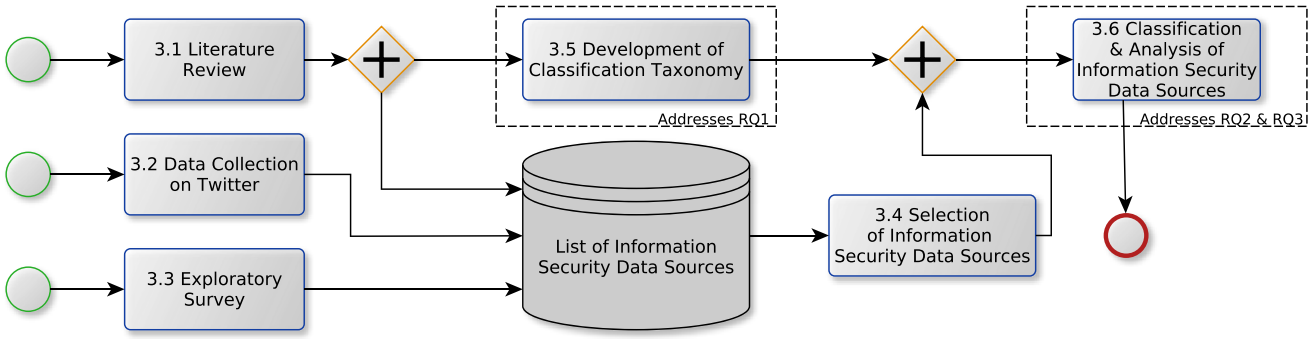

Fig. 1 – Overview of the applied research methodology.

Table 1 – Inclusion and exclusion criteria of the literature study.

Inclusion criteria	*Exclusion criteria*
Peer reviewed articles	*Gray and white literature*
Accessible in Full text	*Non-english articles*
Published between 2008 and 2018	Duplicates
Discussing information security data sources, their characteristics or data formats	*General Information Security Topics*

3.1.1. *Definition of start set*

Since there is no common methodology to define a start set (cf. Wohlin, 2014), we conducted a key word search and kept in mind the five characteristics of a good start set defined by Wohlin (2014): (1) Covering all relevant research communities, (2) containing an adequate number of papers, (3) including highly cited and relevant papers, (4) covering different publishers, years and authors, and (5) including papers based on keywords and corresponding synonyms. Therefore we used the following libraries: ACM Digital Library, IEEE Digital Library, ScienceDirect, Springer Link, and Wiley. These libraries were chosen because they cover the most relevant sources in security engineering (Brereton et al., 2007). In doing so, we used the following two search strings: *'cyber security' AND ('information' OR 'data') AND 'sources'*, and *('analysis' OR 'classification') AND 'of' AND ('cyber security sources' OR 'cyber threat-intelligence')*. We only considered articles that fulfilled our inclusion criteria defined in Table 1.

3.1.2. *Execution of snowballing iterations*

Based on the start set, 30 iterations of forward and backward snowballing were executed until no new papers were identified. In total 653 references were examined with respect to the following selection procedure: (1) Title, abstract and keywords have been analyzed, and (2) partial reading of the selected papers.

In doing so, the selection of papers was performed based on the inclusion and exclusion criteria illustrated in Table 1. We were selecting only peer reviewed articles to ensure that our set consists of high quality publications. Furthermore, only articles that are accessible in full text were considered because it is very difficult to extract complete information and to determine if an article is a match for our set if the full access is not guaranteed. In addition, the search was limited to a period between 2008 and 2018 year. This guarantees that the set we are analyzing consists of recent publications. Finally, the last inclusion criteria implies that the articles should discuss security information sources, their characteristics and formats.

On the other hand, gray and white literature was excluded in order to ensure high quality publications. Furthermore, non-English articles were not included because it is not feasible to translate publications issued in foreign languages. Since our search covered all the major libraries, there were cases when the same article would appear in more than one library. Therefore, the article was included only once and duplicates were removed from the set. The last exclusion criteria ensures that the publications that are covering general information security topics are excluded. This guarantees that the final set is not too larger and allows us to focus on the topics that were aforementioned.

Finally, the snowballing methodology delivered us a set of twelve papers. From eight out of those twelve papers (Brown et al., 2015; Hernandez-Ardieta et al., 2013; Massacci and Nguyen, 2010; Mavroeidis and Bromander, 2017; Sauerwein et al., 2018; Steinberger et al., 2015; Tounsi and Rais, 2018; Tripathi and Singh, 2011), we derived a list of 42 cyber security information sources for our analysis, and from nine (Brown et al., 2015; Johnson et al., 2016; Massacci and Nguyen, 2010; Rader and Wash, 2015; Sauerwein et al., 2018; 2017; Steinberger et al., 2015; Tounsi and Rais, 2018; Zhao and White, 2012), we obtained inputs for the development of our classification taxonomy.

3.2. *Data collection of information security related tweets on Twitter*

In order to identify and collect useful security information shared on Twitter, we implemented a Python script to collect the needed information based on predefined keywords. Therefore, we used the Python library Tweepy (Tweepy python library, 2018), which connects to the Twitter Streaming API and filters the stream of Tweets based on keywords in real time. In doing so, we filtered for Tweets matching Common Vulnerabilities and Exposures (CVE) (Mell and Grance, 2002) identifiers and stored the collected Tweets in our database. CVEs are standardized descriptions for publicly known

information security vulnerabilities and exposures (Mell and Grance, 2002), which enable the tracing of information regarding certain vulnerabilities and exposures on different information channels (e.g. Twitter) during all phases of their lifecycle (Frei et al., 2008). CVE identifiers are unique and correspond to the following pattern: `CVE-\d{4}-\d{4}\d*` (Mell and Grance, 2002). For example, CVE-2016-4117 or CVE-2016-10000 are valid CVE identifiers. In order to eliminate Tweets containing wrong CVE identifiers, we cross validated CVE identifiers included in the obtained data set with CVE identifiers listed on the MITRE's website CVE or National Vulnerability Database (NVD) (National Vulnerability Database, 2018). These two websites include the full list of all available CVE identifiers and manage their assignment. Tweets without any matches were excluded from further analysis. Furthermore, our data collection approach guaranteed that only information security relevant data on Twitter were collected and analyzed. In total, there were 709,880 Tweets collected between May 23, 2016 and March 27, 2018. The analysis showed that 0.8 URLs referencing external websites are contained per Tweet, which might be traced back to a lack of space for detailed descriptions since Twitter limits the content per Tweet to now 280 (and previously 140) characters. In total 567,904 Tweets contained a referenced website whereof 11,437 different websites could be identified. This number of different websites might be traced back, that more than the half of the identified websites are mentioned at most three times in the data dump. In order to guarantee that the analyzed websites are the most common in our data dump and to make them classifiable we decided to analyze the top 50 most referenced websites in the data dump. Therefore, the analyzed websites appeared more than 500 times in the data dump and represent 34% of the Tweets containing a referenced website.

3.3. *Exploratory survey on public available information security data sources used in practice*

The exploratory survey (Fowler Jr, 2013) on information security data sources was part of a more comprehensive survey on information security risk management methods and processes used in practice. An exploratory survey was chosen as it is a suitable approach to investigate the usage of tools or techniques in real world settings (Pfleeger, 1995). In total 29 participants from medium to large-sized organizations in Europe answered the questionnaire. Since the information was collected in anonymized form any further details regarding the participants are not available. It is worth mentioning that the participants were contacted via two security mailing lists. The participants voluntarily participated in the exploratory survey and had to answer the following survey question: *What public available information security data sources are you using as input to information security risk management processes?* In order to answer this survey question, the participants were asked to state the three most relevant public information security data sources which resulted in a list of 87 information security data sources. It is worth mentioning, that there was a large number of overlaps between the stated information security data sources in the filled out surveys. After elimination of duplicates a final list of 32 different public information security data sources was obtained.

3.4. *Selection of information security data sources*

The literature study ($n = 42$), data collection on Twitter ($n = 50$) and exploratory survey ($n = 32$) resulted in a preliminary list of 124 information security data sources. In order to get a final list of information security data sources for the subsequent classification, the following inclusion and exclusion criteria were applied. All information security data sources which were open-source or freely available in English were included, and only provided information regarding vulnerabilities, threats, attacks, risks, affected assets or available countermeasures (as defined in Section 1). The sources which were deprecated, duplicated, closed-source, commercial, not available in English and did not match the inclusion criteria discussed before were excluded. Finally, this procedure resulted in a list of 68 information security data sources for subsequent classification.

3.5. *Development of classification taxonomy*

In order to classify the identified information security data sources, a classification taxonomy was developed. Therefore, the methodology by Usman et al. was applied to systematically develop a classification taxonomy (Usman et al., 2017). The methodology consists of the the following four phases: (a) Planning, (b) identification and extraction, (c) design and construction and (d) validation. At first (cf. Phase (a)) we decided to build a taxonomy to classify the identified information security data sources in order to address our research questions. Therefore (cf. Phase (b)), different categories and dimensions were drawn from literature, based on the papers identified during the systematic literature review (see Section 3.1), the ISO Standard 15408-1 (ISO and Std, 2009) and the data quality dimensions defined in Batini et al. (2009). In doing so, all the authors of this contribution discussed the different quality dimensions described (Batini et al., 2009) and how they can be used to classify the information security data sources. Moreover, a preliminary look on the identified information security data sources indicated which dimensions are worth considering to answer our research questions. In this context, it is worth mentioning that the *Type of Source* category (e.g. News website, expert blog,...) was iteratively defined based on the analysis. In other words, whenever we identified at least two sources of the same type, a new category was created, otherwise (i.e. if only one source of a particular type was found) it was classified as *Other*. Finally, a taxonomy (cf. Phase (c)) was constructed and validated (cf. Phase (d)) through classification of the identified information security data sources (see Section 4).

3.6. *Classification and analysis of information security data sources*

Based on the taxonomy, the final list of 68 information security data sources was classified based on the classification procedure used and described in Carver et al. (2016). In doing so, the research team of this paper consisted of two PhD students and 2 faculty members from two universities in two countries. For the classification of the information security data sources, the following approach was used:

1. One of the faculty members randomly assigned each information security data source to all members of the research team including him. In order to avoid bias, he ensured that each reviewer was paired with each of the other reviewers across the whole set of information security data sources. This means that every information security data source was independently classified by two reviewers.
2. Based on the developed taxonomy (cf. Section 3.5) each reviewer independently classified the assigned information security data sources. In order to avoid misunderstandings each reviewer received a comprehensive explanation of the taxonomy's different dimensions and corresponding classification categories.
3. In order to prevent bias from the other reviewer's classification, each reviewer documented his analysis results in an own spreadsheet.
4. The classification results were analyzed and if there were classification discrepancies among the reviewers the corresponding results were marked.
5. Finally, for any information security data sources where classification discrepancies were detected, all authors discussed these disagreements and re-classified the corresponding source. Based on a majority vote (including all authors) a final classification for each source was made.

In order to avoid false or biased classification we carried out a cross-validation approach. In doing so, each author of this research paper was given a subset of information security data sources to classify that intersected with another author's set. Therefore, we ensured that each information security data source was classified by more than one author. Each author had to fill out a spreadsheet where each row contained a information security data source and each column represented a classification criteria according to our defined taxonomy. In a further step we merged the filled out spreadsheets together and if classification discrepancies were discovered, we resolved them through re-classification or discussion. It is worth mentioning that that 5 out of the 68 information security data sources showed classification discrepancies. We were able to solve them according to aforementioned classification approach (cf. step 5).

Finally, we used R Project (Gentleman et al., 2009) to statistically analyze the resulting classification.

4. Results

In this section, we introduce the developed taxonomy and present the results of the analysis of the identified information security data sources. In doing so, Section 4.1 addresses research question RQ1 and Section 4.2 focuses on research question RQ2 with corresponding sub research questions RQ2.1–RQ2.6. Finally, research question RQ3 with corresponding sub research questions RQ3.1 and RQ3.2 are addressed in Sections 4.4 and 4.5, respectively.

4.1. *Classification taxonomy*

As mentioned in Section 3, we developed a taxonomy to classify information security data sources. Fig. 2 illustrates the taxonomy consisting of the following six dimensions: (1) *Type of information*, (2) *Integrability*, (3) *Timeliness*, (4) *Originality*, (5) *Type of source*, and (6) *Trustworthiness*. In the following, we want to discuss each of these dimensions in detail.

4.1.1. *Type of information*

In order to classify the type of provided information of information security data sources, we distinguish among the following *types of information* These types were derived from IEC (2014) since they represent the most common information types in information security:

- *Vulnerability*: Information regarding a weakness of an asset which might be exploited by a threat.
- *Threat*: Information regarding the potential cause on an unwanted incident.
- *Countermeasure*: Information regarding any administrative, managerial, technical or legal control that is used to counteract an information security risk.
- *Attack*: Information regarding any unauthorized attempt to access, alter or destroy an asset.
- *Risk*: Information describing the consequences of a potential event, such as an attack.
- *Asset*: Information regarding any object or characteristic that has value to an organization.

An information source might provide more than one type of information. Consequently, multiple classifications regarding the *type of information* would be possible. For example, a vulnerability database might provide information on vulnerabilities and resulting risks.

4.1.2. *Integrability*

In order to automate information security risk management processes, such as described in the IEC/ISO 27005 ISO/IEC (2011), the Integrability of information is inevitable. In our context integrability describes to which extent information security data sources and the provided information can be (automatically) integrated into an organization's information security tool landscape and processes (cf. Batini et al., 2009). To decide about the Integrability of information security data sources the taxonomy classifies the (1) *format* of the available information and the provided (2) *interfaces*. In order to classify the format of the available information the taxonomy distinguishes between the following two types:

- *Structured*: The provided security information is available in an standardized and structured data format, such as the Structured Threat Information Expression (STIX) format (Barnum, 2012).
- *Unstructured*: The provided security information is available in unstructured form without following a common data representation format.

In order to classify the provided interfaces the taxonomy distinguishes between the following four types:

- *No interfaces*: The information security data source doesn't provide any interface to access the information.

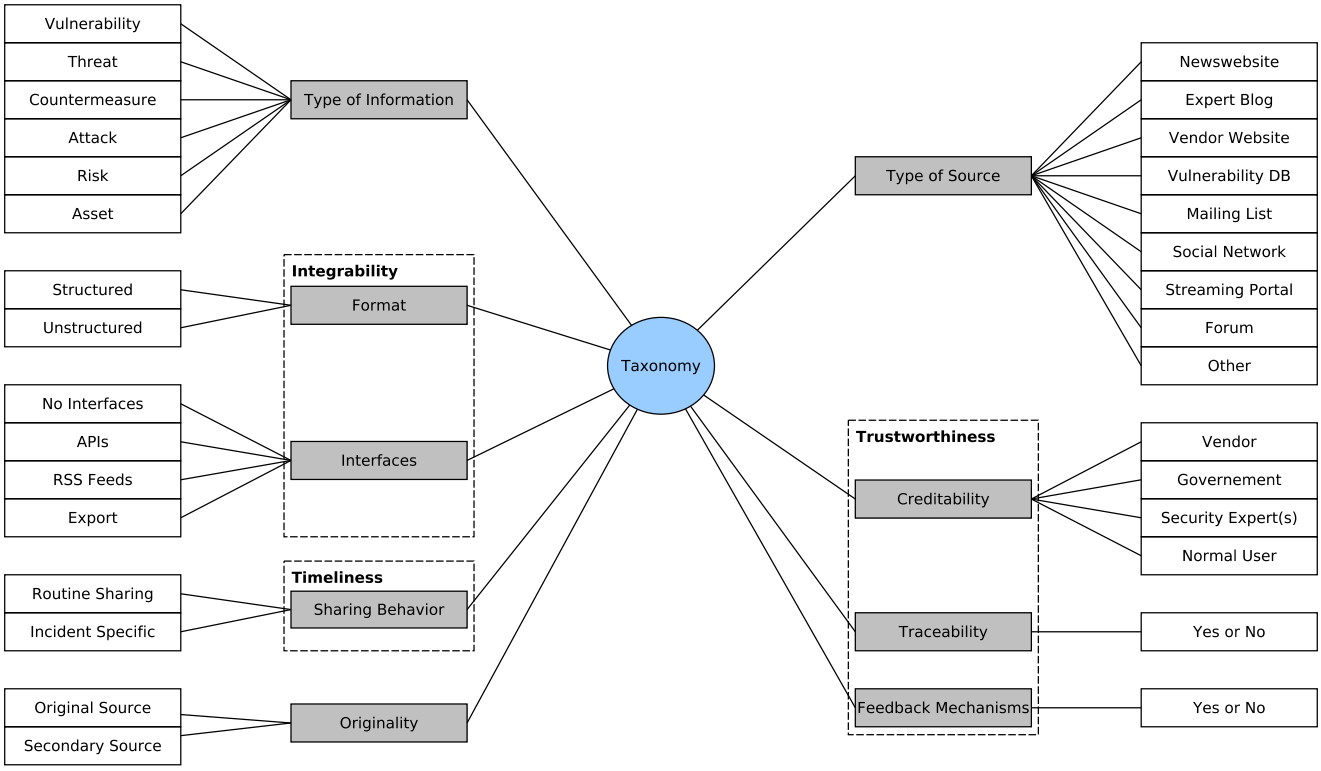

Fig. 2 – Taxonomy to classify the information security data sources.

- *APIs*: The information security data source provides an application programming interface (APIs) to obtain the provided information.
- *RSS Feed*: The information security data source provides an RSS Feed to keep track of the provided information.
- *Export*: The information security data source provides an interface to export contents as XML, JSON or plain text.

It might be possible, that an information source might provide more than one interface. Consequently, multiple classifications regarding the provided *interfaces* would be possible.

4.1.3. *Timeliness*

Timeliness plays an important role in information security risk management since the needed information (e.g. regarding occurring threats) should occur at a suitable time in order to take appropriate actions (cf. Arbaugh et al., 2000; Frei et al., 2008). For example, if an organization receives information regarding an occurring threat as early as possible it can put countermeasures in place and fend off an attacker early. Therefore, our second dimensions provides a closer look on the *sharing behavior* with respect to the point in time. In order to classify the *sharing behavior* the taxonomy distinguishes between the following two categories derived from Zhao and White (2012):

- *Routine*: Information is published at a specific point in time on a regular basis, such as daily, weakly or monthly reports.
- *Incident Specific*: Information is published whenever news are available or a new incident occurs.

4.1.4. *Originality*

In order to decide about the originality and novelty (cf. Batini et al., 2009) of the provided information the taxonomy classifies their origin. In doing so it distinguishes between the following two types of sources:

- *Original source*: Information originates from information security data source, such as (National Vulnerability Database, 2018) which publish their own information.
- *Secondary source*: Information is integrated or copied from another information security data source, such as a meta-source integrating information from multiple original sources.

4.1.5. *Type of source*

The taxonomy classifies different types of information security data sources. Accordingly, we distinguish between the following nine types: (1) *News website*, (2) *Expert Blog*, (3) *(Security Product) Vendor website*, (4) *Vulnerability Database*, (5) *Mailing List* archive, (6) *Social Network*, (7) *Streaming Portal*, (8) *Forum* or (9) *Other*. As discussed in Section 3.5 these nine types were generated based on an incremental process during classification

4.1.6. *Trustworthiness*

Trustworthiness plays an important role in the field of information security when it comes to information sharing since false information might have immense impact on organizational processes (Steinberger et al., 2016). Therefore we analyzed the (1) *creditability*, (2) *traceability* (a definition of these two terms can be found in Batini et al., 2009), and (3) *feedback mechanisms* of information security data sources.

	Types of provided Information						Integrability						Time-liness		Origi-nality		Trustworthiness					
	Vulnerabilities	Threats	Countermeasures	Attacks	Risk	Assets	Structured	Unstructured	No interfaces	APIs	Feeds	Export	Routine Information Sharing	Incident-Specific	Secondary source	Original source	Vendor	Government	Security Expert(s)	Normal User	Feedback Mechanism (Yes/No)	Traceability of Information (Yes/No)
Newswebsite (21%)	100	73	67	93	53	53	7	93	93	0	7	0	93	53	27	73	13	20	87	13	20	80
Blogs (20%)	92	46	38	77	15	38	0	100	100	0	0	0	69	62	0	100	46	0	54	23	38	85
Vendor Website (13%)	100	33	22	67	33	33	11	89	78	11	22	11	89	100	0	100	89	0	11	0	89	22
Vulnerability Databases (13%)	100	11	22	33	56	11	33	67	22	44	44	33	89	44	67	33	22	22	100	0	67	78
Mailinglists (4%)	100	100	67	100	33	33	0	100	100	0	0	0	67	67	67	33	0	67	100	67	0	33
Social Network (3%)	100	100	100	100	100	100	0	100	50	50	0	0	100	100	50	50	50	50	100	100	100	50
Streaming Portal (3%)	100	50	50	100	0	50	0	100	50	50	0	0	50	50	0	100	50	50	100	50	50	100
Forums (3%)	100	50	50	50	0	50	50	50	50	50	0	50	50	50	0	100	50	0	50	50	0	50
Other (20%)	31	31	54	31	15	8	85	15	23	31	15	31	54	46	15	85	38	8	85	38	54	46
Average percentage	90	53	50	70	32	40	22	78	59	30	10	16	71	65	25	75	43	25	75	41	50	58

Fig. 3 – Overview of classification results per type of information security data source in percentage.

In order to classify (1) the taxonomy differentiates between the following four different types of publishers: (a) *Vendor*, (b) *Government*, (c) *Security Expert(s)* and (d) *Normal users*. It is worth mentioning, that the type of publisher is classified as (a), (b) or (c) if enough background information regarding the publisher is available. Otherwise the publisher would be classified as (d). Furthermore, it might be possible that an information source can have more than one type of publisher.

Secondly, the taxonomy classifies if the information is traceable (cf. (2)). In this context security information is classified as traceable if it can be traced back, based on meta-data, to a specific publisher and a publishing date. Otherwise the information is classified as untraced.

Finally, we analyze if the information is validated (cf. (3)). In this context, information is classified as (3) if feedback mechanisms are provided. These mechanisms include user ratings or comments regarding the usefulness of the provided information.

4.2. *Classification of public information security data sources*

We classified and analyzed the list of 68 information security data sources (see Appendix A) based on the taxonomy presented in Section 4.1. We grouped the information security data sources together according to their type (i.e. news website, blog, vendor website, vulnerability database, mailing list archive, social network, streaming portal, forum or other). Fig. 3 gives an overview of the resulting classification per type of information security data source. The numbers in the cells show the percentage of a specific type of source following a certain characteristic. Additionally, the last row shows the average percentages per characteristic considering all 68 sources.

In this section we discuss the classification, depicted in Fig. 3, in detail and answer RQ2 with corresponding sub research questions (RQ2.1 - 2.6).

4.2.1. *Types of information security data sources*

Fig. 4outlines the actual number of information security data sources per type. The majority of identified sources are news websites and blogs, whereof news websites represent 21% and blogs 20%. Furthermore, 13% of the sources are vendor websites or vulnerability databases, and 4% are mailing lists. Streaming portals, social networks and forums represent 3% each. The remaining, 20% of sources can be classified as others. A detailed listing and classification of the information security data sources can be found in Table 2 in Appendix A.

Moreover, 79% of the identified information security data sources are open source and 21% are semi-open source. The latter are those sources which are free of charge up to a certain limit such as providing a limited number of free request. As mentioned in Section 3 closed sources are not considered

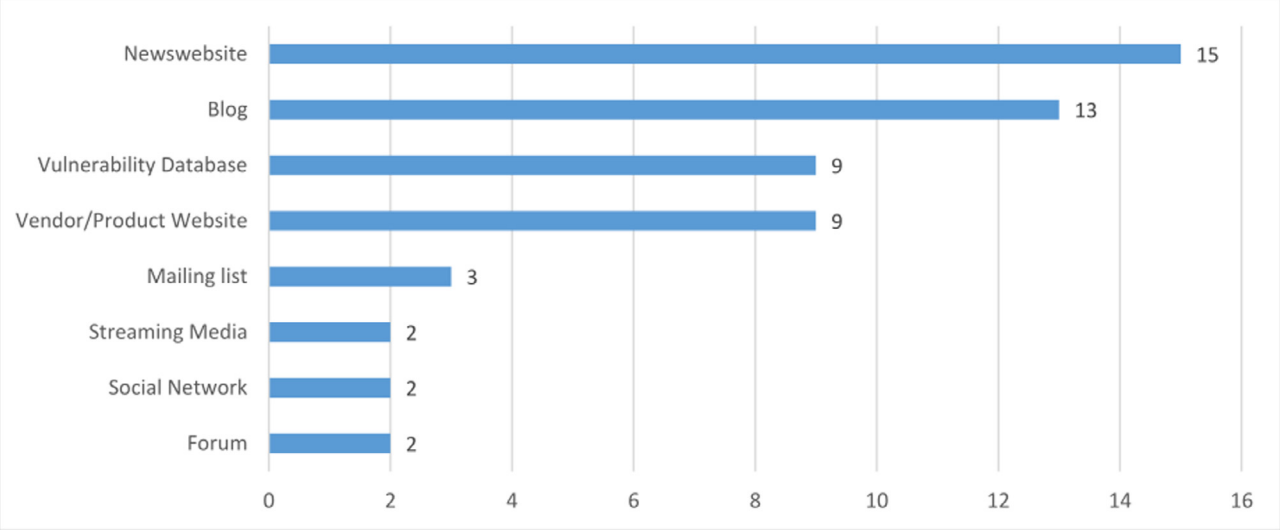

Fig. 4 – Number of information security data sources per type.

since a classification would not be feasible because of access restrictions and biased descriptions of data providers.

In order to get a better understanding regarding the identified information security data sources, in the following we want to illustrate a few examples: The 'IT Security News' (IT Security News, 2018) website was classified as a news website. It provides news and articles about vulnerabilities, exploits, patches, hacks, viruses, malware and trojans. A well known blog in the field of information security is the one of Bruce Schneier (Schneier on security, 2018). He is an expert in the field of security and writes about security topics of interest. The most common vulnerability database is the National Vulnerability Database (NVD) (National Vulnerability Database, 2018) which is maintained by the U.S. government and holds more than 100.000 unique CVE (Mell and Grance, 2002) identifiers with corresponding descriptions. Moreover, it provides a severity rating (Mell et al., 2006) on vulnerabilities which can be used for risk assessment. The 'Symantec Internet Security Threat Report' website (Symantec threat report, 2018) is an example for a vendor web site which releases an Internet security threat report annually. A security mailing list archive is provided by the 'SecList' (SecList, 2018) website which stores and manages multiple well-known mailing lists such as the 'Full Disclosure' mailing list.

4.2.2. *Structure of the provided security information*

As depicted in Fig. 3, the majority of information security data sources, including 78% of the sources, are not structured based on a standard. Accordingly, they have been classified as unstructured. The remaining 22% are structured based on a standard. However, they rely only on two standard namely the Structured Threat Information Expression (STIX) (Barnum, 2012) or Common Vulnerability and Exposures (CVE) (Mell and Grance, 2002) description format.

Moreover, we investigated if standardized enumerators can be found in structured and unstructured sources. Examples are CVE Mell and Grance (2002) identifiers, Common Weakness Enumerators (CWE) (Martin and Barnum, 2008), Common Platform Enumerators (CPE) (Cheikes et al., 2011), Common Attack Pattern Enumeration and Classification (CAPEC) (Barnum, 2008) and Common Vulnerability Scoring System (CVSS) scores (Mell et al., 2006). We found out that CVE enumerators appear in most of the cases (47%), while CVSS and CWE appear only in 20% or 13% of the cases. The other enumerators are only present in one or two information security data sources.

4.2.3. *Integrability of information security data sources*

In order to investigate the Integrability of information security data sources, we analyzed the provided interfaces. As depicted in Fig. 3, the majority, including 59% of the sources do not provide any interface to automatically export data. This might be caused by the fact that the majority of public information security data sources, we considered for classification, are news websites or blogs. Merely 7% of the news websites provide a functionality to receive data in form of an RSS Feed. Furthermore, we observed that mailing list archives do not provide any interfaces.

Information security data sources that provide interfaces to export data are vendor websites, vulnerability databases, social networks, streaming portals, forums and sources classified as others. In this context 30% of the identified sources provide an API, 10% an RSS Feed and 16% an export functionality.

In addition, we analyzed the user interfaces of all ($n = 68$) identified information security data sources. Our investigation showed that 66% of all sources provide basic search functionalities and only 13% enable the statistical analysis of the provided information. However, in most cases these functionalities are kept simple and that is why they are limited.

4.2.4. *Trustworthiness of security information and sources*

Trustworthiness plays a crucial role when it comes to security information sharing (Steinberger et al., 2016). Firstly, we analyzed who has published the information. As depicted in Fig. 3 the majority, including 75% of the security information are published by security expert(s). It is worth mentioning, that a publisher was considered as security expert

if background information regarding the publisher, indicating expert knowledge in the field of information security, was available. Furthermore, 43% of the information are published by vendors and 25% by governmental institutions, such as the National Institute of Standards and Technology (NIST). Finally, 41% of the information are provided by normal users, which we cannot assign to the categories security expert(s), vendor, or government due to their limited background information. In this context, it is worth mentioning that a information security data source can have more than one type of publisher.

Secondly, we analyzed the traceability of the provided information. In doing so our analysis showed that 58% of the identified sources provide meta data which enable the traceability of the provided information. For example, they provide meta data such as contact information of publishers, release dates or change histories.

Finally, we analyzed if an information source provides user feedback mechanisms to validate or assess the provided information. Possible feedback mechanisms are comment or rating functions. These feedback mechanisms are provided by 50% of the identified information security data sources. However, the quality and usefulness of theses mechanisms can not be assessed since it bases on subjective judgments of users.

4.2.5. *Timeliness of security information*

In order to decide about the timeliness of security information, we analyzed the sharing behavior (i.e. when information is shared). Our investigations showed that 71% of the information is shared on routine information sharing and 65% of the information is shared immediately after an security event (e.g. attack) occurred.

4.3. *Originality of provided information*

As depicted in Fig. 3, the majority of sources were classified as primary sources, including 75% of all identified sources. A source classified as primary source provides original content without referencing another source. The remaining 25% of sources were classified as secondary sources, which obtain their contents from primary sources. For example, these meta-sources are vulnerability databases including information from other sources. This might be explained by the fact that a large number of vulnerability sources reference and obtain their data from the National Vulnerability Database (NVD) (National Vulnerability Database, 2018) and Mitre's CVE website (CVE, 2018).

4.4. *Different types of security information and their co-occurrence*

Fig. 5 outlines the co-occurrence of different types of provided security information, including information regarding vulnerabilities, threats, countermeasures, attacks, risks and assets. The size of nodes in Fig. 5 describe the frequency of a certain type of information. It is obvious that the most occurring type of information focuses on vulnerabilities, included in 90% of the identified sources. Secondly, security information regarding attacks can be found in 70% of the sources. Information regarding threats can be found in 53%, countermeasures

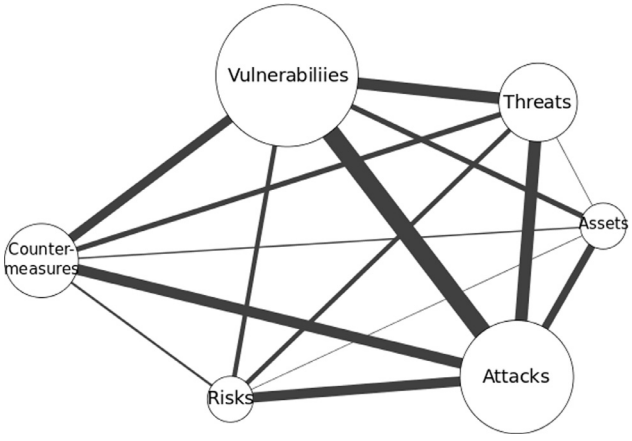

Fig. 5 – Co-occurrence of types of provided information.

in 50%, assets in 40% and risks in 32% of the identified information security data sources. Moreover, the intensities of the edges between the nodes in Fig. 5 describe the co-occurrence relationship of different types of security information. Fig. 5 shows that information security data sources often contain information regarding vulnerabilities and attacks. Moreover, a strong connection between vulnerabilities and threats or rather attacks and threats can be observed. Accordingly, it can be assumed that there are many cases where information regarding vulnerabilities or attacks occur together with threat information. A weaker connection between countermeasures and vulnerabilities or attacks can be observed. Links that connect risks, assets and countermeasure show the weakest relationships.

4.5. *Interfaces to different types of information*

Fig. 6 shows the relationship between the types of provided security information and the availability of interfaces by illustrating the available interfaces per information type. It is obvious that all types of information can be obtained through all types of interfaces. However, the number of sources that provide a certain type of information over a specific interface differs. As depicted in Fig. 6, the majority of sources do not provide any interface and if they provide an interface they mostly focus on APIs. A closer look on the different types of information shows that vulnerability and attack information are the most common information types that can be obtained over interfaces. This can be traced back to the fact that they provide a higher number of sources with interfaces compared to other types of information.

5. Discussion and limitations

The main contribution of this paper is twofold. Firstly, we introduced a taxonomy to classify information security data sources and their contents. Secondly, the taxonomy was used to classify the 68 identified information security data sources. In this section we discuss the classification results and limitations of the research at hand.

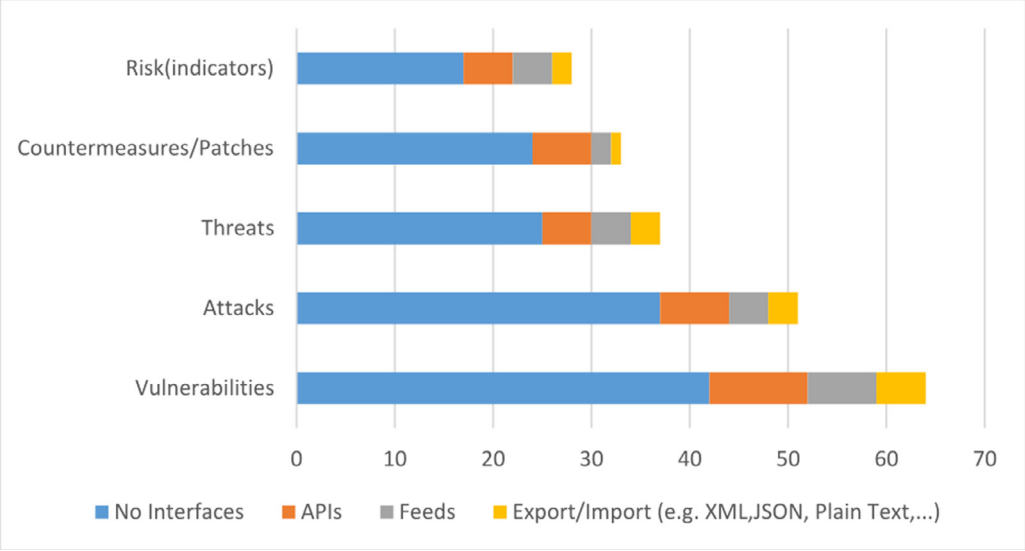

Fig. 6 – Available interfaces per information type.

5.1. *Information security data sources and their Vulnerability Focus*

The analysis of the identified information security data sources indicated, that the primary focus is on information regarding vulnerabilities and corresponding attacks. This strong vulnerability focus might be traced back to our data collection approach on Twitter which bases on CVE identifiers. Under certain circumstances this search strategy might embody some sort of bias compared to other methods that we used during our triangulation study. A classification of the identified websites in the Tweets showed the following distribution of information types: 60% Vulnerability, 25% Countermeasure, 8% Exploit information and the remaining 7% were not classifiable. In addition, there was a considerable amount of sources that we identified through the literature study and exploratory survey. Taking the results of all applied methodologies in consideration, we can conclude that public information security data sources have a strong vulnerability focus. In addition, a comprehensive analysis of the identified information security data sources showed that a considerable number of sources reference or integrate information from official approved vulnerability databases such as the National Vulnerability Database (National Vulnerability Database, 2018). This leads to the conclusion that vulnerability databases serve as primary source for secondary sources, where the contents are extended or adopted for their own purposes.

Moreover, we figured out that the strong vulnerability focus of the identified information security data sources enables the mapping of the majority of their contents to the vulnerability life cycle model (Arbaugh et al., 2000; Frei et al., 2008), which can be consolidated for risk assessment during information security risk management. Thereby, information regarding the discovery and disclosure of vulnerabilities or the availability of countermeasures or threats are of potential interest to information security stakeholders. Moreover, our analysis showed that these information types are available over the identified information security data sources and to a certain extend they appear together with vulnerability information.

5.1.1. *Degree of standardization*

Many recent publications analyzed and discussed a plethora of standards for describing and sharing security information (Johnson et al., 2016; Kampanakis, 2014; Skopik et al., 2016; Steinberger et al., 2015). However, our analysis showed that only 22% of the identified security sources rely on standardized representations of security information. On closer examination it emerged that they primarily rely on two standardized representation formats, namely Common Vulnerability Exposures (CVEs) (Mell and Grance, 2002) and Structured Threat Information eXpressions (STIX) (Barnum, 2012). Therefore, it can be concluded that in this context standards are scarcely used. This might be traced back to the fact that most of the information security data sources are informal security information where information is generated from a heterogeneous group of users in a crowdsourcing manner without using a standardized data format or procedure.

However, we identified several information security data sources that contain standardized identifiers regarding security information, such as CVE identifiers (47% of identified sources). This is an interesting fact as these identifiers enable the (automated) collection, tracking and aggregation of security information regarding the same topics from different sources (Mell and Grance, 2002). For example, discussions on security expert blogs and news websites referencing or containing the same CVE identifier can be easily linked together and made available to as security knowledge to various stakeholders like software developers (Felderer and Pekaric, 2017).

5.1.2. *Limited automated integration functionalities*

Security information obtained from public sources are mostly unstructured which makes it difficult to automatically integrate the available information. In addition, the majority of sources do not provide any interfaces (e.g. APIs) which makes it much more difficult to integrate the provided information.

In order to automatically extract and obtain information from these sources data crawling is required.

However, we identified sources that provide integration functionalities, such as APIs, export functionalities or RSS Feeds. However, most of the interfaces provide limited search functionalities, which results in difficulties to obtain the needed information and might cause an information overload. For example, some sources provide an XML export function, where only the whole data dump of a respective security information source can be exported. Consequently, the information must be processed and analyzed by the consumer.

In order to analyze the crawled unstructured information and handle the information overload caused by the provided interface, big data analytic techniques, such as machine learning or natural language processing, are needed (Dua and Du, 2016). For example, machine learning techniques can be used to cluster the obtained information according to security topics or remove useless security information. Moreover, these techniques might enable the automatic processing, aggregation and elimination of duplicates in the collected information.

5.1.3. *Trustworthiness of public available information security data sources*

In the field of security information sharing trust plays a crucial role (Steinberger et al., 2016), since information obtained from public information security data sourcess might be consolidated for information security decisions affecting critical business processes. Consequently, trust in the publisher or provider of security information is inevitable (Sillaber et al., 2016). Moreover, there might be the problem of untrustworthy providers spreading false information regarding security issues. For example, an attacker might spread false information regarding a vulnerability that leads a victim to desired actions, such as putting a countermeasure in place to open a back door for an attacker.

According to the nature of security information on public information channels it is barely feasible to evaluate the trustworthiness of the provided information. During our investigations we analyzed if the credibility, traceability or quality of the provided security information can be evaluated. We found out that 75% of the information on public information security data sources is provided by security experts. However, this fact doesn't exclude that security experts are spreading false information. Therefore, we analyzed validation functionalities, such as user ratings or comment fields, as feedback mechanisms for users. For example, comments or ratings regarding the usefulness of the provided information from users, independent from the information provider, would be helpful to increase trust in the provided security information. Last but not least, we can state that information regarding the traceability of the provided information increase trust as well. In this context our investigation showed that in nearly 60% of the cases traceability of the provided information is ensured.

5.1.4. *Integration of information security data to information security and risk management processes*

Information Security and Risk Management Processes, such as described in the ISO/IEC 27005 on information security risk management (ISO/IEC, 2011), use security information as input for risk assessment processes. Apart from the aforementioned challenges to automatically integrate largely unstructured information through rudimentary or inexistent interfaces, a formal process or workflow how these types of information can be used is missing. Hence, it can be assumed that the information from these sources is accessed and used in an unstructured manner. Moreover, this phenomenon is comparable to Shadow Cyber Threat Intelligence which undermines official approved information security data sources and, thus, leads to a lack of documentation and traceability (Sauerwein et al., 2018). In this context a comprehensive framework to specify all relevant information security data sources for information security risk management and a subsequent collection and processing of data would be needed.

5.2. *Threats to validity*

The research at hand might be limited by certain threats to validity that have been acknowledged. In order to minimize them from the very beginning, we carried out a triangulation study consisting of a systematic literature study, a quantitative analysis of security expert discussions on Twitter and a exploratory survey with experts in the field. Limitations that have to be acknowledged and accounted for are (i) threats to the identification of information security data sources, (ii) threats to the development of the classification taxonomy and (iii) threats to the classification of information security data sources.

In order to counteract (i) we carried out a triangulation study consisting of quantitative and qualitative methods. In doing so, we included information security data sources described in scientific literature, shared on Twitter and used in practice by security experts responsible for information security processes.

Firstly, for the identification of the initial set of papers for the snowballing methodology we used a search string. The strategy to construct the search string aimed to identify as many relevant publications as possible discussing information security data sources and their classification. Therefore, we decided to use the term "cyber security" within the search terms. However, it is impossible to identify all relevant publications with the described initial search. Therefore, we applied the snowballing methodology until no new papers were identified. Therefore, the risk of missing publications seem to be a moderate threat to validity.

Secondly, it may be argued that the CVE-based data collection on Twitter might bias the study results since the cross validation and elimination of Tweets containing wrong CVE identifiers strongly depends on the MITRE's database and the National Vulnerability Database. This threat seem to be moderate, since the aim of the collection and described cross-validation approach was to identify only information security relevant contents discussed on Twitter between May 23, 2016 and March 27, 2018. Moreover, CVE identifiers are well-established and strongly represented on Twitter. Therefore, our approach represents a systematic way of identifying information security contents without generating false-positives like machine learning techniques.

Thirdly, there might be a selection bias of participants for our exploratory study. This threat can be seen as moder-

ate since participation was voluntarily and we contacted two mailing lists with a representative set of medium to large-sized organizations located in Europe.

We merged the lists of identified information security data sources obtained from the systematic literature study, quantitative analysis of security expert discussions on Twitter and exploratory survey with experts in the field. The different methods delivered us similar information security data sources which we included in our final list. Therefore, the risk of missing relevant information security sources and the selection bias can be classified as moderate.

Additionally, we defined inclusion and exclusion criteria for information security data sources that enabled a systematic selection. It is worth mentioning that a few identified information security data sources were not publicly available and we were not able to classify them. Fortunately, this concerned only 4% of the total number of identified sources and we decided to exclude them from further investigations.

In order to counteract (ii), we based our classification on the results of the systematic literature study, standards in the field and metrics for data quality. Moreover, we adjusted the taxonomy during classification. For example, we added a new category to the dimension *Type of Source* whenever we found two cyber security information sources of the same type.

Finally, in order to inhibit (iii) we have chosen a type of cross validation approach in which each contributor to this research was given a subset of cyber security information sources to analyze and classify that intersected with another contributor's set. This procedure enabled us to identify classification discrepancies and resolve them through re-classification, discussion and and a subsequent majority vote. The issues that were encountered relate to classification of sources based on standards, originality of a source, duplicate entries and trustworthiness.

6. Conclusion

In this paper, we provide a comprehensive analysis and classification of public information security data sources used in research an practice. Therefore, we conducted a triangulation study consisting of a systematic literature review, a quantitative analysis of security expert discussions on Twitter and an exploratory survey on information security data sources used in practice. Based on our research methodology we compiled a list of 68 information security data sources for analysis. In order to classify these sources, we introduced a taxonomy including the following six dimensions: (1) Type of provide information, (2) Integrability, (3) Timeliness, (4) Originality, (5) Type of Source and (6) Trustworthiness. Secondly, we used the taxonomy to classify our identified 68 information security data sources. Our investigations showed that most of the sources focus on information regarding vulnerabilities, the information is available in unstructured form, automatic integration is limited feasible according to missing or insufficient interfaces, and some of the sources obtain and duplicate the information from well-known information security data sources. Moreover, we stated that a framework to systematically and automatically integrate public information sources to Information Security Risk Management Processes would be needed. Future work, will focus on the development of such a framework in order to extract valuable information based on our compiled list of information security data sources.

Acknowledgment

This work was partially supported by the Austrian Science Fund (FWF) through the research project FWF P 26194-N15: Model-Based Security Testing of Clouds (MOBSTECO).

Appendix A

Table 2 outlines the 68 identified public information security data sources with a corresponding mapping to the type of information source. In doing so, we distinguish among security news websites, blogs, vulnerability databases, vendor websites, mailing lists, streaming portals, social networks and forums. The remaining information sources are classified as others. As described in Section 3.5, if only one source of a particular type was found, it was classified as other.

Table 2 – Overview of 68 information security data sources.

		News website	Blog	Vulnerability Database	Vendor Website	Mailing List	Streaming Portal	Social Network	Forum	Other
1	https://cve.mitre.org/			X						
2	https://www.bugtrack.net/									X
3	https://www.phishtank.com/									X
4	https://otx.alienvault.com/								X	
5	https://nvd.nist.gov/			X						
6	https://www.cvedetails.com/			X						
7	https://lists.sans.org/					X				
8	https://www.enisa.europa.eu/topics/threat-risk-management/									X
9	https://www.virustotal.com/de/									X
10	https://www.us-cert.gov/	X								
11	https://www.welivesecurity.com/	X								
12	https://inteltechniques.com/									X

(*continued on next page*)

Table 2 (*continued*)

		News website	Blog	Vulnerability Database	Vendor Website	Mailing List	Streaming Portal	Social Network	Forum	Other
13	https://www.reddit.com/							X		
14	https://seclists.org/fulldisclosure/					X				
15	https://www.fsisac.com/	X								
16	https://thehackernews.com/	X								
17	https://secuniaresearch.flexerasoftware.com/			X						
18	https://www.securityfocus.com/			X						
19	http://www.security-database.com/			X						
20	http://cxsecurity.com/			X						
21	https://www.itsecuritynews.info/	X								
22	https://github.com/									X
23	https://twitter.com/							X		
24	https://securityaffairs.co	X								
25	https://seclists.org/					X				
26	https://tsecurity.de/	X								
27	https://vulners.com/l			X						
28	https://vuldb.com/			X						
29	https://www.symantec.com/				X					
30	https://www.akaoma.com/									X
31	https://bugs.chromium.org/								X	
32	https://www.openwall.com/									X
33	http://legalhackers.com/		X							
34	https://www.youtube.com/						X			
35	https://www.debian.org/Bugs/				X					
36	https://www.fortinet.com/blog		X							
37	https://blog.0patch.com/		X							
38	https://www.schneier.com/		X							
39	https://blog.trendmicro.com/		X							
40	https://thehackernews.com/	X								
41	https://www.kitploit.com/		X							
42	https://nakedsecurity.sophos.com/	X								
43	https://blog.quarkslab.com/		X							
44	https://malware.dontneedcoffee.com/blog/	X								
45	https://esecpro.blogspot.com/		X							
46	https://blog.quttera.com/		X							
47	https://www.ibm.com/security				X					
48	https://www.openssl.org/				X					
49	https://alephsecurity.com/		X							
50	http://boosterok.com/blog/		X							
51	https://www.dailymotion.com/						X			
52	http://www.misp-project.org/									X
53	https://crits.github.io/									X
54	https://www.openbsd.org/				X					
55	https://docs.microsoft.com/en-us/security-updates/				X					
56	https://www.mozilla.org/en-US/security/advisories/				X					
57	https://www.bugzilla.org/									X
58	https://www.sans.org/									X
59	https://www.symantec.com/security-center/threat-report				X					
60	https://www.trendmicro.com/vinfo/us/security/research-and-analysis/threat-reports				X					
61	https://security.googleblog.com/		X							
62	https://krebsonsecurity.com/		X							
63	https://threatpost.com/	X								
64	https://cyberwarzone.com/	X								
65	https://www.darkreading.com/	X								
66	https://www.dshield.org/									X
67	https://www.securitynow.com/	X								
68	https://securityweek.com/	X								
	Total Number	15	13	9	9	3	2	2	2	13

REFERENCES

Abu MS, Selamat SR, Ariffin A, Yusof R. Cyber threat intelligence–issue and challenges. Indones J Electr Eng Comput Sci 2018;10(1).

Arbaugh WA, Fithen WL, McHugh J. Windows of vulnerability: a case study analysis. Computer 2000;33(12):52–9.

Barnum S. Common attack pattern enumeration and classification (CAPEC) schema description. Cigital Inc, http://capec.mitre.org/documents/documentation/CAPEC_Schema_Description_v1, 2008; 3.

Barnum S. Standardizing cyber threat intelligence information with the structured threat information expression (stix). MITRE Corp 2012;11:1–22.
Batini C, Cappiello C, Francalanci C, Maurino A. Methodologies for data quality assessment and improvement. ACM Comput Surv 2009;41(3):16.
Brereton P, Kitchenham BA, Budgen D, Turner M, Khalil M. Lessons from applying the systematic literature review process within the software engineering domain. J Syst Softw 2007;80(4):571–83.
Brown S, Gommers J, Serrano O. From cyber security information sharing to threat management. In: Proceedings of the 2nd ACM workshop on information sharing and collaborative security. ACM; 2015. p. 43–9.
Burger EW, Goodman MD, Kampanakis P, Zhu KA. Taxonomy model for cyber threat intelligence information exchange technologies. In: Proceedings of the 2014 ACM workshop on information sharing & collaborative security. ACM; 2014. p. 51–60.
Carver JC, Burcham M, Kocak SA, Bener A, Felderer M, Gander M, King J, Markkula J, Oivo M, Sauerwein C, et al. Establishing a baseline for measuring advancement in the science of security: an analysis of the 2015 ieee security & privacy proceedings. In: Proceedings of the symposium and bootcamp on the science of security. ACM; 2016. p. 38–51.
Cheikes B.A., Waltermire D., Scarfone K.. Common platform enumeration: naming specification version 2.3. NIST Interagency Report 7695, NIST-IR2011; 7695.
Common vulnerability and exposures (CVE) mitre https://cve.mitre.org/. Accessed: 2018-04-09.
Dua S, Du X. Data mining and machine learning in cybersecurity. CRC press; 2016.
Felderer M, Carver JC. Guidelines for systematic mapping studies in security engineering. In: Empirical research for software security. CRC Press; 2017. p. 67–88.
Felderer M, Fourneret E. A systematic classification of security regression testing approaches. Int J Softw Tools Technol Transf 2015;17(3):305–19.
Felderer M, Pekaric I. Research challenges in empowering agile teams with security knowledge based on public and private information sources. Proceedings of the international workshop on secure software engineering in DevOps and Agile Development (SecSE 2017), 2017.
Felderer M, Zech P, Breu R, Büchler M, Pretschner A. Model-based security testing: a taxonomy and systematic classification. Softw Test Verif Reliab 2016;26(2):119–48.
Fenz S, Heurix J, Neubauer T, Pechstein F. Current challenges in information security risk management. Inf Manag Comput Secur 2014;22(5):410–30.
Fowler Jr FJ. Survey research methods. Sage Publications; 2013.
Frei S, Tellenbach B, Plattner B. 0-day patch-exposing vendors (in) security performance. BlackHat Europe, 2008.
Gentleman R., Ihaka R., Bates D., Chambers J., Dalgaard J., Hornik K.. The r project for statistical computing. http://www.r-project.org/254, 2009.
Harel Y, Gal IB, Elovici Y. Cyber security and the role of intelligent systems in addressing its challenges. ACM Trans Intell Syst Technol 2017;8(4):49.
Hernandez-Ardieta JL, Tapiador JE, Suarez-Tangil G. Information sharing models for cooperative cyber defence. In: Proceedings of the 2013 5th international conference on cyber conflict (CyCon). IEEE; 2013. p. 1–28.
IEC I. 27000 2014. Inf Secur Defin 2014;32.
ISO I., Std I.. Iso 15408-1: 2009. Information technology-Security techniques-Evaluation criteria for IT security-Part; 1.
ISO/IEC. ISO/IEC 27005:2013: Information technology - security techniques - information security risk management 2011.
It security news, https://www.itsecuritynews.info/. Accessed: 2018-04-09.
Jalali S, Wohlin C. Systematic literature studies: database searches vs. backward snowballing. In: Proceedings of the ACM-IEEE international symposium on empirical software engineering and measurement. ACM; 2012. p. 29–38.
Jang-Jaccard J, Nepal S. A survey of emerging threats in cybersecurity. J Comput Syst Sci 2014;80(5):973–93.
Jick TD. Mixing qualitative and quantitative methods: triangulation in action. Admin Sci Q 1979;24(4):602–11.
Johnson C, Badger L, Waltermire D, Snyder J, Skorupka C. Guide to cyber threat information sharing. NIST Spec Publ 2016;800:150.
Kampanakis P. Security automation and threat information-sharing options. IEEE Secur Priv 2014;12(5):42–51.
Making security measurable, https://makingsecuritymeasurable.mitre.org/. Accessed: 2018-04-09.
Martin RA, Barnum S. Common weakness enumeration (CWE) status update. ACM SIGAda Ada Lett 2008;28(1):88–91.
Massacci F, Nguyen VH. Which is the right source for vulnerability studies?: An empirical analysis on Mozilla Firefox. In: Proceedings of the 6th international workshop on security measurements and metrics. ACM; 2010. p. 4.
Mavroeidis V, Bromander S. Cyber threat intelligence model: an evaluation of taxonomies, sharing standards, and ontologies within cyber threat intelligence. Proc IEEE 2017.
Mell P, Grance T. In: Technical Report. Use of the common vulnerabilities and exposures (CVE) vulnerability naming scheme. DTIC Document; 2002.
Mell P, Scarfone K, Romanosky S. Common vulnerability scoring system. IEEE Secur Priv 2006;4(6).
Menges F, Pernul G. A comparative analysis of incident reporting formats. Comput Secur 2018;73:87–101.
Mittal S, Das PK, Mulwad V, Joshi A, Finin T. Cybertwitter: using twitter to generate alerts for cybersecurity threats and vulnerabilities. In: Proceedings of the 2016 IEEE/ACM international conference on advances in social networks analysis and mining (ASONAM). IEEE; 2016. p. 860–7.
National vulnerability database (NVD), https://nvd.nist.gov/. Accessed: 2018-04-09.
Pfleeger SL. Experimental design and analysis in software engineering. Ann Softw Eng 1995;1(1):219–53.
Puhakainen P, Siponen M. Improving employees' compliance through information systems security training: an action research study. Mis Q 2010:757–78.
Qamar S, Anwar Z, Rahman MA, Al-Shaer E, Chu BT. Data-driven analytics for cyber-threat intelligence and information sharing. Comput Secur 2017;67:35–58.
Rader E, Wash R. Identifying patterns in informal sources of security information. J Cybersecur 2015;1(1):121–44.
Sauerwein C, Gander M, Felderer M, Breu R. A systematic literature review of crowdsourcing-based research in information security. In: Proceedings of the 2016 IEEE symposium on service-oriented system engineering (SOSE). IEEE; 2016. p. 364–71.
Sauerwein C, Sillaber C, Breu R. Shadow cyber threat intelligence and its use in information security and risk management processes. Multikonferenz Wirtschaftsinformatik (MKWI), 2018.
Sauerwein C, Sillaber C, Mussmann A, Breu R. Threat intelligence sharing platforms: an exploratory study of software vendors and research perspectives. In: Procceedings of the 13. Internationale Tagung Wirtschaftsinformatik 2017 (WI 2017). AIS; 2017. p. 837–51.
Schneier on security, https://www.schneier.com/. Accessed: 2018-04-09.
Seclist, http://seclists.org/. Accessed: 2018-04-09.
Sillaber C, Sauerwein C, Mussmann A, Breu R. 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. ACM; 2016. p. 65–70.
Skopik F, Settanni G, Fiedler R. A problem shared is a problem halved: A survey on the dimensions of collective cyber defense through security information sharing. Comput Secur 2016;60:154–76.
Soomro ZA, Shah MH, Ahmed J. Information security management needs more holistic approach: a literature review. Int J Inf Manag 2016;36(2):215–25.
Steinberger J, Kuhnert B, Sperotto A, Baier H, Pras A. In whom do we trust-sharing security events. In: Proceedings of the IFIP international conference on autonomous infrastructure, management and security. Springer; 2016. p. 111–24.
Steinberger J, Sperotto A, Golling M, Baier H. How to exchange security events? Overview and evaluation of formats and protocols. In: Proceedings of the 2015 IFIP/IEEE international symposium on integrated network management (IM). IEEE; 2015. p. 261–9.
Symantec threat report, https://www.symantec.com/security-center/threat-report. Accessed: 2018-04-09.
Tounsi W, Rais H. A survey on technical threat intelligence in the age of sophisticated cyber attacks. Computers & Security, Elsevier. Elsevier; 2018. p. 212–33.
Tripathi A, Singh UK. Taxonomic analysis of classification schemes in vulnerability databases. In: Proceedings of the 2011 6th international conference on computer sciences and convergence information technology (ICCIT). IEEE; 2011. p. 686–91.
Tweepy python library, http://www.tweepy.org/, note = Accessed: 2018-04-09.
Usman M, Britto R, Börstler J, Mendes E. Taxonomies in software engineering: a systematic mapping study and a revised taxonomy development method. Inf Softw Technol 2017.
Webster J, Watson RT. Analyzing the past to prepare for the future: writing a literature review. MIS Q 2002 xiii xxiii.
Wohlin C. 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. ACM; 2014. p. 38.
Zhao W, White G. A collaborative information sharing framework for community cyber security. In: Proceedings of the 2012 IEEE conference on technologies for homeland security (HST). IEEE; 2012. p. 457–62.

Clemens Sauerwein is a Ph.D. student and researcher at the Institute of Computer Science at the University of Innsbruck, austria. His research interests include information security risk management, cyber threat intelligence sharing, empirical studies in the field of information security risk management and information systems. He works in close collaboration with industry and transfers his results into practice as a consultant and a member of a security interest group.

Irdin Pekaric is a Ph.D. student and a researcher at the University of Innsbruck, Austria. He is part of the QE (Quality Engineering) research group since 2016. His research areas are in the field of information security. Specifically, he conducts a research on attack model generation for integrated security and safety analysis. His prior work focused on user anomaly detection on both host and network level.

Michael Felderer is a professor in software engineering at the Institute of Computer Science at the University of Innsbruck, Austria and a guest professor at the Blekinge Institute of Technology, Sweden. He holds a Ph.D. and a habilitation degree in computer science. His research interests include in software and security engineering comprise software and security, empirical methods in software and security engineering, software and security processes, software analytics, risk management, requirements engineering, model engineering, and improving industry-academia collaboration. He works in close collaboration with industry and transfers his research results into practice as a consultant and speaker on industrial conferences.

Ruth Breu is head of the Institute of Computer Science at the University of Innsbruck and head of the research group Quality Engineering. She is expert in the areas of Requirements Engineering, Security Engineering and Enterprise Architecture Management. Together with her team she develops tool-based methods for information security management and IT asset documentation with a high degree of automation, collaboration support and situation-awareness. Ruth is co-author of more than 150 international publications and contributor to the scientific community as editor, conference and PC chair.