Data Quality Challenges and Future Research Directions in Threat Intelligence Sharing Practice

Christian Sillaber
Institute of Computer Science
University of Innsbruck
christian.sillaber@uibk.ac.at

Clemens Sauerwein
Institute of Computer Science
University of Innsbruck
clemens.sauerwein@uibk.ac.at

Andrea Mussmann
Institute of Computer Science
University of Innsbruck
andrea.mussmann@uibk.ac.at

Ruth Breu
Institute of Computer Science
University of Innsbruck
ruth.breu@uibk.ac.at

ABSTRACT

In the last couple of years, organizations have demonstrated an increased willingness to participate in threat intelligence sharing platforms. The open exchange of information and knowledge regarding threats, vulnerabilities, incidents and mitigation strategies results from the organizations' growing need to protect against today's sophisticated cyber attacks. To investigate data quality challenges that might arise in threat intelligence sharing, we conducted focus group discussions with ten expert stakeholders from security operations centers of various globally operating organizations. The study addresses several factors affecting shared threat intelligence data quality at multiple levels, including collecting, processing, sharing and storing data. As expected, the study finds that the main factors that affect shared threat intelligence data stem from the limitations and complexities associated with integrating and consolidating shared threat intelligence from different sources while ensuring the data's usefulness for an inhomogeneous group of participants.Data quality is extremely important for shared threat intelligence. As our study has shown, there are no fundamentally new data quality issues in threat intelligence sharing. However, as threat intelligence sharing is an emerging domain and a large number of threat intelligence sharing tools are currently being rushed to market, several data quality issues – particularly related to scalability and data source integration – deserve particular attention.

Keywords

Threat Intelligence Sharing Data Quality; Threat Intelligence Data; Data Quality Challenges

WISCS'16, October 24 2016, Vienna, Austria

ISBN 978-1-4503-4565-1/16/10...$15.00

DOI: http://dx.doi.org/10.1145/2994539.2994546

1. INTRODUCTION

The increasing complexity, heterogeneity, and interconnectedness of information systems has lead to a significant increase in the number of security related attacks [11, 14]. Recent prominent security incidents have shown that a successful cyber attack can quickly lead to a devastating loss of intellectual property, productivity, and money, and can instantly diminish an organization's reputation [11, 17]. Research and practice have shown that there is a strong need for the exchange of data, information and knowledge between organizations to aid the management of vulnerabilities, threats and to mitigate incidents [4, 5].

In practice, an increased willingness of organizations to exchange threat intelligence among each other can be observed and several government-sponsored projects have gained traction in recent years [6]. For example, in the Netherlands the government has introduced a national detection, response and expertise network [7] and NATO has initiated the Cyber Security Data Exchange and Collaboration Infrastructure (CDXI) project [4]. Additionally, several standardization efforts (e.g. CybOX, STIX and TAXII) to support automated threat intelligence sharing gained attention in recent years [9, 8, 19, 12].

The data quality of shared threat intelligence plays an important role as inaccurate data can result in undesired effects [18]. The assumption underlying our research is that threat intelligence sharing efforts are prone to the same data quality problems that plague traditional data sets, namely accuracy, completeness, consistency, timeliness, and relevance. However, to the best of our knowledge, it is not very well understood how these quality dimensions influence threat intelligence data. As research has yet to be conducted in this area, shared threat intelligence may require completely different strategies and tools for data quality management.

The goal of our research is, therefore, to investigate data quality in shared threat intelligence data sets through a series of focus group discussions with ten threat intelligence stakeholders working at security operation centers of ten globally operating organizations. We probed both traditional data quality problems as well as data quality issues relevant to early adopters, data quality issues relevant to specific stakeholder groups and end users of threat intelligence sharing platforms.

While the number of focus group interviews was not sufficient to draw any strong statistical conclusions, the study provides a foundation from which to draw valid, generalizable conclusions about data quality challenges in shared threat intelligence, as well as inform future data quality research initiatives in this emerging domain.

The remainder of this paper is structured as follows. Section 2 provides related work regarding threat intelligence sharing platforms and corresponding challenges. Section 3 outlines the underlying methodology and procedure carried out. Section 4 outlines the results and key findings of the expert discussions. Section 5 gives recommendations and future research perspectives elicited from the key findings. Finally, Section 6 concludes the paper and provides an outlook on future research.

2. RELATED WORK

While organizations have traditionally shared threat intelligence using ad-hoc solutions such as email exchange, phone calls or ticketing systems, a trend to building interconnected communities with associated platforms for the (semi-)automated exchange of threat intelligence could be observed in recent years [2]. An exchange of security knowledge between experts across organizations is desirable as not every organization has the resources to develop adequate security programs independently and organizations can learn from other organizations' mistakes [5].

For example, the Netherlands have introduced the National Detection Network (NDN) [7] to support Dutch organizations in the exchange of relevant threat intelligence. Similarly, NATO has initiated the Cyber Security Data Exchange and Collaboration Infrastructure (CDXI) [4], that provides a knowledge management tool which facilitates the sharing of information on cyber threats.

Several efforts have been made to facilitate threat intelligence sharing in a standardized manner [2]. As a result, a number of compatible and incompatible data formats, protocols, and frameworks have been introduced, including the Common Vulnerability Exposure (CVE), Structured Threat Information Expression (STIX), Trusted Automated eXchange of Indicator Information (TAXII), Open Vulnerability and Assessment Language (OVAL) and the Common Attack Pattern Enumeration and Classification (CAPEC) [8, 19, 12].

In recent years, several authors have analyzed the requirements and challenges for threat intelligence sharing platforms.

Dandurand and Serrano [4] were the first who formulate 11 high-level requirements. While the authors state that customizable quality control of shared threat intelligence is one of the core requirements for such a platform, they do not propose any applicable quality requirements and controls.

Based on the main technology enablement and efficiency challenges experienced in threat intelligence sharing practice, Brown et al. [2] derived several requirements to support threat intelligence sharing. The requirements are general requirements and the authors do not present any tool-based or directly applicable solutions to them.

Serrano et al. [18] outlined key problems that should be addressed by threat intelligence sharing platforms. They primarily focus on fundamental aspects, e.g. legal issues, a threat intelligence sharing platform struggles with, and how they can be addressed.

To the best of our knowledge, no prior research has been conducted that analyzes data quality in threat intelligence sharing platforms.

3. PROCEDURE & STUDY DESIGN

The addressed research problem and research plan were formulated in close collaboration with two partners from industry, namely a security expert group developing a nationwide threat intelligence sharing platform and community as well as a group of potential members of such a community from industry. As already outlined, our goal was to obtain an understanding of the particular data quality challenges threat intelligence sharing faces. For this reason, expert focus group discussions were conducted.

The goal of the expert focus group discussion was to gain a deeper understanding of how organizations currently share threat intelligence, how it is disseminated and which data quality issues can be observed. We conducted two group discussions with ten domain experts at the managerial level with at least five years experience in the domain. The majority of participants had a university degree in computer science (CS) or information systems management (IS). All companies are based in a central European country and are predominately in finance (cf. Table 1). While all companies operate globally, one of them can be considered a small enterprise, two can be considered medium enterprises and the rest are large enterprises with vastly more than 1000 employees each.

The group interviews were conducted in March 2016 at a neutral premise and lasted for roughly one and half hours each. To prepare the group interviews, a telephone interview with developers of a threat intelligence sharing platform was conducted in order to clarify terminology and threat intelligence sharing concepts. During the expert focus group discussion, participants were asked to describe their organization's approaches to threat intelligence sharing and how shared threat intelligence data is used in their organization's processes. Furthermore, they were specifically instructed to discuss data quality issues. Most of the discussions were recorded and transcribed (some parts of the discussions could not be recorded due to confidentiality concerns raised by participants). We took notes during the telephone interviews and the expert focus group discussions. Written summaries were produced immediately after the interviews were completed.

The participants were instructed by two researchers before each session began. Ahead of the second session, short summaries of the previous discussion were given to the participants to encourage new inputs and speed up the new discussion [1]. The sessions were recorded and researchers asked questions during the discussion to clarify any outstanding issues. After completing the data collection, interview transcripts, the written interview summaries and the recordings were analyzed. Qualitative summaries were produced [13] from the collected data in order to extract the facets relevant to the research [3]. We created categories inductively by reducing, paraphrasing and generalizing relevant text passages.

The application of focus group discussions in empirical research might be limited by certain threats to validity [20, 10]. Limitations that have to be acknowledged and accounted for are a selection bias when selecting the participants, unwanted influences of the moderators during the discussions,

Table 1: Participants in the group discussions

ID	Organizational Role	Qualifications		Type of Org.	# of Employees
		Security Specific Certifications	*University Degree in CS or IS*		
1	Security Operations Team Leader		x		<150
2	Security Operations Officer	CISSP, CCSP		Insurance	>1000
3	IT Security Analyst	CISSP, CEH	x	Finance	>1000
4	Cyber Security Incident Response Team			Finance	>1000
5	Managed Security Service Provider		x		>1000
6	Security Specialist		x	Finance	150 - 1000
7	Information Security Officer	CISM		Finance	>1000
8	Head of Security Operations Center	CISA	x	Finance	>1000
9	Security	CISM, CISSP		Finance	>1000
10	Cyber Security Incident Response Team	CISSP, CEH	x	Production	150 - 1000

off-topic discussions, or language barriers. In order to minimize limitations, during group discussion and evaluation we followed the suggestions stated in [20]. According to these suggestions [20], (i) participants of a security expert group meeting were asked to voluntarily participate in our group discussion session, (ii) discussions were lead by skillful, empathetic, and reluctant moderators, (iii) moderators refocused the discussion as soon as it got off track, and (iv) the group discussions were held in English as a common language. Interview partners were managerial-level employees of medium-to-large enterprises and the majority of them were working in the financial sector, which might make results biased towards this particular type of sharing community.

4. FINDINGS

The analysis of the expert group discussions yielded insights into shared threat intelligence data quality which were condensed to a set of findings. These findings are listed and discussed in the following subsections.

Finding 1: Integration of threat intelligence sources amplifies preexisting data quality problems

We gathered from the interviews that stakeholders' trust in the quality of the data provided by the platform is of paramount importance. Interviewees state that traceability and provenance of the threat intelligence must be established and visible, especially as data is curated from different sources and by different stakeholders.

While we could observe that stakeholders trust threat intelligence data from known sources more than from unknown sources, we found that most interviewees agreed that threat intelligence sharing platform providers should act as a trusted mediator that ensures proper integration of different sources. Some types of data quality problems will get worse as as the number of participants and integrated data sources increases. As security data is shared between stakeholders, aggregated from different sources and linked to other data already present in the data sets, the number of base errors also increases.

Not only will the cost and required effort of tracking and fixing these data quality issues increase, operators of threat intelligence sharing platforms are challenged with the herculean task of managing data quality in an unobtrusive fashion - without the help of participating data consumers.

Finding 2: Combining short-lived shared threat intelligence from disjunct industries makes the important intelligence hard to find

Those interviewees already utilizing a threat intelligence sharing platform in their organization stated that it becomes increasingly difficult to find the right intelligence/information in time as data relevant to other industries is included in the platform. In this context, interviewee 3 for example, stated: "...I don't want to read threat intelligence from Company X [from another type of industry] the entire day, because I obviously don't have the time...".

We found that the data quality dimensions of timeliness and relevance were important to the security decision makers. In order to provide a valuable service to organizations participating in the threat intelligence sharing platform, the platform must both ensure that new intelligence is distributed in time and that outdated intelligence is hidden as soon as it loses relevance.

Finding 3: Existing threat intelligence sharing tools often limit data accessibility

The interviewees stated that in order to get the big picture of the current threat landscape, better dashboard like structures that can be freely configured are required. For example, interviewee 2 stated that "... when I access the platform, I need to know what is going on immediately.". As the daily activities of threat intelligence stakeholders can often be described as "looking for the needle in the haystack", stakeholders require qualitative and quantitative information about current incidents that are organized in a news-stream like fashion. However, as stakeholders reported, currently available threat intelligence sharing tools lack the customization, filters, news stream aggregation capabilities and search capabilities required for their daily work.

To manage the high influx of data, several interviewees suggested additional filters according to the geographic and language dimensions of ongoing attacks. For example, interviewee 3 stated that "I only care about phishing campaigns in the languages that are spoken in my organization. If the mails come in Swedish, my employees are smart enough to

ignore them. Therefore, the costs of including them [and manually filtering them] outweighs the potential risks.".

Further missing filters and analysis functionality that have been mentioned were source of intelligence, category (e.g. data regarding incidents), methods of detection as well as time related information. In order to minimize the effort required by the participants to improve the data quality, it was deemed important to share enriched data with other stakeholders.

In summary, as threat intelligence sharing platforms face the challenge of not overwhelming end-users with unnecessary noise, it is important to provide convenient mechanisms for stakeholders to distill the signal quickly.

Finding 4: Manually generated quality errors are difficult to find and often occur due to a lack of common data entry rules

All interviewees not agreed that it is of utmost importance to not only use a common standard (e.g. STIX, TAXII) for threat intelligence sharing but also agreed on the importance of using and enforcing a common vocabulary for data entry to prevent data quality issues. For instance, while most threat intelligence sharing standards provide ample syntax for specifying threat intelligence, they do not provide a common language for full-text data fields such as description or title of entries.

Results from interviews also show a need for a common understanding of flags, tags and meta-data (e.g. risk rating results). To prevent quality errors due to e.g. industry peculiarities, it is necessary to establish and enforce a common understanding of what data and the circumstances under which to share (and when and not to share) them.

Current standardization efforts of threat intelligence sharing form the basis for building standardized exchange platforms but still fail to ensure interpretability of the data across company and industry boundaries.

Manual data export and import (from tools that provide no direct integration) introduce hard to find quality errors that are frequently semantic in nature. They are immune to most existing approaches of error detection and correction.

These errors result from missing standardization of data exchange between different tools, organizations' dependency on low-tech exchange of threat intelligence (e.g. via paper reports), limited access to required sources and occur very frequently at all the organizations the were represented.

Finding 5: Automated integration of external sources can improve data quality

The adoption of automated integration of threat intelligence and data sources will dramatically improve data quality while at the same time making it much easier and more efficient to work with shared threat intelligence by removing manual entry steps and eliminating human generated data entry errors.

Although benefits are obvious for public and anonymized sources, participants expressed restraint when discussing the automatic integration of internal sources due to privacy, security and compliance concerns.

5. RECOMMENDATIONS AND FUTURE RESEARCH

As data quality is extremely important for all cybersecurity decisions, the goal of our research was to discern whether new or different data quality issues exist in threat intelligence sharing. To summarize our findings, there are no fundamentally new data quality issues in shared threat intelligence. However some data quality challenges are more or less pronounced than others and definitely warrant future research - especially due to scalability effects as the number of sources and volume of data processed by threat intelligence sharing platforms might be significantly higher than in the past.

The recommendations presented in this section are drawn mainly from the findings in the previous section. The recommendations are of observational nature, practically oriented and derived from the interviews' results.

Recommendation 1: Ensure that the threat intelligence sharing standard and meta-model fit the stakeholders' needs

Utilize a threat intelligence sharing platform that either supports the meta-model best reflecting your stakeholders' needs, or ensure that the provided meta-model can be adapted accordingly.

Furthermore – also in terms of schema completeness – ensure that the used data model does not rely on too many free-text fields and agree on a common language and workflow for them. This is important to prevent data quality errors that result from manual data entry operations.

In terms of data completeness, make sure that existing data sources are integrated automatically and manual data fusion operations are limited. Especially, the integration of publicly available sources (e.g. DNS-lookups, WHOIS-lookups) should be directly mediated via the intelligence sharing platform.

Recommendation 2: Minimize complexity by unifying intelligence according to vulnerabilities

As stakeholders require realtime support for important security decisions, we found that grouping intelligence according to vulnerabilities is a viable strategy. Vulnerabilities are linked to artifacts (e.g. software) that is either used or not used in organizations. This allows to quickly filter for those vulnerabilities (and linked threats and risks) that truly affect the stakeholder. To further improve the data retrieval process it is worthwhile to automatically synchronize the asset landscape of an organization with the threat intelligence sharing platform to tailor data queries to a relevant subset.

Recommendation 3: Correct data at the main source and link intelligence at the borders

If data has to be corrected (either automatically or through manual intervention), this should happen at the main source of the data to minimize the impact. As shared threat intelligence is often aggregated from disparate sources that only push data into the platform, this is often a difficult task. However, we found that many data fusion activities (e.g. linking attacks to the same attacker, attacks exploiting the same vulnerability) can be performed better outside of the main platform. This also decreases the number of sources and associated complexity.

Recommendation 4: Trust in data is of utmost importance: inform users about data quality

As shown during the expert discussions, the four trust dimensions of provenance, verifiability, reputation and believability are highly relevant to users of threat intelligence sharing platforms.

Provenance, referring to the contextual meta-data representing the origin of the source, was determined to be important in two ways. The first perspective focuses on assessing the trustworthiness of the intelligence itself as an attribute derived from the trustworthiness of the intelligence and its source. The second perspective focuses on determining trust values further tailored to the needs of security decision makers. As current cybersecurity intelligence sharing standards do not account for any of these dimensions, this is a gap that needs to be addressed in future research.

The trust dimension of verifiability is particularly challenging, as participants of threat intelligence sharing platforms are faced with the dichotomy of wanting to share as much about attacks as possible, while still keeping within organizational and legal constraints. Therefore, it might not always be possible to verify every entry shared via such platforms. Moreover, Murdoch and Leaver analyzed this issue and concluded that there is always a conflict between the need for anonymity versus the need to trust the shared information [15].

Therefore, it is required to inform the user of the platform about the quality of the data. For example, if anonymous cybersecurity intelligence is shared (e.g. due to legal reasons), the platform must act as a trusted party and ensure that the users are informed about the limited precision of the available data.

If possible, the platform should ensure that only validated data can be entered into the platform in order to guarantee authenticity of shared information.

Recommendation 5: Crowdsource data quality management

Closely related to the provenance dimension [16], the reputation dimension is concerned with the reputation and – indirectly – the quality of the data. Provide ranking functionality so that publishers' subjective quality can be ranked by other participants.

If possible, implement ranking capabilities for data sets and include indicators such as rankings of CVEs.

The believability dimension [21], describing the extent to which the intelligence is regarded as true and credible, was rarely explicitly mentioned during the expert group discussions. If possible, inform stakeholders about the source of the provided intelligence, including information about the size of the security operating center of the reporting entity, the number of accesses requests to intelligence by large security operating centers and the number of participants of the threat intelligence sharing platform that have taken remediating steps.

Recommendation 6: Automate data quality error detection

The intrinsic dimension of accuracy [21], referring to the degree to which the data correctly represents the real world, was mostly discussed by the interviewees from the viewpoint of unusable data attributes, which can be – to a large degree – automatically detected.

Therefore, if possible, support the automatic detection of attributes that do not conform to the required syntax.

The dimension of validity-of-documents, referring to the valid usage of the underlying taxonomies and syntactic accuracy was extensively discussed (and already mentioned in preceding recommendations) and related quality errors can be often automatically detected.

If possible, implement mechanisms that ensure that words used in free-text fields conform to agreed-upon language and that values used in restricted fields conform to agreed-upon threat intelligence sharing taxonomy. More research is necessary to (semi-) automatically detect semantic data quality errors.

Recommendation 7: Focus on current threat intelligence

As the main interest of participants is to be able to quickly react to emerging threats, data quality improvement efforts should primarily focus on new and volatile data. Supported by appropriate mechanisms, entries that are no longer valid can often be removed or hidden.

If possible, threat intelligence with (technologically) outdated entries (e.g. vulnerabilities of tools no longer used) should be automatically hidden to minimize complexity. However, hidden entries should never be deleted as old data and intelligence related to obsolete systems might still be useful in the future.

Future Research Directions:

There are a number of follow up actions and research directions that can be pursued. Some of these represent unaddressed areas that require a long term investigation of a living threat intelligence sharing platform. Others represent ideas that were revealed during the research as interesting topics that could be explored much further with minimal effort in existing threat intelligence sharing tools.

- More empirical research: As cybersecurity intelligence sharing platforms are an emerging issue, little empirical research has been conducted so far. Due to this, little is known about the "true" value they provide to organizations – and the impact data quality has on this value.

- Empirical evaluation of data formats and taxonomies: While several standardization efforts strive to establish common formats of exchange, fusion and language, they lack empirical validation and are often grounded in the needs of a specific industry or company that pushes their respective standard. Also, competing standards make generalizable research difficult. Current industry independent standardization efforts (e.g. by MITRE, OASIS, NIST) should be further pursued and the trade-off between generalization and specialization investigated.

- Empirical evaluation of shared threat intelligence: Since data quality assurance and control is associated with real data, more empirical investigations on the type of information shared via a threat intelligence sharing platform should be conducted. Based the results ade-

quate quality assurance measures and controls can be developed.

- Identification of additional data sources: Since data errors and inconsistencies are difficult to detect, additional data sources to validate threat intelligence against can be helpful. Therefore, the identification of valuable additional information security data sources which should be integrated into a threat intelligence sharing platform and the development of supporting quality assurance algorithms is desirable.
- Trust and communities: It is worthwhile to investigate data quality and its impact on stakeholder's trust in the data and how communities of "data quality excellence" form within such platforms.
- Threat intelligence use: Little is known about the inter- and intra-organizational use of data extracted from threat intelligence platforms by stakeholders and how the data is used in security decision making processes. Therefore, little is known about the precise impact data quality has on such processes and future research is needed.

6. CONCLUSION

As threat intelligence sharing platforms will gain widespread adoption in the foreseeable future, data quality management mechanisms are required to ensure that threat intelligence sharing platforms bring the promised benefits. As shown by our research, several data quality challenges in shared threat intelligence data exist, but they are not fundamentally new data quality challenges. However, due to the emerging nature of these data sharing platforms, their high data velocity as well as the immediate impact data quality has on organizational decision making processes, data quality in this context is an important issue and several unsolved areas warrant future research.

Acknowledgment: This work was supported by the Austrian Federal Ministry of Science, Research and Economics (BMWFW), QE LaB - Living Models for Open Systems (FFG 822740).

7. REFERENCES

[1] J. K. Brilhart and G. J. Galanes. *Effective group discussion*. McGraw-Hill Humanities, Social Sciences & World Languages, 1992.

[2] S. Brown, J. Gommers, and O. Serrano. From cyber security information sharing to threat management. In *Proceedings of the 2nd ACM Workshop on Information Sharing and Collaborative Security*, pages 43–49. ACM, 2015.

[3] J. L. Campbell, C. Quincy, J. Osserman, and O. K. Pedersen. Coding in-depth semistructured interviews problems of unitization and intercoder reliability and agreement. *Sociological Methods & Research*, 2013.

[4] L. Dandurand and O. S. Serrano. Towards improved cyber security information sharing. In *Cyber Conflict (CyCon), 2013 5th International Conference on*, pages 1–16. IEEE, 2013.

[5] S. Fenz, J. Heurix, T. Neubauer, and F. Pechstein. Current challenges in information security risk management. *Information Management & Computer Security*, 22(5):410–430, 2014.

[6] F. Fransen, A. Smulders, and R. Kerkdijk. Cyber security information exchange to gain insight into the effects of cyber threats and incidents. *e & i Elektrotechnik und Informationstechnik*, 132(2):106–112, 2015.

[7] E. V. D. HEUVEL and G. K. Baltink. Coordination and cooperation in cyber network defense: the dutch efforts to prevent and respond. *Best Practices in Computer Network Defense: Incident Detection and Response*, 35:121, 2014.

[8] P. Kampanakis. Security automation and threat information-sharing options. *Security & Privacy, IEEE*, 12(5):42–51, 2014.

[9] M. Kert, J. Lopez, M. Evangelos, and B. Preneel. State-of-the-art of secure ict landscape. Technical report, ENISA - NIS Platform - Working Group 3, 2014.

[10] K. Louise Barriball and A. While. Collecting data using a semi-structured interview: a discussion paper. *Journal of advanced nursing*, 19(2):328–335, 1994.

[11] L. Marinos and A. Sfakianakis. Enisa threat landscape-responding to the evolving threat environment. *ENISA (The European Network and Information Security Agency)(September 2012)*, 2012.

[12] R. A. Martin. Making security measurable and manageable. In *Military Communications Conference, 2008. MILCOM 2008. IEEE*, pages 1–9. IEEE, 2008.

[13] P. Mayring and M. Gläser-Zikuda. *Die Praxis der Qualitativen Inhaltsanalyse*. Beltz Weinheim, 2008.

[14] A. Miller, R. Horne, and C. Porter. 2015 information security breaches survey. Technical report, PWC, 2015.

[15] S. Murdoch and N. Leaver. Anonymity vs. trust in cyber-security collaboration. In *Proceedings of the 2nd ACM Workshop on Information Sharing and Collaborative Security*, pages 27–29. ACM, 2015.

[16] L. L. Pipino, Y. W. Lee, and R. Y. Wang. Data quality assessment. *Commun. ACM*, 45(4):211–218, Apr. 2002.

[17] PWC. The global state of information securityÂő survey 2016. Technical report, PWC, 2016.

[18] O. Serrano, L. Dandurand, and S. Brown. On the design of a cyber security data sharing system. In *Proceedings of the 2014 ACM Workshop on Information Sharing & Collaborative Security*, pages 61–69. ACM, 2014.

[19] J. Steinberger, A. Sperotto, M. Golling, and H. Baier. How to exchange security events? overview and evaluation of formats and protocols. In *Integrated Network Management (IM), 2015 IFIP/IEEE International Symposium on*, pages 261–269. IEEE, 2015.

[20] D. S. Vogt, D. W. King, and L. A. King. Focus groups in psychological assessment: enhancing content validity by consulting members of the target population. *Psychological assessment*, 16(3):231, 2004.

[21] A. Zaveri, A. Rula, A. Maurino, R. Pietrobon, J. Lehmann, S. Auer, and P. Hitzler. Quality assessment methodologies for linked open data. *Submitted to Semantic Web Journal*, 2013.