Towards a foundational ontology of cybersecurity

Giacomo De Colle[1][0000-0002-3600-6506]

[1] University at Buffalo, Buffalo NY 14260, USA
gdecolle@buffalo.edu

Abstract. In the following paper, I discuss the current state of the art of ontologies in the domain of cybersecurity, I motivate the need for a foundational ontology of cybersecurity, and I present a plan to develop one such ontology. Many projects have attempted to achieve data interoperability in the domain of cybersecurity and to allow for exchange of knowledge in practices of data analysis and digital forensics. Nevertheless, most of these projects are plagued by a series of common issues, like covering a very narrow scope and not adopting a top-level ontology architecture to allow for their interoperability with data coming from neighboring domains. I argue that these issues hinder the success and adoption of these ontologies, and that they can be remedied by developing a foundational ontology that is rooted in a top-level architecture. Such an ontology would also be able to map into it already existing resources such as ontologies, terminologies and taxonomies in order to allow for interoperability of data structured according to them. I present a plan to develop such an ontology departing from the Cyber ontology, itself rooted in the Basic Formal Ontology (BFO) and the Common Core Ontologies (CCO).

Keywords: Foundational Ontology, Cybersecurity, BFO, CCO, Cyber Ontology.

1 Introduction and motivation

The development of a foundational ontology of cybersecurity is a neglected topic. The necessity for a formal representation of the knowledge surrounding cybersecurity has been discussed, for example, by Maathuis et al. (2018), Casey et al. (2018) and Oltramari et al. (2017). In the next session I will present the relatively large number of efforts in the field of cybersecurity ontologies and in the development of tools for coherent representation such as taxonomies and vocabularies. Despite these efforts, the state of the field remains extremely fragmented. Almost all the existing ontology projects focus on some narrow aspect of the cybersecurity domain, such as risk evaluation or insider threat detection. There is almost no attempt at bridging these ontologies with other existing ontologies in the cybersecurity domain, or with ontologies in neighboring fields such as military operations, law, privacy, etc. On the other hand of the spectrum, some of the existing cybersecurity ontology projects suffer from being extremely generic in their terminology and lack any proper representation of the more technical terms used by cybersecurity experts on a day-to-day basis. These issues result in a lack of interoperability and reusability of the data structured with these ontologies, thus replicating instead of solving the data silo problem that ontologies are developed to address.

In order to remedy these issues, I propose the development of a foundational suite of ontologies that acts as an exhaustive ontological representation of the cybersecurity domain. Such an ontology addresses the issue of interoperability between different domains by making use of the ISO/IEC 21838-2:2021 Basic Formal Ontology (BFO) as a top-level ontology architecture (see Arp et al. 2015). Moreover, this suite would develop starting from the most successful existing project in the field, the Cyber ontology (see Donohue et al. 2018), itself developed starting from the Common Core Ontologies (CCO) suite (see CUBRC 2019). By using the Cyber ontology as a departing point, the project will avoid replicating already existing efforts, as well as providing a way to coherently integrate information with the neighboring domains represented by BFO and CCO. Ultimately, this ontology would also serve as a starting basis to map and integrate the already existing ontology projects I will present in the next section. Such mappings will be developed following the principles of translation definitions adopted by Grüninger et al. (2017).

There are multiple benefits that can be obtained by creating a foundational ontology, achieving data interoperability and creating structured knowledge bases in the domain of cybersecurity. A cybersecurity ontology could be used to integrate heterogeneous data coming from cyberspace, conduct analytics and provide structured content to Security Information and Event Management (SIEM) tools. Being organized according to a common ontological representation, this data could also reveal implicit information through standard semantics web tools such as reasoners or the SPARQL query language. That is, a cybersecurity ontology could be used to explain and reason over large bodies of data, as well as efficiently retrieve information from such data. A cybersecurity ontology could moreover be used to encode domain-specific knowledge to automate cybersecurity operations, which is crucial to enact responses when time is a limiting factor, as it is often the case with cyber-attacks and security breaches. For example, ontologies could be used to encode playbooks which illustrate the responses that an organization goes through in case of an attack, and partially automate their enactment. Finally, a foundational cybersecurity ontology could be used to conduct simulations informed by past data on attacks and defense mechanisms, similarly to what is described by Rajesh et al. (2022).

2 State of the Art

Ontologies and related knowledge graph applications in the field of cybersecurity have been developed for quite some time. The review conducted by Martins et al. (2022) lists almost 40 ontologies that are related to the cybersecurity domain. Preliminary projects were undertaken by Obrst et al. (2012) on behalf of MITRE. MITRE is also responsible for developing ATT&CK and D3FEND, vocabularies that respectively document cyberattack and cyberdefensive techniques and which are extremely valuable as data sources for the cybersecurity community. Casey et al. (2018) present the development of CASE and its ontology UCO (Unified Cyber Ontology), which is not tied to any top-level ontology. Akbar et al. (2023) have developed an ontology to represent ATT&CK data, but the project is not complete, and the ontology is similarly not tied to any top-level ontology. Oliveira et al. (2024) identify ontological issues in D3FEND, as well as possible solutions to them, but they do not develop further on integrating

D3FEND with other ontologies. Other notable ontology projects are the insider threat project developed by Costa et al. (2014), STUCCO developed by Iannacone et al. (2015) as a general ontology for cybersecurity knowledge graphs, COoVR developed by Sales TP et al. (2018) as an ontology of value and risk, the ontology of ISO 27005 developed by Agrawal (2016) for risk management, the general ontology of cybersecurity and cyberwarfare developed by Maathuis et al. (2018), the ontology for human factors in cybersecurity developed by Oltramari et al. (2017), the ontology for adaptive systems developed by Ben-Asher et al. (2015).

The issue common to almost all these projects is the lack of a foundational framework that the respective ontologies can use in order to integrate data in a coherent way. For example, let's say that we have data that we want to query to look for insider threat presence, and moreover that we want to relate this information to the presence of cyberattack techniques described by ATT&CK. This is currently impossible because the ontologies developed by Costa et al. (2014) and by Akbar et al. (2023) do not share the same terminology, and there is no way to tell whether their vocabularies refer to the same kind of entities and whether they can be used to structure the same database. This issue could only be solved if the two ontologies shared the same top-level ontological architecture. In addition to this, most of the cybersecurity ontology projects are extremely parochial, and focus only on some specific aspects of the cybersecurity domain and its operations (e.g. risk management). Even when the ontologies are developed to be broader in scope, they end up including only few and very general terms that do not include reference to the more technical terminology that a cybersecurity expert uses on a day-to-day basis.

The few exceptions to the scenario presented above are CRATELO, developed by Oltramari et al. (2014), and the Cyber ontology, developed by Donohue et al. (2018). CRATELO is an ontology developed starting from DOLCE (see Gangemi et al., 2002). CRATELO includes SECCO (a middle level ontology of security) and OSCO (an ontology of cyber operations). CRATELO is thus one of the few projects that include a proper axiomatization, and which is related to a top-level ontology. Similarly, the Cyber ontology is developed departing from the Common Core Ontologies (CCO), which are themselves a mid-level architecture that departs from the Basic Formal Ontology (BFO). Both CRATELO and the Cyber ontology seem to be good departing points for developing proper ontologies of cybersecurity. The Cyber ontology in particular has been recently introducing a large number of terms related to the cybersecurity domain. Nevertheless, both CRATELO and the Cyber ontology have wide possibilities for further development. Both can be expanded with a richer terminology coming from the day-to-day vocabulary of cybersecurity experts. Even if the community surrounding the Cyber ontology started to integrate MITRE's taxonomies, neither the Cyber ontology nor CRATELO have been entirely or systematically linked to ATT&CK and D3FEND, which are the two main terminology sources for cybersecurity analytics. And none of the two, as far as I know, has thoroughly been tested in their capability to answer competency questions in the form of queries, especially if related to neighboring fields such as privacy infringement, cyberwarfare or intelligence analysis operations.

3 Problem Statement and Contributions

The literature review presented in the section above shows that despite the continuous efforts in the development of ontologies in the cybersecurity domain, a lack of a unified framework to represent knowledge in the cybersecurity domain still persists. How can a foundational ontology for cybersecurity be built, so that it's well axiomatized, well grounded in an existing top-level ontology, and capable of answering useful competency questions? Such an ontology, or suite of ontologies, would be able to represent the main types of entities in the domain of cybersecurity operations - malicious actors, data and its properties such as levels of privacy, accesses to data, operative systems, login credentials, etc. as well as relating these elements to data in neighboring fields. For example, laws, regulations, state actors, roles such as enemy and friend, military operations, etc.

When this representation is built, it could answer different use cases. The first one is running queries and data analysis over knowledge bases which integrate different and heterogeneous data sources. For example, an ontology of cybersecurity could be used to investigate datasets containing information about cyber infrastructure in order to identify weaknesses and possible vulnerabilities. The second use case is employing the ontology in close connection with tools of the trade that cybersecurity experts use in order to automate tasks such as pattern detection, data analysis, real-time response to threats, support to decision making in situations where time is a limiting factor, etc. Building a structured and well-grounded ontological representation of the cybersecurity domain is already an extremely valuable effort and one that fixes a large gap in the literature. Moreover, integrating an ontological representation of the kind described with tools of cybersecurity analysis is a novel use case for an ontology that could push towards further development of interdisciplinary research at the border between ontology and other areas of computer science.

4 Research Methodology and Approach

The first step to develop an ontology, or suite of ontologies, devoted to cybersecurity is to review already existing ontology projects and frameworks, as well as studying already existing neighboring non-ontological projects which can be used as data sources or to which the ontology could be applied. For example, MITRE's taxonomies DEF3ND and ATT&CK are the main resources that are used to represent knowledge of cybersecurity experts and that are already employed as terminological standards in the field. As such, an ontological representation of the two taxonomies is the main priority of the project. Other similar projects to be properly investigated are the Structured Threat Information eXpression (STIX), developed by MITRE for the DHS, and the already existing STUCCO ontology as applied to simulations of cyberattacks. Already existing ontology projects should also be reviewed to assess which ones are useful efforts that need to be integrated in a well-developed foundational ontology for cybersecurity. These ontologies would eventually be mapped into the foundational ontology by making use of the translation definition method discussed by Grüninger et al. (2017).

The second step is to identify a set of competency questions that an ontology of cybersecurity needs to be able to answer in order to be successfully and usefully employed. These competency questions will take the form of SPARQL queries and will be one of the evaluation tools used to assess the success or failure of the research. Said questions will be identified through interviews and collaboration with domain experts, as well as study of foundational texts in the field and other authoritative resources (see for example Sikorski and Honig (2012), Katz and Lindell (2014), Stamp (2005) and NIST (2018)). This stage of research will also bring to identification of core terminology used in the field, as well as broader understanding of the type of entities and relations that need to be represented in order to capture knowledge of the domain experts. The competency questions that are identified will also guide the development of object properties and axioms that will be used in reasoning over data tagged with the ontology for extraction of implicit knowledge.

Once these competency questions are identified, proper development of the ontology can begin. In order to develop a foundational ontology of cybersecurity that is capable of bridging the gap between already existing terminologies, a suitable starting place must be found in some already existing and well-developed ontology architecture. As discussed in the second section of this paper, the only two options in the field are CRATELO and the Cyber ontology. For several reasons, the Cyber ontology seems to be a more suitable starting place. Being integrated in the environment of the widely adopted CCO and BFO, it is already directly interoperable with a larger number of domains, especially those that could be relevantly tied to cybersecurity. For instance, CCO and BFO have already been mandated for information sharing practices in the US Intelligence Community and the DHS. As such, they represent the natural tools to be used with other cybersecurity initiatives of the US government such as STIDS. Neighboring domains that have been already represented by BFO and CCO are the domain of information artifacts (see Ceusters (2012)), software (see Malone et al. (2014)), military operations (see Morosoff et al. (2015)) and agents (see CUBRC (2019)).

Given that the Cyber ontology represents the most well-developed ontology in the field, and the one that most extensively covers domain-specific terms, it will be taken as a starting place for horizontal expansion and downwards population of new terms and classes. The development of the ontology will be modularized depending on the subdomains discovered during the exploratory step 2 discussed in this section. It is entirely possible that the Cyber ontology itself could act as a foundational ontology for cybersecurity once expanded and linked to MITRE's projects. Building the ontology will also take into consideration, during the process of class population and axiomatization, of the competency questions identified in step 2. In this way, high-quality reasoning over data and extraction of implicit knowledge from a database will be possible and will also be tested with data tagged with MITRE's ATT&CK and D3FEND, as described in the next section. The structure of the data collected by MITRE should serve as a guiding principle for the development of the classes in the ontology, alongside with best practices taken from the successful experiences of other projects in the BFO community, such as the OBO foundry. As a final step, integration with the tools of the trade of cybersecurity identified during step 2 will also be explored.

5 Evaluation Plan

After the ontology or set of modularized ontologies have been built according to the steps previously discussed, its capabilities will be tested through real life data taken by MITRE or other similar data sources identified in the second step, as discussed in the previous section. The first step of the testing analysis will then be to ingest instance-level data into the ontology, thus effectively creating a knowledge base. If this step is accomplished, it will mark a first success for the project, and allow me to proceed to the second testing step. This will be executed by running the queries identified and developed in the previous step of the project against the data structured by the ontology. If the queries can be successfully applied to the knowledge base thus constructed, reasoning and implicit knowledge extraction will successfully prove the quality of the ontology and its applicability. The ontology and the queries will also be integrated with analytics tools proper of the cybersecurity sector, as identified in the previous step. Success in this third step would prove extremely fruitful. Not only because it would provide a helpful tool for cybersecurity analysis, but also because it would be a new application of ontologies to a neighboring computer science discipline.

6 Results

The stage of the research is at the moment at its beginnings. Preliminary results include an extensive literature review and evaluation of the existing semantic web projects in cybersecurity, which has narrowed down the ontological projects to two main candidates to depart from in order to develop a foundational cybersecurity ontology. As a part of these preliminary studies, I have also started identifying data sources and already adopted terminological standards in the field, such as MITRE's ATT&CK and D3FEND, NIST recommendations, ISO standards such as ISO 27005, and STIMS. A study of foundational notions of cybersecurity has also begun, as well as contacting cybersecurity experts. Future immediate efforts will include continuing studies of foundational concepts of cybersecurity and contacts with domain experts. After this is done, I will follow with a development of a first set of competency questions to act as a benchmark for the ontology. At the same time, I will begin partitioning the subdomains of cybersecurity needed for establishing the modules of the ontology. A preliminary analysis has so far identified the need for a reference ontology of cyberspace (which the Cyber ontology itself seems to already be), one for operative systems, one for network operations and one for malware. All these terms are already present in the Cyber ontology itself and respective ontology modules will likely be built by downward population starting from parent terms in the Cyber ontology.

7 Conclusions

The preliminary studies undergone so far show that ontological efforts in the cybersecurity domain remain fragmented. The need for a unified framework to relate the existing projects and to ground them in the practice of cybersecurity experts is made clear by the narrow scope of the existing ontologies and by their low level of accuracy. In

order to remedy these issues, I have proposed the development of a foundational ontology of cybersecurity, starting from the Cyber ontology and divided in modules corresponding to different cybersecurity subdomains. Such a project will be able to achieve interoperability between heterogeneous data sources from the cybersecurity realm and related fields, thus allowing for more precise and more extensive data analysis. Given the difficulties proper of the cybersecurity expert and the intelligence analyst in data integration (see Mandrick and Smith 2022), this is an issue of primary concern in an era where finding vulnerabilities in huge informatic infrastructure is a priority. Without tools to handle big data in a structured way, cybersecurity defense experts will have a hard time automating tasks and identifying meaningful content in the web of growing interactions we create in cyberspace.

Acknowledgments. The author wishes to acknowledge the help of comments and suggestions from Barry Smith and John Beverley.

Disclosure of Interests. The author has no competing interests to declare that are relevant to the content of this article.

References

1. Agrawal, V. Towards the Ontology of ISO/IEC 27005: 2011 Risk Management Standard. *International Symposium on Human Aspects of Information Security and Assurance.* (2016)
2. Akbar, K.A., Halim, S.M., Singhal, A., Abdeen, B., Khan, L., & Thuraisingham, B.M. The Design of an Ontology for ATT&CK and its Application to Cybersecurity*, Proceedings of the Thirteenth ACM Conference on Data and Application Security and Privacy* (2023)
3. Arp, R., Smith, B., & Spear, A.D. Building Ontologies with Basic Formal Ontology, MIT Press. (2015)
4. Ben-Asher, N., Oltramari, A., Erbacher, R. F., & Gonzalez, C. Ontology-based Adaptive Systems of Cyber Defense, *Proceedings of the Conference on Semantic Technology for Intelligence, Defense, and Security, STIDS 2015,* (2015)
5. Casey, E., Barnum, S., Griffith, R., Snyder, J., Beek, H.V., & Nelson, A. The Evolution of Expressing and Exchanging Cyber-investigation Information in a Standardized Form*.* in: Biasiotti, M., Mifsud Bonnici, J., Cannataci, J., Turchi, F. (eds) *Handling and Exchanging Electronic Evidence Across Europe. Law, Governance and Technology Series,* vol 39. Springer, Cham. (2018)
6. Ceusters W. An information artifact ontology perspective on data collections and associated representational artifacts. *Stud Health Technol Inform*, pp. 68-72. PMID: 22874154. (2012)
7. CUBRC. An Overview of the Common Core Ontologies. (2019)
8. Costa, D., Collins, M., Perl, S.J., Albrethsen, M., Silowash, G., & Spooner, D. An Ontology for Insider Threat Indicators: Development and Application. *Semantic Technologies for Intelligence, Defense, and Security*. (2014)
9. Donohue, B., Jensen, M., Cox, A.P., & Rudnicki, R. A common core-based cyber ontology in support of cross-domain situational awareness. *Defense + Security*. (2018)

10. Gangemi, A., Guarino, N., Masolo, C., Oltramari, A., & Schneider, L. Sweetening Ontologies with DOLCE. *International Conference Knowledge Engineering and Knowledge Management*. (2002)
11. Grüninger, Michael, Carmen S. Chui and Megan Katsumi. "Upper Ontologies in COLORE." *Joint Ontology Workshops* (2017)
12. Iannacone, M., Bohn, S., Nakamura, G., Gerth, J., Huffer, K., Bridges, R., Ferragut, E., & Goodall, J. Developing an Ontology for Cyber Security Knowledge Graphs. *Proceedings of the 10th Annual Cyber and Information Security Research Conference*, 1–4. https://doi.org/10.1145/2746266.2746278. (2015)
13. Katz, J., & Lindell, Y. Introduction to Modern Cryptography, Second Edition, CRC. (2014)
14. Maathuis, C., Pieters, W., & van den Berg, J. Developing a Cyber Operations Computational Ontology. *Journal of Information Warfare*, *17*(3), 32–49. https://www.jstor.org/stable/26633164. (2018)
15. Malone, J., Brown, A., Lister, A.L., Ison, J.C., Hull, D., Parkinson, H.E., & Stevens, R. The Software Ontology (SWO): a resource for reproducibility in biomedical data analysis, curation and digital preservation. *Journal of Biomedical Semantics, 5*, 25 - 25. (2014)
16. Mandrick, B., & Smith, B. Philosophical foundations of intelligence collection and analysis: a defense of ontological realism. *Intelligence and National Security, 37*, 809 - 819. (2022)
17. Martins, B. F., Serrano Gil, L. J., Reyes Román, J. F., Panach, J. I., Pastor, O., Hadad, M., & Rochwerger, B. A framework for conceptual characterization of ontologies and its application in the cybersecurity domain. *Software and Systems Modeling*, *21*(4), 1437–1464. https://doi.org/10.1007/s10270-022-01013-0. (2022)
18. Morosoff, P., Rudnicki, R., Bryant, J., Farrell, R., & Smith, B. Joint Doctrine Ontology: A Benchmark for Military Information Systems Interoperability. *Semantic Technologies for Intelligence, Defense, and Security*. (2015)
19. National Institute of Standards and Technology. Framework for Improving Critical Infrastructure Cybersecurity, Version 1.1 (NIST CSWP 04162018; p. NIST CSWP 04162018). National Institute of Standards and Technology. https://doi.org/10.6028/NIST.CSWP.04162018. (2018)
20. Obrst, L., Chase, P., & Markeloff, R. Developing an Ontology of the Cyber Security Domain. *Semantic Technologies for Intelligence, Defense, and Security*. (2012)
21. Oliveira, I.J., Engelberg, G., Barcelos, P.P., Sales, T.P., Fumagalli, M., Baratella, R., Klein, D., & Guizzardi, G. Boosting D3FEND: Ontological Analysis and Recommendations. *Formal Ontology in Information Systems*. (2023)
22. Oltramari, A., Cranor, L.F., Walls, R.J., & Mcdaniel, P. Building an Ontology of Cyber Security. *Semantic Technologies for Intelligence, Defense, and Security*. (2014)
23. Oltramari, A., Henshel, D.S., Cains, M., & Hoffman, B. Towards a Human Factors Ontology for Cyber Security. *Semantic Technologies for Intelligence, Defense, and Security*. (2015)
24. Rajesh, P., Alam, M., Tahernezhadi, M., Monika, A., & Chanakya, G. Analysis Of Cyber Threat Detection And Emulation Using MITRE Attack Framework. *2022 International Conference on Intelligent Data Science Technologies and Applications (IDSTA)*, 4–12. https://doi.org/10.1109/IDSTA55301.2022.9923170. (2022)
25. Stamp, M. (2005). Information Security: Principle and Practice, Wiley.