The Cybersecurity Ontology Network: the first building block for a comprehensive Cybersecurity Knowledge Graph

Mikel Egaña Aranguren[1,*], Jesualdo Tomás Fernandez Breis[2], Alexander Garcia Castro[3] and Markus Rompe[3]

[1]University of Basque Country (UPV/EHU), Spain
[2]University of Murcia (UM), Spain
[3]Siemens Energy, Germany

Abstract

The escalating complexity and dynamism of the cybersecurity landscape necessitates a robust, adaptive framework for the integration, analysis, and dissemination of cybersecurity knowledge. This paper introduces the Cybersecurity Ontology Network (CON), a foundational step towards establishing a comprehensive Cybersecurity Knowledge Graph (CKG). The CON framework is designed to semantically organize and interconnect diverse cybersecurity information, focusing on software components, thus including vulnerabilities, libraries, and projects. By leveraging ontological structures, CON enables a more nuanced, context-aware synthesis of cybersecurity data from disparate sources, facilitating advanced analytical capabilities.

Keywords

Cybersecurity, Ontology, Knowledge Graph, Data integration

1. Introduction

Data-centric architectures in enterprises emphasize the role of data as first class citizens, as opposed to the traditional "one application one database" architectures, that emphasized applications over data [1, 2]. Companies have realized that data is their most important asset, whereas applications are ephemeral.

Knowledge Graphs (KGs) [3] generally combine data from different sources of varying structure and granularity. From an industrial perspective, the concept of enterprise KG (EKG) [4] would facilitate the development of a KG that contains and connects all the relevant data for a company. The industrial relevance of EKGs is supported by the existence of the OMG EKG Forum[1]. Despite graph-based representations do not require a schema, KGs are usually structured by the content of ontologies, which also provide a precise meaning to the data represented in the KG.

OK4I - Ontologies and KGs for Industry, July 15–19, 2024, Enschede, Netherlands
*Corresponding author.
mikel.egana@ehu.eus (M. E. Aranguren)
https://mikel-egana-aranguren.github.io/ (M. E. Aranguren)
0000-0001-8081-1839 (M. E. Aranguren)

[1]https://www.ekgf.org/

Cybersecurity is one of the most salient domains in this transition to EKGs, since it requires up to date data from heterogeneus sources that needs to be integrated. We present the Cybersecurity Ontology Network (CON), a first step towards the creation of a Cybersecurity KG at Siemens Energy (SE).

2. The need for data integration in Cybersecurity

Currently, information related to Cybersecurity at SE is scattered in data silos: software projects, software library data, and vulnerabilities. Figure 1 shows that (1) the software project data stores information about the organization that owns the project, the authors of the software project, the name, the components, etc.; (2) the software library data includes information about software libraries, their versions, dependencies with other libraries, etc.; and (3) the vulnerability data includes information about the name, effect, level of risk and mitigation of risk, and the libraries and versions affected by the vulnerabilities.

From this description it can be easily noticed that there is an overlap in the data, but that the heterogeneity of the data makes their interoperability difficult. The lack of interoperability implies that answering critical questions such as the next ones becomes cumbersome:

- Which components are affected by the Vulnerability CVE-2021-33430?
- If I update a package then how is that updating going to affect other packages?
- Which components can be updated in the projects affected by the vulnerability?

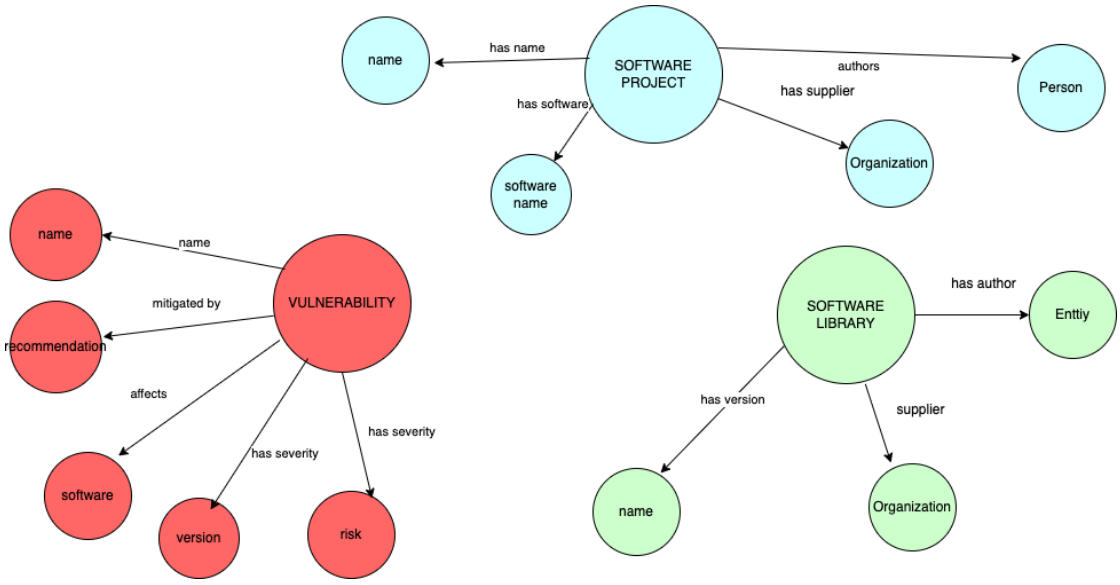

Figure 1: Current data silos pertaining to cybersecurity at SE.

The KG approach would speed up the process of getting the answers to those questions through meaningful data integration and exploitation. The first step towards such graph is the

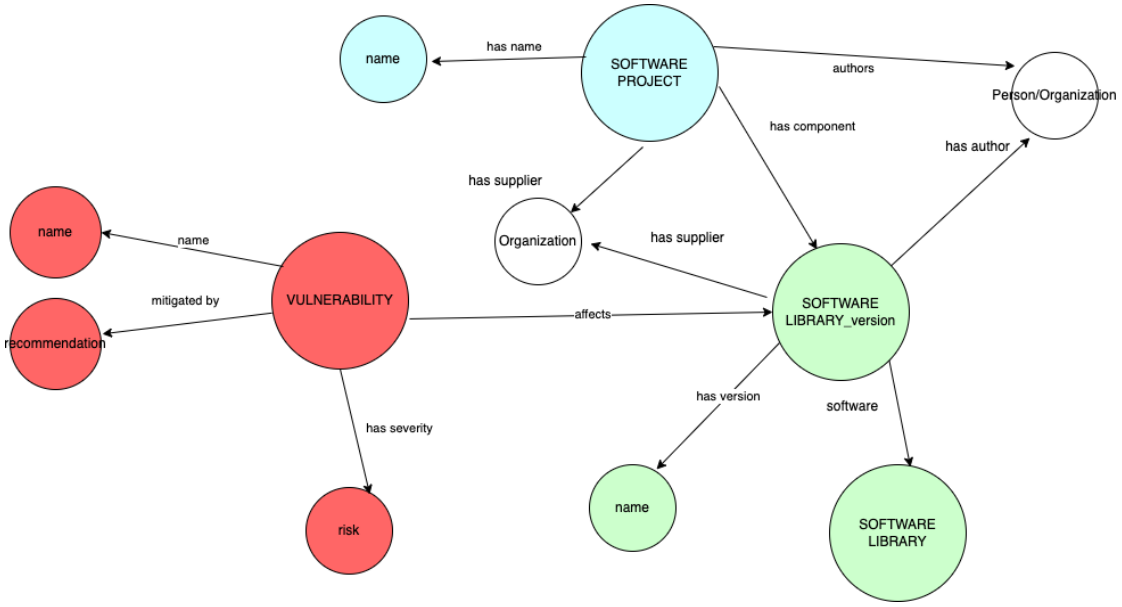

Figure 2: By relating vulnerabilities to software libraries and software libraries to software projetcs, a comprehensive picture of cybersecurity is obtained.

ontology that provides the structure and meaning to the data, that is, the set of classes and properties of the entities involved, and hence facilitate integration. Figure 2 depicts a possible graph-based representation and connection of the entities described in Figure 1. There, we can see how overlapping entities are not represented in a redundant way and that there are links connecting the previous data silos. In the next section, we describe how we are approaching the development of the ontology network that will enable data interoperability for the effective management of software vulnerabilities.

3. The Cybersecurity Ontology Network

The Cybersecurity Ontology Network (CON) is an OWL[2] ontology developed with the ontology editor Protégé [5]. CON is based on the CycloneDX specification[3], and all the elements of the specification can be found in the ontology (Figure 3). Therefore, the current CON is focused on the central element of CycloneDX, that is, the Software Bill of Materials (SBOM).

An SBOM is essentially a comprehensive inventory of all the components that make up a piece of software. This includes not only the direct code written by the software developers but also any third-party components, libraries, frameworks, and other dependencies that the software uses to function. The component class presents a rich axiomization as seen in Figure 4. For example, a common pattern used in different parts of the ontology is the inclusion of a covering axiom when the specification presents a given set of values (Figure 4). The *mapping*

[2]https://www.w3.org/TR/owl2-overview/
[3]https://cyclonedx.org/specification/overview/

annotation property has been used to provide best practices when the ontology is used to map data from external sources to RDF (Figure 4).

Figure 3: Main entities of CON: classes (left), object properties (center), data properties (right).

4. Conclusions

In this work we have presented the initial steps towards the development of a KG for handling cybersecurity information. The CON offers a vocabulary to model vulnerabilities and affected libraries and projects. Its design has been driven by the practicalities of KG generation. In future extensions, the vulnerability and the software library modules will be further developed.

References

[1] D. McComb, Software Wasteland: How the Application-centric Mindset is Hobbling Our Enterprises, Technics Publications, 2018. URL: https://books.google.es/books?id=_6JutAEACAAJ.

[2] D. McComb, The Data-centric Revolution: Restoring Sanity to Enterprise Information Systems, Technics Publications, 2019. URL: https://books.google.es/books?id=5XuIxwEACAAJ.

[3] A. Hogan, E. Blomqvist, M. Cochez, C. d'Amato, G. D. Melo, C. Gutierrez, S. Kirrane, J. E. L.

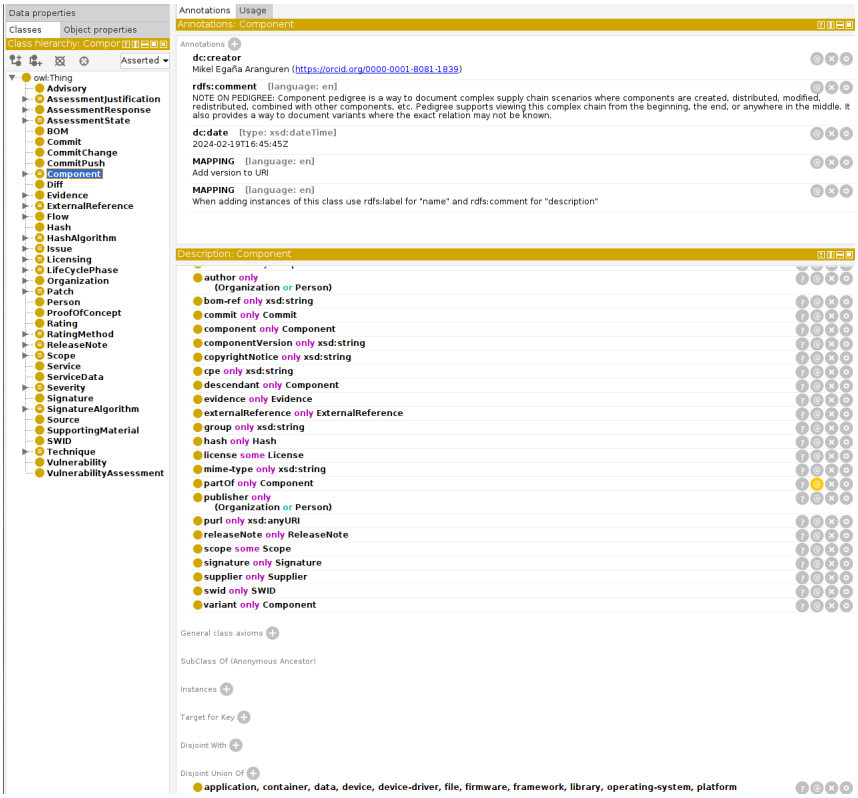

Figure 4: Axioms of the class Component. The covering axiom can be appreciated in the bottom right, under Disjoint Union Of: application, container, etc. There are two MAPPING annotation properties, in the annotation pane (Top right), presenting information on how to use the ontology to map information to RDF: how to create the appropriate URI and what values to use for rdfs:label and rdfs:comment properties.

Gayo, R. Navigli, S. Neumaier, et al., Knowledge graphs, ACM Computing Surveys (Csur) 54 (2021) 1–37.

[4] J. M. Gomez-Perez, J. Z. Pan, G. Vetere, H. Wu, Enterprise knowledge graph: An introduction, in: Exploiting linked data and knowledge graphs in large organisations, Springer, 2017, pp. 1–14.

[5] M. A. Musen, The protégé project: a look back and a look forward, AI Matters 1 (2015) 4–12. URL: https://doi.org/10.1145/2757001.2757003. doi:10.1145/2757001.2757003.