**Informe Técnico / Technical Report**

**Towards the Consolidation of Cybersecurity Standardized Definitions**
Beatriz F. Martins, Lenin J. S.Gil, José F. R. Román, José Ignacio Panach, and Óscar Pastor López

<table>
<thead>
<tr>
<th>Ref. #:</th>
<th>PROS-TR-2021-I</th>
</tr>
</thead>
<tbody>
<tr>
<td>Title:</td>
<td>Towards the Consolidation of Cybersecurity Standardized Definitions</td>
</tr>
<tr>
<td>Author (s):</td>
<td>Beatriz Franco. Martins, Lenin Javier Serrano Gil, José Fabián. Reyes. Román, José Ignacio Panach, and Óscar Pastor López</td>
</tr>
<tr>
<td>Corresponding autor (s):</td>
<td><a href="mailto:bmartins@pros.upv.es">bmartins@pros.upv.es</a>, <a href="mailto:lseranno@pros.upv.es">lseranno@pros.upv.es</a>, <a href="mailto:jreyesg@pros.upv.es">jreyesg@pros.upv.es</a>, <a href="mailto:joigpana@uv.es">joigpana@uv.es</a>, <a href="mailto:opastor@dsic.upv.es">opastor@dsic.upv.es</a></td>
</tr>
<tr>
<td>Document version number:</td>
<td>1</td>
</tr>
<tr>
<td>Final version:</td>
<td>-</td>
</tr>
<tr>
<td>Release date:</td>
<td>-</td>
</tr>
<tr>
<td>Keywords:</td>
<td>Conceptualization, Cybersecurity, Knowledge Graphs, Cybersecurity Ontology, Ontology</td>
</tr>
</tbody>
</table>
Towards the Consolidation of Cybersecurity Standardized Definitions

Beatriz Franco Martins\textsuperscript{1} [0000–0001–9198–1047],
Lenin Javier Serrano Gil\textsuperscript{1,2} [0000–0002–1631–7139],
José Fabián Reyes Román\textsuperscript{1} [0000–0002–9598–1301],
José Ignacio Panach\textsuperscript{3} [0000–0002–7043–6227], and
Oscar Pastor\textsuperscript{1} [0000–0002–1320–8471]

\textsuperscript{1} PROS Research Center, Universitat Politècnica de València, Camino de Vera s/n, 46022 Valencia, Spain, \{bmartins, lserrano, jreyes\}@pros.upv.es, opastor@dsic.upv.es
\textsuperscript{2} Ingeniería de Sistemas e Informática, Universidad Pontificia Bolivariana Km 7 via Bucaramanga - Piedecuesta, Santander, Colombia
\textsuperscript{3} Escola Tècnica Superior d’Enginyeria, Universitat de València, Avinguda de l’Universitat, 46100 Burjassot, Valencia joigpana@uv.es

Cybersecurity is a vast and complex domain, therefore enterprises are actively seeking efficient solutions in this matter. Knowledge Graphs (KG) are one of the mechanisms that organizations use to explore the security among assets and possible attacks. However, the great amount of information can create misinterpretation of concepts represented in these structures of conceptualizations. As a KG may be considered an implementation of a conceptualization, the grounding of concepts is fundamental. Therefore, the support of Conceptual Modeling best-practices, especially regarding the branch of Ontologies. We made a pilot study that finds out the state-of-art in "Cybersecurity Ontologies”. From this study, we propose a survey to extend our terminological approach. The survey produced a huge amount of data, thus we develop a REST API for data manipulation and a NoSQL database to store them which is the main contribution of this document. Our goal is to provide an ontological analysis tool to help stakeholders avoid misinterpretations during KGs development and implementation.

\textbf{Keywords:} Conceptualization, Cybersecurity, Knowledge Graphs, Cybersecurity Ontology, Ontology

1 Introduction

Nowadays, organizations are focused on the active search for solutions that ensure efficient and safe management and protection of their assets. An application context, especially for large companies, is that of Cybersecurity, which is a broad/extensive and quite complex domain that requires an interdisciplinary approach. One of the mechanisms by which organizations bet to explore security between assets and possible attacks is “Knowledge Graphs” (KG) [58]. Concerning the Conceptual Modeling standpoint, the
grounding of concepts is fundamental to implement KG, and it is one of the most relevant ontology applications [17]. That is why the application of ontologies in the cybersecurity domain emerges today as a research topic of great importance and interest. The main objective of this research work is to facilitate a pragmatic and iterative solution that meets the needs of organizations in terms of Cybersecurity, and in this way contribute to Ontology Engineering research.

However, before providing a proposal to achieve this problem, we look for the solution proposals that exist in the state-of-art. Previously, we conducted a pilot study [44] looking for existing works that deal with cybersecurity requirements from an ontological perspective. As the results we took from this research provided a huge amount of data, we develop a Representational State Transfer Application Programming Interface (REST API) for data manipulation and a Not Only SQL (NoSQL) database to store these data. Our goal is to provide data analytics and reasoning using these data and in future work provide a tool to facilitate the process of Ontological Analysis [17]. Through this document, we present the REST API we develop and some initial results these approaches provide.

We have organized the rest of this document in the following way: Section 2 presents the pilot study that supports this work. Section 3 details the proposal of an API to support ontological analysis in complex fields, like cybersecurity. Section 4 depicts the actual state of the proposal with some further research directions.

2 The Pilot Study

There is not a definitive architectural solution for the design and development of KGs supported by ontologies yet. This problem is mainly due to the complexity and interdisciplinarity of the domain. Therefore, we made a pilot study [44] to identify proposals in the cross-field of Cybersecurity and Ontologies, evaluate the existing Cybersecurity Ontologies’ level of applicability, and identify the possible data sources of cybersecurity information. In this initial research, we found that the knowledge base for cybersecurity is extensive and context-dependent.

In the pilot study, we support our cybersecurity perspective using the ISO/IEC 27032:2012 [25] and ISO/IEC 27000:2018 [27] standards. These standards make up the knowledge base to identify and detect the most used terms cybersecurity definitions in the presented ontologies in the articles that we found. However, we observe the need to compare the definitions contained in these ontologies with the different definitions in a broad amount of cybersecurity standards. Therefore, we use a NoSQL database to store the standards’ definitions and a REST API to analyze them, Section 3 detail our tool proposal.

From the ISO/IEC standards, we extract 156 terms and their definitions, complying with ontological concepts, and we count the number of its citations in the papers found. To do this, we applied to the articles a semi-automatic technique (a regular expression search cycle) through a sequence of steps.

**Automatic Search:** We develop a script in Python [4] to obtain the clear text of the documents. Then, we search for terms from the ISO/IEC selected definitions in each

---

Towards the Consolidation of Cybersecurity Standardized Definitions

of the documents by executing queries with regular expressions over an algorithm we developed;

**Context Validation:** We execute another Python algorithm – from the Automatic Keyword Extraction from Individual Documents [57] – to provide context validation. Next, we extract the key phrases using the “RAKE short for Rapid Automatic Keyword Extraction algorithm” implementation do validate. Then, we perform a second round of reading the documents to verify if all terms comply with cybersecurity's context.

**Filtering:** Lastly, we filter and eliminate the deviation of terms before summarizing the citations from the total of ISO/IEC terms that we got in our sample papers.

This terminological reference base usually presents concepts (or entities) used in ontologies and is mostly supported by all consecrated cybersecurity standards (beyond ISO/IEC used). However, it is out of our scope to guarantee and verify if all terms mean the same conceptual thing (in terms of ontological grounding). This semantic adequacy of the conceptualization is future research that is part of the Ontological Engineering process during the course of the project.

Table 1 shows the total number of occurrences of cybersecurity terminology in our pilot study. We use these terms to clarify the semantics of these terms by cross-examining their definitions at the most relevant Cybersecurity standards available. We used the outcomes of our previous pilot study to extract the found terms and use them in our survey, which is also a contribution of this paper.

<table>
<thead>
<tr>
<th>Term</th>
<th>Total of citations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Access Control</td>
<td>98</td>
</tr>
<tr>
<td>Application</td>
<td>208</td>
</tr>
<tr>
<td>Asset</td>
<td>348</td>
</tr>
<tr>
<td>Attack</td>
<td>942</td>
</tr>
<tr>
<td>Authentication</td>
<td>14</td>
</tr>
<tr>
<td>Bot</td>
<td>121</td>
</tr>
<tr>
<td>Availability</td>
<td>61</td>
</tr>
<tr>
<td>Compromise</td>
<td>2</td>
</tr>
<tr>
<td>Confidentiality</td>
<td>57</td>
</tr>
<tr>
<td>Companionship</td>
<td>61</td>
</tr>
<tr>
<td>Control</td>
<td>184</td>
</tr>
<tr>
<td>Component</td>
<td>75</td>
</tr>
<tr>
<td>Event</td>
<td>333</td>
</tr>
<tr>
<td>Indicator</td>
<td>8</td>
</tr>
<tr>
<td>Information Need</td>
<td>6</td>
</tr>
<tr>
<td>Information Security</td>
<td>40</td>
</tr>
<tr>
<td>Integrity</td>
<td>45</td>
</tr>
<tr>
<td>Internet</td>
<td>60</td>
</tr>
<tr>
<td>Likelihood</td>
<td>14</td>
</tr>
<tr>
<td>Malicious Software</td>
<td>5</td>
</tr>
<tr>
<td>Malware</td>
<td>218</td>
</tr>
<tr>
<td>Measure</td>
<td>117</td>
</tr>
<tr>
<td>Measurement</td>
<td>6</td>
</tr>
<tr>
<td>Measuring</td>
<td>62</td>
</tr>
<tr>
<td>Objective</td>
<td>29</td>
</tr>
<tr>
<td>Organization</td>
<td>271</td>
</tr>
<tr>
<td>Performance</td>
<td>53</td>
</tr>
<tr>
<td>Plagiarism</td>
<td>5</td>
</tr>
<tr>
<td>Policy</td>
<td>117</td>
</tr>
<tr>
<td>Process</td>
<td>401</td>
</tr>
<tr>
<td>Provider</td>
<td>75</td>
</tr>
<tr>
<td>Reliability</td>
<td>11</td>
</tr>
<tr>
<td>Requirement</td>
<td>93</td>
</tr>
<tr>
<td>Review</td>
<td>42</td>
</tr>
<tr>
<td>Risk</td>
<td>259</td>
</tr>
<tr>
<td>Risk Assessment</td>
<td>10</td>
</tr>
<tr>
<td>Risk-Management</td>
<td>7</td>
</tr>
<tr>
<td>Stakeholder</td>
<td>50</td>
</tr>
<tr>
<td>Trojan</td>
<td>366</td>
</tr>
<tr>
<td>Trojan Horse</td>
<td>12</td>
</tr>
<tr>
<td>Vulnerability</td>
<td>775</td>
</tr>
</tbody>
</table>

### 3 Terminological Investigation

Next we describe the details of the terminological investigation we conduct.

### 3.1 Objective

Our main goal is to identify the existence of definitions for the terms contained in the ISO/IEC 27032:2012 and ISO/IEC 27000:2018 standards in a broad set of other documents accepted by cybersecurity community. These terms are present in the primary
studies that describe the design and implementation of ontologies for the domain of Cybersecurity. Therefore, we expect to consolidate the definitions of each term and identify the context of the use of them based on the standards they belong to. Lastly, we can identify possible misinterpretations on cybersecurity ontologies concerning the terminology used by them.

In summary, our goal is to identify and evaluate the existing Cybersecurity Ontologies’ terminology, their context, and use.

### 3.2 Cybersecurity Standards

Definitions used by standards such as those in ISO/IEC exist to clarify the interpretation of terms present in the knowledge domain of those standards. However, the standards use natural (or technical) language that leaves room for more diverse interpretations by the community. In other words, well-known standards may provide conflicting definitions for the same term, depending on the point of view taken. Thus, we also need to know the meanings, the context of use, and the importance of these terms. Therefore, we expand our cybersecurity perspective, providing a terminological investigation based on the verification we made at the pilot study. We use the terms previously found at the studies’ verification to look for definitions of these terms in additional recognized standards by the cybersecurity community. Table 2 shows the standards we use.

<table>
<thead>
<tr>
<th>Institution</th>
<th>Standard</th>
</tr>
</thead>
<tbody>
<tr>
<td>ITU-T</td>
<td>Data Communication Networks - Open Systems Interconnection (OSI) [7]</td>
</tr>
<tr>
<td>CCMB</td>
<td>CCMB-2017-04-001 [8]</td>
</tr>
<tr>
<td>MAEC 50</td>
<td>MAEC™ Specification - Core Concepts [42], MAEC™ Specification - Vocabulary [43]</td>
</tr>
<tr>
<td>OASTM Committee Specification</td>
<td>STIX™ Version 2.1 [5], TDX™ Version 2.1 [65]</td>
</tr>
<tr>
<td>MITRE Corporation</td>
<td>CVE-9999-9999 [14], MITRE ATT &amp; CK: Design and Philosophy [82], Ten Strategies of a World-Class Cybersecurity Operations Center [84], Science of Cyber Security [48], Standardizing Cyber Threat Intelligence Information with the STIX™ [22], The neutral automated exchange of indicators information (TAXII™) [11]</td>
</tr>
<tr>
<td>NERC</td>
<td>Glossary of Terms Used in NERC Reliability Standards [50], CIP Control Systems Security Working Group (NERC-CIP-3-0) [49]</td>
</tr>
<tr>
<td>CCRA</td>
<td>Common Criteria Portal (CCv3.1-Release 5) [12]</td>
</tr>
<tr>
<td>Spain Government</td>
<td>Security Guide (CCN-STR-400) [16]</td>
</tr>
<tr>
<td>Spanish National Cybersecurity Institute</td>
<td>Cybersecurity Terms Glossary [21]</td>
</tr>
<tr>
<td>Common Criteria</td>
<td>Standard 1508 – Cyber Security [86]</td>
</tr>
</tbody>
</table>
3.3 Consolidating Definitions

To consolidate the definitions of the terms previously found in the studies, we propose a survey because the amount of standards is vast as well as the number of terms. We invite 21 cybersecurity students to participate in this survey. It is important to note that the survey is part of a collaboration with the Department of Systems Engineering and Informatics of the Universidad Pontificia Bolivariana (UPB, Colombia)\(^5\).

The students searched for each term one or more definitions in all these standards. We define a questionnaire with a spreadsheet template in which the students present their impressions about the meaning, context, and use of each definition depending on which source it is. We divided the terms among the students, so each student worked with only two different terms. Therefore, we cover 42 of the terms found in the papers pilot study search, and we complete the survey looking for the last definition, covering the remaining term. The students had two weeks to present their results.

Meanwhile, we developed a NoSQL database\(^6\) and the REST API\(^7\) to store and manipulate the resulting survey data. Then, we consolidate all standards (sources), terms, and definitions of the survey through the API developed. Below we present an API code fragment responsible to query definitions by regular expressions (RegEx).

```javascript
// Get definitions list by regex
function getDefinitionsByRegEx(req,res){
    var definition = new Definition();
    definition.regex = req.params.regex;

    Definition.aggregate([
        {$match: { regex: definition.regex }},
        {$lookup: {
            from: "sources",
            localField: "source",
            foreignField: "_id",
            as: "source"
        }}
    ]).exec((err,definitions) => {
        if(err) return res.status(500).send({message: 'Incorrect request.'});
        return res.status(200).send({definitions});
    });
}
```

We can see one example of the results produced through this code with the term *Confidentiality* that has several definitions. The code below shows a fragment of this

---

\(^5\) [https://www.upb.edu.co/es/home](https://www.upb.edu.co/es/home)

\(^6\) Stored thought a MongoDB (https://www.mongodb.com/) database

\(^7\) Implemented with NodeJS (https://nodejs.org/en/)
Confidentiality is enforced by preventing unauthorised disclosure of user data in transit between the two end points. The end points may be a TSF or a user. Confidentiality of TSF Data during transmission is necessary to protect such information from disclosure. Some possible implementations that could provide confidentiality include the use of cryptographic algorithms as well as spread spectrum techniques.

An evaluator may have access to sponsor and developer commercially-sensitive information (e.g. TOE design information, specialist tools), and may have access to nationally-sensitive information during the course of an evaluation. Schemes may wish to impose requirements for the evaluator to maintain the confidentiality of the evaluation evidence. The sponsor and evaluator may mutually agree to additional requirements as long as these are consistent with the scheme’s confidentiality requirements affect many aspects of evaluation work, including the receipt, handling, storage and disposal of evaluation evidence.

Confidentiality

- The property that information is not made available or disclosed to unauthorized individuals, entities, or processes (3.54).
- The property that information is not made available or disclosed to unauthorized individuals, entities, or processes.

The JSON file was edited suppressing, the surplus of data. The objective is to provide a better presentation and reduce size.
Towards the Consolidation of Cybersecurity Standardized Definitions

Firstly, the very same definitions appear in different sources: line 31 [12] is the same as 20 [11], line 51 [7] is the same as 41 [27], and lines 81 [40], 91 [41] and 101 [61] are same as 71 [52]. However only one of those is the primary source while the others are references to it, in this case, the primary sources are at the lines, respectively the 31 [12] in the previous release, 51 [7], and 91 [41]. With this tool we intend to gather all the considering domain terminology definitions according to their sources, to facilitate our analysis.

In common, all definitions consider the term Confidentiality a Property that can be assigned to many different Individuals. Some of the aforementioned refer to

\[ Property \text{ and Individuals in the ontological sense } [18] \]
kinds of Information like Proprietary, Sensitive or Personal, others refers to User Data. Indeed it is important to see that the Data term’s meaning is not the same as Information since not all data refers to information. Moreover, in a step forward we need to determine if the property of some individual being Confidential is quantified or not; and if it is, what is its quality structure the and how to measure if (it is possible) [20]. This kind of analysis is an example of how terminological validation is important, indeed this is part of an ontological analysis concerning the cybersecurity domain.

Another example of the use of the API refers to the ontologies we found. In this case, we intend to cross the ontology analysis results, including the definitions it uses, with the standards’ definitions. The code below shows a fragment of the information we collect about the SECCO ontology, which is a sub-ontology of CRATELO [55,56,54,3].

```javascript
// Get ontology
function getOntology(req,res){
    var ontology = new Ontology();
    ontology._id = req.params.id;
    Ontology.aggregate([
        { $match: { _id : ontology._id } },
        { $lookup : {
            from : "definitions",
            localField : "definitions",
            foreignField : "_id",
            as : "definitions" }
        },
        { $lookup: {
            from: "regexes",
            localField: "definitions.regex",
            foreignField: "_id",
            as: "regex" }
        },
        { $lookup : {
            from : "terms",
            localField : "regex.term",
            foreignField : "_id",
            as : "term" }
        },
        { $graphLookup : {
            from : "regexes",
            startWith : "$regex.next",
            connectFromField : "regex.next",
            connectToField : "_id",
            as : "next" }
        },
        { $lookup: {
            from: "ontologies",
            localField: "subOntologyOf",
            foreignField: "_id",
```
Towards the Consolidation of Cybersecurity Standardized Definitions

This code result presents the information we catch about the SECCO ontology, as below (http://localhost:3800/api/ontology/). We can see that the result also shows the definitions this ontology use and from which source these definitions came. The source can be any standard or document. Here we reduce file results showing only one definition since the file is large.

```
{ "ontology": {
  "_id": "600f1eaa10370e2e78c743d8",
  "definitions": [
    {
      "_id": "600f5a13d289480c60440184",
      "source": "600f59aed289480c60440183",
      "regex": "5eee523ad541e23b1e3855cb",
      "text": "(Risk). The risk is the probability that a successful attack occurs.",
      "locale": "pag 94"
    }
  ],
  "cqs": [],
  "name": "SECCO",
  "domain": "Security",
  "subOntologyOf": [
    {
      "_id": "600d7f5af2b31f1bb0080d7c",
      "definitions": [
        {
          "_id": "600f7ddef53a1f166000415f",
          "regex": [],
          "text": "(Risk). The risk is the probability that a successful attack occurs.",
          "locale": "en-US"
        }
      ],
      "domain": "Cybersecurity",
      "language": "OWL-Lite"
    }
  ],
  "language": "OWL-Lite"
}
```

...
All of this denotes how huge and complex is to provide a conceptualization of the cybersecurity domain. Therefore, one of the goals of the survey we made and its resulting API is to get together domain terminology definitions according to their sources, to facilitate our analysis. Then, we are cross comparing the result of this analysis with the definitions used in the ontologies we found in the pilot study, as a next step.

4 Conclusions

In this document, we present our proposal for an API in which we can consolidate definitions of the terms used in the cybersecurity domain. We present an example showing how complex is the set of definitions for a single concept, indeed this complexity gets increased concerning the vast amount of concepts, their relations, and the context in which they are applied. Our intention is also to analyze the standard support that provides the grounding for the concepts over the cybersecurity domain.

The API using a NoSQL database sounds a relevant contribution to help Ontology Engineers on ontological analysis where complex domains are the scenario. The objective of this kind of approach is to identify the semantics of the concepts used, their similarities, and differences. From this initial step, we aim to provide a link between the domain terminology, its context with its representations in ontologies, following the approach of [19]. Besides, the control of this information allows us to do reasoning and present results from a friendly interface, both are future research works preceding a final solution proposal to provide interoperability among ontologies implemented as KGs.

Acknowledgments. This work has been developed under the project Digital Knowledge Graph – Adaptable Analytics API with the financial support of Accenture LTD.
References

6. CCDB (ed.): CC and CEM addenda Exact Conformance , Selection-Based SFRs , Optional SFRs, vol. V0.5. CCDB, Geneva - Switzerland (May 2017)
42. MAECT™ Specification - Core Concepts (2017)
43. MAECT™ Specification - Vocabularies (2017)
53. NIST (ed.): Generally accepted principles and practices for securing information technology systems. NIST (2018)
14 Martins, B.et al.