A Meta-Ontological Approach to Securing the Semantic Web Data

Giampaolo Bella[1], Domenico Cantone[1], Gianpietro Castiglione[1], Marianna Nicolosi-Asmundo[1] and Daniele Francesco Santamaria[1]

[1]*Department of Mathematics and Computer Science, University of Catania, Viale Andrea Doria 6 - 95125 - Catania, Italy*

Abstract

The Semantic Web is a family of technologies that in the last decade increased its diffusion outside the academia, especially in business and industry contexts. As a natural consequence, the related security issues deriving from the adoption of such technologies has become urgent. However, even though the Semantic Web technological stack provides some layers to secure semantic-based applications, the proposed layers are only defined conceptually. Hence, providing to developers new approaches for defining security policies for accessing semantic data in a general and clear way becomes necessary, as demanded by the FAIR (Future Artificial Intelligence Research) principles.

The work presented in this paper aims to preliminarily anticipate and define a general and verticalisable ontological meta-model for securing semantic data, and specifically for implementing security properties and policies. Such meta-model can be leveraged by knowledge engineers to establish at low level how software and users should consume data, in order to lighten the applications from the management of security risks.

Keywords

RDF, Semantic Web, Ontologies, Security, OWL, SWRL.

1. Introduction

The Semantic Web is a disruptive technology with a huge impact on industry, thanks to its adoption in Artificial Intelligence (AI) [1]. The substantial expansion of the Semantic Web is driven by the innovative approach of web ontologies in defining machine-intelligible, interconnected, and shared data. This data can be feed to specific automatic theorem provers to infer new knowledge. Despite being an urgent issue given its adoption in AI, data security in the Semantic Web has never received adequate attention. The World Wide Web Consortium (W3C) proposed three security layers (Trust, Signature, and Encryption) for the infrastructure of Semantic Web

Semantic Shields I: 1st International Workshop on Modeling for Security - 15 July 2024 Twente, Netherlands
Corresponding Author: Daniele Francesco Santamaria

giampaolo.bella@unict.it (G. Bella); domenico.cantone@unict.it (D. Cantone); gianpietro.castiglione@phd.unict.it (G. Castiglione); marianna.nicolosiasmundo@unict.it (M. Nicolosi-Asmundo); daniele.santamaria@unict.it (D. F. Santamaria)

https://www.dmi.unict.it/giamp/ (G. Bella); https://www.dmi.unict.it/nicolosi/ (M. Nicolosi-Asmundo); https://www.dmi.unict.it/santamaria/ (D. F. Santamaria)

0000-0002-7615-8643 (G. Bella); 0000-0002-1306-1166 (D. Cantone); 0000-0003-2215-0416 (G. Castiglione); 0000-0003-4456-5110 (M. Nicolosi-Asmundo); 0000-0002-4273-6521 (D. F. Santamaria)

© 2023 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

CEUR Workshop Proceedings (CEUR-WS.org)

technologies. However, W3C has never provided clear guidelines on their implementation. As a result, developers of the Semantic Web have adopted *ad hoc* strategies for securing data, which are neither shareable nor generalisable. This has led to the creation of back-end software solutions that act as middle-ware for data access and modification. While these techniques may be effective, they lack a general and reusable approach, resulting in a multitude of solutions each with its own set of advantages and disadvantages. Moreover, knowledge engineers should share the security properties and policies adopted on their knowledge base, with other developers to build applications with a higher level of granularity in the management of knowledge graphs. For instance, consider a knowledge base describing citizens' daily life, including topics such as car insurance, healthcare, and taxes. It would be particularly meaningful if users could control how data is shared among authorities, such as allowing only selected insurance companies access to car information. Developing applications with such a level of control can turn into a nightmare. We can ease the application development by defining data management policies at the ontologies' design level.

In the context of relational databases, security policies are ensured at the back-end level by a family of strategies known as (Multi-)Tenancy, aiming to group portions of data based on their adherence to the same set of security policies. Considering this, the main goal of our work is to adapt the concept of data-tenancy at a deeper level by defining security policies on ontological data. To do so, an approach could consists in defining ontologies that describe who and how can manage ontological data, hence we need ontologies describing ontologies that are meta-ontologies.

The approach of this work-in-progress is oriented towards designing meta-ontologies to define and implement security policies in such a way as knowledge engineers can define how data is consumed with a higher level of abstraction. This minimises development efforts required to code Semantic Web applications, thus entrusting knowledge engineers with the responsibility of defining security policies and reducing the risk of systematic errors or bugs derived from their implementation.

We propose MOSS - *A Meta-Ontology for Securing the Semantic Web*, a meta-ontology that aims at applying standard approaches of data tenancy from relational databases to the realm of the Semantic Web. This step is done through the definition and implementation of meta-models for web ontologies, characterising security policies for the semantic data. The goal is to translate semantic data tenancy strategies at the ontological level, going beyond the application back-end layer, thus allowing knowledge engineers to define data security policies during the design phase. The policies defined by data engineers are subsequently enforced by a semantic database management system (DBMS), hence providing a general approach that eases the subsequent back-end development from data security concerns, thus reducing the risk of errors. Therefore, data tenancy, especially at the ontological level, represents a further step towards the openness and automation of security strategies, as they can be encoded through open and understandable formats for software agents, specifically those of the Semantic Web.

In the proposed approach, semantic data tenancy will be achieved by identifying the most relevant security properties for accessing and modifying data, and by encoding the identified properties into an ontological meta-model, constructed both by leveraging a combination of OWL annotations and axioms.

2. Related Work

The World Wide Web Consortium (W3C), namely the agency regulating Web standards, addressed only superficially the problem of securing the architecture [2], although it is evident, especially in what concerns the definition of privacy/security properties and policies [3]. The work by Thuraisingham [4] provided a general overview of the approaches on securing the Semantic Web. Through an analysis of various current policy languages, the work by Olmedilla et al. [5] offers an introduction to policy-based security and privacy protection, demonstrating how these languages can be applied to various applications. Halpin [6] described a semantic attacker that targets inference processes and their alternatives. These include the use of contemporary cryptography, which thwarts attacks by means of Transport Layer Security (TLS), and the ways in which W3C standards like the Internet Engineering Task Force (IETF) OAuth and W3C Web Cryptography Application Programming Interface (API) can address the use-cases required by the Semantic Web.

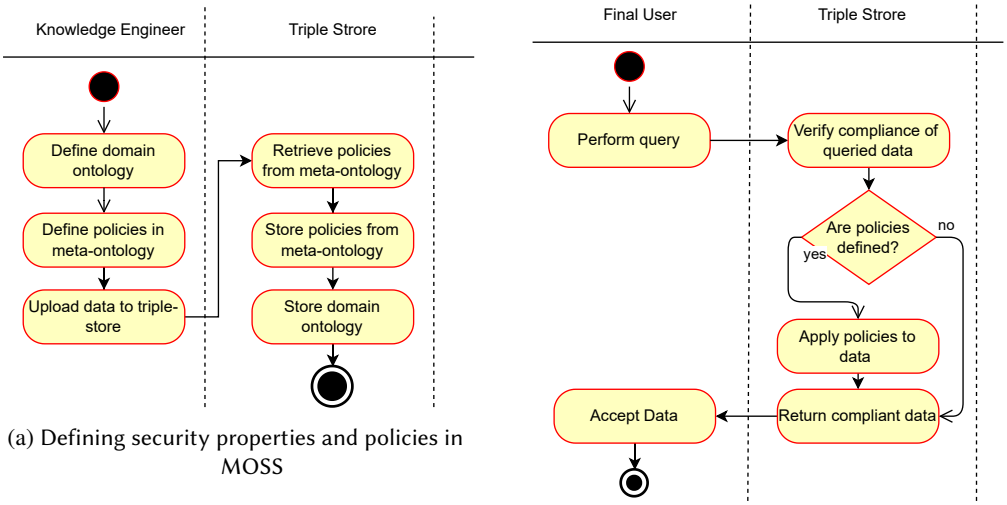
Some other works moved towards the principles of security in web ontologies. Denker et al. [7] summarised an ontological approach to enhancing the Semantic Web with security, providing an ontological description of some security measures. Kirrane et al. [8] noticed that Semantic Web technologies are not being properly used in the context of security. Castiglione et al. [9–11] presented an ontology mapping (automated and manual) for the legal language of security, particularly focusing on the European Network and Information Security (NIS) 2 Directive. Kagal et al. [12] presented a policy language that allows properties and policies to be described in terms of deontic concepts and models speech acts. Finally, Lima et al. [13] investigated the adoption of the Semantic Web for securing health data.

3. The MOSS approach

MOSS aims at realising Semantic Web Data tenancy through the following steps.

- **(a)** Analyse the most prominent security properties and policies for accessing and modifying data, for the subsequent ontological representation. The goal is to identify the security terminology and the related semantic relationships, particularly the ones involving the terms adopted in the context of securing web data.
- **(b)** Encode the definitions of the identified security properties and policies into ontological meta-models. The goal is to design the meta-models for security on the Semantic Web by way of a suitable terminology.
- **(c)** Implement the designed ontological meta-models through the Semantic Web standard languages. The goal is to build from the designed models effective OWL ontologies to be adopted.

The purpose of building the MOSS ontology is to provide a set of semantic models to secure ontological data (e.g., a crucial task for preventing the risks caused by data breaches) by following the security properties and policies the knowledge engineers encode. Security properties and policies should be designed at an ontological level and guaranteed by the overlying DBMS. This feature enables a general, yet practical approach that simplifies the development of Semantic Web



(a) Defining security properties and policies in MOSS

(b) Ensuring MOSS's security properties and policies at DBMS level

Figure 1: Flow diagrams describing the usage of MOSS

applications while ensuring security and scalability. Hence, thanks to **(a)** we ensure to adopt a standard terminology from the state of the art that guarantees the meta-ontology results compliant with the standard security terms; the goal of **(b)** is to design the meta-models for the security properties and policies leveraging the standard language pointed out in **(a)**. Finally, with **(c)** we encode the developed meta-model by leveraging OWL annotations. Even though RDF [14] can be used to annotate OWL ontologies by leveraging reification, RDF reasoning capabilities are extremely limited: for instance, SWRL [15] does not support reification as well, hence it cannot be adopted for this purpose. For this reason, it is feasible to foresee a suitable extension of SWRL to include annotations both for RDF entities and, more conveniently, for RDF statements. As an alternative to SWRL extension, it is feasible to adopt SPARQL Construct query form to generalise the application of security policy: such solution is however more cumbersome than defining suitable SWRL rules.

Once developers have defined the ontology specifying the data manipulation properties and policies, these can be uploaded to the semantic DBMS (in our case, OpenLink Virtuoso [16]) as illustrated in Figure 1(a). Security properties and policies are retrieved and uploaded into specific graphs by the DBMS. They are ultimately ensured by the DBMS, which is purposefully extended either through *ad hoc* plugins or Virtuoso/PL stored procedures (see Figure 1(b)). These extensions are dynamically applied at query time before the results are returned to users. The development of these tools is one of the areas for future work. Data are hence compliant with the ontology security properties and policies defined by the knowledge engineers.

From some of the most authoritative resources in the security field [17–19], we extract at least the following security properties and policies:

P1. **Confidentiality**. Confidentiality aims to preserve authorised restrictions on information

access and disclosure, including the protection of personal privacy and proprietary information. Therefore, a loss of confidentiality results in the unauthorised disclosure of information.

P2. **Access Control**. Access control implements a security policy that specifies who or what (e.g., in the case of a process) is allowed to access each specific system resource and the type of access that is permitted in each instance.

P3. **Authentication**. Authentication refers to the property of being genuine, verifiable, and trustworthy. It involves having confidence in the validity of a transmission, message, or its originator. This includes to verify that users are who they claim to be and that every input received by the system originates from a trusted source.

P4. **Authorisation**. Authorisation concerns the granting of a right or permission to an entity to access a system resource. This function determines who is trusted for a given purpose.

P5. **Privacy**. Privacy assures that individuals control or influence what information related to them may be collected and stored and by whom and to whom that information may be disclosed.

P6. **Anonymity**. Anonymity enables users to conceal or alter their identifying information, providing privacy and protection for their identity. However, it also presents challenges in holding them accountable for their actions and statements.

P7. **Availability**. Availability ensures timely and reliable access to and use of information. Therefore, a loss of availability is the disruption of access to or use of information or an information system.

P8. **Integrity**. Integrity involves the guarding against improper information modification or destruction, including ensuring information non-repudiation and authenticity. Therefore, a loss of integrity is the unauthorised modification or destruction of information.

We can assume that P7 and P8 are guaranteed by DBMSs. The goal is to encode P1-P6 into ontological models to be adopted for securing semantic data through properties and policies defined by knowledge engineers. As an example, we can consider the subset of data regarding a physical person, namely *user_1* (see Figure 2), organised in three main tenancies. The first tenancy regards the medical information about the person, including the medical history, whose access should be limited to the selected medical insurance company. A second tenancy involves the car insurance information, limited to the selected car insurance company selected by the person. A third tenancy concerns public information, such as the car model and related plate number. Other information, such as the user's personal data, can be anonymised and limited to the owner user.

To realise the mentioned tenancies, we first model the user's data domain into OWL statements, then the security properties and policies, and finally we connect them through specific OWL annotations. For instance, concerning the considered case study in Figure 2, we can introduce the following OWL statements describing a subset of user's data:

O1: user_1 hasLastName Doe.
O2: user_1 hasMedicalInsuranceCompany med_company.
O3: user_1 hasMedicalHistory user_1MH.
O4: user_1MH hasMedicalCase xxyyzz_t.

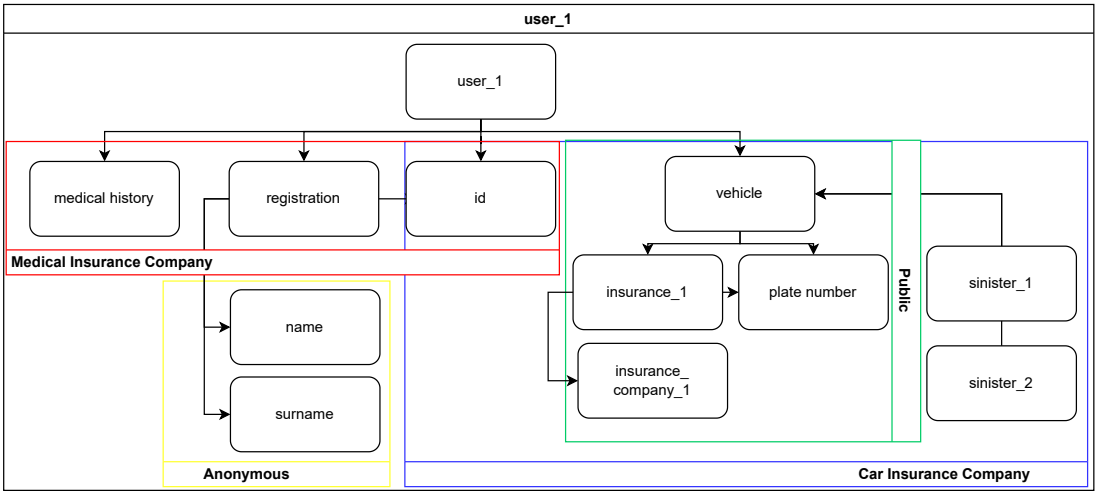


Figure 2: Example of data tenancy concerning a physical person

In statement O1, the user is identified by his/her last name; in O2, by a medical insurance company; and in O3, by his/her medical history, which contains the specific case outlined in O4.

We can now combine the defined statements within OWL annotations, for modelling the following security policies (the remaining policies can be modelled in an analogous way):

S1: O1 moss:securityPolicy p1. p1 rdf:type moss:Anonymity.
S2: O1 moss:securityPolicy p2. p2 rdf:type moss:Authorisation; moss:authorised :user_1.

In the previous OWL annotations, *securityPolicy* is an annotation property, *Anonymity* and *Authorisation* are OWL classes, and *p1*, *p2* are OWL named individuals. Figure 3 shows through the editor Protégé [20] the definition of the security policies concerning O1.

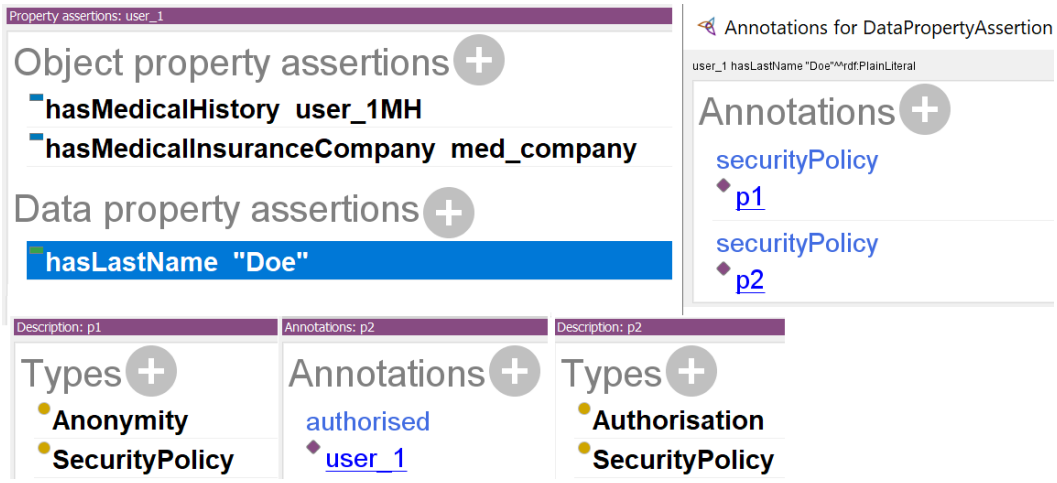


Figure 3: Security policies concerning statement O1.

The statements in S1 guarantee the anonymity of statement O1, while those in S2 grant the access only to *user_1*; in an analogous way, the statements in S3 ensure anonymity to statement O3, while those in S4 limit to *med_company* the access to O4.

SPARQL can be now exploited to generalise the application of security policies, for instance, limiting to medical insurance companies the access to any subject of the object-property *hasMedicalCase*. In this case, the construct query, which ensures that any medical insurance company is allowed to access statement O4, is the following:

```
CONSTRUCT { ?p :authorized ?m }
WHERE { ?x owl:annotatedProperty :hasMedicalCase.  ?x :securityPolicy ?p.
   ?p a :Anonymity.   ?x owl:annotatedProperty ?z.   ?m a :InsuranceCompany. }
```

However, it is more convenient to adopt SWRL to define this kind of rules, since it is tight integrated with OWL in such a way as to allow semantic reasoners to conjoin their inference capabilities. Nevertheless, a suitable extension of SWRL admitting RDF statements is required and it is one of the future extension concerning the MOSS approach. Another advantage of utilising reified annotations is the ability to use classes as objects of object-properties modelling security properties (e.g., *authorized*), thereby extending the associated security policy to all instances of these classes.

4. Conclusions and future work

Despite the relevance of Semantic Web technologies in industry, securing data in such context is still an open issue. Whenever Semantic Web applications are secured, this happens at back-end level by way of *ad hoc* measures. On the contrary, security properties and policies should be defined at ontological level by knowledge engineers, and applied by the overlying DBMS through general mechanisms; this implies that front-end and back-end developers are relieved from the security concerns, thus delivering more robust and reliable semantic applications. *A Meta-Ontology for Securing the Semantic Web*, in short MOSS, moves towards achieving such goal. This is achieved by defining security policies at the ontological level and applying them to the ontologies developed by knowledge engineers, who can now specify how data should be accessed and utilized.

Future goals are clearly stated. We shall finalise the ontology by modelling all the defined security properties. Next, we need an extension of SWRL that admits RDF statements. Subsequently, we shall provide OpenLink Virtuoso with all the means to apply the ontological policies defined by MOSS. Moreover, MOSS will be extended so as to be included in the OASIS ontology [21–23], thus bringing MOSS's approach security data in multi-agent systems. MOSS will be also applied to the case study concerning ontologies for historical buildings [24] and for archaeological findings cataloguing [25]. Finally, representing MOSS through the decidable fragments of set-theory as in [26–28] is one of our future commitments.

Acknowledgements

Giampaolo Bella acknowledges the project "FuSeCar" funded by the MUR Progetti di Ricerca di

Rilevante Interesse Nazionale (PRIN) Bando 2022 - grant 2022W3EPEP - CUP: E53D23008210006.

Domenico Cantone acknowledges partial support from project "STORAGE–Università degli Studi di Catania, Piano della Ricerca 2020/2022, Linea di intervento 2" and from the "Naples Dante Project" funded by the MUR Progetti di Ricerca di Rilevante Interesse Nazionale (PRIN) Bando 2022, grant 2022ZJ4N9E. Domenico Cantone is member of the Gruppo Nazionale Calcolo Scientifico-Istituto Nazionale di Alta Matematica (GNCS-INdAM).

Gianpietro Castiglione acknowledges a studentship by Intrapresa S.r.l. and Italian "Ministero dell'Università e della Ricerca" (D.M. n. 352/2022).

Marianna Nicolosi Asmundo acknowledges partial support from project "STORAGE–Università degli Studi di Catania, Piano della Ricerca 2020/2022, Linea di intervento 2", from the "Naples Dante Project" funded by the MUR Progetti di Ricerca di Rilevante Interesse Nazionale (PRIN) Bando 2022, grant 2022ZJ4N9E, and from the "Contact-induced change and sociolinguistics: an experimental study on the Gallo-Italic dialects of Sicily" by the MUR PRIN Piano Nazionale di Ripresa e Resilienza (PNRR), Bando 2022, grant P2022YWS8T. Marianna Nicolosi Asmundo is member of the Gruppo Nazionale Calcolo Scientifico-Istituto Nazionale di Alta Matematica (GNCS-INdAM).

Daniele Francesco Santamaria acknowledges the Research Program PIAno di inCEntivi per la Ricerca di Ateneo 2020/2022 — Linea di Intervento 3 "Starting Grant" - University of Catania.

References

[1] A. Breit, L. Waltersdorfer, F. J. Ekaputra, M. Sabou, A. Ekelhart, A. Iana, H. Paulheim, J. Portisch, A. Revenko, A. T. Teije, F. Van Harmelen, Combining machine learning and semantic web: A systematic mapping study, ACM Comput. Surv. 55 (2023). doi:10.1145/3586163.

[2] World Wide Web Consortium (W3C), The security of the semantic web - secrecy, trust and rationality, 2003.

[3] S. A. Azwari, Privacy, security and policies of the semantic web: A review, Journal of Advances in Information Technology (2022). URL: https://api.semanticscholar.org/CorpusID:247199940.

[4] B. Thuraisingham, Security standards for the semantic web, Computer Standard & Interfaces 27 (2005) 257–268. doi:10.1016/j.csi.2004.07.002.

[5] D. Olmedilla, Security and privacy on the semantic web, in: Security, Privacy, and Trust in Modern Data Management, 2007, pp. 399–415. URL: https://api.semanticscholar.org/CorpusID:14132471.

[6] H. Halpin, Semantic insecurity: Security and the semantic web, in: International Workshop on Semantic Web Technologies, 2017, pp. 187–202.

[7] G. Denker, L. Kagal, T. Finin, Security in the semantic web using owl, Information Security Technical Report 10 (2005) 51–58. doi:10.1016/j.istr.2004.11.002.

[8] S. Kirrane, S. Villata, M. d'Aquin, M. d'Aquin, S. Kirrane, S. Villata, Privacy, security and policies: A review of problems and solutions with semantic web technologies, Semant. Web 9 (2018) 153–161. doi:10.3233/SW-180289.

[9] G. Castiglione, G. Bella, D. F. Santamaria, Towards grammatical tagging for the legal language of cybersecurity, in: Proceedings of the 18th International Conference on Availability, Reliability and Security, ARES '23, Association for Computing Machinery, New York, NY, USA, 2023, pp. 1–9. doi:10.1145/3600160.3605069.

[10] G. Bella, G. Castiglione, D. F. Santamaria, An automated method for the ontological representation of security directives, in: Proceedings of the Joint Ontology Workshops 2023, Episode IX: The Quebec Summer of Ontology, co-located with the 13th International Conference on Formal Ontology

in Information Systems (FOIS 2023), Sherbrooke, Québec, Canada, July 19–20, 2023, volume 3637, CEUR Workshop Proceedings, 2023, pp. 1 – 17.

[11] G. Bella, G. Castiglione, D. F. Santamaria, An ontological approach to compliance verification of the NIS 2 directive, in: Proceedings of the Joint Ontology Workshops 2023, Episode IX: The Quebec Summer of Ontology, Sherbrooke, Québec, Canada, July 19–20, 2023, volume 3637, CEUR Workshop Proceedings, 2023, pp. 1 – 12.

[12] L. Kagal, T. Finin, A. Joshi, A policy based approach to security for the semantic web, in: D. Fensel, K. Sycara, J. Mylopoulos (Eds.), The Semantic Web - ISWC 2003, Springer Berlin Heidelberg, Berlin, Heidelberg, 2003, pp. 402–418.

[13] V. Lima, D. Alves, F. Bernardi, R. Rijo, Security approaches for electronic health data handling through the semantic web: A scoping review, Semantic Web 14 (2022) 1–14.

[14] G. Klyne, J. J. Carroll, Resource description framework (rdf): Concepts and abstract syntax, W3C Recommendation, 2004. URL: http://www.w3.org/TR/2004/REC-rdf-concepts-20040210/.

[15] I. Horrocks, P. F. Patel-Schneider, H. Boley, S. Tabet, B. Grosofand, M. Dean, SWRL: A semantic web rule language combining OWL and RuleML, 2004. URL: http://www.w3.org/Submission/SWRL/.

[16] OpenLink, Virtuoso Universal Server, 2020. Link: https://docs.openlinksw.com/virtuoso/.

[17] W. Stallings, L. Brown, Computer Security: Principles and Practice, 3rd ed., Prentice Hall Press, USA, 2014.

[18] J. Nin, J. Herranz, Privacy and Anonymity in Information Management Systems: New Techniques for New Practical Problems, Springer Publishing Company, Incorporated, 2012.

[19] N. Ferguson, B. Schneier, T. Kohno, The Context of Cryptography, John Wiley & Sons, Ltd, 2015, pp. 1–22. doi:10.1002/9781118722367.ch1.

[20] 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.

[21] G. Bella, G. Castiglione, D. F. Santamaria, A behaviouristic approach to representing processes and procedures in the OASIS 2 ontology, in: Proceedings of the Joint Ontology Workshops 2023, Episode IX: The Quebec Summer of Ontology, Sherbrooke, Québec, Canada, July 19–20, 2023, volume 3637, CEUR Workshop Proceedings, 2023, pp. 1 – 17.

[22] G. Bella, D. Cantone, M. Nicolosi Asmundo, D. F. Santamaria, The ontology for agents, systems and integration of services: recent advancements of OASIS, in: 23rd Workshop From Objects to Agents, WOA 2022, Genova 1–3 September 2022, volume 3261, CEUR-WS, 2022, pp. 176 – 193.

[23] D. Cantone, C. F. Longo, M. Nicolosi-Asmundo, D. F. Santamaria, C. Santoro, Towards an ontology-based framework for a behavior-oriented integration of the iot, in: 20th Workshop From Objects to Agents, WOA 2019, Parma 26–28 June 2019, volume 2404, CEUR-WS, 2019, pp. 119 – 126.

[24] C. Cantale, D. Cantone, M. Nicolosi-Asmundo, D. F. Santamaria, Distant reading through ontologies: The case study of Catania's benedictines monastery, JLIS.it 8 (2017) 205 – 219. doi:10.4403/jlis.it-12342.

[25] D. Cantone, M. Nicolosi-Asmundo, D. F. Santamaria, S. Cristofaro, D. Spampinato, F. Prado, An EPIDOC ontological perspective: The epigraphs of the castello ursino civic museum of Catania via CIDOC CRM, Archeologia e Calcolatori 30 (2019) 139 – 157. doi:10.19282/ac.30.2019.10.

[26] D. Cantone, M. Nicolosi-Asmundo, D. F. Santamaria, A set-theoretic approach to reasoning services for the description logic $\mathscr{D}\mathscr{L}_{\mathbf{D}}^{4,\times}$, Fundamenta Informaticae 176 (2020) 349 – 384. doi:10.3233/FI-2020-1977.

[27] D. Cantone, M. Nicolosi-Asmundo, D. F. Santamaria, A set-based reasoner for the description logic dl4,xd, in: 3rd International Workshop on Sets and Tools, SETS 2018, Southamptom, 5 June 2018, volume 2199, 2018, pp. 52 – 66.

[28] D. Cantone, M. Nicolosi-Asmundo, D. F. Santamaria, Conjunctive query answering via a fragment of set theory, in: 17th Italian Conference on Theoretical Computer Science, ICTCS 2016, Lecce, 7–9 September 2016, volume 1720, CEUR-WS, 2016, pp. 23–35.