Security approaches for electronic health data handling through the Semantic Web: a scoping review

Vinícius Costa Lima[a,b,*], Filipe Andrade Bernardi[a,b], Domingos Alves[c], Rui Pedro Charters Lopes Rijo[d,e,f]

[a] *Health Intelligence Laboratory, Ribeirão Preto Medical School, University of São Paulo, Brazil*
[b] *Bioengineering Post Graduate Program, School of Engineering of São Carlos, University of São Paulo, Brazil*
[c] *Department of Social Medicine, Ribeirão Preto Medical School, University of São Paulo, Brazil*
[d] *School of Technology and Management, Polytechnic Institute of Leiria, Portugal*
[e] *Institute for Systems Engineering and Computers at Coimbra, Coimbra, Portugal*
[f] *Center for Health Technology and Services Research, Porto, Portugal*

Abstract. Integration of health information systems are crucial to advance the effective delivery of healthcare for individuals and communities across organizational boundaries. Semantic Web technologies may be used to connect, correlate, and integrate heterogeneous datasets spread over the internet. However, when working with sensitive data, such as health data, security mechanisms are needed. A scoping review of the literature was undertaken to provide a broad view of security mechanisms applied to, or along with, Semantic Web technologies that could allow its use with health data. Searches were conducted in the most relevant databases for the scope of the present work. The findings were classified according to the main objective and features presented by each solution. Twenty studies were selected for the review. They introduced mechanisms that addressed several security attributes, such as authentication, authorization, integrity, availability, confidentiality, privacy, and provenance. These mechanisms support access control frameworks, semantic and functional interoperability infrastructures, and privacy compliance solutions. The findings suggest that the application and use of Semantic Web technologies is still growing, with the healthcare area being particularly interested. The main security mechanisms for Semantic Web technologies, the key security attributes and properties, and the main gaps in the literature were identified, helping to understand the technical needs to mitigate the risks of handling personal health information over the Semantic Web. Also, this research has shown that complex and robust solutions are available to successfully address several security properties and features, depending on the context that the electronic health data is being managed.

Keywords: Semantic Web, Health Information Systems, Electronic Health Records, Computer Security, Interoperability

1. Introduction

In the current World Wide Web (WWW), most content is not easily accessible by machines, since it was made for human interpretation. The Semantic Web (SW) term was coined by Tim Berners-Lee and established by the World Wide Web Consortium (W3C), understood as an extension of the WWW that, besides of linking hypertext documents, can also recognize the information meaning and, through inference rules and ontologies, assist in knowledge management [5,53]. SW aims to express the meaning of a given information, i.e., its properties and the complex relationships between different types of data, in a way that enables the interpretation of its meaning without worrying about its form of representation [49].

Integration of health information systems are crucial to advance the effective delivery of healthcare for individuals and communities across organizational boundaries [25]. SW technologies can be used in open and sensitive contexts to connect, correlate, and integrate heterogeneous datasets spread over the Internet. However, when working with sensitive data, such as health data, security mechanisms are needed to protect it from unauthorized access.

Data leaks can cause harm in a variety of ways. A breach of security can result lives damage and in financial and legal consequences. For instance, improper handling of confidential data can violate government regulations, resulting in fines and other sanctions. Also, it may be a strong disincentive to data sharing initiatives among organizations [48]. In the case of health records, it usually represents the physical and/or mental condition of a patient. If security is breached, the disclosure of personal information may cause economic, social, and psychological potential harms [40].

The main goal of this scoping review is to provide a broad view of security mechanisms applied to, or along with, SW technologies that could allow its use with health data, as well as to identify possible research gaps in existing literature, and key characteristics or factors related to security in the SW.

The next section explains the research methodology. In the third section, the review will be reported. The fourth section presents a discussion based on the findings. Finally, conclusions will be drawn in the last section.

2. Materials and Methods

Scoping reviews are ideal to provide an overview of a given topic, as well as to determine its coverage and give a clear indication of the volume of literature and studies available [39]. The methodology is based on Arksey and O'Malley [3] and guidelines provided by the Joanna Briggs Institute [44], which recommend a five-stage framework for scoping review. In the following subsections, each stage will be detailed.

2.1. Stage 1: Identifying the objective and research questions

This scoping review aims to verify the existing contributions in the literature to answer the following research questions:

Q1. Which are the main security mechanisms being applied to, or along with, SW technologies to protect health data?

Q2. Which key security properties are being addressed?

Q3. What are the knowledge gaps regarding security for health data handling through the SW?

2.2. Stage 2: Identifying relevant studies

In this stage, the search strategy, i.e., the selected databases and keywords, and the eligibility criteria for assessing each primary study were defined to carry out the search to narrow the studies.

Search strategy. The following databases were searched: PubMed/MEDLINE, IEEE Xplore Digital Library, Scopus, Embase, Web of Science, ProQuest, and Cochrane Database of Systematic Reviews. These databases are most relevant in the scope of the present work, covering health and technology-related topics. The search string was defined as follows:

health AND "semantic web" AND (security OR privacy OR "access control" OR integrity OR confidentiality OR cryptography)

Eligibility criteria. Research papers, among others, full papers, reviews, and conference papers, and non-research studies, e.g., editorials, letters, in English language were included. The established time frame was from May 2001, when the term Semantic Web was first coined [5], to July 2021. Publications that do not refer to the keywords "*health*", "*semantic web*", and at least one security related term in the title or the abstract were excluded, as well as papers that present only proposals/non-implemented models or solutions not applied to the Semantic Web.

2.3. Stage 3: Study Selection

Two investigators have independently screened each retrieved article based on title and abstract for eligibility. Then, the full text was retrieved, and the investigators have performed another round of review. Fruitful discussions with the research team resolved the two reviewers' disagreements. Reviewers were not blinded to the journal's title, study authors, or associated institutions.

Although not being a systematic review, the Preferred Reporting Items for Systematic Review and Meta-Analysis (PRISMA) [38] flow diagram was used to better comprehend the study selection.

2.4. Stage 4: Charting the data

An extraction strategy was defined to capture relevant data from the selected studies. Data extracted must be enough to answer the research questions established in Stage 1. Table 1 indicates the type of data that were extracted from each included article.

Table 1. Data extraction strategy

Scope	Data to be extracted
Summary	Title, authors, publication type, year of publication, periodic/journal, aims/objectives
Q1	Results and security mechanisms
Q2	Applicability and related security properties
Q3	Advantages, features and limitations

2.5. Stage 5: Collating, summarizing and reporting the results

Since scoping studies seek to present an overview of all material reviewed [3], data were classified and presented in a table ordered by the category and by the year of publication. The table contains narrative content with data obtained in Stage 4. The narrative synthesis will seek to investigate similarities and differences between studies to explore patterns, themes, and relationships.

3. Results

Initially, 303 articles were found in the selected databases of which 189 articles were included in the study after removing 114 duplicates. After screening the titles and abstracts, the number of articles was reduced to 39. However, a full-text assessment for eligibility excluded 19 additional articles because 2 studies were inaccessible (no free or institutional access) and 17 did not meet the eligibility criteria. Finally, 20 (39-19) studies were selected to be included in this research.

The Figure 1 shows the PRISMA flow diagram. Details of the selected studies are presented in Table 2.

According to the explored literature, research about security approaches to handle electronic health data through the Semantic Web has increased gradually since 2010. The results showed that the selected articles were published between 2005 and 2021, but the volume was higher between 2014 and 2019. Of the 20 included studies, there are 14 research articles, 5 conference papers and 1 editorial.

The findings were classified according to the main objective of each solution, namely Access Control (11 articles), Interoperability Infrastructure (2 articles), and Privacy Compliance (6 articles). The following security attributes and features were addressed by the papers: authentication, authorization, integrity, availability, confidentiality, privacy, and provenance. Depending on the purpose of the solution, each study typically provides mechanisms to implement one or more of each attribute or features.

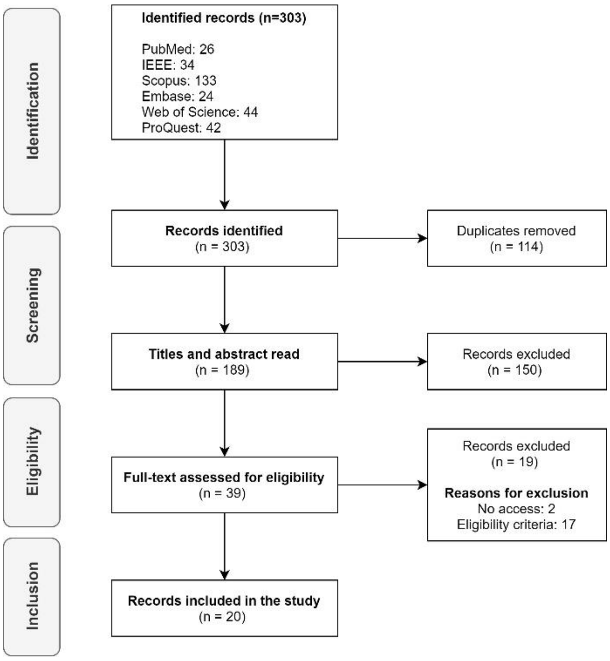

Fig 1. PRISMA flow diagram

Authentication is the mechanism used to make sure users or external systems are correctly identified, that is, that their identity is valid [42]. Authorization refers to permissions that a user must hold to access specific resources, usually after a successful authentication process [50]. Data integrity must be protected against malicious deletion, modification or fabrication. Therefore Information must be accurate, complete, consistent, so only selected individuals with the appropriate permissions can change it [8]. The availability property refers to the system being able to be used by a user or machine whenever necessary and without disruptions [42]. Confidentiality considers that only authorized individuals can have access to data, systems and services through a set of rules and restrictions [54], [7]. Privacy is related to the right of no intrusion over one’s personal information [22], avoiding the disclosure of any identifiable data. Provenance refers to the origination or source of specified data, such as requests to personal information, supporting privacy requirements and traceability of data flows [41].

Table 2. Details of the selected studies

#	Title	Authors	Publication type	Year	Periodic	Categories	Applicability and related security properties	Results and security mechanisms	Advantages and features
1	Privacy compliance and enforcement on European healthgrids: An approach through ontology	Rahmouni H.B., Solomonides T., Mont M.C., Simon Shiu	Article	2010	Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences	Access Control	Authorization, Privacy	Authors suggest a direct mapping from high-level legislation on privacy and data protection to operational-level privacy-aware controls. In this case, Web Ontology Language and Semantic Web Rule Language are used for the specification and reasoning of access control policies, as well as user and data categories (as defined in legislations) and data disclosure contexts (e.g., maps the purposes of a needed consent). The paper defines an architecture for the enforcement of these controls on access control models adopted in healthgrid security infrastructures. A unique privacy context mode is involved when only one record of data is subject to a sharing request. A multi-privacy context mode is involved when large amounts of data need to be shared.	The proposed solution allows the specification and evaluation of access policies to preserve patient's privacy in data sharing contexts. Semantic Web technologies provide flexibility to adapt the architecture based on different regulations from each country in Europe.
2	Using OWL and SWRL to represent and reason with situation-based access control policies	Beimel, D; Peleg, M	Editorial	2011	Data & Knowledge Engineering	Access Control	Authorization	Authors use ontologies defined in Web Ontology language (OWL) to model Situation (scenarios) classes, formulating data-access rule classes. A set of data-access rule classes makes up the organization's data-access policy. The SWRL engine is used to infer new knowledge and relations. Then, the DL reasoner is used for knowledge classification and for real-time realization of the incoming data-access request as a member of an existing Situation class to infer the appropriate response. The inferred response type can be approved or denied. The OWL-based SitBAC knowledge framework complies with the need-to-know principle for data disclosure, which means that data is disclosed only when it is strictly necessary for someone to conduct official duties.	A context-based access-control solution is useful due to it's flexibility to model organizations' policies and the different scenarios of incoming data access requests that a healthcare information system may receive. Semantic Web technologies give the capability to reason and evaluate requests based on a specific scenario, considering who, when, where, and why the data is needed.

3	A SWRL bridge to XACML for clouds privacy compliant policies	Rahmouni H.B., Mont M.C., Munir K., Solomonides T.	Conference paper	2014	CLOSER 2014 - Proceedings of the 4th International Conference on Cloud Computing and Services Science	Access Control	Authentication, Authorization, Privacy	Authors used Semantic Web technologies, such as OWL and SWRL, to model privacy requirements defined in European and national data protection laws as privacy aware access control policies. Through mathematical formalism, semi-automated mapping templates were established to transform the Semantic Web access control policies in XACML policies, a highly portable standard of access control.	The use of ontologies and semantic technologies could provide relatively easy interpretation of legislation at an operational level and the mapping of these ontologies to standard XACML policies could facilitate the implementation of SWRL rules in existing systems and complex environments.
4	Access control management for e-Healthcare in cloud environment	Sun, Lili; Yong, Jianming; Soar, Jeffrey	Article	2014	EAI Endorsed Transactions on Scalable Information Systems	Access Control	Authentication, Authorization	SAC extends RBAC by considering the semantics of objects and associates permission with concepts instead of objects. Authors use ontologies for the RBAC security model and implement access control system in semantic web environment. An infrastructure was designed to enforce and evaluate authenticated requests. The SAC performs queries to the semantic knowledge base in order to find attributes associated with subjects and objects and translates the request to the XACML format. The XACML evaluates the request against an access control policy and sends the response to the requester.	This work supports the access in heterogeneous, distributed and large environments. It also provides the cross organizational access control, which is crucial in healthcare applications.
5	A semantic authorization model for pervasive healthcare	Li Z., Chu C.-H., Yao W.	Article	2014	Journal of Network and Computer Applications	Access Control	Authorization	The authorization model is composed of 4 layers, namely data/user, ontology, authorization, and application layers. Each layer handles a specific mechanism, such as data resources and organizations, concepts and relationships of objects, policies and rules, and the views, respectively. Authorizations can be specified at different levels of the predefined concept hierarchies and be propagated to lower-levels. Relying on ontology reasoning tools, the context dynamics must be encoded to enforce context-aware authorizations. Therefore, considering the source of the user's request, the data resource, the concept trees, the mapped relationships, the concept-level policies, and the security rules, semantic reasoning is	Large-scale, distributed and heterogeneous systems demand support for fine-grained authorizations on different data records and portions for complex objects such as EHRs. However, the proposed model can be easily adapted to other large-scale distributed information systems where complex organizations and data resources are involved.

								conducted to obtain context-aware authorizations.	
6	COC: An ontology for capturing semantics of circle of care	Dong X., Samavi R., Topaloglou T.	Conference paper	2015	Procedia Computer Science	Access Control	Authorization, Privacy, Provenance	The proposed ontology and easy integration with FHIR supported EHRs for explicit and implicit access consent. Access logs are annotated with the COC ontology and converted into a RDF dataset that can be queried using SPARQL queries to investigate if an individual is in the circle of care of a patient.	Overcome shortcomings of RBAC systems (e.g. capturing the consent of patients) and patient-centric approaches (e.g., patients may not be computer-savvy enough or have the necessary knowledge to be able to set permissions)
7	A fine-grained context-aware access control model for health care and life science linked data	Liu, ZT; Wang, JD	Article	2016	Multimedia Tools and Applications	Access Control	Authentication, Authorization	Authors use Semantic Web technologies to allow publishers of Linked Data to define access conditions for their data by extending XACML with semantics. XACML rules are used to define the policies and SWRL rules to express semantic relations and inference problems. Automated decision making to permit or deny an access request is accomplished through inference processes based on the semantic relations among entities.	Considering that the Semantic Web is widely used for publishing open Linked Data, defining an approach to make available sensitive linked datasets is crucial to allow an administrator to take advantage of Semantic Web technologies for heterogeneous data integration.
8	Towards a semantic medical internet of things	Dridi A., Sassi S., Faiz S.	Conference paper	2017	Proceedings of IEEE/ACS International Conference on Computer Systems and Applications, AICCSA	Access Control	Authentication, Authorization, Confidentiality, Privacy	The Semantic Medical IoT platform is capable of receiving data from medical equipment, IoT devices and electronic health records through a semantic interoperability layer, which performs semantic annotation and data integration. To ensure the security and confidentiality of information, the platform defines new contract-based security policies and provides a set of mechanisms for user's authentication and privacy control of personal health personal data.	It is a comprehensive platform that deals with raw health data and transforms it into interpretable information, while approaching security and privacy concerns.
9	Establishment of access levels for health sensitive data exchange through semantic web	Lima V.C., Alves D., Pellison F.C., Yoshiura V.T., Crepaldi N.Y., Rijo R.P.C.L.	Conference paper	2018	Procedia Computer Science	Access Control	Authentication, Authorization	A simple-but-robust mechanism to control which pieces of data an pre-authorized external system can retrieve from a semantic API.	The solution allows granular access control to semantic annotated data, taking advantage of ontologies properties.

10	Semantic privacy-preserving framework for electronic health record linkage	Lu, Yang; Sinnott, Richard O	Article	2018	Telematics and Informatics	Access control	Authorization, Privacy	Authors combine techniques of data anonymity, access control (XACML), and data semantics to show how privacy preservation can be satisfied through specifying background knowledge and further restricting the access to certain data. The lack of semantic expressiveness is a barrier for finer-grained authorization. Therefore, the semantic framework includes policy formalization, compliance checking and knowledge discovery to prevent privacy risks with arbitrary linkages. To support semantic reasoning, policy vocabulary, domain knowledge and internal logic are mapped into ontological concepts. The mechanism of XACML systems are extended with Semantic Rules to achieve compliance checking between access requests and security policies.	The proposed semantic framework for privacy-preserving and records linkage has the potential to promote patient-centered healthcare, due to the possibility of accessing the complete historical information of the patient, including electronic health records stored in different health facilities.
11	Multi Authority Access Control in a Cloud EHR System with MA-ABE	Dixit, S; Joshi, KP; Choi, SG	Conference paper	2019	IEEE International Conference on Edge Computing (IEEE EDGE)	Access Control	Authentication, Authorization, Confidentiality, Privacy	The solution implements a secure access control mechanism for user authentication and a robust crypto module for data encryption in order to tighten security and privacy before moving the data out of the organization. Attribute Based Access Control (ABAC) and a Multi-Authority EHR Ontology are used to carry out an access decision by matching the extracted attributes against the confidential access policies defined by an organization stored within the Policy unit in the form of Semantic Web Rule Language (SWRL) rules.	The use of a multi-authority ontology and SWRL rules provide flexibility for a complex environment where attributes differ among organizations. Therefore, reasoning is useful to evaluate an access request.
12	ARTEMIS: towards a secure interoperability infrastructure for healthcare information systems.	Boniface M., Wilken P.	Article	2005	Studies in health technology and informatics	Interoperability Infrastructure	Authentication, Authorization, Confidentiality, Integrity, Privacy	The infrastructure works as a broker, or a middleware, that can enable the communication of standalone health information systems through mediation between semantic security and privacy policies, abstracting the differences in security requirements and capabilities of each system. The compatibility is achieved by using semantic web services and ontologies for reasoning of roles, clinical concepts and security policies.	The ARTEMIS project delivers a complete architecture for functional, semantic and organizational interoperability, including security mechanisms for secure data exchange across organizations boundaries.

13	Secure semantic smart healthcare (S3HC)	Tiwari S.M., Jain S., Abraham A., Shandilya S.	Article	2019	Journal of Web Engineering	Interoperability Infrastructure	Authentication, Authorization, Confidentiality, Integrity, Privacy, Availability	A healthcare ontology named HCIoTO was designed for transferring the collected data from the device to the knowledge base and vice-versa. SWRL rules and SPARQL queries are used to represent the accuracy and correct semantic reasoning between patients and doctors. Several security layers are available, such as RDF Security, XML Security and secure communication protocols. The framework collects, integrates and stores data from several connected devices. To protect the data, it addresses several security properties, such as confidentiality (encryption), integrity (hash functions), authentication (device identity verification), authorization (access policies), and availability.	For patients, the main advantage is that the S3HC framework supports doctors in analyzing the collected vital signs and in providing an appropriate and secure service for patients. From a technical perspective, it is a robust framework that delivers a semantic infrastructure for data storing, integration, and querying.
14	Security Framework for Tuberculosis Health Data Interoperability Through the Semantic Web	Lima, VC; Pellison, FC; Bernardi, FA; Alves, D; Rijo, RPCL	Article	2021	International Journal of Web Portals	Interoperability Infrastructure	Authentication, Authorization, Confidentiality, Integrity, Availability	The framework implements a Security Layer to support authentication and authorization for semantic web services and a query endpoint, ensuring confidentiality, integrity and availability of the data during exchange events. In addition, endpoints for SPARQL and GraphQL queries are available, enabling the extraction of tuberculosis health data from a regional health information system. The solution is based on hybrid cryptography (a combination of symmetric and asymmetric techniques for encryption and transmission of big data), hash functions and ontologies.	The solution relies on ontologies and a virtual triple store database to convert, in real-time, legacy data stored in relational databases to semantic formats, so only responses in semantic formats (e.g. JSON-LD and RDF) are sent to authorized requesters. The final user can extract data using different technologies (SPARQL, GraphQL or APIs).
15	A Model-driven Privacy Compliance Decision Support for Medical Data Sharing in Europe	Boussi Rahmouni H., Solomonides T., Casassa Mont M., Shiu S., Rahmouni M.	Article	2011	Methods of information in medicine	Privacy Compliance	Privacy	Authors use ontologies to model the required domain and context information about data sharing and privacy requirements and a set of Semantic Web Rule Language rules to reason about legal privacy requirements that are applicable to a specific context of data disclosure. A semantic web application is also available to provide decision support for clinicians to enhance privacy compliance. The application allows users to obtain privacy management guidelines for different entities in European	The solution delivers is appropriate for non-technical users (e.g. clinicians), due to the availability of a graphical interface to obtain privacy management guidelines from distinct entities in Europe. The use of ontologies and SWRL rules allow the modelling of requirements and legal aspects of organizations, as well reasoning about a

								countries involved in a data sharing process.	specific context of data disclosure.
16	Knowledge-based personalized search engine for the Web-based Human Musculoskeletal System Resources (HMSR) in biomechanics	Dao T.T., Hoang T.N., Ta X.H., Ho Ba Tho M.C.	Article	2013	Journal of Biomedical Informatics	Privacy Compliance	Confidentiality	Semantic web services process requests and a multi-agent crawler searches the world wide web based on user-defined keywords. The search engine encrypts the results to protect medical information using a cryptography algorithm and a pair of keys. The user must use a private key to decrypt and read the search result.	The confidentiality provided by the semantic-based search engine is a feature that may stimulate the use of the solution in sensitive contexts, i.e., when dealing with personal health information.
17	Preserving patients' privacy in health scenarios through a multi-context-aware system	Celdran, AH; Perez, MG; Clemente, FJG; Perez, GM	Article	2017	Annals of Telecommunications	Privacy Compliance	Privacy	h-MAS allows users to choose profiles (e.g. privacy policies) to manage when, where, how, and to whom their private information can be revealed. These profiles are specific to the context in which users are located, aimed to protect the privacy of their personal information, which can be modified by adding, modifying, or deleting its policies according to his/her interests. Information about users and contexts is represented by ontologies defined in OWL 2 and privacy policies are expressed in SWRL. Reasoning is performed to decide if a given information can be disclosed. The reasoner receives the ontological models generated by the Jena API and applies SPARQL queries to obtain the requested information.	The multicontext-aware system is a patient-centered solution for privacy-preserving of personal health information. h-MAS allows the user to disclose only specific information needed in a given context (intra or inter contexts).
18	Ontology for Attack Detection: Semantic-Based Approach for Genomic Data Security	Noor, S; Ahmed, M; Saqib, MN; Abdullah-Al-Wadud, M; Islam, MS; Fazal-e-Amin	Article	2017	Journal of Medical Imaging and Health Informatics	Privacy Compliance	Privacy	The DAG is the main contribution of the paper. The system is able to analyze and validate incoming requests through inference rules. Incoming requests are parsed and potential attacks are compared with the information stored in the knowledge-base by inferring over the rules, and the system generates alerts upon detecting an attack.	Genome sequencing technologies offer great promises of medical advances, but individuals privacy stands at risk because of outsourcing sensitive information. The ontology captures the context of important attacks and threats on genomic data as well as potential consequences and the vulnerabilities exploited by these attacks for further analysis and to conduct mitigation actions.
19	Improving privacy in health care with an	Ozgu Can; Yilmazer, Dilek	Article	2019	Expert Systems	Privacy Compliance	Privacy, Provenance	To preserve patients' privacy and provide traceability of historical data, the authors defined a healthcare	Through an Ontology-Based Access Control, the implemented privacy-

	ontology-based provenance management system							provenance information system able to search for security violations based on access permissions defined by patients for their medical data. The solution relies on domain ontologies from different health fields to query, trace and protect sensitive data, as well as for the definition of access permissions.	aware provenance management system allows to trace patients' medical record and to define permission and prohibition rules on patients' data. The most important feature of the proposed system is detecting privacy violations by using the access history of data and the defined access permissions and prohibitions.
20	An integrated framework for privacy protection in IoT â€" Applied to smart healthcare	Mansour, NA; Hanadi, B; Amjad, R; Mahmoud, B	Article	2021	Computers and Electrical Engineering	Privacy Compliance	Privacy	The authors use an inference model, based on the Semantic Web and its supporting technologies (e.g., domain ontologies expressed using OWL) to allow the user to determine the privacy risks incurred when some personal data elements are shared with a data consumer. The framework provides useful information about the data request, such as the type of data being requested, the data consumer, and the context of the patient. Also, other data elements and accessible information about the same user are gathered (for example, from public data sources) and combined. Finally, a list of risks and recommendations is provided for the data owner.	The feasibility and the utility of the solution is demonstrated by applying it to a case-study from healthcare and real patients. Based on the inference capabilities of the Semantic Web, the user can take an informed decision about the risks and benefits of sharing the personal data. It is a simple and powerful framework for privacy protection in IoT environments.

3.1. Access Control

Studies in this category mainly present authentication and/or authorization mechanisms. However, other features for confidentiality, privacy and provenance control are frequently available. All studies presented some authorization mechanism for health data integration through semantic web technologies [4,16,51,17,18,31,32,34,35,46,47]. However, only a few features have implemented additional features for authentication [16,18,32,34,46,51], confidentiality [16,18], privacy [16–18,35,46,47], and provenance [17].

Rahmouni et. al. presented an ontology-based approach to tackle conflicting privacy and ethical requirements between national regulations in European countries by relying on Semantic Web technologies, such as Semantic Web Rule Language (SWRL) rules for specification and reasoning of access control policies [47]. Beimel and Peleg developed the SitBAC knowledge framework, a formal healthcare-oriented, context-based access-control framework able to represent patient's data-access scenario and perform inferences to either approve or deny access to data, based on Web Ontology Language (OWL), a Description Logics (DL) reasoner and a SWRL engine [4].

Rahmouni et. al. described a mathematical formalism for mapping SWRL privacy rules to standard access control based on eXtensible Access Control Markup Language (XACML) policies to avoid runtime overheads related to the enforcement of SWRL rules on complex and heterogeneous architectures [46]. Sun et. al. defined a Semantic based Access Control model (SAC) for e-Healthcare, which considers semantic relations among different entities in cloud computing environments, with the XACML standard to support description and management of distributed policies [51]. Lu and Sinnott presented a semantic based access control framework to extend the XACML with semantic capabilities to support fine-grained access control and ensure that privacy leakage can be detected and prevented [35].

Li et. al. proposed a multi-layer authorization model that supports specifying and enforcing authorizations for pervasive healthcare delivery using ontologies and semantic web technologies to conceptualize data and explicitly express the relationships among concepts and instances involved in information sharing [31]. Dong et. al. defined a Circle of Care (CoC) ontology that specifies concepts and relations necessary to capture a patient's circle of care and allows one to make inferences about who is in a patient's circle of care and, therefore, can access a patient's health records [17]. Liu and Wang introduced a fine-grained context-aware access model for Health Care and Life Sciences (HCLS) Linked Data [34].

Dridi et. al. developed a platform for the semantization of the Internet of Things (IoT) in the medical and healthcare field, regarding interoperability and integration of heterogeneous data, data visualization and access controls mechanisms [18]. Lima et. al. introduced the implementation of authentication and authorization mechanisms in a semantic Application Programming Interface (API) - an API that produces responses in semantic formats, such as JSON-LD - based on access levels mapped to ontology properties, delivering granular access to a semantic tagged dataset [32]. Dixit et. al. established a framework that enables policy based multi-authority access authorization to Electronic Health Records (HER) systems using the Multi-Authority Attribute Based Encryption (MA-ABE) and Semantic Web technologies to provide a secure, semantically rich approach to facilitate secure data sharing among organizations who manage different attributes of end users using a shared dataset [16].

3.2. Interoperability Infrastructure

Studies have proposed more comprehensive architectures, capable of dealing with several of the security criteria mentioned above. These are generally more complex solutions, but they can support the functional and semantic interoperability of health data within a single framework.

Boniface et. al. presented the ARTEMIS interoperability infrastructure for health information systems based on semantic web services and ontologies to broker between organizational policies [9]. Working as middleware, it can abstract the differences in security requirements (roles, clinical concepts and policies) and capabilities of each system through reasoning. More recently, Tiwari et. al. proposed the Secure Semantic Healthcare (S3HC) framework to represent, integrate and securely exchange data collected by healthcare devices [52]. The authors use an ontology designed for transferring the collected data from the device to the knowledge base and vice versa.

Lima et.al. propose a framework for securing health data in a real case scenario focused on tuberculosis data exchange over the Semantic Web [33]. The solution is flexible and provides several endpoints (SPARQL, GraphQL and APIs) for functional and semantic interoperability of tuberculosis data. The framework is based on hybrid cryptography - a combination of symmetric and asymmetric techniques for encryption and transmission of big data -, hash functions and ontologies.

3.3. Privacy Compliance

The concern with data confidentiality and personal privacy issues have been growing in the last years, driven by new regulations and laws. For patient health information, the Health Insurance Portability and Accountability Act (HIPAA) requires the creation of national standards to safeguard sensitive data from being disclosed without the patient's consent or knowledge [1]. Also, the European General Data Protection Regulation (GDPR) aims to protect natural persons with regard to the processing of personal data [20]. SW technologies can underpin existing solutions to comply with these legal requirements in distinct contexts through their mapping and modelling.

Rahmouni et. al. presented an ontology-based approach for decision support regarding privacy in the sharing of patient data across European platforms through a SW application that can obtain privacy management guidelines for different entities in European countries involved in a data sharing process [10]. Dao T.T. et.al. presented an approach to protect medical information using an asymmetric encryption algorithm with public and private keys, providing confidentiality for a semantic search engine to obtain Human Musculoskeletal System Resources (HMSR) information [14]. Celdrán et. al. proposed the h-MAS tool, a privacy-preserving multi context-aware solution, which allows users to choose profiles (e.g., privacy policies) to manage when, where, how, and to whom their private information can be revealed [29].

Noor et. al. defined an Ontology for Detection of Attack on Genomic data (DAG) using semantic web technologies and a knowledge base of threats [43]. Relying on domain ontologies, Can and Yilmazer introduced a privacy-aware provenance management model to detect privacy violations and query provenance data, enabling traceability of historical data [11]. Lastly, Aljara et.al. developed an integrated solution for users (data owners) of IoT applications to enhance privacy protection in events of private data sharing with a data consumer by calculating privacy risks associated with that specific sharing and comparing them to the benefits to-be received, providing a list of risks and recommendations to allow the user to take a pragmatic and informed decision [2].

4. Discussion

In the articles in the Access Control category, although some works involve privacy control, this is done in the context of a broader mechanism that, through access policies based on rules and semantics, allows inferring whether a given agent has the necessary access rights, thus guaranteeing the privacy of individuals. On the other hand, articles in the Privacy Compliance category propose approaches focused on guaranteeing privacy by mapping privacy requirements (e.g., laws and regulations) through models and ontologies, and they do not present a complete set of tools for access control. The studies classified as Interoperability Infrastructure are more comprehensive as they provide tools that span the previous two categories through a complete infrastructure for functional and semantic health data interoperability.

In health, protecting data confidentiality is crucial. In Dao et.al [14], asymmetric encryption is performed, which deals with the problem of safely distributing the decryption key. Although the authors did not define access control mechanisms, they have demonstrated that using cryptography in sensitive contexts is an efficient way to ensure confidentiality and protect data from non-authorized readers, because only those in the possession of a decryption key are able to read the message. However, it is not clear how the authors deal with the amount of data that can be encrypted, due to limitations in the asymmetric cryptography [36], which may be a concerning factor to deal with big data.

Ideally, confidentiality, privacy and provenance mechanisms should coexist in favour of an in-depth privacy compliance solution for health data handling. However, no study has presented a solution that considers all these mechanisms together.

Several studies share the same technologies to deliver their solutions, including OWL/ontologies and SWRL rules [2,6,28,37,45,10,12,13,15,19,21,26,27], RDF [17,52], SPARQL [17,29,52], XACML [34,35,46,51], and Internet of Things (IoT) devices [2,18,52]. It demonstrates the flexibility of SW tools to allow the implementation of security mechanisms to protect sensitive data and still enable the interoperability and integration of such data. The semantic web may act as a bridge for the joint use of semantic technologies with classical ones, such as web services/APIs, XACML, and MA-ABE. Furthermore, it was observed that the use of these technologies in IoT devices is feasible to allow the traffic of sensitive data (e.g., personal health information), captured by these devices and safely transmitted over the internet.

Finally, the findings of this work show that the main gaps in the literature refer to the absence of a complete computational architecture able to cover all the desired security properties in environments that handle sensitive data through the SW, and the high complexity of implementing existing solutions that, in most of the times, demand non-trivial changes in health information systems that were not initially developed considering the need of semantics.

Although the articles classified in the Interoperability Infrastructure category in this research may touch on those issues, they may represent a disincentive for the use of SW technologies to promote interoperability and semantic integration of health data. However, the development of a plug-and-play solution that could be offered as a service by trusted and secure cloud computing providers, as suggested by Alraja et.al [2], could be a way to simplify the implementation of security mechanisms in SW contexts.

Although not designed for the Semantic Web, industry standards may reduce the impact of implementing security and semantic features. For instance, the Integrating the Healthcare Enterprise (IHE) non-profit organization offers several profiles for security and privacy control (e.g. Document Encryption and Audit Trails) [30], while the Health Level Seven International (HL7) FHIR defines exchange protocols and content models to be used with security protocols (e.g. Digital Signatures and Authentication) [23,24].

The findings of this scoping review satisfactorily answer the research questions previously defined. The main security mechanisms for SW technologies, the key security attributes and properties, and the main gaps in the literature were identified, helping to understand the technical needs to mitigate the risks of handling personal health information over the SW, and, ultimately, enabling the semantic interoperability and integration of such data.

5. Conclusion

In this research, the literature was explored to obtain a broad view of security approaches applied to the SW for electronic health data handling. The findings have shown that complex and robust solutions are available to successfully address several security properties and features, depending on the context that the electronic health data is being managed.

Although the Semantic Web paradigm was coined in 2001, the results suggest that the application and use of SW technologies is still growing, with the healthcare area being particularly interesting due to the SW inference capabilities for records linking and derivation of access policies. However, the complexity of a given solution tends to increase as more SW technologies and tools are incorporated, which can be seen as a disadvantage mainly for existing solutions not initially designed for the SW.

Even though this research was motivated by the healthcare scenario and involved the management of sensitive data, the SW is domain independent, since an adequate basis is available for its implementation (e.g., specific ontologies). The publishing of open linked data through the SW usually does not demand complex security mechanisms, but the security approaches identified in this study may be adapted to other scenarios that intend to use SW with sensitive data.

5.1. Limitations of the study

This study presents some limitations. Although a satisfactory number of articles were selected, the inclusion and exclusion criteria restricted the scope and the possible applicability of Semantic Web technologies and the associated security mechanisms, due to the initial interest of seeking for solutions applied into the health field. Therefore, future work can include research conducted in other areas, as well as to carry out comparison between them.

Additionally, no quality appraisal was performed, which can result in the inclusion of low-quality papers.

Acknowledgements

This study was supported by the São Paulo Research Foundation (FAPESP) [grant number 2020/01975-9].

Disclosure of any conflict of interest

The authors declare no conflicts of interests.

References

[1] 104th United States Congress, T., HIPAA. Health Insurance Portability and Accountability Act of 1996., 1996.

[2] Alraja, M.N., H. Barhamgi, A. Rattrout, and M. Barhamgi, An integrated framework for privacy protection in IoT — Applied to smart healthcare, *Comput. Electr. Eng.* 91 (2021) 107060. doi:10.1016/j.compeleceng.2021.107060.

[3] Arksey, H., and L. O'Malley, Scoping studies: Towards a

methodological framework, *Int. J. Soc. Res. Methodol. Theory Pract.* 8 (2005) 19–32. doi:10.1080/1364557032000119616.

[4] Beimel, D., and M. Peleg, Using OWL and SWRL to represent and reason with situation-based access control policies, *Data Knowl. Eng.* 70 (2011) 596–615. doi:10.1016/j.datak.2011.03.006.

[5] Berners-Lee, T., J. Hendler, and O. Lassila, The Semantic Web. A new form of Web content that is meaningful to computers will unleash a revolution of new possibilities, *Sci. Am.* 284 (2001) 34–43. doi:10.1038/scientificamerican0501-34.

[6] Bhandary, M., M. Parmar, and D. Ambawade, A Blockchain Solution based on Directed Acyclic Graph for IoT Data Security using IoTA Tangle, *Proc. Fifth Int. Conf. Commun. Electron. Syst. (ICCES 2020)*. (2020) 827–832.

[7] Biennier, F., and J. Favrel, Collaborative business and data privacy: Toward a cyber-control?, *Comput. Ind.* 56 (2005) 361–370. doi:10.1016/j.compind.2005.01.004.

[8] Bodin, L.D., L.A. Gordon, and M.P. Loeb, Evaluating information security investments using the analytic hierarchy process, *Commun. ACM*. 48 (2005) 78–83. doi:10.1145/1042091.1042094.

[9] Boniface, M., and P. Wilken, ARTEMIS: Towards a secure interoperability infrastructure for healthcare information systems, *Stud. Health Technol. Inform.* 112 (2005) 181–189.

[10] Boussi Rahmouni, H., T. Solomonides, M. Casassa Mont, S. Shiu, and M. Rahmouni, A model-driven privacy compliance decision support for medical data sharing in Europe, *Methods Inf. Med.* 50 (2011) 326–336. doi:10.3414/ME10-01-0075.

[11] Can, O., and D. Yilmazer, Improving privacy in health care with an ontology-based provenance management system, *Expert Syst.* 37 (2020) 1–18. doi:10.1111/exsy.12427.

[12] Chen, F.C., and C.L. Tsai, A light fingertip touch reduces postural sway in children with autism spectrum disorders, *Gait Posture*. 43 (2016) 137–140. doi:10.1016/j.gaitpost.2015.09.012.

[13] D'Ambrosio, L., M. Tadolini, R. Centis, J.D. Chalmers, and G.B. Migliori, A new free-cost e-service supporting clinicians to manage their difficult-to treat tuberculosis cases: The European Respiratory Society-World Health Organization tuberculosis Consilium, (2017). doi:10.4103/ijmr.IJMR_37_17.

[14] Dao, T.T., T.N. Hoang, X.H. Ta, and M.C. Ho Ba Tho, Knowledge-based personalized search engine for the Web-based Human Musculoskeletal System Resources (HMSR) in biomechanics, *J. Biomed. Inform.* 46 (2013) 160–173. doi:10.1016/j.jbi.2012.11.001.

[15] Davies, J., J. Welch, D. Milward, and S. Harris, A formal, scalable approach to semantic interoperability, *Sci. Comput. Program.* 192 (2020) 102426. doi:10.1016/j.scico.2020.102426.

[16] Dixit, S., K.P. Joshi, and S. Geol Choi, Multi Authority Access Control in a Cloud EHR System with MA-ABE, *Proc. - 2019 IEEE Int. Conf. Edge Comput. EDGE 2019 - Part 2019 IEEE World Congr. Serv.* (2019) 107–109. doi:10.1109/EDGE.2019.00032.

[17] Dong, X., R. Samavi, and T. Topaloglou, COC: An ontology for capturing semantics of circle of care, *Procedia Comput. Sci.* 63 (2015) 589–594. doi:10.1016/j.procs.2015.08.389.

[18] Dridi, A., S. Sassi, and S. Faiz, Towards a semantic medical internet of things, *Proc. IEEE/ACS Int. Conf. Comput. Syst. Appl. AICCSA*. 2017-Octob (2018) 1421–1428. doi:10.1109/AICCSA.2017.194.

[19] Edafe, A., O.O. Mumini, and S. Oluwarotimi, A Genetic-Neuro-Fuzzy Inferential Technique for Diagnosis of Tuberculosis, in: Proc. 2015 Work. Pervasive Wirel. Healthc., ACM, New York, NY, USA, 2015: pp. 39–44. doi:10.1145/2757290.2757299.

[20] EUROPEAN PARLIAMENT, T., and T. COUNCIL OF THE EUROPEAN UNION, EU General Data Protection Regulation, 2016.

[21] Gerald, L.B., S. Tang, F. Bruce, D. Redden, M.E. Kimerling, N. Brook, N. Dunlap, and W.C. Bailey, A decision tree for tuberculosis contact investigation, *Am. J. Respir. Crit. Care Med.* 166 (2002) 1122–1127. doi:10.1164/rccm.200202-124OC.

[22] Grodzinsky, F.S., and H.T. Tavani, P2P networks and the Verizon v. RIAA case: Implications for personal privacy and intellectual property, *Ethics Inf. Technol.* 7 (2005) 243–250. doi:10.1007/s10676-006-0012-4.

[23] Health Level Seven International, HL7 FHIR Specification, *HL7 FHIR Specif.* (2019). http://hl7.org/fhir/%0Ahttp://hl7.org/implement/standards/fhir/.

[24] Health Level Seven International, HL7 FHIR Security, *HL7 FHIR Secur.* (2019). https://www.hl7.org/fhir/security.html (accessed April 19, 2021).

[25] HIMSS, H.I. and M.S.S., Definition of Interoperability, (2013) 2013.

[26] Hossain, M.S., F. Ahmed, Fatema-Tuj-Johora, and K. Andersson, A Belief Rule Based Expert System to Assess Tuberculosis under Uncertainty, *J. Med. Syst.* 41 (2017). doi:10.1007/s10916-017-0685-8.

[27] Howles, T., C.J. Romanowski, S. Mishra, and R.K. Raj, A Holistic, Modular Approach to Infuse Cybersecurity into Undergraduate Computing Degree Programs, *6th Annu. Symp. Inf. Assur.* (2011) 67–70. http://www.albany.edu/iasymposium/proceedings/2011/ASIA11Proceedings.pdf#page=76.

[28] Hripcsak, G., C.A. Knirsch, N.L. Jain, R.C. Stazesky, A. Pablos-mendez, and T. Fulmer, A Health Information Network for Managing Innercity Tuberculosis : Bridging Clinical Care , Public Health , and Home Care Nurse Service of New York — teamed up to create the information infrastructure necessary to provide coordinated , effective care to p, 76 (1999) 67–76.

[29] Huertas Celdrán, A., M. Gil Pérez, F.J. García Clemente, and G. Martínez Pérez, Preserving patients' privacy in health scenarios through a multicontext-aware system, *Ann. Des Telecommun. Telecommun.* 72 (2017) 577–587. doi:10.1007/s12243-017-0582-7.

[30] IHE, IHE ITI TF-1 (rev 17), *Int. J. Healthc. Technol. Manag.* 1 (2020) 1–177.

[31] Li, Z., C.H. Chu, and W. Yao, A semantic authorization model for pervasive healthcare, *J. Netw. Comput. Appl.* 38 (2014) 76–87. doi:10.1016/j.jnca.2013.06.006.

[32] Lima, V.C., D. Alves, F.C. Pellison, V.T. Yoshiura, N.Y. Crepaldi, and R.P.C.L. Rijo, Establishment of access levels for health sensitive data exchange through semantic web, *Procedia Comput. Sci.* 138 (2018) 191–196. doi:10.1016/j.procs.2018.10.027.

[33] Lima, V.C., F.C. Pellison, F.A. Bernardi, D. Alves, R. Pedro, and C. Lopes, Security Framework for Tuberculosis Health Data Interoperability Through the Semantic Web, *Int. J. Web Portals.* 13 (2021) 36–57.

doi:10.4018/IJWP.2021070103.

[34] Liu, Z., and J. Wang, A fine-grained context-aware access control model for health care and life science linked data, *Multimed. Tools Appl.* 75 (2016) 14263–14280. doi:10.1007/s11042-016-3269-6.

[35] Lu, Y., and R.O. Sinnott, Semantic privacy-preserving framework for electronic health record linkage, *Telemat. Informatics*. 35 (2018) 737–752. doi:10.1016/j.tele.2017.06.007.

[36] Mamun, A. Al, K. Salah, S. Al-maadeed, and T.R. Sheltami, BigCrypt for Big Data Encryption, (2017) 93–99.

[37] Mertz, L., (Block) Chain Reaction: A Blockchain Revolution Sweeps into Health Care, Offering the Possibility for a Much-Needed Data Solution, *IEEE Pulse*. 9 (2018) 4–7. doi:10.1109/MPUL.2018.2814879.

[38] Moher, D., A. Liberati, J. Tetzlaff, and D.G. Altman, Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement (Reprinted from Annals of Internal Medicine), *Ann. Intern. Med.* 151 (2009) 264–269. doi:10.1371/journal.pmed.1000097.

[39] Munn, Z., M.D.J. Peters, C. Stern, C. Tufanaru, A. McArthur, and E. Aromataris, Systematic review or scoping review? Guidance for authors when choosing between a systematic or scoping review approach, *BMC Med. Res. Methodol.* 18 (2018) 1–7. doi:10.1186/s12874-018-0611-x.

[40] Nass, S.J., L.A. Levit, and O.L. Gostin, Beyond the HIPAA Privacy Rule: Enhancing Privacy, Improving Health Through Research., Washington, D.C., 2009. http://www.ncbi.nlm.nih.gov/books/NBK9571/ accessed 01/08/2016.

[41] National Institute of Standards and Technolgy, NIST Privacy Framework - a tool for improving privacy through enterprise risk management, 2020. doi:10.6028/NIST.CSWP.01162020.

[42] National Institute of Standards and Technology, Minimum Security Requirements for Federal Information and Information Systems, 2006. http://csrc.nist.gov/publications/fips/fips200/FIPS-200-final-march.pdf.

[43] Noor, S., M. Ahmed, M.N. Saqib, M. Abdullah-Al-Wadud, M.S. Islam, and Fazal-E-Amin, Ontology for attack detection: Semantic-based approach for genomic data security, *J. Med. Imaging Heal. Informatics*. 7 (2017) 1309–1323. doi:10.1166/jmihi.2017.2221.

[44] Peters, M.D.J., C.M. Godfrey, H. Khalil, P. McInerney, D. Parker, and C.B. Soares, Guidance for conducting systematic scoping reviews, *Int. J. Evid. Based. Healthc.* 13 (2015) 141–146. doi:10.1097/XEB.0000000000000050.

[45] Peterson, K., R. Deeduvanu, P. Kanjamala, and K. Boles, A Blockchain-Based Approach to Health Information Exchange Networks, *Mayo Clin.* (2016) 10. doi:10.1016/j.procs.2015.08.363.

[46] Rahmouni, H.B., M.C. Mont, K. Munir, and T. Solomonides, A SWRL bridge to XACML for clouds privacy compliant policies, *CLOSER 2014 - Proc. 4th Int. Conf. Cloud Comput. Serv. Sci.* (2014) 27–37. doi:10.5220/0004853900270037.

[47] Rahmouni, H.B., T. Solomonides, M.C. Mont, and Simon Shiu, Privacy compliance and enforcement on European healthgrids: An approach through ontology, *Philos. Trans. R. Soc. A Math. Phys. Eng. Sci.* 368 (2010) 4057–4072. doi:10.1098/rsta.2010.0169.

[48] Raman, P., H.G.H. Kayacık, and A. Somayaji, Understanding Data Leak Prevention, *Annu. Symp. Inf. Assur.* 2016 (2011) 27–31. doi:10.1109/IConAC.2015.7313979.

[49] Robu, I., V. Robu, and B. Thirion, An introduction to the Semantic Web for health sciences librarians., *J. Med. Libr. Assoc.* 94 (2006) 198–205.

[50] Stouffer, K., V. Pillitteri, S. Lightman, M. Abrams, and A. Hahn, Guide to Industrial Control Systems (ICS) Security, 2015. doi:10.6028/NIST.SP.800-82r2.

[51] Sun, L., J. Yong, and J. Soar, Access control management for e-Healthcare in cloud environment, *ICST Trans. Scalable Inf. Syst.* 1 (2014) e3. doi:10.4108/sis.1.2.e3.

[52] Tiwari, S.M., S. Jain, A. Abraham, and S. Shandilya, Secure semantic smart healthcare (S3HC), *J. Web Eng.* 17 (2019) 617–646. doi:10.13052/jwe1540-9589.1782.

[53] Zenuni, X., B. Raufi, F. Ismaili, and J. Ajdari, State of the Art of Semantic Web for Healthcare, *Procedia - Soc. Behav. Sci.* 195 (2015) 1990–1998. doi:10.1016/j.sbspro.2015.06.213.

[54] Zissis, D., and D. Lekkas, Addressing cloud computing security issues, *Futur. Gener. Comput. Syst.* 28 (2012) 583–592. doi:10.1016/j.future.2010.12.006.