Privacy, Security and Policies of the Semantic Web: A Review

Sana Al Azwari
Dept. Information Technology, Taif University, Taif, Saudi Arabia
Email: alazwari.s@tu.edu.sa

***Abstract*—Issues of privacy, security of the semantic Web are interrelated as they contribute to the general usability of the semantic Web. Issues regarding privacy have shown lack in coverage over emergent technologies which create a need to support the new technologies and incorporate them into improved privacy policies. This article reviews the privacy, security and policies of the semantic Web with the aim to analyze each entity and identify problems and potential solutions that lie within each domain. The research methodology applied in collection and analysis of information follows the PRISMA methodology which defines the steps needed for qualitative collection of articles that are relevant to the research. Key findings from the research reveals the current policies implemented in the semantic Web and the gaps that to be filled. Finally, the article gives recommendation based on different research to improve the semantic Web. A review of the privacy, security and policies of the semantic Web reveals the current policies and how they suit the semantic Web and recommendations on how to improve the policies to cover more areas.**

***Index Terms*—security, privacy, policy, access control, malware**

I. Introduction

Standards set by the World Wide Web consortium (W3C) seek to expand the vast world internet's capability to be data machine-readable. With this consideration in mind, the semantic Web creates the opportunity for the information to be in a machine-readable format. However, a challenge arises in setting the semantics with data. To solve this problem, the principles of metadata are applied in technologies such as the Resource Description Framework (RDF) and the Web Ontology Language (OWL) [1]. Previous work on security and privacy policy issues have focused on proper handling of the information that is produced within the semantic Web. However, the gaps on the existing knowledge fail to emphasize on the challenges of disruptive technology on existing security and privacy policies [1]. This study aims to address the potential challenges and solutions to changes introduced by the growing technology while building up on the existing knowledge. A critical analysis of the privacy, security, and data policies concerning the semantic Web shows that the semantic Web can support information integration by observing the regulations and policies.

Manuscript received July 16, 2021; revised December 2, 2021.

The semantic Web technologies' primary aim is to simplify distributing, sharing, and safeguarding knowledge across multiple points in the worldwide Web. The semantic Web can further enhance existing data, privacy, and security policies governing the Worldwide Web (WWW). The semantic Web can improve the policies by intelligently and flexibly handling privacy and security issues [2]. For example, the semantic Web can support information thinking and make sense of information through its robust capabilities [1]. The authors in [3] further notice that apart from constructing details for extraction of information from the semantic Web, it is essential to capitalize on security of the digital infrastructure to ensure a fundamental and dynamic semantic Web is created. The policy terms allow description of policies in deontic concepts and a distributed security control infrastructure. Furthermore, the semantic Web principles allow deeper exploration into topics such as; malware detection, fraud detection, and data validation, which have long been swept under the table.

Despite the fact the discoveries in the field of semantic Web and Linked Data create an opportunity for further privacy and security issues to arise, solutions on how to face the emergent issues can be drawn from the semantic Web. The unique ability of semantic Web technologies is to create models that can be simulated to show how they react in real-world applications. The models can be analyzed and tailored to a particular problem and provide innovative solutions to solve them [4]. For example, the semantic Web can be used to create accurate models simulating the data security environment. Therefore, the developed model can detect security issues that can be built using semantic analysis of the information. Furthermore, the significant analysis of the information from the created data model can empower internet users to control their interactions with the Web [1]. The resulting ability given to Web users implies that they can manage their privacy while online and reduce privacy issues in the process.

The topics of privacy, security, and correct handling of the information-related policies cut across the technological world but have been neglected in semantic Web technologies. Recent research on the semantic Web technologies and linked Data domain has primarily focused on enabling the sharing of open datasets. For example, recent research has been done by LL Wang on the novel viruses disease to attempt an available research

doi: 10.12720/jait.13.1.85-94

dataset [4]. The study aims to facilitate text mining and information retrieval systems over the massive collection of metadata. Such research shows the strides taken to push the abilities of the semantic Web much further. However, as the semantic Web technologies and concepts continue to gain ground in issues that deal with sensitive information and applications in an industrial context, there is an inherent need to analyze the potential privacy and security problems that might result from the discoveries of the various researches being undertaken.

Previous work on the privacy, security and policy issues explores the current measures that have been put to govern the semantic Web. For example, the authors in [4] focused on the importance of security considerations in achieving a secure communication network between computers. The research sheds light on cyber security challenges and measures that have been put in place to oversee a secure semantic Web infrastructure. Additionally, the research in [5] focuses on the need for more privacy and confidentiality of personal information. The study recommends necessary steps to enforce privacy and cede more control to the user to establish trust. Despite the numerous existing articles on the privacy, security and policies of the semantic Web, there are still gaps that have not been covered. For example, the articles do not incorporate the changes that have taken in the technological sector [6]. Disruptive technology has created new issues in the semantic Web that were not covered by previous policies. The purpose of this research study is to investigate existing literature on the privacy and security of the semantic Web while trying to recommend new regulations that include current changes in the semantic Web technology.

II. Research Methodology

The research methodology used for this study was qualitative research. The qualitative research required collection of data through searching previous literary work on issues regarding privacy, policies and security. The data collection method used was through the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) methodology where each stage follows a specific guide for information review [7]. The PRISMA framework allows for qualitative analysis of the collected information and ascertaining the relevancy of the articles to the research question [8]. The qualitative analysis process is therefore improved since there room for systemic reviews of the collected data. A process analysis of the synthesized information follows the data analysis process. The first step is to analyze the data against the research question to see if it provides relevant information. The second step involves analysis of the data and further interpretation of the results.

The underlying guidelines and principles of the PRISMA frameworks provide a step-by-step approach on how to apply for systemic reviews on data gathered from qualitative research. The advantage of applying this method is that it allows for transparent communication of the results and the criteria used in selecting the relevant material necessary for the research. The role of performing systemic reviews in literature is to synthesize the collected data in a way that communicates the discovered insights [9]. The first step of applying the PRISMA methodology in this research was to search quality information from online libraries and databases that can be used to inform the study. For instance, the systemic review for this study applied data searched from online databases such as JSTOR and EBSCO. The primary reason for using this database in the research was their ability to provide a pool of quality information on the semantic Web, which helped in providing an extensive review of the semantic Web taxonomies. Additionally, these online databases provided search fields which allowed the refinement of data to specific criteria relevant to the study [10]. The search fields were advantageous because they narrowed down information to pertinent only topics. For example, searching the term “semantic Web taxonomies” provided relevant qualitative research applied in this study. Furthermore, the online databases provided features such as drop-down boxes that allowed a combination of significant terms for a more specific search of information.

The inclusion-exclusion criteria were essential in creating an effective search statement that eased searching for relevant information that was up to date. The requirements give room to conduct a customized search for literature. For example, the inclusion-exclusion standards for this research searching materials were dated from 2014 to ensure the relevancy of the material to the study. Additionally, information was searched on topics on privacy, semantic Web, and security. The inclusion of the relevant keywords ensured that the search query gives as many pertinent articles to the study as possible [11]. Based on the inclusion-exclusion criteria search, 50 records were identified through database searching, while additional 10 sources were obtained from other sources. The next step included removing duplicate records in which a further 20 articles were released. The remaining 40 records were screened, which led to the exclusion of 5 journals based on their relevancy. Finally, after assessing the remaining articles for eligibility, only 20 articles were applied to review the semantic Web. The application of the inclusion-exclusion criteria speeds up the searching process by the quick provision of relevant material.

III. Taxonomies of Privacy, Security and Policies of the Semantic Web

There exists a complex relationship between privacy, security, and data policies that might not be obvious from the first glimpse. Even though the three fields are interrelated, each is complex and multidisciplinary in its own right. Each category of the issues and policies tends to represent a wide array of challenges that result from the use of semantic Web technologies and the possible solutions that can be created to solve the resulting issues [5]. By separating the aspects of the semantic Web into an independent unit, it is possible to classify the elements into taxonomies that allow more profound analysis and synthesis of each facet and the possible solutions derived from assessing emergent issues.

A. Taxonomy of Privacy

Privacy seems to be an ambiguous and polysemic term that is often context-based, meaning it cannot be simplified to a less complex concept. There is a need to focus on the existing privacy threats and issues that enable a proper analysis of the privacy policies [12]. Privacy policies are expected to educate clients about the assortment and utilization of their information by sites, portable applications, and different administrations or apparatuses they associate with [13]. This likewise incorporates educating clients about any decisions they may have concerned such information rehearses [6]. Be that as it may, barely any clients read these frequently extended protection arrangements. The individuals who do experience issues getting them since they are written in tangled and vague language. A promising way to deal with assistance defeat the present circumstance spins around semi-consequently commenting on strategies, utilizing blends of publicly supporting AI, and standard language handling [11]. Recent annotated privacy policies have been created about the semantic Web. For instance, annotated policies such as PrivOnto have been designed to represent this concept [14].

1) Protecting user privacy

The issue of privacy has been given pronounced importance in countries all over the world. Nearly every country globally has formulated statutes and regulations that seek to uphold the virtue of privacy. Furthermore, most countries have integrated privacy as a crucial part of the constitution and fundamental right to every citizen. For example, the Supreme Court in the United States concluded that the Fourth amendment has the power to protect against government-related searches if a person has a reasonable expectation for privacy [6]. Such laws provide clear evidence of the importance of confidentiality in governing technology and its use. Other countries in the world have also protected privacy by including it in their respective constitutions. For instance, Brazil declares that confidentiality and the right to a private life are non-negotiable, meaning privacy is paramount [11]. Therefore, the concept of privacy covers a wide range of issues that governments and other authoritative bodies protect aspects relating to privacy.

The theoretical framework of confidentiality should be analysed using a bottom-up structure to grasp the concept of privacy [6]. The bottom-up design is essential because it removes the possibility of abstractness while defining privacy in context with the semantic Web technologies. The idea of confidentiality should identify the perspectives that people hold on what privacy means to them and what they deem private information [6]. However, the theories describing the concepts of privacy have to less variable and contingent to avoid having a short lifespan of being applicable [6]. The taxonomy of confidentiality's primary focus is to uncover the intricate complexities of privacy in a very consistent way. By failing to approach privacy consistently, the analysis risks being a discordant mess, leading to wrong interpretations of the semantic Web.

The taxonomy of privacy gives a theoretical framework that allows for understanding the multifaceted nature of privacy. The foundation for the taxonomical classification of confidentiality is based on processes that infringe on privacy. Taxonomy is essential since it allows for identifying and analysing the varying and recognized privacy violations in the semantic Web technologies industry. The classification of confidentiality in the semantic Web classifies privacy in terms of four broad classes that are varying in terms of the underlying concept [6]. Each of the identified ideas allows the researchers to view privacy in different perspectives and contexts applicable to the particular case of concern. The first categorization of privacy focuses on information collection, whereby issues regarding surveillance and interrogation are considered. Other classification groups include the information processes in the semantic Web domain. Various information processing procedures are investigated to find the problems that compromise privacy. For instance, information processing focuses on the aggregation, identification and exclusion of information for privacy purposes [11]. The taxonomical classifications of privacy provide the appropriate framework to navigate the multidisciplinary field properly.

In today's world, privacy has become a significant concern to all parties that use the semantic Web. Managing privacy policies and knowing sensitive information has slowly turned into a de facto practice in the technology domain, especially in semantic Web technology. The growing importance of privacy has been evident in other parts of the world, such as Europe. For example, in Europe since May 25th, regulations about data protection have been formed; such limitations include the General Data Protection Regulation that came into force to enforce privacy policies [15]. The general data protection regulation is a statute that the vendors using the semantic Web should comply with to assure the users that their private information is kept safe and further assurances of their privacy. However, there have been challenges in holding vendors accountable and confirming if they follow regulations. Challenges in upholding the user's privacy rights and storing their data have prompted the need for interoperability.

Given the rising number of transactions and communications made over the semantic Web in recent years, the level of privacy concern has risen. The increasing level of trust concerns has reached the point where various worldwide Web and concerned citizens have begun taking action. The semantic Web's network architecture is often very complex, relying on a complex algorithm to perform tasks on the Web [12]. The communications channels used are usually multiplex in nature, taking into account multiple users and processing massive data per second. Big data algorithms' sheer number of processing introduces another level of complexity to the semantic Web architecture, which leads to recent privacy issues. For this kind of technology to fully develop, there needs to be a deep level of trust between the technology producers and the targeted users [15]. The eventual erosion of this confidence level can lead to disasters, such as the digital economy and the Web itself. The onus is on the stakeholders to portray transparency to

the users to develop trust. The clarity can be achieved by implementing privacy policies.

Due to the fast pace at which technological advancements have been made in the semantic Web, it is safe to say that there is a lack of required techniques and practices to provide interoperable privacy controls. Designing the privacy controls covers a comprehensive scope that has not been covered extensively due to the lack of proper semantic Web practices. Current privacy control policies have covered more minor aspects covering a smaller spectrum; these aspects include permission granting and tracking protection [12]. However, the covered ranges do not directly relate with the more extensive taxonomy of privacy due to the inability of existing technologies and policies lacking interoperability. Further arguments have been that their standard vocabularies used in the semantic Web are lacking. For instance, Polleri argues that currently, there is a shortage of tongues that explain and exchange personal information, which is crucial in supporting the user's right to personal privacy [15]. Standards have to be set that describe the processes undertaken to exchange information to ensure more current aspects such as interoperability have been factored in.

For proper implementation of the privacy control measures that govern the semantic Web, a consensus has to be reached on the standards to be used to manage privacy. As seen earlier, the previous policies to cover privacy are not well extensive since the emergent technologies have created new issues that need to be included. It is important to have an agreed uniform code across all semantic Web users [16]. The uniformity allows every user to abide by the same policies, ensuring no loopholes in privacy. Efforts have been made to find a final agreement of the standards regarding confidentiality. For instance, forty experts participated in the W3C workshop on data privacy controls and Vocabularies in Vienna on the 17th and 18th of April 2018 [15]. The workshop's main agenda was to tackle problems related to privacy in the modern semantic Web environment. Such efforts by recognized authoritative bodies show why privacy is an important consideration in the semantic Web architecture and, therefore, an important consideration.

Interoperability of systems in the semantic Web is an important consideration in more recent technology since it helps solve market concentration problems. The advantage of interoperability is that it allows users to have control of the presence in a platform. By controlling their online presence, users can protect their sensitive information due ceding of total control by technology vendors [5]. Different market concentration cases have been witnessed in previous periods, which has led to privacy rights violations. Social media giants Facebook have been at the forefront of many news outlets for all the wrong reasons. The social media giants have been constantly accused of spying on their users and violating their privacy rights. Such actions have led to a social and ethical backlash that has since seen the company face multiple lawsuits. Facebook's case is not a stand-alone as numerous other vendors have come under the spotlight for violating the privacy policies they are obligated to uphold. Introducing the aspect of interoperability helps in reducing such cases and improving privacy across the semantic Web platforms.

Access control is one of the significant issues associated with the privacy of the semantic Web. Research reveals that a good number of access control issues originate from the use of outdated technology apparatus such as the keycards, which, when it lands in the wrong hands, can compromise the privacy of the semantic Web due to unauthorized access to accounts. Additionally, access control issues originate from a lack of integration with the central system infrastructure, which creates a gap that malicious parties can take advantage of. Therefore, it is vital to provide an effective access control strategy to ensure such privacy issues are safeguarded. With significant developments in technology such as the cloud technology, privacy and access control issues have become more persistent since there is a lack of a framework to support the emergent technology and provide necessary guidelines to prevent access control issues. Password management is considered a critical issue that affects the privacy of the semantic Web. Poor management is considered to be one of the main contributors to privacy breaches. Many people typically use weak passwords that provide an easy time to crack, making access control significant privacy issues in the semantic Web. Therefore, modern solutions need to be implemented to prevent similar issues that might be avoided.

2) Solutions to privacy challenges

Standard solutions have to be developed to cater to interoperable privacy. The process of created standardized interoperability policies is usually intensive since it covers a wide range of stakeholders. The Institute of Electrical and Electronic Engineers (IEEE) is responsible for overseeing such a process, which helps project a standardized and unified approach to handling the problem. The IEEE is fundamentally responsible for creating the high-level interoperability standards that can be implemented on the semantic Web to improve privacy control measures [5]. Another aspect that should be considered in developing such policies is adversarial interoperability. Adversarial interoperability applies when a new business offers a product or service that works with other vendors' existing products. However, there are challenges to implementing the principles advocated by interoperability. The challenges arise because interoperability is not developed equally. The inequality creates a loophole that allows companies to subvert the statutes that will enable intercommunication, ultimately compromising privacy [12]. Alternatives should be investigated on how to standardize all interoperability policies.

Implementing adversarial interoperability creates a privacy advantage in that it allows the online communities to be self-governed. Self-governance is important since it grants autonomy to online communities. Independence means that sensitive personal information is less susceptible to online hacks, which improves privacy. The semantic Web will have enhanced capabilities once the implementation has been adopted by vendors operating in

the environment [6]. For instance, the software vendors will have the ability to create tools that allow users to hold private conversations in their respective communities. Privacy is enabled through block encryption codes that make it hard for anyone to decipher the message without the special decryption key. The encryption technology will also help prevent personal data mining by the mining bots that have been exponentially on the rise [5]. Legal reforms have to be enacted to change laws that can impede the realization of adversarial interoperability. Previous standards give the most power to the most renowned companies, which disadvantage smaller enterprises since their data is still exposed to leaks and hack, which threatens privacy.

Controls need to establish to cover data privacy vocabularies and rules. Their lack of enough privacy vocabulary means that the less content of the privacy aspect is protected. To create data privacy vocabularies, a concerted effort should be put in place by the recognized bodies to avoid ambiguity in describing privacy controls. The W3C recently held a seminar made up of a community group called 'Data privacy vocabularies and controls CG' (DPVCG) to bring together people for a common cause [5]. The co-unity group's objective the community group is to fit related endeavours and unite partners with advanced recommendations to create vocabularies to empower semantic interoperability and trade of straightforwardness logs about close-to-home information handling, empower information compactness for information subjects, and so forth. The specific degree of utilization cases identified with making individual information preparing interoperable by particular guidelines to ease evidence of consistency with the General Data Protection Regulation (GDPR) and related security insurance guidelines will be the principal deliverable of the community group [12]. The community can create deliverables and steps that can be followed to improve privacy policies.

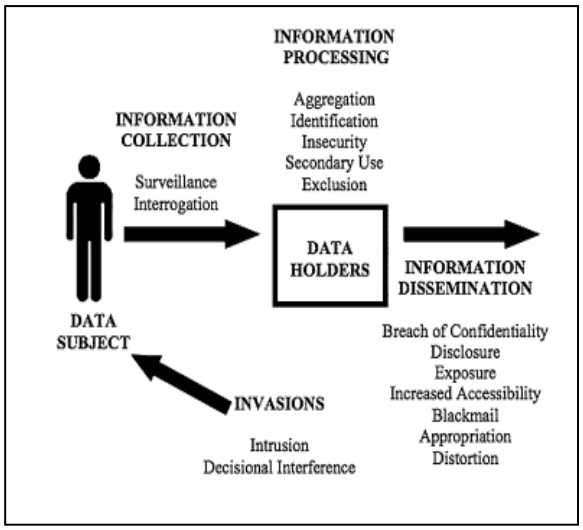

Figure 1. Classification of taxonomies creating privacy problems, from [6].

Fig. 1 shows the classification of taxonomies of the semantic Web and the activities that are undertaken at each stage. The phases include information collection whereby it is processed through aggregation and identification. Each stage presents a potential challenge that needs to be solved. During the information dissemination stage, challenges of privacy such as the breach of confidentiality occurs whereby information might be disclosed to unauthorized parties which can lead to blackmail and distortion. System invasions might compromise the security of the semantic Web through cyber-attacks.

B. Taxonomy of Security

Cybersecurity issues in the semantic Web are directly related to the security issues and threats associated with technological devices such as phones and computers. Cyber-security has many classifications that deal with different aspects of security. For example, organizations such as the software engineering institute and the European Union Agency for Network and Information Security (ENISA) have developed their customized cybersecurity classes. Their classifications have shared similarities since they overlap each other and cover similar topics [16]. These classifications seem to overlap because they seek to solve persistent issues in the cybersecurity field. Different taxonomies have also been created that form a subset of security, allowing in-depth analysis of each subgroup to provide holistic solutions. For instance, the European cybercrime Centre (EUROPOL) emphasizes problems that threaten the semantic Web security architecture [1].

The scope of the EUROPOL classification of security incidents covers areas that affect the overlying security measures of the semantic Web network architecture and the related information systems. The range covers security cases that have a significant impact on the essential digital services that are offered in the semantic Web. They identified many issues that are usually noted and reported to the recognized national authoritative bodies. The obligation to inform the incidents falls under article 14 and Article 16 of the National Intelligence Service (NIS) directive [17]. Furthermore, cybersecurity incidents that have a huge impact on electronic communications have to be reported to the competent national authorities; this statute falls under Article 13a of the framework directive. In the case of security breaches to the network that greatly impacts the breached device's security trust, an early notification has to be given to the supervisory bodies under Article 19 of the EIDAS regulation [16]. These identified scopes created by EUROPOL form an important foundation on which security policies and measures can be based upon.

1) Security challenges

The taxonomy of security is largely divided into two main parts: the security incidents' nature and the impact of a security breach. Classification of security into these two broad spectrums enables analysis of the security breaches cause and effect. Analysing the nature of the security incident that has occurred, analysts can determine the security breach’s trigger. Furthermore, the violation's impact can be measured by looking at the affected sector and its effects since the incident [17]. For instance, when analysing the cause of a security breach in the semantic Web, various factors such as system failure, natural

phenomena, and third-party failures can lead to a security incident. These incidents may compromise the structure of a device that uses the semantic Web as a platform to provide services [17]. Furthermore, the impact of such incidents can be assessed through factors such as the sector impacted, the scale of the effects, and the general outlook of events after the incident. Analysis of security through the core parts allows extensive research to provide insightful solutions.

Identifying the root cause of the security incident is important in knowing the type of event that might have led to the security incident. EUROPOL classifies the root categories of security incidents mutually exclusive in that they all occur through different triggers in the semantic Web. The type of nature of the incident is system failures. System failures occur very often in the semantic Web, whereby various functions that compose the whole system may crash, leading to a system failure [17]. System failures are usually characterized by the trigger event being exclusively internal. Exclusively internal events are only triggered by the system itself and not subject to external interference. For example, system failures may occur through hardware failures and software bugs. These occurrences can be referred to as the trigger of the security event. They may help the cybersecurity experts identify the cause and develop solutions to remedy the security breach [18]. The nature of the root causes is subject to changes over time as more information about the incident is revealed, which sheds further knowledge.

The nature of the incident is further broken down by assessing the severity of the identified threat. Assessment of the severity of the security incident shows the extent to which the system has been affected. Risk assessment is crucial with regards to analysing the severity of the attacks since it informs the necessary measures that can be undertaken to mitigate the risk. For instance, an attack can be classified as of a high impact if it affects a large part of the semantic Web system [17]. High impact attacks may include Denial of Service (DOS) attacks since they flood the servers with fake requesting leading to a system crash. Denial of services attacks is considered one of the most high-impact security attacks in the semantic Web since it takes a while before regaining control of the whole system. Further classification of attacks due to severity includes medium and low impact [19]. Medium severity attacks usually have fewer repercussions on the system than high impact attacks and can be mitigated easily without compromising the main system functionalities.

The semantic Web's type of reasoning mainly depends on the exchange of trusted information between communicating parties. Therefore, Transport Layer Security (TLS) is an essential feature that has been used extensively in providing a secure transfer of data. The goal of the TLS is to use resource identifiers for identifying online users with the primary objective of supporting a decentralized semantic Web network. Each user has the own unique identifier, making it easier to provide a robust security network [10]. Additionally, a person can maintain a record of activities that allow for easier retrieval in future returns. Despite the advantages of using TLS to add an extra layer of security to the semantic Web, there have been criticisms on the security challenges it presents to the semantic Web. For instance, the security framework violates the security limits of the semantic Web. Additionally, the technology is based on outdated technology, implying that it can no longer support modern frameworks hence creating new security issues [20]. The possession of a supercookie allows any person to access unauthorized sites due to its capability to access various sites with the same certificate information. As a result, semantic Web security is compromised because a hacker can access a user's data by merely requesting their Web certificate.

Regardless of whether the client effectively distinguishes themselves with a customer certificate, current programs utilize the unreliable MD5 hash feature in the marked customer certificate. MD5 has demonstrated not to be impact safe, which implies that an aggressor can create a fake customer certificate whose mark can be confirmed utilizing the public key even though the hacker doesn't have the client's private key [10]. Like this, a client can be imitated by a hacker. Due to these security and protection issues, program vendors are now censuring keygen from HTML and customer authentications dealing with the application layer, which will mean WebID+TLS will quit working. Albeit the Semantic Web local area presently cannot seem to draw in mind it, the W3C Web Cryptography API is currently giving current cryptographic functions. Disposing of passwords should be possible through validation by equipment tokens or different authenticators by the W3C Web Authentication API, which is planned to not abuse the equivalent beginning strategy, for example, authentication keys that vary based on their origin current cryptographic natives, for instance, ECDSA [21]. Like the rest of the Web, the Semantic Web can operate unequivocal approval of individual information to move using IETF guidelines like OAuth [21]. These techniques will ensure that the semantic Web can effectively separate identities to prevent compromise of the semantic Web security.

2) Addressing security issues

Various factors should be taken into consideration when assessing the possible severity of an attack. The reviews eliminate any form of biasness or misinformation that might occur during the analysis period. The organization's potential risk should be taken into account to avoid taking huge risks that offer small returns. Various incidents have a varying impact on the semantic Web's security structure, and as such, the cost of fixing each risk should be calculated [19]. The amount needed to fix the incident or to protect the device from an attack is considered when analysing cost structure because it informs the viability of taking a specific solution over a list of other possible solutions that might exist. Furthermore, different types of attacks occur at different speeds and frequencies, which challenge how to face these challenges since each challenge is unique in its own right. Some attacks are considered more aggressive, such as the DOS attacks, which affect the semantic Web's critical services [17].

Cybersecurity experts should give such an aggressive form of aggression more prominence since they carry impactful consequences.

Cybersecurity further analyses the sectors that might be affected by an attack on the semantic Web. Various sectors in the modern world today have an online presence and therefore use services that are provided by the semantic Web. Having services online means that the system is subject to attack at any moment since everything can be accessed through the network. Sectors that might be affected include; energy, transport, communication, and the digital infrastructure sectors [17]. For instance, the energy sector's impact might affect essential services such as oil and gas. Such an attack on this sector would have a more significant global impact due to the wide range of consequences. The digital infrastructure sector is vital regarding the Semantic Web since it provides the platform and necessary architecture to perform end point-to endpoint communication. Attacks that might be carried out include exchange domain systems and interconnection points [22]. Identifying the affected sector is essential in creating a customized set of policies that address the security issues that might pop up in the respective industry.

The cybersecurity group goes on further to recommend a framework for ascertaining the severity level of an attack. Various classifications of the case are important because they give categories that are easy to identify and remedy if an attack occurs. The EUROPOL categorizes the severity of the impacts into red, yellow, green, and white [17]. A red level impact means a significant impact on the semantic Web; a yellow level indicates a slightly lower result than the red level. The green level shows a low impact security breach, while the white level shows no significant effect was detected [22]. The above categories are essential in security classification in the semantic Web since each type has its guidelines on how to resolve issues. However, several factors have to be considered when classifying security instances since some might be more severe than others. For example, in emergency services, the general well-being of the people should be considered. Such considerations are made to safeguard the lives of people.

There is a need to understand how TLS works since it is a network-level configuration, not an application-level technology. For instance, intruding on an organization-level TLS handshake to begin a user-driven identity also, validation convention in WebID+TLS is a terrible plan to the extent that it blends the application-level idea of a user's profile with the network level that sends bits around [23]. Somewhat, the issues with the utilization of TLS on the Semantic Web is that network-level data if an HTTP association is scrambled utilizing TLS is uncovered through the level of URI used in Semantic Web applications, including yet not restricted to WebID+TLS. URIs is additionally presented to HTML connections. Subsequently, research stresses that a change to HTTP disregards the rule that pleasant URIs do not change; thus, the appropriation of HTTPS would break existing connections.

Speculation of the possible outlook of the impact is essential in developing corrective strategies. The security attack outlook can give insightful information to the cybersecurity team, which might help create problems beforehand. Different levels of perspectives can be considered when developing security policies. The outlook of an impact may seem to be improving concerning a specified period [19]. Specifying the period is crucial in analysing the view because it helps gauge the severity of the case and how long it might recover. For instance, when considering the severity of an impact, EUROPOL suggests that an outlook can be classified as improving if the level of effects decreases in the next six hours. The shorter time an impact takes to be less severe, the better the outlook of the breach [17]. A stable outlook means that for the next six hours, it is expected that the severity of the case will remain the same. However, worsening impact only acts to show the system defence mechanisms' deteriorating state hence an increasing seriousness. Issues regarding cyber security should be considered with careful thought in order to reduce risks concerning cyber security.

It is recommended that when it comes to the security of the semantic Web, there should be correct labelling of the technical taxonomies. The ENISA body recently created a reference taxonomy used for Computer Security Incidence Report Team (CSIRTs) and based on the e-CSIRT taxonomy. The labels can be identified as abusive content in which the range of the issues may cause harm to the intended target. For example, vocabulary classified as offensive content includes; spam, hate speech, and hurling insults [17]. The security architecture of the semantic Web faces daily attempts and breach through abusive content that users receive. Labelling the different taxonomies is essential since it enables creating algorithms to filter out such content and improve the semantic Web's security. Furthermore, malicious code can also be interpreted as security issues since it affects the semantic Web's functionalities [17]. The injection of malicious code into the system affects how the services are rendered and might lead to system failures. Categories that fall into the adversarial code classification are the Trojan horses and worms.

It is essential to define the machine tags and namespaces in the technical sector of security. It is necessary to develop the nametags since it makes room for easy integration of the semantic Web security tools. Using nametags makes the content more machine-readable; hence, it is easier to decode and process it [22]. According to [22], it is vital to include the security policies in the early development stages of the semantic Web infrastructure. The development of these policies in the early stages is important for building confidence to promote adaption of recent systems such as the internet of things (IoT). The additional advantage of using namespaces and tags is that it makes it easier to update previous versions of named taxonomical vocabularies [17]. The consistency in the naming conventions proves to be essential because it gives uniformity and simplicity in updating new information. For instance, the technical namespace has made it possible

for a person to fetch automatically updates namespaces such as Github.

C. Taxonomy of Policy

Policies are essential in defining guidelines to be followed when solving issues related to the semantic Web. Policies concerning privacy and security are usually overlapping each other, showing a close relationship between the two concepts of the semantic Web. Tasks need to be formulated that associate with management and observance of the formulated policies regarding sharing knowledge across the semantic Web. It is essential to develop policies that cover the semantic Web's intellectual properties [24]. The guidelines will ensure that the intellectual properties are used for the proper purpose to avoid misappropriation. Intellectual properties include elements such as software and data licenses. Developing taxonomy of privacy is imperative because it forms the basis for the data and software technologies in the semantic Web.

1) Policy considerations

Data-oriented frameworks and applicant the focus of current improvements of the World Wide Web (WWW). Arising undertakings centre their business model on offering some benefit from information assortment, reconciliation, handling, and reallocation. These sorts of frameworks are not recent, as the Web has empowered for a long-time apparatus. For example, news aggregators, which gather articles from different suppliers, republish them as assortments of short readings, frequently zeroing in on explicit subjects such as legislative issues and sport [25]. These days, the extraction, distribution, and reuse of information on the Web is setup training, and countless Application Programming Interfaces (APIs) give admittance to JavaScript Object Notation (JSON) records, information tables, or Linked Data for an assortment of utilization cases, traversing from content and media linkage to science and instruction [25]. The significant aspect of focus is the publication of licenses and terms and conditions associated with Application Programming Interfaces (APIs) and semantic Web architecture.

Data warehouses gather an enormous assortment of information sources and interact with them to execute the work process that associates information in their unique sources to applications that should abuse this information. Proposed frameworks make new difficulties as far as the volume of data to be put away. Furthermore, they require novel handling methods, such as stream-based investigation [25]. However, more critically, the interest in more modern ways to deal with information administration. In the Web of open information, designers can access an enormous assortment of data and frequently distribute the consequences of their handling [20]. Accordingly, they need to realize the use requirements connected to an information source they need to adventure. They need to uphold the request to distribute the right strategies close to the information they disseminate.

Challenges arise when analysing the kind of policies that emanate from the licenses related to a given output that is data intensive. Policies and information streams can be portrayed inside the Semantic Web, depending on principles like the W3C PROV model2 to portray measure executions and the Open Digital Rights Language3, which can be misused for formalization and approval [25]. Primarily, it is conceivable to indicate Policy Propagation Rules (PPR) [16] by partner arrangements with information stream steps, albeit this movement brings about an enormous number of rules to be put away [14]. It is conceivable to pack a PPRs data set by utilizing a metaphysics of the potential relations between information protests, the Data node ontology4, applying the (A)AAAA philosophy, and information designing methodology misses Formal Concept Analysis (FCA).

2) Policy recommendations

Rule-based associations and policies are needed to empower secure information access and utilization of unacceptable conditions, especially in the Semantic Web. The defeasible rationale is utilized to prevail upon deontic articulations, for instance, to check the similarity of licenses or to approve requirements connected to parts on multi-specialist frameworks [24]. The issue of licenses' similarity has been widely concentrated in writing, and devices that can perform such evaluation do exist. Past work presents a type of strategic thinking, to be specific arrangement engendering [4]. A Policy Propagation Rule (PPR) is a Horn statement characterized by a partner a Data node connection with an Open Digital Rights Language (ODRL) strategy. Dissuading Horn rules is a successful managing arrangement method, mainly because Horn rules permit manageable defeasible thinking [11]. While in this article, we shed a spotlight on arrangements spread, PPRs can, on a fundamental level, be incorporated with rule-based reasoners for strategy approval.

Cooperative policy enforcement should be encouraged since it includes both aspects of computer-to-computer communication and human-computer interactions. It is significant to implement this policy since it identifies the information necessary to gain access to the needed resource. Implementing a cooperative policy will help reduce the instance of bias towards the aspect of functionality instead of protection, which is considerably less secure than contextualized policies. Additionally, most users lack a clear understanding of how the policy guidelines are implemented; therefore, they do not realize the risks involved in the breach of these policies [9]. Therefore, there is a need to clearly understand the semantic Web policy so that first-time Web users can have a good experience while interacting with the semantic Web. Cooperative policy enforcement will provide the opportunity to eliminate the negative scenarios that users commonly encounter while using the semantic Web. The negative instances can be provided with counter-responses that do not compromise user confidentiality.

Process executions can be portrayed in the Semantic Web utilizing the Provenance Ontology (PROV-O). PROV-O describes work process executions as specialists, activities, and resources included [25]. The Data node metaphysics has been intended to depict Semantic Web applications by methods for the relations between the information associated with their cycles. The cosmology is

a tax9W3C ODRL Community Group taxonomy of potential ties between information objects, which may be necessary for an interaction execution, such as those depicted with PROV-O [25]. It can along these lines be utilized to additionally qualify the ramifications of the activities acted in such an interaction. Data nodes can depict measure suggestions in an information arranged way, particularly as the organization of information objects. While approaches and interaction executions can be addressed, in the present paper, we target contemplating the path toward thinking upon the spread of systems across an information stream.

IV. Conclusion

By analysing the interrelation between privacy, security, and policies, these concepts overlap each other. The three aspects of the semantic Web share fundamental principles that apply in each domain. The current growth rate of the semantic Web and its related technologies means that it is hard to keep up with innovation's fast pace. However, the problems currently faced in the semantic Web can be solved by the robust formulation of integrated policies. Once implemented, these policies and frameworks ensure consistent improvement in how privacy and security issues are handled. An integrated approach to semantic ways provides holistic solutions that offer end-to-end solutions. The key findings from the research has a profound impact on the global scale because the finding can be used to develop a more robust semantic Web infrastructure that integrates the taxonomies of polices, privacy and security. Major limitations of the study arose from insufficient sample size for statistical measurement. However this research provides more room for further research on how the network security of the semantic Web can be improved for secure communication.

Conflict of Interest

The author declares no conflict of interest.

References

[1] L. Kagal, T. Finin, and A. Joshi., "A policy based approach to security for the semantic web," in *Proc. International Semantic Web Conference*, 2003, pp. 402-418.

[2] P. Pranav, S. Dutta, and S. Chakraborty, "Security issues for the semantic web," in *Web Semantics*, Academic Press, January 2021, pp. 253-267.

[3] K. Joshi, A. Gupta, S. Mittal, C. Pearce, A. Joshi, and T. Finin, "Semantic approach to automating management of big data privacy policies," in *Proc. IEEE International Conference on Big Data*, Washington D.C., 2016, pp. 482-491.

[4] H. Brar and G. Kumar, "Cybercrimes: A proposed taxonomy and challenges," *Journal of Computer Networks and Communications*, January 2018.

[5] C. Doctorow. (2021). Interoperability and privacy: Squaring the circle. Electronic Frontier Foundation. [Online]. Available: https://www.eff.org/deeplinks/2019/08/interoperability-and-privacy-squaring-circle

[6] R. Indra and M. Thangaraj, "An integrated recommender system using semantic web with social tagging system," *International Journal on Semantic Web and Information Systems*, vol. 15, no. 2, pp. 47-67, April 2019.

[7] A. C. Tricco, *et al.*, "PRISMA extension for scoping reviews (PRISMA-ScR): Checklist and explanation," *Annals of Internal Medicine*, vol. 169, no. 7, pp. 467-473, October 1993.

[8] A. Oltramari, *et al.*, "PrivOnto: A semantic framework for the analysis of privacy policies," *Semantic Web*, vol. 9, no. 2, pp. 185-203, January 2018.

[9] W. M. Bramer, M. L. Rethlefsen, J. Kleijnen, and O. H. Franco, "Optimal database combinations for literature searches in systematic reviews: A prospective exploratory study," *Systematic Reviews*, vol. 6, no. 1, pp. 1-12, December 2017.

[10] A. A. Selçuk, "A guide for systematic reviews: PRISMA," *Turkish Archives of Otorhinolaryngology*, vol. 57, no. 1, p. 57, March 2019.

[11] S. Kirrane, S. Villata, and M. D'Aquin, "Privacy, security and policies: A review of problems and solutions with semantic web technologies," *Semantic Web*, vol. 9, no. 2, pp. 153-161, January 2018.

[12] R. Kalaiprasath, R. Elankavi, and R. Udayakumar, "Cloud security and compliance - A semantic approach in end to end security," *International Journal on Smart Sensing & Intelligent Systems*, vol. 10, pp. 482-494, September 2017.

[13] H. J. Pandit, D. O'Sullivan, and D. Lewis, "An ontology design pattern for describing personal data in privacy policies," in *Proc. WOP@ ISWC*, Monterey, California, 2018, pp. 29-39.

[14] S. Niksefat, P. Kaghazgaran, and B. Sadeghiyan, "Privacy issues in intrusion detection systems: A taxonomy, survey and future directions," *Computer Science Review*, vol. 25, pp. 69-78, August 2017.

[15] P. A. Bonatti, *et al.*, "Data privacy vocabularies and controls: Semantic web for transparency and privacy," in *Proc. SW4SG@ ISWC*, Monterey, California, 2018.

[16] A. Hendre and K. P. Joshi, "A semantic approach to cloud security and compliance," in *Proc. IEEE 8th International Conference on Cloud Computing*, 2015, pp. 1081-1084.

[17] C. Group. (2018). Cybersecurity incident taxonomy. CG Publication. [Online]. Available: https://ec.europa.eu/information_society/newsroom/image/document/2018-30/cybersecurity_incident_taxonomy_00CD828C-F851-AFC4-0B1B416696B5F710_53646.pdf

[18] C. Duma, A. Herzog, and N. Shahmehri, "Privacy in the semantic web: What policy languages have to offer," in *Proc. Eighth IEEE International Workshop on Policies for Distributed Systems and Networks*, Bologna, Italy, 2007, pp. 109-118.

[19] I. Agrafiotis, J. R. Nurse, M. Goldsmith, S. Creese, and D. Upton, "A taxonomy of cyber-harms: Defining the impacts of cyber-attacks and understanding how they propagate," *Journal of Cybersecurity*, vol. 4, no. 1, p. tyy006, 2018.

[20] L. Kagal, T. Finin, M. Paolucci, N. Srinivasan, K. Sycara, and G. Denker, "Authorization and privacy for semantic web services," *IEEE Intelligent Systems*, vol. 19, no. 4, pp. 50-56, July 2004.

[21] B. Thuraisingham, "Security standards for the semantic web," *Computer Standards & Interfaces*, vol. 27, no. 3, pp. 257-268, March 2005.

[22] I. Alqassem and S. Davor, "A taxonomy of security and privacy requirements for the Internet of Things (IoT)," in *Proc. IEEE International Conference on Industrial Engineering and Engineering Management*, Malaysia, 2014, pp. 1244-1248.

[23] S. Wilson, *et al.*, "Analyzing privacy policies at scale: From crowdsourcing to automated annotations," *ACM Transactions on the Web*, vol. 13, no. 1, pp. 1-29, December 2018.

[24] T. Meline, "Selecting studies for systemic review: Inclusion and exclusion criteria," *Contemporary Issues in Communication Science and Disorders*, vol. 33, pp. 21-27, March 2006.

[25] E. Daga, A. Gangemi, and E. Motta, "Reasoning with data flows and policy propagation rules," *Semantic Web*, vol. 9, no. 2, pp. 163-183, January 2018.

Sana M. Al Azwari received the BS in computer science from Taif University, Taif, Saudi Arabia, in 2004, and the MS and PhD in information sciences from Strathclyde University, Glasgow, UK, in 2010 and 2016, respectively.

She joined the Information Technology Department, Taif University, Taif, Saudi Arabia, as an assistant professor in 2017. At present, she is the Vice Dean of the college of Computer and Information Technology, Taif University. Her current research interests include data science, big data, machine learning, data mining, ontologies and the semantic Web.
Dr. Al Azwari is an ambassador of Women in Data Science committee and is an international science ambassador for Strathclyde University, Glasgow, UK. She awarded the King Abdulaziz and his Companions Foundation for the Gifted Award, Saudi Arabia, in 2006.