Int. J. Advanced Networking and Applications
Volume: 15 Issue: 04 Pages:100-200 (2023) ISSN: 0975-0290 (Special Issue – ASCIS 2023)

A FAIRification Framework for the Findability, Accessibility, Interoperability and Reusability of Cyber Security Ontologies Using FAIR Data Principles

Tshepiso Larona Mokgetse
Department of Computer Science and Information Systems, Botswana International University of Science and Technology, Palapye, Botswana
Email: tmokgetse@gmail.com /mt21100029@studentmail.biust.ac.bw
Hlomani Hlomani
Department of Computer Science and Information Systems, Botswana International University of Science and Technology, Palapye, Botswana
Email: hlomanihb@biust.ac.bw
Sridaran Rajagopal
Department of Computer Applications, Marwadi University, Rajkot, Gujarat, India
Email: sridaran.rajagopal@marwadieducation.edu.in

----------ABSTRACT----------

This paper proposes a FAIR-inspired framework for integrated cyber security ontologies that offers a systematic approach to applying FAIR data principles in the field of cyber security. The framework is developed based on a combination of adopted research and incorporates the FAIRification process. This FAIRification process is instrumental in guiding the development of cyber security ontology models, which are essential for realizing the principles of Findability, Accessibility, Interoperability, and Reusability. The framework addresses the absence of applied FAIR standards in existing cyber security ontology models, utilizing sampled cyber security models as reference systems for source data. The resulting cyber security ontology models create a virtual data warehouse, integrating data from domain experts and existing models. The FAIRification process emphasizes collaboration among cyber security stakeholders, including private and public entities, analysts, engineers, and administrators, to ensure the meaningful interpretation of data based on cyber security ontologies. Detailed discussions within the framework cover the FAIR principles, outlining specific criteria for Findability, Accessibility, Interoperability, and Reusability. Notably, the framework emphasizes the importance of formal, accessible, shared, and broadly applicable language for knowledge representation in achieving Interoperability. Further components of the framework involve the curation and validation of data, development of cyber security ontology data models, and the creation of ontology metadata models. These models play a crucial role in making cyber security data Findable, Accessible, and Reusable, aligning with the FAIR principles. Additionally, data management systems, including the dataset integration system and resource metadata system, are introduced to facilitate the integration and annotation of ontologized data. The framework's use of a data triplestore, FAIR data, and the grlc server enhances the Accessibility of cyber security data on the web. The triplestore represents knowledge graphs in RDF, allowing for efficient querying through SPARQL. The grlc server further facilitates the generation of web APIs, making cyber security data Accessible in an open, free, and universally implementable manner.

Keywords – **Cyber Security, FAIR, Framework, Ontologies.**

Date of Submission: 08 October 2023 ` Date of Acceptance: 23 November 2023

I. INTRODUCTION

Cyber security threats and attacks have been a revenue gobbler for many businesses globally from time immortal. We have seen these cyber threats evolve in their complexity throughout the years. Recently cyber security attacks like ransomware and cryptojacking have become the new black-market money-making schemes. [1] states that ransomware attacks have gained billions of dollars from their victims each year. Ransomware is where the hacker uses technology that allow them to kidnap the organization or country's computer files then demand for large sums of money as ransom for the release of the files. For example, countries

Int. J. Advanced Networking and Applications
Volume: 15 Issue: 04 Pages:100-200 (2023) ISSN: 0975-0290 (Special Issue – ASCIS 2023)

like Germany, Belgium, and the Netherlands have fallen victim to the ransomware attacks [2]. Researcher Treleaven [3] believes that the current cybercrime explosion affecting the society is caused by emerging technologies and artificial intelligence, in an environment where financial regulators, law enforcement, and institutes are not prepared.
The number of breached records across the world in October 2023 has been 3.8 billion records [4]. This tragic information makes apparent the critical importance of mitigation solutions to cyber security.

The objective of this study is to develop a framework that integrates cyber security ontology models, cyber security

Int. J. Advanced Networking and Applications
Volume: 15 Issue: 04 Pages:100-200 (2023) ISSN: 0975-0290 (Special Issue – ASCIS 2023)

domain knowledge, and incorporates the 'F'- Findability, 'A' – Accessibility, 'I'- Interoperability, and 'R'- Reusability (FAIR) data principles. The ultimate goal of FAIR is to facilitate the reuse of data. To achieve this, different technologies are used to annotate resource and thereby creating metadata ensuring that resources are well-described so that they can be replicated and/or combined in different settings. The framework proposed by the study will facilitate the adoption of the FAIR data principles in the cyber security domain through ontologies. Ontologies are important in increasing the interoperability and machine readability of cyber security datasets.

Ontologies are arguably the best way to manage data, in terms of information integration and knowledge management. They address the "I" - Interoperable part of FAIR and also the "R" – Reusable part of FAIR since by definition they are machine readable formal representation of a domain of discourse [5, 6]. Amid the technologies being used, Ontologies have gained attention in knowledge management [7]. Kaewboonma [7] further adds that for knowledge management, ontologies proved to be very useful and effective and thus have been applied in system modelling, information retrieval, and information systems. They have also showed to be better than other techniques in storing and retrieving knowledge semantically.

We posit that a FAIR-inspired cyber security framework will be more effective in guiding handling of cyber threats. Operating within the auspices of Open-Data, the "A" – Accessible part of FAIR posits that data should be "as open as possible and as closed as necessary" [8]. It should be "open" in order to enable the reusability and to accelerate research, but at the same time it should be "closed" to protect the privacy of the subjects. This would then complete the FAIR data principles of: 'Findable', 'Accessible', 'Interoperable', and 'Reusable'. Brewster [9] deems the movement of open data as a utopian vision of various data sharing possibilities because of the respect of realistic practical data sharing across various parties provided by FAIR.

The rest of the paper is organized as follows: Section I presents the introduction and background to the proposed framework. Section II on the other hand presents a synthesis of prior and related work, while section III is the proposed FAIR-inspired Framework for Cyber Security ontologies where a depiction of the framework is given with a detailed discussion of each component. This is then followed by section IV which presents the limitations of the study. Lastly, the conclusions of the study are presented in section V.

II. LITERATURE REVIEW

It is important to look into the research that has been done in the four main areas that are involved in designing the framework, which are Open Data, Cyber Security, Ontology, and FAIR data principles. Table 1 gives a summary of the main concepts of this paper. Much research has been done on ontologies and cyber security individually but not together. FAIR data principles are a fairly new concept that started in 2016 and has been encouraged to be implemented within the design of ontologies to provide good quality data. The three have not been merged. Merging Open Data, Ontology, FAIR principles and cyber security will produce advanced knowledge representation and information integration with quality usable data.

Table 1: Concepts Summary

Concept	Objective
Cyber Security	Cyber security is a safeguarding mechanism for systems connected though the internet from cyber attacks. Revolutionization of technological aspects affects all areas of life from agriculture, shopping, communication, banking, and education.
Ontology	Framework that is common allowing sharing and the reuse of data across multiple applications and communities, which will be automatically processed by tools to reveal possible relations between the data and manually process where needed. The vision will be more successful with the use of the semantic web for the extension of web principles from documents to that of data.
FAIR data principles	The fair principles are intended as a guide to enable digital resources to become more Findable, Accessible, Interoperable and Reusable for machines and thus also for humans. Using FAIR and Ontology together will counter some of the limitations of each of them.

2.1 Cyber security

The existence of the internet came to being in the 1960s only accessed by limited groups of researchers, scientists and the US Defense Force. Since the 1960s there has been tremendous growth in the internet. According to [10] computer crime was initially known as causing physical damage to the computers infrastructure. However, in the 1980s a new trend appeared where computers were caused to malfunction through the use of a virus or malicious code. The exponential growth of the Internet and its use further moved computer crime from physical damage (as known in the 1960s) to viruses and malware (as known in the 1980s) to personal data manipulation for financial benefit (as exemplified by things like ransomware in 2000s). A report by Pande [10] shows that close to 25 computers are

Int. J. Advanced Networking and Applications
Volume: 15 Issue: 04 Pages:100-200 (2023) ISSN: 0975-0290 (Special Issue – ASCIS 2023)

victimized by cyber-attacks every second. Additionally, since 2013 over 800 million people have been affected by cyber-attacks [10].

The proliferation of cyber threats and attacks and their constantly evolving technology requires an investment in cyber security measures. Cyber security is defined as a process of protecting devices, computer systems, and networks from unauthorized access through malware and hacking [11]. [12] acknowledges that cyber security is a safeguarding mechanism for systems connected though the internet from cyber attacks. [11] explains that the revolutionization of technological aspects affects all areas of life from agriculture, shopping, communication, banking, and education. This means that via the internet all needs are easily available. An increase in cyber-crimes is more eminent - this emphasizes the importance of cyber security in the new digital age.

According to [13] the main cyber security attacks that people and organizations should be vigilant of are malware, phishing, and ransomware. A malware attacker obtains unauthorized access to secret information by causing damage to the target system through an injected software that is malicious [13]. For example, a perpetrator fronts as an employee of a target organization to deceive the stakeholders into providing information that is confidential. The most damaging attack thus far has been ransomware which is a malicious software that intrudes a target system to encrypt the organization files. The attacker will then use sensitive information they have encrypted to demand large ransom to give the victim the decryption key, for user to regain access to their encrypted data [13]. [12] says Organizations and people have started to adopt cyber security techniques for the prevention of access that is illegal in their digital systems and data centers.
Guly´as [14] list strategies to help in fighting cybercrimes as follows:
1. Information sharing that is intensified,
2. Cyber attach simulators,
3. Vaulting encrypted data,
4. Familiarization of the jurisdiction and norms for international crime fighting.

SESAMO security model project was presented by [15] to integrate safety and security assessments to develop embedded systems. Another project by MITRE Corporation was developed by [15] to guide in the evaluation of cyber attacks impact. [12] advice that a the growing cyber attacks need a robust cyber security mitigation strategy to provide great security posture from hostile cyber attacks trying to gain access, delete, alter, extort, or destroy data. [11] agrees that a robust cyber security system is important to ensure data remains confidential and available. Given the foregoing, this study proposes the application of FAIR data principles and the integration of Ontologies to cyber security to greatly improve the mitigation of cyber threats and attacks.

2.2 Ontology

Ontology is defined [6] as a set of representative primitives with which to model a domain of knowledge. These primitives are classes, attributes and relationships. [7] discusses the importance of creating a framework that is common allowing sharing and the reuse of data across multiple applications and communities, which will be automatically processed by tools to reveal possible relations between the data and manually process where needed. Moreover, this vision will be more successful with the use of the semantic web for the extension of web principles from documents to that of data. The data would have the possibility of being accessed from a web architecture like URIs, provided the document relations are discovered.

The use of ontologies as knowledge management tools has been advocated for by [7], [16] who state that ontologies provide automatic provision in acquiring, accessing, and maintaining data. Their suggestions have had an influence in the choice of using ontologies in this research. Web semantics can provide major benefits to the use and management of data.

Ontologies are not however, a “silver bullet” solution as they can have challenges associated with them as suggested [17]. They point out a challenge of ontology domains resisting accurate formalization. They say that each individual domain has points at which formalization develops into a ‘straightjacket’ rather than a ‘liberating force’. This shows that the challenge is not deciding a better approach but developing techniques for the multiple approaches to seamlessly work together closely [17]. With regards to the cyber security approach, an ontology is not being developed on its own but with the guidelines of FAIR data principles. Just as [7], [18] had mentioned that one ontology methodology is not enough to create a framework that is seamless, adding another one based on the requirements will give better chances of success.

2.3 FAIR Data Principles

A group of 30 open data pioneers led by Larry Lessig in 2007 came up with a way in which data provided by the government could be reviewed and analyzed by other people thus government operations improved [19]. This was done to improve the use of data. The data could be accessible, complete, license-free, and machine processable says [19]. The group made a set of values and throughout the years tested and improved them. In 2013 according to [19], an executive order was signed by the former US President, Barack Obama resulting in Open Data Policy being established. The examples of open data that can be given are academic sources like survey data, open-access journals, and results from scientific experiments. Janssen [20] explains that having open access to data that is funded publicly will offer big returns from public investments, wealth can be generated from outputs, complex problems are solved through the data and policies can be made, also data sets of large quantities are analyzed.

FAIR data principles were first introduced in 2014 as small practices and principles for guiding research data stewardship for the field of life science [21]. Brewster states that FAIR data principles originally started from the research of life sciences, where there is a custom for metadata principles that is shared [9]. The data collected in any organization should have the assurance of quality and

Int. J. Advanced Networking and Applications
Volume: 15 Issue: 04 Pages:100-200 (2023) ISSN: 0975-0290 (Special Issue – ASCIS 2023)

fitness for an organizations data asset. "The fair principles are intended as a guide to enable digital resources to become more Findable, Accessible, Interoperable and Reusable for machines and thus also for humans" [22]. Since the FAIR data principles inception, they have tremendously grown in research and have been adopted in many fields like computer science, medicine, biology, and agriculture. Making the choice to implement FAIR principles should lead to results that are machine actionable meaning personalized solutions can be defined by stakeholder communities then [22, 6].

2.4 Ontology and FAIR Data

The World Wide Web was created over twenty plus years ago and since that time we have migrated from a "web of documents" to a "web of data" vision – giving birth to concepts like linked data and open data. Boeckhout [5] deems the movement of open data as an utopian vision of various data sharing possibilities because of the respect of realistic practical data sharing across various parties provided by FAIR data.

Private and public organizations worldwide have adopted to the importance of using FAIR principles in data management (e.g., the Research Data Alliance (RDA) and the European Open Science Cloud) [6]. Poveda Villal´on [6] further states that ontologies are a major contributor in some of the FAIR data principles as they mostly address the Interoperability – "I" and Reusability – "R" part of FAIR. [9] also agrees that in order to achieve fine-grained and secure access to data that is FAIR, ontologies need to be used. [6] demonstrates the technical and social methods in defining a roadmap for the generation and publishing of FAIR ontologies on the web in comparison with other existing and ongoing frameworks. This research will adopt a similar method of comparing the already existing FAIR, cyber security, and Ontology frameworks to identify the successes and shortcomings and guide the development of the proposed framework.

According to [9], FAIR principles are not used regularly to data that originates from sensitive sources there are mostly used in open data, which is a cause of concern for privacy and confidentiality of data. It is the norm to apply FAIR principles to open data that is available publicly and to academic research data sets that are carefully defined. Based on the fact that law enforcement agencies collect data from a multitude of sources, then annotation, linking and enrichment may bring to rise new issues of confidentiality [9]. [18] also expressed a concern of the challenges on privacy of data and making the system interoperable in a harmonized way with the fact that they will be too many brains with different ideas on the project. It may also slow down the process of merging the various community ideas together and editing and changing along the way.

Brewster [9] underlines the '-A"- in FAIR that stand for Accessibility as '- Accessible under well-defined conditions"- meaning that there are great reasons for shielding access to some data. For example, looking at national security, personal privacy, and competitiveness. Other critical challenges of ontologies are making data findable and interoperable. [23] says this is because the vocabularies are dynamic, some of the terms belong to multiple ontologies, and others simply lack the terms that are needed. Using FAIR and Ontology together will counter some of the limitations of each of them.

Various ontologies have been designed and developed in different domains (e.g., medical, to physics, and computer Science) but finding them has proven to be a difficult process. Garijo [5] found that this could be due to lack of documentation provided, version problems, and URI resolution problems amongst many. Based on this, [5] created guidelines to tackle these problems. Still with their significant efforts for making data FAIR, they encountered issue relating to the fact that vocabularies were problematic in accessing, understanding, and reusing. They said that this was because of ontology classes and properties lacking definitions, ontology URIs being non-resolvable, imported ontologies being unavailable or deprecated, the documentation lacking examples and diagrams, and lastly scientific publications providing descriptions to ontologies with no implementation reference describes [6], [4]. Even though the scientific communities have acknowledged properly documenting ontologies, versioned, published and maintained and the adaptation to FAIR principles the recommendations come with no guideline for their implementation to target vocabulary. These are some of the challenges this research will be tackling when designing the framework.

III. PROPOSED FAIR-INSPIRED FRAMEWORK FOR CYBER SECURITY ONTOLOGIES

The framework for integrated cyber security ontologies is proposed as a guiding principle for the application of FAIR data principles to cyber security as seen in Fig 1. The development of the framework was a combination of adopted research that included work from [24]. In addition, the FAIRification process by [25] was adopted to provide guidance for the cyber security ontologies framework. The implementation of FAIR data principles is conducted in the form of cyber security ontology models. Existing cyber security models were sampled to demonstrate the procedures of the framework. The cyber security models serves as a reference for the source systems data, thereby creating a big virtual data warehouse. The domain expert data also forms part of the data warehouse. Firstly, a discussion of the proposed framework is presented followed by the discussion of each component including the ontology models sampled.

3.1 FAIRification of Cyber Security

Results from the cyber security ontology models from research studied revealed that there were no FAIR standards that were applied for the data and metadata which is globally understood by both humans and machines. The infusion of FAIR to Cyber security ontologies is attainable only by bringing together experts in cyber security and using valid online data. According to [26] FAIRification is possible through FAIR data modelers, producers of data, and consumers of data. The combination of experts is a precursor for the acquisition of the meaning of data based on

Int. J. Advanced Networking and Applications
Volume: 15 Issue: 04 Pages:100-200 (2023) ISSN: 0975-0290 (Special Issue – ASCIS 2023)

cyber security ontologies. The FAIR guiding principles are as detailed by [25]:
1. To be Findable:
– F1. (meta)data are assigned a globally unique and persistent identifier.
– F2. data are described with rich metadata (defined by R1 below).
– F3. metadata clearly and explicitly include the identifier of the data it describes.
– F4. (meta)data are registered or indexed in a searchable resource.

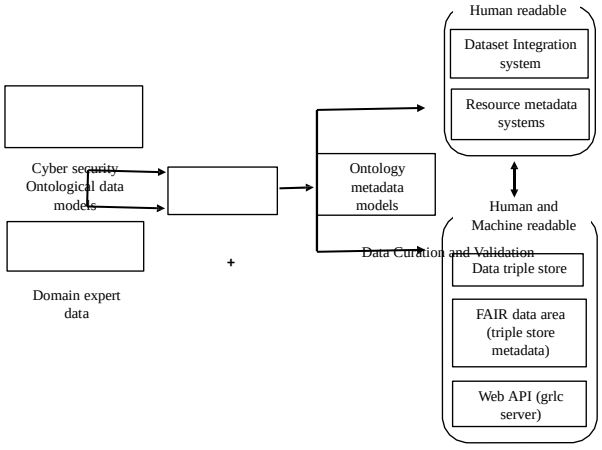

Figure 1. Proposed FAIR-inspired Framework for Cyber Security Ontologies.

2. To be Accessible:
– A1. (meta)data are retrievable by their identifier using a standardized.
communications protocol
– A1.1 the protocol is open, free, and universally implementable.
– A1.2 the protocol allows for an authentication and authorization procedure, where necessary.
– A2. metadata are accessible, even when the data are no longer available.
3. To be Interoperable:
– I1. (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.
– I2. (meta)data use vocabularies that follow FAIR principles.
– I3. (meta)data include qualified references to other (meta)data.
4. To be Reusable:
– R1. meta(data) are richly described with a plurality of accurate and relevant attributes.
– R1.1. (meta)data are released with a clear and accessible data usage license.
– R1.2. (meta)data are associated with detailed provenance
– R1.3. (meta)data meet domain-relevant community standard.

3.2 Cyber security domain expert data

The FAIRification process for the cyber security ontology models begins with the synchronization of cyber security stakeholders. The stakeholders are private businesses and organizations, public businesses and organizations, computer forensic analysts, security engineers, system administrators, digital forensic engineers, and any other professional in the cyber security field. The first activities involve developing ontology models collaborating with data collectors, analysts, managers, and cyber security specialists for recording data. This would typically be followed by system developers and database managers who use the data to develop machine actionable metadata. Both these activities are performed to address the 'F' 'A' 'R', that is Findability, Accessibility, and Reusability in the FAIR data principles.

3.3 Curation and Validation of data

The curation and validation of data as seen in Fig. 2 is aimed at raw dataset coming from the cyber security domain experts and ontology models. It is important in increasing dataset quality of incoming cyber security information. The first step is curating the data, that means values, types, and data fields are given characteristics. Furthermore, cyber security concepts like fields for threats, attacks, mitigation strategies, and vulnerabilities are extracted. When the data curation process is complete the data needs to be validated. The validation process requires predefined quantitative relationships and values that are expected according to the set semantic model. Semantic rules need to be defined prior to curation and validation of the data. The curated data is checked against the predefined rules to see if it conforms.

3.4 Cyber security Ontology data models

The cyber security ontology data models are used to meet the 'I' – Interoperability in FAIR. The development of the user-focused data-driven infrastructures should be developed using cyber security questions for the formation of data models. This is a crucial step for enabling interoperability of data models that is effective among the cyber security community. The models in Fig. 2, Fig. 3, and Fig. 4, are adopted as samples for this paper representing the type of cyber security models that are essential for the FAIR inspired cyber security ontologies.

The FAIR data principle in focus to implementing interoperability in cybersecurity ontologies are as follows:
1. I1 – (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.
2. I2 – (meta)data use vocabularies that follow FAIR principles

To effectively depict and describe the ontology interoperability of the FAIRification process, cyber security models were adopted from various studies. The first model depicted in Fig. 2 was adopted from Silva [27]. It represents a concept model that is necessary for mapping terms that are important from cyber security experts and valid online cyber security resources. The conceptual model is also used for questions that are needed to extend cyber security with valid data. The conceptual model represented the expert data for cyber security. This research only showed a sample model of what it would look like, but FAIRification developers need to dive deeper into research and experts to create a

Int. J. Advanced Networking and Applications
Volume: 15 Issue: 04 Pages:100-200 (2023) ISSN: 0975-0290 (Special Issue – ASCIS 2023)

fully fleshed conceptual model that models as much relevant information in cyber security as possible.

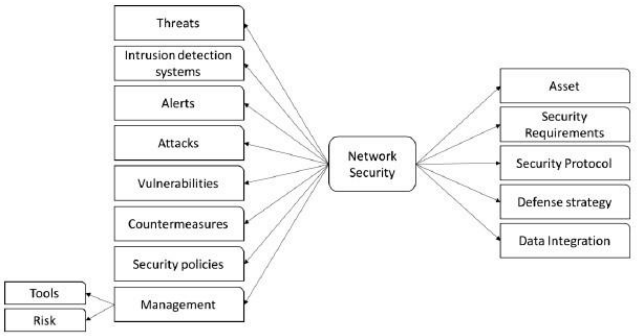

Figure 2. Conceptual model for mapping terms from cyber security experts and valid online cyber security resources [27].

When the conceptual model with all the relevant data is completed, the next step is to create the cyber security ontological model. The cyber security ontological model was created using the data received from the conceptual model. The ontology model should have a description in the Web Ontological Language (OWL) and developed in the Resource Description Framework (RDF) representing web semantics. The cyber security ontological model should have clearly defined core concept of cyber security. The threats mitigation and threats solution procedures should outline that all components in their registries is an outcome of a procedure, thus making the procedures core concepts of the model. Another important core concept of the semantic model that should be captured are the threats and vulnerabilities concepts. The concepts can be captured as measurements, depicting the measurement points of the incoming threats and vulnerabilities.

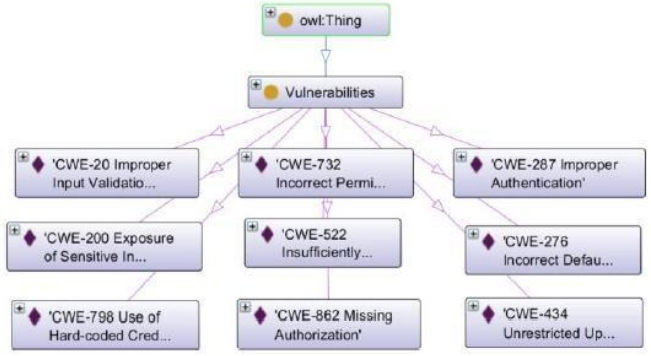

Figure 3. Vulnerability model for cyber security [28]

Fig. 3 adopted from [28] is a model that illustrates the threats and vulnerabilities involved in cyber security. The model is reused for the development of the cyber security ontological model threats and vulnerability classes, along with Fig. 4, also adopted from [28] representing the threats mitigation and threats solution model. The 'R' - Reusability in FAIR is put in practice by reusing these models from previous research relevant to the FAIRification process for cyber security. The developed ontology models not only allow interoperability in cyber security but the multiple cyber security models from different researchers used also allow integration of data of external knowledge.

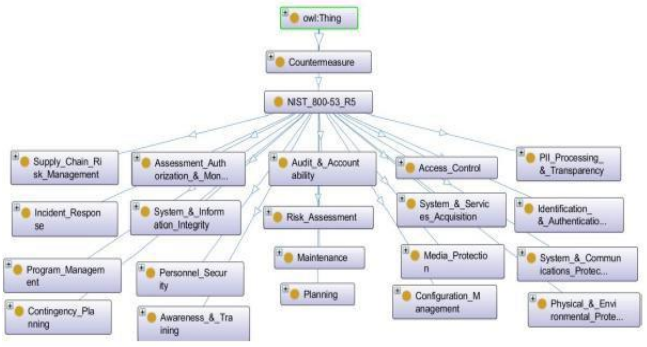

Figure 4. Countermeasure model for cyber security [28]

3.5 Ontology metadata models for Cyber security resources

The ontology metadata models component of the framework is significant for the provision of an ontology model that will be exposed in a machine-readable approach. The ontology metadata models are used to provide metadata of threats, vulnerabilities, mitigations, solutions, and attack resources in the cyber security domain to be 'F'- Findable, 'A'- Accessible, and 'R'- Reusable by machines and humans. The main FAIR principles prioritized and used in this step are used for enabling effective and efficient machine action for an analysis that is evidence based within and across the cyber security domain/community. These are as follows:

For Findable:

– F1 - (meta) data are assigned a globally unique and persistent identifier.

– F2 - data are described with rich metadata.

– F3 - metadata clearly and explicitly include the identifier of the data it describes.

For Reusability:

– R1- (meta) data are richly described with a plurality of accurate and relevant attributes.

An ontology model for cyber security was presented by [29] (see Fig. 5). The model was used in this study to demonstrate the design of a model for a cyber security metadata model for managing metadata of basic cyber security datasets. Metadata elements from the ontology model depicted in Fig. 5 are "Risk", "Vulnerabilities", "Person", "Vulnerable", "Group", "CIA", "Organization", "Likelihood", "Asset", "Event", "Threat", "Objective", "Safe", "Control", and "Consequences". The metadata elements have properties as seen in Fig. 5. This includes "mitigates", "owns", "has", "affects", "causes", "exploits", "leasesTo", "controls", and "modifies". Fig. 5 illustrates that cyber security data resources can be specified as a knowledge base which can describe cyber security resources containing data input of experts. This creates a structured semantics of metadata for cyber security resources that is precise and richer.

Int. J. Advanced Networking and Applications
Volume: 15 Issue: 04 Pages:100-200 (2023) ISSN: 0975-0290 (Special Issue – ASCIS 2023)

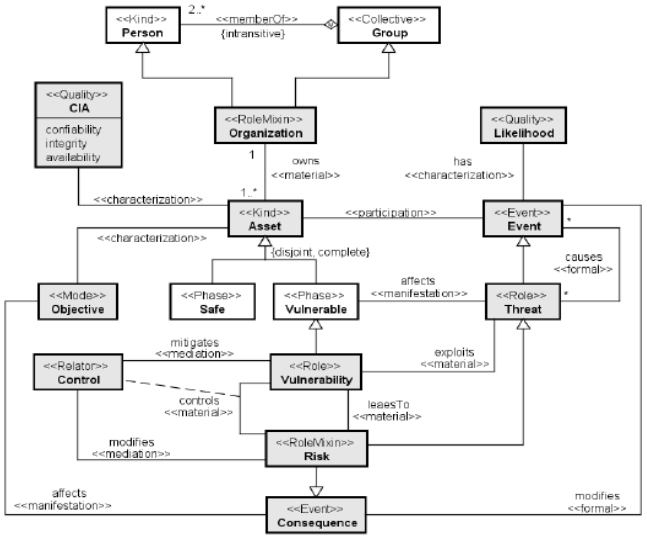

Figure 5. Model for a cyber security metadata model [29]

3.6 Data management systems

The next stage in the framework involves development of two data management systems, the dataset integration system and the resource metadata system. The data management systems are the human readable components of the framework illustrating how data that is ontologized is integrated with the systems. The two data management systems have no provision of providing semantic modelling functionality directly. However, the two systems provide essential annotation functionality used for foundation for the connection of the ontology models. The systems are also used for adding access to cyber security concepts and measurements explained in section E as instances of ontology models.

3.6.1 Dataset Integration system

The dataset integration system operates like a data warehouse system for cyber security data, that complies with workflow standards used in preparation of data for research. The main role of the integration system is to bring datasets from various sources/ systems in the cyber security domain into one warehouse. The stakeholders have to choose one common vocabulary that supports annotations and transformation of the data. APIs are provided that can be incorporated in cyber security workflows. Moreover, the dataset integration system provides developers and researchers with a single access point that is central to data that is syntactically machine-readable. The framework illustrates a relation between the data integration system and the triple datastore. The integration of cyber security datasets is managed by the system to instantiate the ontology linking data model in RDF. Data files from the system (.csv files) is used as inputs for the connection from the integration system to triple store (ontology model), creating knowledge graphs represented as RDF.

3.6.2 Resource metadata systems

The resource metadata system is used to add metadata that is valuable about the cyber security data resources. It functions like metadata extension tool of the integration system. It should be used for providing annotations on dataset level for example where, how, by whom, when, the conditions that data was collected. The information containing the answers to the questions is published online. Cyber security ontology queries are processed and results published through the resource metadata system.

The data management systems apply the ‘Findability’ aspect of the FAIR data principles. The Findability principles concentrate on the metadata of the framework as follows:
– F1. (meta)data are assigned a globally unique and persistent identifier.
– F2. data are described with rich metadata.
– F3. metadata clearly and explicitly include the identifier of the data it de-
scribes.
– F4. (meta)data are registered or indexed in a searchable resource.

The integrated system together with the resource metadata system handle the data metadata that come from the ontology models and domain experts. The globally unique and persistent identifier F1 is assigned by the integrated system. The resource metadata addresses F2 by providing a rich description of cyber security metadata. RDF from the triplestore represents metadata about cyber security web resources like XML files, thus implementing F3. Lastly, F4 is applied in the FAIR data area and API where data indexing is conducted for data that will be exchanged using RDF and shared on the Web.

3.7 Data Triplestore , FAIR Data, and grlc server

The data file from the integration system are converted to knowledge graphs represented as RDF and sent to the triple store. The knowledge graphs are used for the representation of entities and their relationships formed in the cyber security domain. The RDF model is explained by [30, 31] as a graph with a set of subject, predicate, and object triples. Researcher [32] concurs that the RDF base element is the triple resource (subject) linked to other resource (object) through an arc to another resource (predicate) and stored in triplestores. According to [30] triplestores are unique data management systems used to store RDF data. Furthermore, triplestores support declarative queries that are written in languages such as SPARQL. RDF and SPARQL are used for the application of the FAIR data principles; A1. (meta)data are retrievable by their identifier using a standardized communications protocol model that supports integrated and uniform access to information sources and services.

The Proposed FAIRification Framework for Integrated Cyber security ontologies focus specifically on data triplestore for indexing, storing, and the querying of RDF data. SPARQL is the query language that is used for retrieving data that is linked from the triplestore. For users that do not have much knowledge in ontologies or SPARQL it may be challenging to work with data in triplestores. However, solutions such as the grlc server make seamless. [33] says that Web APIs can be generated using the ontology data and grlc server. The grlc server will automatically convert the SPARQL queries to WEB APIs

Int. J. Advanced Networking and Applications
Volume: 15 Issue: 04 Pages:100-200 (2023) ISSN: 0975-0290 (Special Issue – ASCIS 2023)

making the RDF data 'A' – Accessible on the Web. The Grlc server addresses the FAIR principle; A1.1 the protocol is open, free, and universally implementable as well as intelligent allowing for applications for information processing on the Web. It is a free server that can be found on github that allows information processing in the web.

Querying the Web APIs of FAIR cyber security data will be more effective with these tools. The data structure will allow answering cybersecurity questions using queries in the form of the ontology models like it would happen in the real world situations between stakeholders, but much faster, efficient, and effective. Access to the cyber security web API as depicted in the framework will come with permission procedures to check, validate, and confirm that the person trying to gain access has authority and has no malicious intentions. This step is developed as part of the FAIR principles; A1.2 the protocol allows for an authentication and authorization procedure, where necessary and R1.3. (meta)data meet domain-relevant community standard.

The FAIR data principle "I1. (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation", is implemented by the use of RDF language that provides the knowledge representation through knowledge graphs and supporting quick data source integration for bridging the differences in web semantics. The cyber security ontology models use vocabularies that FAIR principles as required by principle; "I2. (meta)data use vocabularies that follow FAIR principles.", they are standardized mechanisms for interchanging data of different semantic data using vocabularies that are FAIR.

IV. Limitations and Future Work

The research implemented majority of the FAIR data principles in the proposed framework but not all of them were covered. Priority was given to the data principles that were critical to the implementation of the principles to cyber security ontologies. The principles that have not yet been used are A2: where metadata are accessible, even when the data are no longer available, I3: where (meta)data include qualified references to other (meta)data, R1.1: where (meta)data are released with a clear and accessible data usage license, and R1.2: where (meta)data are associated with detailed provenance. Future work in the cyber security domain will include these data principles and improve on the proposed framework.

V. Conclusion

In conclusion, the proposed FAIR-inspired framework for integrated cyber security ontologies presents a comprehensive and systematic approach to applying FAIR data principles in the field of cyber security. The framework emphasizes the Findability, Accessibility, Interoperability, and Reusability of cyber security data through the development and implementation of FAIR data principles. The journey begins with the FAIRification of cyber security, addressing the lack of applied FAIR standards in existing ontology models. The infusion of FAIR into cyber security ontologies involves collaboration among experts in the field, using valid online data to enhance the Findability, Accessibility, Interoperability, and Reusability of the information. The guiding principles of FAIR, as detailed in the framework, ensure that data is assigned unique identifiers, described with rich metadata, retrievable through standardized protocols, and usable across various applications. The framework then delves into the synchronization of cyber security stakeholders and the development of ontology models, involving private and public entities, analysts, engineers, administrators, and other professionals. The subsequent steps of curation, validation, and the creation of cyber security ontology data models contribute to achieving Interoperability, using formal, accessible, shared languages for knowledge representation.

Furthermore, the inclusion of ontology metadata models for cyber security resources plays a crucial role in making data Findable, Accessible, and Reusable by machines and humans. The FAIR principles guide the provision of machine-readable metadata for threats, vulnerabilities, mitigations, solutions, and attack resources in the cyber security domain. The framework's data management systems, including the dataset integration system and resource metadata system, facilitate the integration of ontologized data and the addition of valuable metadata. These systems align with the FAIR principles, especially in terms of making data Findable through unique identifiers, rich metadata, and registration in searchable resources.

The incorporation of a data triplestore, FAIR data, and grlc server enhances the framework's capability to index, store, and query RDF data. The use of SPARQL queries and the grlc server ensures the Accessibility of cyber security data on the web. The overall structure promotes the effective application of FAIR principles, providing a foundation for seamless information exchange and interoperability within the cyber security community. In summary, the proposed FAIR-inspired framework serves as a valuable guide for implementing FAIR data principles in cyber security ontologies, fostering a standardized, collaborative, and interoperable approach to managing and utilizing cyber security data.

References

[1] M. Mijwil, O. J. Unogwu, Y. Filali, I. Bala, and H. Al-Shahwani, Exploring the Top Five Evolving Threats in Cybersecurity: An In-Depth Overview, *Mesopotamian Journal of Cyber Security*, 2023, doi: 10.58496/mjcs/2023/010.

[2] M. Horowitz, Cyber security report 2023, Mission Valley ROP, 2023. , https://www.mvrop.org/cms/lib/CA01922720/Centrcity/Domain/59/2023-cyber-security-report.pdf.

[3] P. Treleaven *et al.*, The Future of Cybercrime: AI and Emerging Technologies Are Creating a Cybercrime Tsunami, *SSRN Electronic Journal*, 2023, doi: 10.2139/ssrn.4507244.

[4] F. Neil, List of Data Breaches and Cyber Attacks in 2023, IT Governance.

Int. J. Advanced Networking and Applications
Volume: 15 Issue: 04 Pages:100-200 (2023) ISSN: 0975-0290 (Special Issue – ASCIS 2023)

https://www.itgovernance.co.uk/blog/list-of-data-breaches-and-cyber-attacks-in-2023

[5] D. Garijo and M. Poveda-Villalón, Best Practices for Implementing FAIR Vocabularies and Ontologies on the Web, 2020. doi: 10.3233/ssw200034.

[6] M. Poveda-Villalón, P. Espinoza-Arias, D. Garijo, and O. Corcho, Coming to Terms with FAIR Ontologies, in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2020. doi: 10.1007/978-3-030-61244-3_18.

[7] N. Kaewboonma and K. Tuamsuk, Application of Ontologies for Knowledge Management, *TLA Research Journal*, *10*(2), December 2017, 2017.

[8] A. Landi *et al.*, The 'a' of fair – as open as possible, as closed as necessary, *Data Intell*, *2*(1–2), 2020, doi: 10.1162/dint_a_00027.

[9] C. Brewster, B. Nouwt, S. Raaijmakers, and J. Verhoosel, Ontology-based access control for fair data, *Data Intell*, *2*(1–2), 2020, doi: 10.1162/dint_a_00029.

[10] S. Jadey, S. C. Girish, K. Raghavendra, G. Prasanna Kumar, H. R. Srinidhi, and K. M. Anilkumar, Introduction to cyber security, in *Methods, Implementation, and Application of Cyber Security Intelligence and Analytics*, 2022. doi: 10.4018/978-1-6684-3991-3.ch001.

[11] Bhushan Bharat, The Growing Importance of Cyber Security in the Digital Age, *International Journal for Innovative Research in Multidisciplinary Field*, *9*(5), pp. 234–239, 2023.

[12] G. Abdumalikov, Profound Importance of Cyber security in the Field of Business, *INTERNATIONAL JOURNAL ON HUMAN COMPUTING STUDIES*, *4*(2), 2022.

[13] E. M. Kala, The Impact of Cyber Security on Business: How to Protect Your Business, *Open Journal of Safety Science and Technology*, *13*(02), 2023, doi: 10.4236/ojsst.2023.132003.

[14] O. Gulyas and G. Kiss, Impact of cyber-Attacks on the financial institutions, in *Procedia Computer Science*, 2023. doi: 10.1016/j.procs.2023.01.267.

[15] C. Izuakor, Understanding the impact of cyber security risks on safety, in *ICISSP 2016 - Proceedings of the 2nd International Conference on Information Systems Security and Privacy*, 2016. doi: 10.5220/0005796805090513.

[16] S. Yasunaga, M. Nakatsuka, and K. Kuwabara, Web ontology building system for novice users: A step-by-step approach, in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2010. doi: 10.1007/978-3-642-12101-2_15.

[17] I. Jurisica, J. Mylopoulos, and E. Yu, Using ontologies for knowledge management: An information systems perspective, *Proceedings of the ASIS Annual Meeting*, *36*, 1999, doi: 10.1007/s10115-003-0135-4.

[18] F. Freitas, H. Stuckenschmidt, and N. F. Noy, Ontology Issues and Applications Guest Editors' Introduction, *Journal of the Brazilian Computer Society*, *11*(2), 2005, doi: 10.1007/BF03192372.

[19] L. B. Ayre and J. Craner, Open Data: What It Is and Why You Should Care, *Public Library Quarterly*, *36*(2), 2017, doi: 10.1080/01616846.2017.1313045.

[20] M. Janssen, Y. Charalabidis, and A. Zuiderwijk, Information Systems Management Benefits, Adoption Barriers and Myths of Open Data and Open Government Benefits, Adoption Barriers and Myths of Open Data and Open Government, *Information Systems Management*, *29*, 2012.

[21] M. Boeckhout, G. A. Zielhuis, and A. L. Bredenoord, The FAIR guiding principles for data stewardship: Fair enough?, *European Journal of Human Genetics*, *26*(7). 2018. doi: 10.1038/s41431-018-0160-0.

[22] A. Jacobsen *et al.*, Fair principles: Interpretations and implementation considerations, *Data Intelligence*, *2*(1–2), 2020. doi: 10.1162/dint_r_00024.

[23] S. Spoor *et al.*, Tripal v3: An ontology-based toolkit for construction of FAIR biological community databases, *Database*, *2019*(1), 2019, doi: 10.1093/database/baz077.

[24] A. A. Sinaci *et al.*, From Raw Data to FAIR Data: The FAIRification Workflow for Health Research, *Methods Inf Med*, *59*(6), 2020, doi: 10.1055/s-0040-1713684.

[25] M. D. Wilkinson *et al.*, Comment: The FAIR Guiding Principles for scientific data management and stewardship, *Sci Data*, *3*, 2016, doi: 10.1038/sdata.2016.18.

[26] N. Queralt-Rosinach *et al.*, Applying the FAIR principles to data in a hospital: challenges and opportunities in a pandemic, *J Biomed Semantics*, *13*(1), 2022, doi: 10.1186/s13326-022-00263-7.

[27] D. V. Silva and G. R. Rafael, Ontologies for network security and future challenges, in *Proceedings of the 12th International Conference on Cyber Warfare and Security, ICCWS 2017*, 2017.

[28] L. O. Wicklund, *Cybersecurity Ontology*, Masters Degree, Stockholm University, 2022.

[29] B. F. Martins, L. Serrano, J. F. Reyes, J. I. Panach, O. Pastor, and B. Rochwerger, Conceptual Characterization of Cybersecurity Ontologies, in *Lecture Notes in Business Information Processing*, 2020. doi: 10.1007/978-3-030-63479-7_22.

Int. J. Advanced Networking and Applications
Volume: 15 Issue: 04 Pages:100-200 (2023) ISSN: 0975-0290 (Special Issue – ASCIS 2023)

[30] M. Lissandrini, T. Sagi, T. B. Pedersen, and K. Hose, Understanding RDF Data Representations in Triplestores, in *CEUR Workshop Proceedings*, 2022.

[31] S. Decker, P. Mitra, and S. Melnik, Framework for the semantic web: an RDF tutorial, *IEEE Internet Comput*, *4*(6), 2000, doi: 10.1109/4236.895018.

[32] Champin Pierre-Antoine, RDF Tutorial, *Research gate*, 2022.

[33] A. Meroño-Peñuela and C. Martinez-Ortiz, grlc: the git repository linked data API constructor., *J Open Source Softw*, *6*(67), 2021, doi: 10.21105/joss.02731.

BIOGRAPHIES

Tshepiso Larona Mokgetse is currently pursuing her PhD in Computer Science in the Department of Computer Science and Information Systems at Botswana International University of Science and Technology (BIUST). As well as working as a teaching assistant at BIUST. She graduated in 2019 with an MSc degree in Computer Applications from Gujarat Technological University, Ahmedabad, Gujarat, India. Her PhD thesis is focused on Ontologies, Knowledge management and Animal Traceability, Open data. She is currently working on publications focused on her area of interest. Area of interest being Ontologies, Open data, Knowledge management, Traceability Technologies, and Internet of Things.

R. Sridaran is a Ph.D in Computer Science from Madurai Kamaraj University. He has published 50+ research papers in leading journals and chaired many conferences. He has also guided number of research scholars in the areas of Cloud Computing & Security, IoT, e-learning and Block Chain. He has got more than 30 years of experience with leading educational institutions at senior academic administration levels. He is currently the Dean of Faculty of Computer Applications, Marwadi University Rajkot. He has initiated the formation of Computer Society of India, Rajkot Chapter as a founder Chairman.

Hlomani B. Hlomani is a Senior Lecturer at the Botswana International University of Science and Technology He received his undergraduate degree in Information Technology from the Cape Peninsula University of Technology, Cape Town, South Africa. He also received both his MSc and PhD. degrees in Computer Science from the University of Guelph, Guelph, Ontario, Canada, in 2009 and 2014, respectively. His research interests include Artificial Intelligence, The semantic web, Ontologies, Knowledge management, and Knowledge engineering.