# Concept-based Extension of SKOS Defense Controlled Vocabulary: Techniques and Implications

PERICLES S. GIANNARIS[1], NIKOLAOS DOUKAS[2], NIKOS MASTORAKIS[1]
[1]Sector of Electrical Engineering and Computer Science,
ASEI (Military Institutes of University Education), Hellenic Naval Academy,
Hadjikyriakou Avenue, Piraeus, P.C. 18539,
GREECE

[2]Division of Mathematics and Engineering,
ASEI (Military Institutes of University Education), Hellenic Army Academy,
Leoforos Eyelpidon (Varis – Koropiou) Avenue Vari P.O. 16673,
GREECE

*Abstract:* - A controlled vocabulary is a set of terms that are utilized to represent knowledge in a domain. In the domain of defense, the use of terms such as "command, control, communications, computers, intelligence, surveillance and reconnaissance (C4ISR)", "armored personnel carriers (APC)", or "biological and bioinspired structures for multispectral surveillance" denote the core competencies of domain experts and the depth of diverse knowledge. This paper describes the second phase of the pilot project to create a defense-related controlled vocabulary with a focus on the Russo-Ukraine conflict. Applications for data annotation, SKOS hierarchical vocabulary development, and vocabulary quality analysis are used to identify terms in text, express identified terms in military press releases as SKOS vocabulary, and assess its structure. The preliminary vocabulary is extended by 173 concepts. The quality of the vocabulary is validated against a SKOS checker of twenty-four criteria.

*Key-Words:* - Controlled Vocabulary, Knowledge Representation, SKOS, Semantic Web, Semantic Web, Named Entity Recognition.

Received: June 19, 2023. Revised: February 8, 2024. Accepted: March 15, 2024. Published: May 7, 2024.

## 1 Introduction

A controlled vocabulary is a set of terms that are utilized to represent knowledge in a domain. For example, in the domain of defense are used the following terms: "command, control, communications, computers, intelligence, surveillance and reconnaissance (C4ISR)", "armored personnel carriers (APC)", or "biological and bioinspired structures for multispectral surveillance". The structure that a controlled vocabulary can acquire is that of a simple list of terms or a complex graph structure. The World Wide Web Consortium (W3C) considers controlled vocabularies as the basic element of the Semantic Web, a set of technologies to link disparate data sources. To standardize the creation of a controlled vocabulary in a machine-readable format and to assist with its assessment, W3C recommends the simple knowledge organization (SKOS) schema. SKOS is based on the machine-readable resource description framework (RDF) data model therefore, it can be defined as an OWL ontology, [1], [2].

In the initial phase of our pilot study for the creation of a Russo-Ukraine conflict controlled vocabulary, we extracted terms from a sample of North Atlantic Treaty Organization (NATO) press releases. The terms in the vocabulary are hierarchically structured according to the SKOS data model. In this paper, we describe three steps to extend our initial controlled vocabulary with new concepts. First, we analyze a new batch of press releases to identify terms to express in SKOS. Second, we utilize the friend-of-a-friend (FOAF) language to describe persons and their social links within the context of NATO. FOAF depends on W3C's standards, specifically on extensible markup language (XML), XML Namespaces, RDF, and web ontology language (OWL), [3]. Lastly, we conduct a preliminary assessment of our extended controlled vocabulary.

A couple of published research is noted to demonstrate the utility of controlled vocabularies in

different disciplines. First, [4], emphasize that Gene Ontology (GO) is a controlled vocabulary with three hierarchies of terms for (a) gene functions, (b) larger-scale biological processes, and (c) cellular components to discover information about gene products, [4]. Second, Based on a methodology to semantically encode text data to machine understandable format, [5], propose a controlled vocabulary as the foundation of a thematic thesaurus modelled according to SKOS to streamline the interoperability of knowledge bases at knowledge-intensive institutions such as libraries and universities, [5].

The rest of the paper is organized as follows, section two discusses the methodology for extending our Russo-Ukraine conflict controlled vocabulary based on text analysis and the employment of tools to leverage SKOS and FOAF; section three discusses pilot results, preliminary evaluation of the controlled vocabulary and, limitations; and section four concludes the paper.

# 2 Methodology

Figure 1 displays the tasks involved in creating a controlled vocabulary for the Russo-Ukraine conflict using terms from NATO press releases.

## 2.1 Data Acquisition

Similarly to the initial phase, we obtain press releases from the NATO Press Office News site, [6]. Our query, "Ukraine AND Russia AND Conflict", covers the period from 27 July 2023 to 14 December, 2023. The search engine returns twenty-eight press releases, between 23 August, 2023 and 14 December, 2023, in batches of ten per results page. We manually store and organize the results in a spreadsheet. The data are organized in three columns: date, image, headline, and opening paragraph.

## 2.2 Data Wrangling

This task has two steps. Firstly, we utilize the "select object", "filter", "sort", "macro option for Visual Basic for Application (VBA) code", and "trim" functions in our spreadsheet editor to remove images, standardize word spacing, convert hypertext to plain text, and sort our sample of text data by date. Our text data sample is now organized into three columns: the date, the headline, and the press release. Second, we measure the frequency of each word in the data set and count data according to unique dates, the number of press releases per date, words per headline, and press release, [7], [8] using R base functions, [9]. Text quantification provides a preliminary view of the potential distribution of terms considering our data set.

## 2.3 Text Data Annotation

Text data annotation is the process of tagging information to text to circumstantiate its meaning. As in the first phase of our research, we annotate terms in our sample of press releases to be added toour controlled vocabulary. Specifically, we annotate text similarly to preparing data for named entity recognition (NER), the computational technique to recognize predefined categories of entities in a body of text, [10] for example, "[Secretary General]$_{\text{designation}}$ [Jens Stoltenberg]$_{\text{proper\_name}}$ visited [Kyiv]$_{\text{city}}$". To annotate text, we employ Label Studio [11], a data annotation tool for NER projects that is offered as a feature in DagsHub [12], a platform that supports data analytics projects. Additionally, we leverage the code configuration interface of the tool to customize the annotation tags to our previous manual annotation codebook, [13]. The annotated data are saved in a comma-separated values (CSV) file and supplied in a JavaScript Object Notation (JSON) like structure. Next, we programmatically parse the data file to extract terms and their corresponding annotations, [14], [15], in a data frame, a type of a two-dimensional array, Table 1.

Table 1. Randomly selected terms and corresponding annotation tags

| Term | Annotation tag |
|---|---|
| nato deputy secretary general | "designation" |
| vilnius summit | "defense_or_civil_event" |
| new york | "city" |
| intensive care units in hospitals | "infrastructure" |
| italy | "country" |
| pressing security challenges | "challenges_situation_problem" |
| artillery systems | "aegis_arms_research_instruments" |
| ministry of the interior | "government_authority" |
| medical services to civilians | "defense_process_measures" |

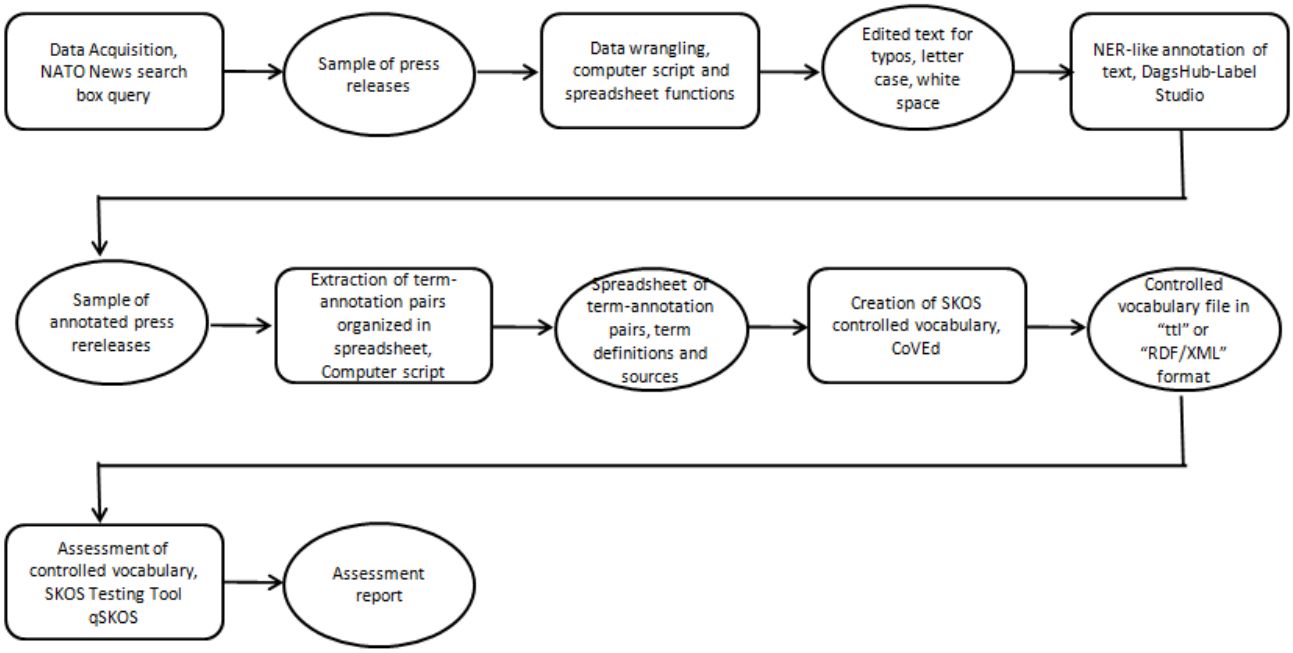


Fig. 1: Flowchart of the process to create a Russo-Ukraine conflict SKOS-based controlled vocabulary

The data frame is stored in a spreadsheet for further analysis. For each annotated term and corresponding tag, we provide definitions and related sources, which are used in the next task.

### 2.4 Creation of Controlled Vocabulary

SKOS has three basic elements that are defined as classes, properties, and relations. Every element begins with the prefix "skos". The elements are distinguished according to the uppercase or lowercase letter after the prefix for example, a class can be encoded as "skos:Concept"; a property can be encoded as skos:prefLabel, [16]. The property skos:related" can be used to express a relation between two concepts. Importantly, SKOS models data according to a concept-centered approach compared to a term-centered approach, [17].

In this phase, we employ the Controlled Vocabulary Editor (CoVEd), which uses the SKOS schema and the Terse RDF Triple Language (TTL) syntax to extend our controlled vocabulary. CoVEd was developed by the Narralive research team of the Athena Research and Innovation Center at the University of Athens, Greece, [18], [19] in [20]. We begin by creating a concept scheme that incorporates the sets of concepts for our vocabulary. Each concept is identified by a uniform resource identifier (URI). The following SKOS concepts and properties are the backbone of our vocabulary: "skos:ConceptScheme", "skos:Concept", "skos:inScheme", "skos:topConceptOf", "skos:hasTopConcept". The "skos:prefLabel" encodes the preferred lexical label for our resource. The use of the "skos:note" and "skos:definition" properties helps us to provide information about our SKOS concepts. Furthermore, the "skos:broader" and "skos:narrower" properties are used to assert a hierarchical connection between two concepts; the "skos:related" property is used to declare semantic relations between concepts in our concept scheme, [21].

To encode data related to persons and their social network in a semantic way, we have the opportunity to use named properties and classes from the friend-of-friend (FOAF) schema that are provided by CoVEd. FOAF is based on RDF and the Web Ontology Language (OWL) to link persons and information using the Web, [22]. In this analysis, we use the properties "foaf:person", "foaf:name", "foaf:knows".

## 3 Results and Discussion

Our search query returns 28 press releases published by the NATO Press Office about the Russo-Ukraine conflict between 27July, 2023, and 14December, 2023. The six most frequent words, as in unigram, in our sample of press releases are the following: "nato" 23 times, "secretary" 14 times, "Ukraine" 14 times, "general" 13 times, "support" 9 times, and "defense" 6 times. We annotate 173 unique terms in the text using twenty different tags, Table 2. The annotation tags correspond to the hierarchy of concepts in the SKOS vocabulary that we developed during the first phase of this study.


Fig. 2: Snippet of the hierarchy of concepts of the SKOS Russo-Ukraine Conflict vocabulary

As mentioned in the methodology, the result of the NER-like annotation process is available in JSON-based structure:

[{"end": 22, "text": "NATO Secretary General", "start": 0, "labels": ["DESIGNATION"]}, [...], {"end": 413, "text": "US-led Ukraine Defence Contact Group meeting", "start": 369, "labels": ["DEFENCE_OR_CIVIL_EVENT"]}, [...], {"end": 136, "text": "Brussels", "start": 128, "labels": ["CITY"]}, [...], {"end": 39, "text": "Jens Stoltenberg", "start": 23, "labels": ["PROPER_NAME"]}]

The above excerpt shows that JSON data are structured as name-value pairs e.g., "text": "Brussels" or "labels": ["CITY"]. Here, “text” is paired with the term we identify in the press release; “labels” is paired with the annotation tag we select for that term.

Table 2. Excerpt of tags that are used to annotate terms in press releases

| Annotation tags |
|---|
| designation |
| proper_name |
| continent_or_state_or_province |
| defence_or_civil_event |
| nato_engagements |
| defense_project |
| biomedical |

SKOS models data in concept schemes. Our SKOS data structure has a five-level hierarchy of concept specificity. The preferred label for our concept scheme is expressed as ‘skos:prefLabel "russo-ukrainian-defence-diplomacy-vocabulary"@en’.
SKOS Play, [23], prints the controlled vocabulary in HTML format that allows to computationally quantify the generated information. We count 224 concepts. Concerning terms, SKOS Play defines “nato” as the top term (TT). Here, TT corresponds to the above-mentioned first level of hierarchy. The second level comprises twelve concepts; the third level has eighteen concepts, and the fourth level has twenty concepts. The above concepts are defined as broad terms (BTs). The rest 173 concepts correspond to terms identified in the sample of press releases. The later are defined as narrow terms (NTs). Figure 2 shows part of the tree structure of our SKOS model. Furthermore, there are 180 relations expressed as “skos:related” between concepts.

Below, excerpts of our SKOS model, in TTL serialization, that models poly-hierarchy about the concept “russell-c-leffingwell-lecture”:

```
ex4:nato rdf:type skos:Concept ;
dcterms:identifier
"http://example.com/periclesRepo/ex4/nato"^^xsd:anyURI ;
skos:inScheme ex4: ;
skos:topConceptOf ex4: ;
skos:narrower ex4:[. . .] ;
skos:narrower ex4:democratic_action ;
skos:narrower ex4:[. . .] ;
skos:prefLabel "nato"@en .
ex4:democratic_action rdf:type skos:Concept ;
dcterms:identifier
"http://example.com/periclesRepo/ex4/democratic_action"^
^xsd:anyURI ;
skos:broader ex4:nato ;
skos:inScheme ex4: ;
skos:narrower ex4:peace_building_event ;
skos:prefLabel "democratic_action"@en .
ex4:peace_building_event rdf:type skos:Concept ;
dcterms:identifier
"http://example.com/periclesRepo/ex4/peace_building_even
t"^^xsd:anyURI ;
skos:broader ex4:democratic_action ;
skos:inScheme ex4: ;
skos:narrower ex4:defence_or_civil_event ;
skos:prefLabel "peace_building_event"@en .
ex4:defence_or_civil_event rdf:type skos:Concept ;
```

```
dcterms:identifier
"http://example.com/periclesRepo/ex4/defence_or_civil_ev
ent"^^xsd:anyURI ;
skos:broader ex4:peace_building_event ;
skos:definition "Event: a thing that happens or takes place,
especially one of importance; a planned public or social
occasion"@en ;
skos:inScheme ex4: ;
skos:narrower ex4:[. . .] ;
skos:narrower ex4:russell-c-leffingwell-lecture ;
skos:narrower ex4:[]. . . ;
skos:note "source: https://languages.oup.com/google-
dictionary-en"@en ;
skos:prefLabel "defence_or_civil_event"@en .
ex4:russell-c-leffingwell-lecture rdf:type skos:Concept ;
dcterms:identifier
"http://example.com/periclesRepo/ex4/russell-c-leffingwell-
lecture"^^xsd:anyURI ;
skos:broader ex4:defence_or_civil_event ;
skos:definition "Russell C. Leffingwell Russell C.
Leffingwell Century Society Archives The Russell C.
Leffingwell Lecture, inaugurated in 1969, was named for a
charter member of CFR who served as its president from
1944 to 1946 and as its chairman from 1946 to 1953. This
lecture is given by a distinguished foreign official who is
invited to address CFR members on a topic of major
international significance. The lectureship was originally
endowed by the Morgan Guaranty Trust Company and by
Edward and Lucy Leffingwell Pulling, and more recently
through the generosity of Thomas Leffingwell Pulling and
his son Edward Leffingwell Pulling"@en ;
skos:inScheme ex4: ;
skos:note "source: https://www.cfr.org/project/russell-c-
leffingwell-lecture-series"@en ;
skos:prefLabel "russell c. leffingwell lecture"@en .
```

Below, example of relations between concepts based on SKOS and FOAF in TTL serialization:

```
ex4:proper_name rdf:type skos:Concept ;
dcterms:identifier
"http://example.com/periclesRepo/ex4/proper_name"^^xsd:
anyURI ;
skos:broader ex4:person ;
skos:definition "A proper noun is the name of a particular
person, place, organization, or thing. Proper nouns begin
with a capital letter"@en ;
skos:inScheme ex4: ;
skos:narrower ex4:[. . .] ;
skos:narrower ex4:jens-stoltenberg ;
skos:narrower ex4:[. . .] ;
skos:note "source:
https://www.collinsdictionary.com/dictionary/english/prope
r-
noun#:~:text=A%20proper%20noun%20is%20the,Compare
%20common%20noun."@en ;
skos:prefLabel "proper_name"@en .
ex4:jens-stoltenberg rdf:type skos:Concept ;
dcterms:identifier
"http://example.com/periclesRepo/ex4/jens-
stoltenberg"^^xsd:anyURI ;
skos:broader ex4:proper_name ;
skos:definition "Jens Stoltenberg is a Norwegian politician
who has served as the 13th secretary general of NATO
since 2014"@en ;
skos:inScheme ex4: ;
skos:note "source:
https://en.wikipedia.org/wiki/Jens_Stoltenberg"@en ;
skos:prefLabel "jens stoltenberg"@en ;
skos:related ex4:[. . .] ;
skos:related ex4:joe-biden ;
skos:related ex4:[. . .] ;
foaf:knows ex4:[. . .] ;
foaf:knows ex4:joe-biden ;
foaf:knows ex4:[. . .] ;
foaf:name ex4:jens-stoltenberg ;
foaf:person ex4:jens-stoltenberg .
```

Interestingly, the Narralive research team extends the CoVEd's graph functionality with a social network-like visualization for relations between concepts expressed as "skos:related", Figure 3.

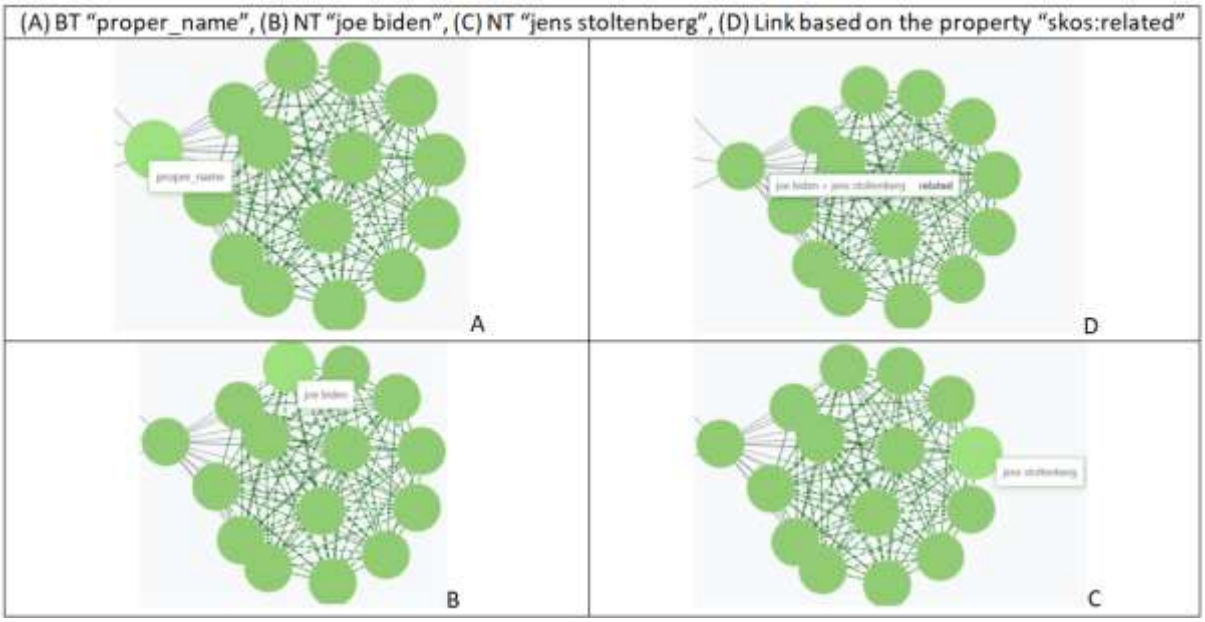


Fig. 3: The matrix represents extracts of relations between concepts. Relations are expressed as "skos:related"

The "graph" feature in CoVEd provides users with an additional functionality to visualize skos encoded relations in a social network-like depiction, Figure 3. Here, "graph" is leveraged to illustrate relations between concepts that are hierarchically below "proper_name". The nodes represent concepts. The dotted threads link two concepts. (A) Node that illustrates the higher level concept "proper_name"; (B) and (C) Nodes that illustrate narrower concepts, here, the proper names of NATO leaders; (D) Link that illustrates relation between two concepts. The network representation is used to demonstrate potential professional relations between individuals within the organizational context of the Alliance.

### 3.1 Controlled Vocabulary Assessment

To assess the quality of our SKOS vocabulary, we use the SKOS Play Testing Tool, which is frontend for qSKOS, a tool for finding quality issues in SKOS vocabularies, [24], [25], [26]. By default, are selected twenty four rules that correspond to quality checking functions to assess the vocabulary, [27]. The output is a report that states that 224 concepts are processed. Of the 24 rules that are checked, three failed to be verified. Two of the three failed rules are warnings such as "uc - Undocumented Concepts" and "urc - Unidirectionally Related Concepts". The warnings indicate that there are thirty one concepts that are found not to use any SKOS documentation properties, and two concepts that do not include reciprocal relations. The one fail found is the "var - Valueless Associative Relations" rule. It refers to finding sibling concept pairs that are also connected by an associative relation. The following are excerpts from warnings:

uc:http://example.com/periclesRepo/ex4/democratic_action
urc:http://example.com/periclesRepo/ex4/farid-safarov
var:http://example.com/periclesRepo/ex4/jane-harman

This assessment of the vocabulary quality is an initial step to determine its suitability for reuse and extensibility in the development processes. Furthermore, the checking functions flag potential quality problems that can be interpreted as structural errors. Or, they can hamper vocabulary integration due to concept incompatibility and inconsistencies

### 3.2 Discussion

The preliminary text analysis of our sample of NATO press releases provides an overview of the volume of information pertaining to civil and military affairs. Yet, text analysis is the groundwork for capturing and structuring knowledge that could be leveraged by computational models. Moreover, text analysis aids in emphasizing the grammatical form of the terms that we are interested in selecting. For example, the vocabulary should contain terms in one of the following grammatical forms: noun or noun phrase for instance, "alliance" or "council on foreign relations"; verbal noun for instance, "hosting a multinational nato battlegroup"; adjective or pre-modified phrase for instance, "humanitarian aid"; and, post-modified noun phrase for instance, "lessons learned on countering hybrid tactics" [28].

There are five levels in the hierarchical structure of this SKOS vocabulary. Conceptually, the last level corresponds to terms annotated during the NER-like procedure. Our polyhierarchical system for classifying terms in press releases is the source of these levels. The report on the analysis of our vocabulary that is generated by SKOS Play distinguishes between broad terms (BT) and narrow terms (NT) with "nato" being the top term (TT). Generally, polyhierarchy denotes that concepts can potentially belong to more than one category.

For relations between concepts, there are two main types of SKOS properties: associative and hierarchical. For example, the property "skos:broader" asserts that a concept has a general meaning and "skos:narrower" is the inverse property. The following excerpt describes hierarchical relations:

```
ex4: rdf:type skos:ConceptScheme ;
dcterms:identifier
"http://example.com/periclesRepo/ex4/"^^xsd:anyURI ;
skos:hasTopConcept ex4:nato ;
skos:prefLabel "russo-ukrainian-defence-diplomacy-
vocabulary"@en .
ex4:nato rdf:type skos:Concept ;
dcterms:identifier
"http://example.com/periclesRepo/ex4/nato"^^xsd:anyURI ;
skos:inScheme ex4: ;
skos:topConceptOf ex4: ;
skos:narrower ex4:[. . .] ;
skos:narrower ex4:democratic_action ;
skos:narrower ex4:[. . .] ;
skos:prefLabel "nato"@en .
ex4:democratic_action rdf:type skos:Concept ;
dcterms:identifier
"http://example.com/periclesRepo/ex4/democratic_action"^
^xsd:anyURI ;
skos:broader ex4:nato ;
skos:inScheme ex4: ;
skos:narrower ex4:peace_building_event ;
skos:prefLabel "democratic_action"@en .
ex4:peace_building_event rdf:type skos:Concept ;
dcterms:identifier
"http://example.com/periclesRepo/ex4/peace_building_even
t"^^xsd:anyURI ;
skos:broader ex4:democratic_action ;
skos:inScheme ex4: ;
skos:narrower ex4:defence_or_civil_event ;
skos:prefLabel "peace_building_event"@en .
ex4:defence_or_civil_event rdf:type skos:Concept ;
dcterms:identifier
"http://example.com/periclesRepo/ex4/defence_or_civil_ev
ent"^^xsd:anyURI ;
skos:broader ex4:peace_building_event ;
```

```
skos:definition "Event: a thing that happens or takes place,
especially one of importance; a planned public or social
occasion"@en ;
skos:inScheme ex4: ;
skos:narrower ex4:[. . .] ;
skos:narrower ex4:russell-c-leffingwell-lecture ;
skos:narrower ex4:[. . .] ;
skos:note "source: https://languages.oup.com/google-
dictionary-en"@en ;
skos:prefLabel "defence_or_civil_event"@en .
ex4:russell-c-leffingwell-lecture rdf:type skos:Concept ;
dcterms:identifier
"http://example.com/periclesRepo/ex4/russell-c-leffingwell-
lecture"^^xsd:anyURI ;
skos:broader ex4:defence_or_civil_event ;
skos:definition "Russell C. Leffingwell Russell C.
Leffingwell Century Society Archives The Russell C.
Leffingwell Lecture, inaugurated in 1969, was named for a
charter member of CFR who served as its president from
1944 to 1946 and as its chairman from 1946 to 1953. This
lecture is given by a distinguished foreign official who is
invited to address CFR members on a topic of major
international significance. The lectureship was originally
endowed by the Morgan Guaranty Trust Company and by
Edward and Lucy Leffingwell Pulling, and more recently
through the generosity of Thomas Leffingwell Pulling and
his son Edward Leffingwell Pulling"@en ;
skos:inScheme ex4: ;
skos:note "source: https://www.cfr.org/project/russell-c-
leffingwell-lecture-series"@en ;
skos:prefLabel "russell c. leffingwell lecture"@en .
```

The property “skos:related” expresses associative connection between concepts. For example, we use associative relations for concepts that are under the proper_name category. However, according to the W3C recommendation for SKOS vocabularies, it is not required to encode associative relations for pairs of concepts that share the same broader or narrower concept categories, [28], [29]. Nevertheless, in this phase, we intend to express links between individuals that are associated with the alliance under various professional capacities. Similarly, we use the “foaf:name” and “foaf:knows” properties to express relations between persons in the vocabulary.

The utilization of the properties “skos:definition” and “skos:note” aims to provide a better understanding of the meaning of a concept through general documentation and the corresponding source, [30]. Concerning limitations to the pilot phase of our research, we underline two top issues. First, the size of our data sample should be increased to full press releases. Potentially, a larger data set will provide additional terms for the vocabulary. Consequently, it provides an increased knowledge organization system (KOS). Second, the polyhierarchy of categories of the SKOS vocabulary is a structure that is under revaluation. The goal is to address potential complicatedness regarding broader-narrower concept relations. The next step is to analyze entire press releases to expand the current vocabulary.

Professional networks within the Treaty connected by FOAF relationships can present insights into characteristics and motifs of social networks in the Semantic Web. Here, the use of FOAF has the potential capacity to derive organizational relations from contextual information or domain knowledge through data mining techniques such as classification clustering, or logical inference.

# 4 Conclusion

This paper describes the second phase of the pilot project to create a Russo-Ukraine conflict controlled vocabulary. At this phase, CoVED and SKOS Play Test are utilized to express 224 terms as SKOS vocabulary and assess its structure. Of the twenty-four rules that test the vocabulary, one rule is marked as failed in the assessment report due to Valueless Associative Relations. The analysis of full press releases will be used as the next phase in this research to potentially increase the number of concepts in the vocabulary.The RDF-based foundation of the proposed vocabulary provides a nascent glimpse of an artificial intelligence (AI) system to define classifications for defense initiatives supporting peace-building civil and diplomatic actions.

*Acknowledgements:*
We thank Dr. N. Mastorakis at the Hellenic Naval Academy and Dr. N. Bardis at the Hellenic Army Academy for facilitating this research. Appreciations are also extended to Myrto Koukouli and the NARRALIVE Research Team for providing access to CoVEd and creating a repository for this project.

*References:*

[1] A. Miles and S. Bechhofer, *“SKOS Simple Knowledge Organization System Reference,”* W3C Recommendation, [Online]. https://www.w3.org/TR/skos-reference/ (Accessed Date: February 15, 2024).

[2] I. Yoo and X. Hu, “Clustering Large Collection of Biomedical Literature Based on Ontology-Enriched Bipartite Graph Representation and Mutual Refinement Strategy,” in *Advances in Knowledge Discovery and Data Mining*, in Lecture Notes in Computer Science. Berlin,

Heidelberg: Springer, 2006, pp. 303–312, https://doi.org/10.1007/11731139_36.

[3] D. Brickley and L. Miller, *"FOAF Vocabulary Specification,"* FOAF Vocabulary Specification, [Online]. http://xmlns.com/foaf/spec/ (Accessed Date: February 15, 2024).

[4] J. H. Gennari, A. Silberfein, and J. C. Wiley, *"Integrating genomic knowledge sources through an anatomy ontology,"* Pac. Symp. Biocomput. Pac. Symp. Biocomput., Fairmont Orchid on the Big Island of Hawaii, USA, vol. 10, pp. 115–126, 2005, ISBN 981-256-046-7.

[5] G. D. Solomou, D. A. Koutsomitropoulos, A. K. Kalou, and S. D. Botsios, "Semantify Educational Resources using SKOS and Learning Object Ontologies," *19th International Conference on Circuits, Systems, Communications and Computers (CSCC 2015)*, Zakynthos Island, Greece, pp. 360–365, 2015.

[6] NATO Press Office, "NATO - News: News," NATO's Newsroom, [Online]. https://www.nato.int/cps/en/natohq/news.htm (Accessed Date: February 15, 2024).

[7] C. Khancome, "String Matching Algorithm Using Multi-Characters Inverted Lists," *WSEAS Transactions on Computers*, Vol. 22, pp. 151-158, 2023, https://doi.org/10.37394/23205.2023.22.18.

[8] S. Oh, J. Jeong, C.-G. Lee, J. Yoo, and G. Nam, "Synergistic Training: Harnessing Active Learning and Pseudo-Labeling for Enhanced Model Performance in Deep Learning," *WSEAS Transactions on Computers*, Vol. 22, pp. 114-119, 2023, https://doi.org/10.37394/23205.2023.22.14.

[9] The R Foundation, "R version 4.2.2 beta," R: The R Project for Statistical Computing, [Online]. https://www.r-project.org/ (Accessed Date: February 15, 2024).

[10] IBM.com, "What is named entity recognition?, IBM," IBM, [Online]. https://www.ibm.com/topics/named-entity-recognition (Accessed Date: February 15, 2024).

[11] Label Studio, "Label Studio Documentation — Overview of Label Studio," Label Studio, [Online]. https://labelstud.io/guide/get_started.html (Accessed Date: February 15, 2024).

[12] DagsHub, "DagsHub: The Home for Machine Learning Collaboration," DagsHub, [Online]. https://dagshub.com/ (Accessed Date: February 15, 2024).

[13] P. S. Giannaris, J. Borges, and N. Doukas, "Design of a Defense Controlled Vocabulary, SKOS-based Encoding of Terms from Military Press Releases," in *2023 13th International Conference on Dependable Systems, Services and Technologies (DESSERT)*, Athens, Greece, Oct. 2023, pp. 1–6, doi: 10.1109/DESSERT61349.2023.10416475.

[14] S. Sivkov, "Residue Number Systems Quantization for Deep Learning Inference," *WSEAS Transactions on Computers*, Vol. 22, pp. 296-301, 2023, https://doi.org/10.37394/23205.2023.22.33.

[15] S. Saranya and G. Usha, "Emotion Classification on Social Media Comments Using Categorical Feature Extraction Along With the Bidirectional Encoder-based Recurrent Neural Network Classification," *WSEAS Transactions on Computers*, Vol. 22, pp. 284-295, 2023, https://doi.org/10.37394/23205.2023.22.32.

[16] Linked Heritage, "Terminology," Linked Heritage, [Online]. http://linkedheritage.cab.unipd.it/training/LO-07/en/02.html (Accessed Date: February 15, 2024).

[17] T. Baker, S. Bechhofer, A. Isaac, A. Miles, G. Schreiber, and E. Summers, "Key choices in the design of Simple Knowledge Organization System (SKOS)," *J. Web Semant.*, vol. 20, pp. 35–49, May 2013, doi: 10.1016/j.websem.2013.05.001.

[18] Koukouli, M., Katifori, A., Roussou, M., & Ioannidis, Y. (2024). Controlled Vocabulary Editor. SKOS vocabularies without SKOS training. In K. Karpouzis, Y. Skarpelos (eds.), *Computational Methods for the Digital Humanities*. Springer. https://doi.org/10.5281/zenodo.11080329.

[19] M. Koukouli, A. Katifori, and Y. Ioannidis, "SKOS vocabulary without SKOS training. A tool for everyone," presented at the *Symbiosis of Tradition and Digital Technology, Tallinn, Estonia*, May 22, 2022.

[20] CIDOC, "Symbiosis of Tradition and Digital Technology CIDOC conference abstracts MAY 23-26, 2022 Tallinn, Estonia." *CIDOC*, May 23, 2022. Accessed: Feb. 15, 2024, [Online]. https://vlaamsekunstcollectie.be/volumes/imported/Esinejate-abstraktid_0505.pdf (Accessed Date: February 15, 2024).

[21] A. Miles and D. Brickley, "SKOS Core Vocabulary Specification," SKOS Core Vocabulary Specification W3C Working Draft 2 November 2005, [Online]. https://www.w3.org/TR/swbp-skos-core-spec/#note (Accessed Date: November 19, 2024).

[22] D. Brickley and L. Miller, "FOAF Vocabulary Specification," FOAF Vocabulary Specification 0.99, [Online]. http://xmlns.com/foaf/spec/ (Accessed Date: February 15, 2024).

[23] T. Francart, "SKOS Play! - Thesaurus & Taxonomies," SKOS Play, [Online]. https://skos-play.sparna.fr/play/home (Accessed Date: February 15, 2024).

[24] T. Francart and C. Maden, "SKOS testing tool," [Online]. https://skos-play.sparna.fr/skos-testing-tool/ (Accessed Date: February 15, 2024).

[25] C. Mader, "cmader/qSKOS." Aug. 15, 2023, [Online]. Available: https://github.com/cmader/qSKOS (Accessed Date: February 15, 2024).

[26] C. Mader, B. Haslhofer, and A. Isaac, "Finding Quality Issues in SKOS Vocabularies." arXiv, Jun. 06, 2012.

[27] O. Suominen and C. Mader, "Assessing and Improving the Quality of SKOS Vocabularies," *J. Data Semant.*, vol. 3, no. 1, pp. 47–73, Mar. 2014, 2024, doi: 10.1007/s13740-013-0026-0.

[28] I. Spasić, D. Schober, S. A. Sansone, D. Rebholz-Schuhmann, D. B. Kell, N. W. Paton, "Facilitating the development of controlled vocabularies for metabolomics technologies with text mining." *BMC Bioinformatics*, 9 (Suppl 5), S5 (2008), pp.1-16, https://doi.org/10.1186/1471-2105-9-S5-S5.

[29] P. P. F. Barcelos, "Quality Issues," GitHub, [Online]. https://github.com/cmader/qSKOS/wiki/Quality-Issues (Accessed Date: February 15, 2024).

[30] E. Summers, A. Isaac, C. Redding, D. Krech, "LCSH, SKOS and Linked Data", Metadata for Semantic and Social Applications, Dublin Core: Innovation and Moving Forward, *DCMI '08: Proceedings of the 2008 International Conference on Dublin Core and Metadata Applications*, Berlin Germany September 22-28, 2008, pp.25-33, [Online]. https://library.oapen.org/bitstream/handle/20.500.12657/37029/1/DC_proceedings.pdf#page=41 (Accessed Date: April 26, 2024).

**Contribution of Individual Authors to the Creation of a Scientific Article (Ghostwriting Policy):**

The authors equally contributed tothe present research, at all stages from the formulation of the problem to the final findings and solution.

**Sources of Funding for Research Presented in a Scientific Article or Scientific Article Itself:**

No funding was received for conducting this study.

**Conflict of Interest:**

The authors have no conflicts of interest to declare.