Creating Proprietary Terms Using Lightweight Ontology: A Case Study on Acquisition Phase in a Cyber Forensic Process

Tamer Fares Gayed, Hakim Lounis
Dépt. d’Informatique
Université du Québec à Montréal
Succursale Centre-ville, Montréal QC H3C 3P8, Montréal, Canada
gayed.tamer@courrier.uqam.ca  lounis.hakim@uqam.ca

Moncef Bari
Dépt. de Didactique
Université du Québec à Montréal
Succursale Centre-ville, Montréal QC H3C 3P8, Montréal, Canada
bari.moncef@uqam.ca

Abstract—Terms and their meaning connections provided by the Resource Description Framework (RDF) present nowadays the standard mechanism for Linking Data (LD) on the web. All the existing terms, whether they are built-in terms (imported from well-known vocabularies on the semantic web) or proprietary terms (custom terms created by data publisher) can be used to describe and link different things in the world through RDF statements, and by applying the general architecture of the World Wide Web known as Linked Data Principles (LDP). Sometimes, these existing terms are not enough and adequate to describe a particular data set; more proprietary terms need to be created and developed in a dedicated vocabulary using lightweight ontology of LD. The latter uses the constructors of Resource Description Framework Schema (RDFS) and little features from Web Ontology Language (OWL) to create new proprietary terms describing such data set. This idea is depicted in this paper through a phase retrieved from a Cyber Forensic (CF) process, called acquisition phase, where different forensic tasks need to be described using new proprietary terms. This paper explains how these new proprietary terms can be created and published using the constructors of the lightweight ontology to describe this forensic phase.

Keywords—Linked Data; Linked Data Principles; Resource Description Framework Schemas; Web Ontology Language; Proprietary Terms; Cyber Forensic; State Preservation.

I. INTRODUCTION

Today, the WWW has radically altered the way to share information [1][3]. The interrelation is not just between documents, but it has evolved to link the data within these documents (i.e., Linked Data-LD) using the same aspects of web architecture (URI [4] and Hyper Text Transfer Protocol-HTTP [5]). This data is described and represented using different terms imported from well-known vocabularies on the semantic web, or from proprietary terms retrieved from custom vocabularies created by the data publisher (i.e., such terms representing real objects and abstract concepts in the world). HTTP URIs are not only used to identify and interlink web documents but also interlink such terms, the fact that allows the latter to be dereferenceable/resolvable (i.e., it means that HTTP clients can look up the URI using the HTTP and retrieve a description of term/resource that is identified by this URI).

Sometimes, the existing terms are not adequate, or there are no existing ontologies (vocabularies) containing terms describing a particular data set. Thus, new proprietary terms need to be developed in a dedicated vocabulary, applying the features of RDFS [9] and OWL [10] to describe this particular data set. However, before creating a new custom term, some aspects (criteria) should be taken into consideration:

1. Search for terms from widely used vocabularies that could be reused to describe the domain in interest. If the widely deployed vocabularies do not provide the required terms to describe such domain, so new terms should be defined as proprietary terms.
2. When you define a new term, you need to have a namespace that you own and control (i.e., unique namespace), in order to mint your new terms to this domain/namespace.
3. When you create new terms, you have to map these terms to those in existing vocabularies.
4. Apply the LDP to your new terms by using the web technology stack (HTTP, URL, and RDF [8]) and this task takes place along the publication process, starting from the identification of terms until their publication.
5. Label and comment each term you create.
6. If your term is a property (predicate), you have to define its domain and range using the constructors of RDFS and do not overload your new term with ontological axioms.
7. If at later time you discover that another term was enough, an RDF link should be set between the new created term and the existing one.

This paper resumes the work published in [7] (see Figure 1). In this work, a Cyber Forensics-Chain of Custody framework (CF-CoC) was proposed to transform the CoC from tangible documents into electronic data (e-CoC). This framework provided several layers to perform such task. The first two layers were about the creation of custom terms (i.e., forensic/victim terms) using lightweight ontology. This paper discusses how such terms are identified and defined using lightweight ontology in order to describe forensic data in order to be published later on the web of data.

The current paper is organized as follows: next section discusses the state of the art and the map between lightweight ontology and CF. Section 3 presents the different constructors of lightweight ontologies; constructors of RDFS and OWL. Section 4 explains how the first two layers of the

1 http://linkeddatabook.com/editions/1.0/
framework are designed and implemented, and how some forensic terms of the acquisition phase are defined. Finally, the last section concludes and summarizes this work.

II. STATE OF THE ART: ONTOLOGICAL CONCEPTS AND FORENSIC PROCESS

Publishing data on the web passes by three phases. It starts with identifying terms in the domain of interest. These terms are the things whose properties and relationships will be used later in the publication of data. Second phase, the identified terms are defined using different constructors of RDFS [9] and OWL [10], and uniquely named by HTTP URIs (i.e., as being explained in last section, a new term should be minted to a domain controlled and owned by the creator of the term). Finally, once terms are identified, defined, and named using HTTP URIs, they are then published on standardized contents formats. This format is the RDF [8] that provides a generic data model composed of a triple containing three slots: subject, predicate, and object, where the defined terms occupy one or more slot(s) in this triple. Because the RDF model does not provide any domain-specific terms for describing classes of things in the world and their interrelation, it allows the combination of schema languages such as RDFS/OWL [9][10], and different mixture of distinct RDF vocabularies. Publication of these created terms takes place through a publication form (i.e., the third layer of the framework).

Although, the process of selecting terms is subjective and depends on the term creator (i.e., we may have two creators selecting and identifying two different terms describing the same concept in the real world), this does not affect the quality of terms being published, because the LDP on the web of data make them self-descriptiveness. The latter advantage is due to two reasons:

- LDP with naming using HTTP/URIs, offer a dereferenceable nature to the term, so that any LD consumption applications can look up the RDFS/OWL definitions and retrieve more information about such term [11].
- LDP with some schema constructors (i.e., OWL) can map a new term to existing terms from well-defined vocabularies in the form of RDF links [12].

In a domain like CF, it is scarce to find forensic terms or well known vocabularies describing it, because it is still in its infancy and development. The most related work to define an ontology in CF, was published in [13], where an ontological model (i.e., with small “o”) was created for outlining CF tracks in the education process. Its aim was only to construct a hierarchical structure for classification of certification domains (i.e., the best convenient vocabulary to be used by the web of data to construct such type of ontology is the Simple Knowledge Organization System - SKOS [1]).

Thus, CF is a domain that requires the definition of new proprietary terms. In the CF-CoC framework, the need of creating new terms emerged from the objective to describe and represent tangible documents of CoC into electronic data.

As shown in Figure 3, each forensic phase will have a corresponding lightweight ontology. Each lightweight ontology has a set of \( n \) categories, which will be equivalent to \( n \) forensic tasks. A category in the vocabulary should be described using a set of \( m \) terms. These terms are the proprietary terms describing a forensic task.

This work considered the acquisition phase as an example to elaborate the idea of creating lightweight ontology with new proprietary forensic terms. [14][15][16][17][18].

---

2 http://www.w3.org/2004/02/skos/
Acquisition Phase: this phase is about acquiring digital evidences from digital suspected devices (e.g., small-scale devices, large-scale devices, etc.). It contains three forensics tasks: state preservation, recovering, and copying. The role player of this phase is the first responder [17][14].

- State preservation: the first task is saving the state of the digital device under question [19], by seizing the machine containing the suspected device.
- Recovery: after seizing the suspected device, the role player tries to recover all deleted files on the device, especially the system files that records valuable details about this suspected device.
- Copy: after recovering the deleted files, the first responder takes copy from the suspected device to avoid tampering and alteration.

From the above forensics tasks, the role player can start to determine different terms to describe his CoC. In this paper, the authors are those who identified some terms of acquisition phase (see Table 1).

<table>
<thead>
<tr>
<th>T-Box</th>
<th>Term name</th>
<th>Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>First_responder</td>
<td>Class</td>
<td></td>
</tr>
<tr>
<td>Role_player</td>
<td>Class</td>
<td></td>
</tr>
<tr>
<td>Acquisition</td>
<td>Ontology</td>
<td></td>
</tr>
<tr>
<td>Digital_media</td>
<td>Class</td>
<td></td>
</tr>
<tr>
<td>preserve</td>
<td>Property</td>
<td></td>
</tr>
<tr>
<td>preservedby</td>
<td>Property</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>A-Box</th>
<th>Term name</th>
<th>Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>ISN</td>
<td>Property</td>
<td></td>
</tr>
<tr>
<td>PDA_device</td>
<td>Subject/Object</td>
<td></td>
</tr>
<tr>
<td>04-4023-32-362</td>
<td>Object</td>
<td></td>
</tr>
</tbody>
</table>

The T-Box, A-Box, and the column type will be discussed in next section.

III. LIGHTWEIGHT ONTOLOGY: RDFS AND OWL CONSTRUCTORS

RDF is the standard format to create LD and it is sufficient to use the constructors of RDFS and a little feature of OWL. Combination of constructors from both vocabularies represents the lightweight ontology of RDF, known as RDFS++. Next subsections highlight all the RDFS constructors and some OWL primitive constructors to construct the first two layers of CF-CoC framework.

The RDFS and OWL constructors are classified according the term type (rdfs:class, or to be a property owl:objectProperty). This definition takes place before the term will be used (i.e., before its publication- T-Box). Later, the defined terms are used to describe and publish different data (A-Box, Assertion Box) [23]. The type of the term also determines its slot position during publication.

A. RDFS Constructors

The RDFS constructors are used to define terms and their relationships. Consider the term in question is named X.

B. OWL Constructors

The primitive selected from the OWL are mainly used to map between class and property terms.

IV. DESIGN AND IMPLEMENTATION: CASE STUDY ON ACQUISITION PHASE

A. Work Environment

The CF-CoC framework is implemented using Php and easyRDF, Graphiz tool, and its graph objects are used within the easyRDF to produce and draw different RDF models.

B. CF-CoC Terms Definitions

As shown in Figure 3, the first step is to create the ontology corresponding to a forensic phase. This ontology will contain all the forensic terms describing the different tasks of acquisition phase. Next figure shows the CF-CoC web application.

As shown in Figure 4, the CF-CoC web application contains several modules. The first module, “Proprietary Terms”, contains two main tasks: the creation of ontology and the creation of terms (Figure 5).

Figure 5 Tasks of Proprietary Terms Module

1) Creation of Ontology (Vocabulary):

This task is about to create the ontology object of the acquisition phase (see Figure 3). The domain name field is required to mint the ontology to a unique domain name owned by the publisher (aspect 2). The second field is about the selection of role player certificate [20]. In addition, the value type of the role player can be a resource or a literal. Next fields are the ontology name and its label description. Last field is the publication date of the acquisition ontology (see Figure 6).

After completing this form, the acquisition ontology is generated by using the Graphiz module [24] (see Figure 7).

Figure 6 Creation of Acquisition Ontology

Figure 7 Creation of Acquisition Ontology

2) Creation of terms (Acquisition Terms):

This task relates to four essential fields. The first field is the term name. The second field is selecting ontology to append the new proprietary term. The third field specifies the category/forensic task (see Figure 3). In our case, the category could be one of the three tasks provided in section 2 (preservation, recovery, or copy). In this field, the user may select ‘New’ to create a new category or select ‘Existing’ to import an existing category, defined in another vocabulary (ontology) created by another role player (see Figure 8) (i.e., two different forensic phase may have a common category/task [17]). Last field is the selection of term type (i.e., a term can be a property or a class).

Let now consider the following tangible CoC:

“The name of the first responder in the acquisition phase is Jean-Pierre. He is the role player of this phase, and he preserved the state of the digital media, PDA device, which has the SN: 0G-4023-32-362. The date he did this task is 5 March 2014”

Figure 8 Creation of a New Term
The first step to create an e-CoC from this tangible CoC is to identify the terms (see Table 1). This case study contains T-Box and A-Box information. Terms of T-Box are of type class and property. The Role_player term is a class that can be defined as a subclass from the class Person in the FOAF (friend of a friend) ontology \(^4\) (see Figure 9). This term will belong to a forensic task called Preservation.

The First_responder term is a class that can be an instance of the Role_player class. Now, the Preservation category will be found under the ‘Existing’ category. Finally, the Digital_media is a subclass of owl:Thing (see Figure 10).

Now, the property terms (owl:objectProperty) will be defined. The domain and range of the term preservedby are defined to be Digital_media and First_responder class, respectively. This property term is defined to be a sub-property from foaf:made property (see Figure 11).

The preserve property is the inverse of preservedby property. Thus, the domain and range of the former will be also the inverse, First_responder and Digital_media respectively. Simply, if a digital media is preserved by a first responder, then this means that the first responder preserved the digital media (see Figure 12). The last property is SN: the serial number of a device is an inverse functional property, because each serial number identifies one and only one subject (see Table 3).

After creating all terms, the role player can generate the acquisition ontology with all the property and class terms of the preservation forensic task (Figure 12).

The created T-Box terms will be used to publish data. Therefore, they will describe the A-Box data. The latter is the e-CoC that will be consumed later by jury in court of law using different consumption application (browsing [22] [21], searching [6] or querying [2]). Now, the data can be described and published using the terms defined in the T-Box and by using the third layer of the framework, the user can publish different triples (i.e., using different vocabularies of the semantic web) and by the support of proprietary vocabularies defined by the role player.

Figure 13 shows the e-CoC (A-Box) of the forensic preservation task. This generated ontology does not answer all the question of CoC. It answers only the Who: Jean-Pierre, What: PDA device, and When: publication date of ontology. In order to have the answers to other questions, more terms need to be determined and defined. In this figure, the cf-coc-Acq is the prefix namespace of the acquisition ontology: Jean-Pierre is an instance from the First_responder class (i.e., which is an instance of the Role_player class), PDA device is an instance of Digital_media (i.e., which is instance from Things class), and preservedby is the inverse property of preserve property. SN is a functional property where its domain is the PDA device and its range is the 0G4023-32-362 (i.e., which is an instance of the Literal class).

V. CONCLUSION AND FUTURE WORK

This paper discusses the idea of creating proprietary terms using lightweight ontology in order to publish and describe forensic information on the web of data. The work illustrates this idea through the preservation task of the acquisition forensic phase. It depicts how the first two layers of the CF-CoC framework are designed and implemented. Future work will discuss the remaining layers of this framework and how an e-CoC will be consumed by juries in a court of law.

\(^4\) http://www.w3.org/TR/owl-ref
http://xmlns.com/foaf/spec/
REFERENCES


---

* Labels and comments are reduced to increase the image clearness; RP: Role_player, FR: First_responder, and DM: Digital_media