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

Tamer Gayed
Hakim Lounis
Moncef Bari

July 2014
Agenda

- Introduction & Definitions
- Research Motivations
- Research Problems
- CF-CoC Framework
- Creating Proprietary terms
- Conclusions
Agenda

• Introduction & Definitions
• Research Motivations
• Research Problems
• CF-CoC Framework
• Creating Proprietary terms
• Conclusions
Introduction

Thesis title:

• Representing and Managing *Chain of Custody in the Cyber Forensics* using *Linked Data Principles*

Today’s discussions:

• Creating Proprietary terms using lightweight ontology: A case study on acquisition phase in a cyber forensic process
Introduction

What is the Cyber Forensics (CF) ?

• Is a technique for identifying, collecting, preserving, analyzing, and presenting digital evidence (DE) in a form useful to the court so that the cybercriminals face justice in the court of law (digital investigation).

• Thus, digital investigation is about investigate digital incidents to determine the root-cause of an incident and successfully prosecute a perpetrator.

• Each forensic phase is accomplished by a role player.
Introduction

What is the Chain of Custody (CoC) – Les chaînes de traçabilité?

• Is a chronological tangible document that accompanies each phase in the forensic process to answer 6 questions:
  • What
  • Why
  • When
  • Where
  • Who
  • How

• This known as the 5Ws and 1H
Introduction

Classical way to publish data on web

Web Aspects:
- URL
- HTTP
- HTML

SW Definitions
Introduction

Before and after 2006

• **Before 2006**, most of ontologies are published in dump files and most of them are not interlinked.

• **2006**: Tim Berners-Lee *underlined* set of rules to follow (Guidelines) for publishing data on the web inspired from the same principles of web aspects.

• Rules are:
  - Use **URIs** as names for things.
  - Use **HTTP** as universal access mechanism.
  - Include **RDF** statements that link to other URIs.
  - A query Language **SPARQL** can be used to provide useful information from the represented data.
Introduction

How this can be realized?

- URL
- HTTP
- HTML
- URI
- HTTP
- RDF
- SPARQL

SW Definitions
Introduction

Emerging of Linked Opened Data (LOD),

- **Oct 2007**, The LOD Project has been started.
Introduction

**LOD (Cont.)**

- *Linked Open Data (LOD) Project*: is the most visible project using the LDP (URLs, HTTP, and RDF).

- This project created a shift in the community of research and development of the semantic web.

- *Nowadays, the web is not* just concentrated for the interrelation between *web documents* but also between the *raw data* within these documents.

- Today, *the semantic web is a web of data*
Introduction

CF-CoC Web Application

Forensic Process

Role players

Jury

e-CoC

CF-CoC
Agenda

• Introduction & Definitions
• Research Motivations
• Research Problems
• CF-CoC Framework
• Creating Proprietary terms
• Conclusions
Agenda

• Introduction & Definitions
• Research Motivations
• Research Problems
• CF-CoC Framework
• Creating Proprietary terms
• Conclusions
Research Motivations

Why LDP to represent and manage CoC in CF?

- Similar Nature between LDP and CF:
  - Each forensic phase can lead to another.
  - LD allows the connection between different resources in different forensic phase.
  - Thus, LDP allow role players & juries to navigate between different forensic phases through the RDF typed links
Research Motivations

Why LDP to represent and manage CoC in CF ? (Cont)

• Linked data consumption applications are able to interpret any data even it is represented with unknown vocabulary :
  • URI dereferenceable
  • Mapping between URIs

⇒ All forensic data will be resolvable
Research Motivations

Why LDP to represent and manage CoC in CF? (Cont.)

- RDFS and OWL vocabularies can be used with RDF model allowing the subsumption and relationships between terms

⇒ Useful for juries to infer more information from the data
Research Motivations

Why LDP to represent and manage CoC in CF? (Cont.)

- Accompanied with different provenance metadata to provide the answer to other six questions, related to the data origin

⇒ Provenance metadata can be used concurrently with the published/forensic data to describe their provenance and complement the missing answers related to the forensics investigation.

5WS and 1H, on the level of data origin
Research Motivations

Why LDP to represent and manage CoC in CF? (Cont.)

• *LDP is a way to represent different forensic concepts and able to realize KR objectives*

⇒ Representation of data allows:

• Surrogate of concepts & Ontological commitments
• Medium of the Role player to express different details about forensic process
• RDF model is a standard language that avoid the ambiguity
Research Motivations

Why LDP to represent and manage CoC in CF? (Cont.)

• Investigation process is a common task between different role players (Social Environment)
  ⇒ LDP allow mapping between different terms in different forensic phases

  • Level of URIs
  • Level of terms
Research Motivations

Why LDP to represent and manage CoC in CF? (Cont.)

- Naming Resources using URI, allows its deferenceability

⇒ Forensic resources will be deferenceable (retrieve a description of term/resource that is identified by this URI), allow the jury to understand the resource in hand
Agenda

• Introduction & Definitions
• Research Motivations
• Research Problems
• CF-CoC Framework
• Creating Proprietary terms
• Conclusions
Agenda

• Introduction & Definitions
• Research Motivations
• Research Problems
• CF-CoC Framework
• Creating Proprietary terms
• Conclusions
Research Problems

Generally

• *role players*: Need to securely record, describe, and manage the results of their forensic investigation

• *Juries*: Need to understand and consume, securely, the digital evidences and take the proper decision about the provided information
Research Problems

We need a solution to solve the following issues:

• CoC need to undergo a radical transformation from tangible document into electronic data to not be only used by human, but also by machine.

• e-CoCs need to be secured since their publication by the role player till their consumption by the juries.

• Provenance of information is crucial to guarantee the trustworthiness and confidence of the information provided.

• Judges’ awareness and understanding the digital evidences are not enough to evaluate and take the proper decisions.
Agenda

- Introduction & Definitions
- Research Motivations
- Research Problems
- CF-CoC Framework
- Creating Proprietary terms
- Conclusions
Agenda

• Introduction & Definitions
• Research Motivations
• Research Problems
• CF-CoC Framework
• Creating Proprietary terms
• Conclusions
CF-CoC Framework

- PKI Approach
- Pattern Consumption Applications
- CF-CoC Web Application Form
  - Forensic Part
  - Victim Part
  - Semantic Web Vocabularies and Domain Light Weight ontologies

Provenance Information
Agenda

- Introduction & Definitions
- Research Motivations
- Research Problems
- *CF-CoC* Framework
- Creating Proprietary terms
- Conclusions
Agenda

- Introduction & Definitions
- Research Motivations
- Research Problems
- CF-CoC Framework
- Creating Proprietary terms
- Conclusions
First Three Layers of **CF-CoC**

*Creating Proprietary (Custom) – Using Lightweight Ontology*

- **Why ?**
  - Terms of the semantic web are not enough/adequate to describe certain data set or a new domain context.
  - => New Proprietary terms need to be defined

- Lightweight ontology of LD (Linked Data) is the RDFS++.

- RDFS++ combines the RDFS constructors and some primitives constructors from OWL.

- The primitives constructors imported from OWL are those which are used to equivalent and map between different class and property terms
First Three Layers of CF-CoC

Creating Proprietary (Custom) - Forensic Term

• 7 Commandments to create new terms on the LD:
  – Don’t create a term if an existing one will suffice.
  – When you define a new term, you need to have a namespace that you own and control.
  – When you create new terms, it is recommended to map these terms to those in existing vocabularies.
  – Apply all the LDP (HTTP, URL, and RDF) to the term.
  – If your term is a property (predicate), you have to define its **domain** and **range** using the constructors of RDFS++ (RDFS, OWL Primitives) and do not overload your new term with ontological axioms
  – 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.
  – Label and comment each term you create
First Three Layers of \textit{CF-CoC}

\textit{Creating Proprietary (Custom) - Forensic Term}

- 7 \textit{Commandments to create new terms on the LD}:
  - Don’t create a term if an existing one will suffice.
  - When you define a new term, you need to have a namespace that you own and control.
  - When you create new terms, it is recommended to map these terms to those in existing vocabularies.
  - Apply all the LDP (HTTP, URL, and RDF) to the term.
  - If your term is a property (predicate), you have to define its \textit{domain} and \textit{range} using the constructors of \texttt{RDFS++ (RDFS, OWL Primitives)} and do not overload your new term with ontological axioms.
  - 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.
  - Label and comment each term you create.
### RDFS++ - RDFS Constructors

#### Common Constructors between Property and Class terms

<table>
<thead>
<tr>
<th>Property Term</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>rdfs : comment</td>
<td>Any term should have a comment. A comment is used to provide a human-readable description of a resource. Comment is an instance of rdf : Property</td>
</tr>
<tr>
<td>rdfs : label</td>
<td>Any term should have a label. A label is used to provide a human-readable name for a resource. Label is an instance of rdf : Property</td>
</tr>
</tbody>
</table>

#### If X is a term of type ( rdf : type ) Property ( rdfs : Property / owl : ObjectProperty )

<table>
<thead>
<tr>
<th>Property Term</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>rdfs : subPropertyOf</td>
<td>When the term X is of type property it can be also a sub property of another property term. The subPropertyOf of a property term is a term of type Property</td>
</tr>
<tr>
<td>rdfs : range</td>
<td>The range of a property term is always a Class. A range of a property term X states that the object slot of the X (i.e., where X is a predicate, because X is a property), interpreted by a reasoners as an instance of said range of X</td>
</tr>
<tr>
<td>rdfs : domain</td>
<td>The domain of a property term is always a Class. A domain of a property term X states that the subject slot of the X (i.e., where X is a predicate, because X is a property), interpreted by a reasoners as an instance of said domain of X</td>
</tr>
</tbody>
</table>

#### If X is a term of type ( rdf : type ) Class ( rdfs : Class )

<table>
<thead>
<tr>
<th>Property Term</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>rdfs : subClassOf</td>
<td>When the term X is of type Class, it can bee also a sub class of another Class term. The subClassOf of a property term is a term of type Class</td>
</tr>
</tbody>
</table>
When the type (\(\text{rdf:type}\)) of a property term \(X\) is defined to be of \(\text{InverseFuntionalProperty}\), Whenever \(X\) property is used as a predicate in a triple, its object will have one and only one subject. Thus, each object should be able to uniquely identify a subject. This constructor is a sub class of \(\text{owl : objectProperty}\).

Two URI terms can be mapped together using the \(\text{sameas}\) constructor. This constructor indicates that these two terms actually refer to the same thing. It can be used as well to map between two ontologies.
Creating Proprietary terms

- Creating the forensic phase ontology
- Determine the forensic tasks
- Identification of terms
- Creating/Defining of terms
Creating Proprietary terms

**Forensic Process**

<table>
<thead>
<tr>
<th>Example : Kruse Model</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Acquisition</strong> : acquiring evidence from suspect storage devices</td>
</tr>
<tr>
<td><strong>Authentication</strong> : ensuring that the acquired evidence has not been altered and kept its integrity</td>
</tr>
<tr>
<td><strong>Analysis</strong> : examining the data in order to identify pieces of evidence and determine their significance</td>
</tr>
</tbody>
</table>

**Acquisition Phase** containing 3 forensics tasks:

- State Preservation
- Recovery
- Copying
Creating Proprietary terms

• Example: The Tangible CoC of a state preservation

“The first responder name of 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 21 Feb 2014”
Creating Proprietary terms

- **Identification of terms**

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

<table>
<thead>
<tr>
<th>A- Box (publication)</th>
<th>Term name</th>
<th>Type</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Jean-Pierre</td>
<td>Subject/Object</td>
</tr>
<tr>
<td></td>
<td>PDA-device</td>
<td>Subject/Object</td>
</tr>
<tr>
<td></td>
<td>0G-4023-32-362</td>
<td>Object</td>
</tr>
</tbody>
</table>
Creating Proprietary terms

• **Creation of Acquisition Object**

**Definition of Acquisition Ontology**
• **Creation of terms**

**Lightweight ontology of Forensic Preservation task**
Lightweight ontology of Forensic Preservation task (Cont)
Lightweight ontology of Forensic Preservation task (Cont)
https://127.0.0.1/roleplayer/Acquisition#Jean-Pierre

PDA device

0G-4023-32-362

e-CoC of Forensic Preservation
State
Agenda

- Introduction & Definitions
- Research Motivations
- Research Problems
- CF-CoC Framework
- Creating Proprietary terms
- Conclusions
Agenda

- Introduction & Definitions
- Research Motivations
- Research Problems
- CF-CoC Framework
- Creating Proprietary terms
- Conclusions
Conclusions

1. This work guarantees the construction of a complete e-CoC, where forensic data and documentary data can be integrated under a unified framework.

2. PKI approach is used to ensure the identity of each role player participating in the forensics process and bending the LDP to a small scale (Notion of LCD).

3. Adding provenance metadata to each e-CoC NG, answers the questions related to the origin of information

4. This framework will foster the subject matter content and provide descriptive chains of custody
Thanks

Any Question?