Deliverables 5.2

Proposals for standardization

DEFRAUDify - Detecting Fraudulent activities on the internet

Author: F.J. Rumph (Web-IQ)

Date: September 2023

Status: Final

Version: 1.1

Table of Contents

[1 Table of Contents 2](#_Toc145881407)

[2 Acronyms 3](#_Toc145881408)

[3 Introduction 5](#_Toc145881409)

[4 Concerns of Cyber Threat & Fraud Intelligence standards 6](#_Toc145881410)

[4.1 Composability 6](#_Toc145881411)

[4.2 Data modelling aspects 6](#_Toc145881412)

[4.2.1 Information domain coverage 6](#_Toc145881413)

[4.2.2 Provenance 6](#_Toc145881414)

[4.2.3 Time dimension 6](#_Toc145881415)

[5 Relevant cyber security standards 7](#_Toc145881416)

[5.1 Cyber threat intelligence standards 7](#_Toc145881417)

[5.1.1 STIX and TAXII 7](#_Toc145881418)

[5.1.2 MISP 7](#_Toc145881419)

[5.2 Linked data and semantic web standards 7](#_Toc145881420)

[5.2.1 RDF and related standards 7](#_Toc145881421)

[5.2.2 UCO and CASE 8](#_Toc145881422)

[5.2.3 ANITA 8](#_Toc145881423)

[5.2.4 Lifting CTI standards to RDF/OWL 8](#_Toc145881424)

[5.2.5 Interpol DW-VA 9](#_Toc145881425)

[6 Semantic web and RDF in DEFRAUDify with UCO 10](#_Toc145881426)

[6.1 Extensions for the Unified Cybersecurity Ontology 10](#_Toc145881427)

[6.1.1 Identity 10](#_Toc145881428)

[6.1.2 Location 11](#_Toc145881429)

[6.1.3 Employment 12](#_Toc145881430)

[6.1.4 Social media and accounts/profiles 12](#_Toc145881431)

[6.1.5 Risk scores 12](#_Toc145881432)

[6.1.6 Honey tokens 14](#_Toc145881433)

[6.2 Unified Cybersecurity Ontology usability in DEFRAUDify 18](#_Toc145881434)

[6.2.1 Information domain coverage 18](#_Toc145881435)

[6.2.2 Composability 18](#_Toc145881436)

[6.2.3 Provenance 18](#_Toc145881437)

[6.2.4 Time dimension 19](#_Toc145881438)

[6.2.5 General usability observations 19](#_Toc145881439)

Acronyms

|  |  |
| --- | --- |
| **Acronym** | **Explanation** |
| CASE | Cyber-investigation Analysis Standard Expression |
| CTD | Cyber Threat Dashboard |
| CTI | Cyber Threat Intelligence |
| CybOX | Cyber Observable eXpression |
| EMCDDA | European Monitoring Centre for Drugs and Drug Addiction |
| FIBO | Financial Industry Business Ontology |
| FIBO FND | FIBO Foundations |
| JRC | EU Joint Research Centre |
| MISP | Malware Information Sharing Platform |
| OMG  | Object Management Group |
| OWL | Web Ontology Language |
| RDF | Resource Description Language |
| RDFS | RDF Schema |
| SHACL | Shapes Constraint Language |
| SKOS | Simple Knowledge Organization System |
| STIX | Structured Threat Information Expression |
| TAXII | Trusted Automated Exchange of Intelligence Information |
| UCO | Unified Cybersecurity Ontology |
| W3C | World Wide Web Consortium |
| XML | Extensible Markup Language |
| XSD | XML Schema Definition |

Introduction

As illustrated within deliverable D4.1, a diverse set of tools has been developed over various use cases around Cyber Threat & Fraud Intelligence. The breadth of the use cases and the number of tools to be connected calls for interoperability beyond ad-hoc point-to-point connections. This document describes relevant standards and provides considerations for further standardization from the DEFRAUDify project.

The need for interoperability between Cyber Threat & Fraud Intelligence tools furthermore transcends the needs of individual businesses. E.g., in relation to the financial use case within DEFRAUDify, Interpol et al identified several measures to combat cyber-enabled financial crimes around virtual assets and darknet service providers including intelligence sharing, having OSINT capabilities as well as transaction monitoring.

Also, the European Commission has identified the need for EU-coordinated Darknet monitoring to counter criminal activities with the goal to reinforce the capacities in this area of EMCDDA and Europol. For it, a preparatory action was initiated executed by the EU Joint Research Centre (JRC). Members of the DEFRAUDify consortium attended a data model workshop in April 2023 hosted by the JRC on this topic.

This document provides an overview of concerns for Cyber Threat & Fraud Intelligence information exchange, in particular pertaining to conceptual data modelling, in section 4. Next, a relevant standardization is described in section 5. For the Cyber Threat Dashboard, an RDF based architecture was selected as was described in DEFRAUDify deliverable 3.2. Section 6 details how its initial ontology can be mapped to the Unified Cybersecurity Ontology (UCO) and how it can be extended. Also, it provides a more general discussion on its usability from the DEFRAUDify perspective.

Concerns of Cyber Threat & Fraud Intelligence standards

* 1. Composability

As mentioned in the introduction, use cases in Cyber Threat & Fraud Intelligence can cover a plethora of interrelated domains and may need to be supported by multiple tools. For example, the Cyber Threat Dashboard combines various data-sources and tools:

* Darknet data
	+ Web-IQ – Voyager
	+ CFLW – Dark Web Monitor
* Clear-web data
	+ Web-IQ – IRIS
	+ TNO – Web OSINT deduplication
* Honey-token data
	+ Almende – Honeypot/-token
* Attack surface intelligence
	+ TU/e
* Corporate data

And even within these data sources several domains can be identified as described in deliverable D3.2. Due to the diversity of these domains, it is necessary to use standards that are not only extensible but also are composable.

* 1. Data modelling aspects

Information domain coverage

First and foremost, integration of Cyber Threat & Fraud Intelligence tools requires the ability to express information about e.g., attack patterns, identities, targets, threat actors, etc.

Provenance

Next to support for the various concepts related to cyber threats and fraud, in this context provenance metadata is of importance. In particular when fusing data from various (external) sources including darknet and clear-web data, data from other organizations, etc. the origins of the data is of importance for auditing but also analysis purposes. Also, chain of evidence/custody is an important element in the cyber security domain.

Time dimension

Related to provenance is time metadata. Various time dimensions can be discerned:

* Event time – when a particular event happened.
* Observation time – when a particular data point has been observed.

Relevant cyber security standards

* 1. Cyber threat intelligence standards

STIX and TAXII

Structured Threat Information Expression[[1]](#footnote-2) (STIX) is a standard used for cyber threat intelligence (CTI) exchange. In version 2, it integrated Cyber Observable eXpression[[2]](#footnote-3) (CybOX) for expressing information about cyber observables. STIX covers 18 domain object types:

* Attack Pattern
* Campaign
* Course of Action
* Grouping
* Identity
* Indicator
* Infrastructure
* Intrusion Set
* Location
* Malware
* Malware Analysis
* Note
* Observed Data
* Opinion
* Report
* Threat Actor
* Tool
* Vulnerability

TAXII[[3]](#footnote-4) is a protocol for suited for transporting STIX information in a variety of architectures. This set of standards is adopted in the CTI industry by various vendors.

MISP

MISP[[4]](#footnote-5) is a similar standard as STIX/TAXII as well as a software package for sharing CTI. With MISP users can share events (e.g., from an incident) and various communities exist. MISP can be extended using taxonomies.

A notable taxonomy is the Dark Web taxonomy[[5]](#footnote-6) that allows expressing information in various domains:

* Topic – such as hacking, weapons, violence, etc.
* Motivation – such as wiki, forum, file-sharing, etc.
* Structure – such as incomplete, captcha, police-notice, etc.
* Service – such as URL, content-type, port, etc.
* Content – html, page-title, credit card numbers, etc.
	1. Linked data and semantic web standards

RDF and related standards

The CTI standards and taxonomies mentioned above have a limited scope whereas e.g., the DEFRAUDify Cyber Threat Dashboard aims to provide a much wider domain of information to be fused. The semantic web and linked data technologies standardized at W3C provide mechanisms to share information and build knowledge graphs.

Resource Description Framework (RDF) and related technologies such as RDF schema, OWL and SHACL provide a data model that is extensible / flexible to allow for evolution of individual tools and the data that they process and produce whilst at the same time provide a strong platform for interoperability for concepts shared between tools.

UCO and CASE

At the start of the DEFRAUDify project, the Unified Cybersecurity Ontology (UCO) and Cyber-investigation Analysis Standard Expression (CASE) were identified as relevant standards. This was supported also by the call from Europol’s European Cybercrime Centre (EC3) to industry to adopt these standards[[6]](#footnote-7). UCO and CASE can be considered an evolution on other standards such as CybOX.

The UCO ontology supports representing information across the cyber security domain including concepts such as person and organisation, victim, location, device, file, account, message, etc. It aims to be a foundational ontology covering concepts that are shared across various application sub-domains. UCO uses RDFS (RDF schema), OWL (Web Ontology Language) and SHACL (Shapes Constraint Language) for defining the ontology.

UCO strives to support extension both at the level of the ontology itself, but also allows expression of information using open ‘shapes’. Aspects that deal with an individual application domain are to be dealt with in separate ontologies, such as CASE. The CASE ontology allows expression of information on cyber investigations covering concepts such as time and status information on an investigation.

Recently, UCO and CASE have been brought together. CASE is now part of the Linux foundation.

ANITA

The Horizon 2020 project ANITA[[7]](#footnote-8) (Advanced Tools for fighting online illegal trafficking) has developed an ontology focused on online marketplaces for illicit goods. It describes various general marketplace concepts such as actors, shops, products, sales, production, etc. Also, it describes more specific concepts such as substances, precursors and weapons and provides a set of instances for them.

Lifting CTI standards to RDF/OWL

Various initiatives exist to ‘lift’ CTI standards to RDF and/or OWL. E.g., Ulicny et al[[8]](#footnote-9) describe how an OWL ontology can developed based on other standards such as STIX. Also, members of the OASIS Threat Actor Context technical committee have developed the TAC ontology[[9]](#footnote-10), also mapping STIX to OWL. No practical use of such standards is known within the DEFRAUDify project.

Interpol DW-VA

Interpol has developed taxonomies related to dark web and virtual assets[[10]](#footnote-11). It is comprised of various concepts. The terms in the taxonomy are related through ‘broader concept’ relationships. The top-level taxonomy is:

* Actor – authorities, shop owner, etc.
* Asset – cryptocurrency, credentials, etc.
* Infrastructure – dark web, decentralized apps, relay node, wallet, etc.
* Process – multi sig, exit scam, yield farming, etc.
* Service – bulletproof hosting, coin swapping, decentralized exchange, etc.
* Technology – NFT, proof of stake, metadata, smart contract

Interpol indicates the taxonomy is intended to extend the UCO and CASE standards. The taxonomy is provided as Simple Knowledge Organization System (SKOS) schema.

Semantic web and RDF in DEFRAUDify with UCO

The adoption of semantic web and RDF standards in DEFRAUDify was explored for the Cyber Threat Dashboard use case. A project specific ontology was developed to allow for experimentation with RDF without the weight of a standards such as UCO and CASE. The DEFRAUDify ontology describes various types of entities (e.g., persons and organizations), their properties (their names, screen names from their social media profiles, etc.) as well as their relationships (e.g., a person being employed by an organization). This ontology was initially described in D3.2.

* 1. Extensions for the Unified Cybersecurity Ontology

Many of the concepts required for the CTD use case were found in the UCO ontology. In this section concepts and other elements for DEFRAUDify are discussed that had no direct mapping and extensions for UCO are provided.

This section contains code listings using the Turtle syntax[[11]](#footnote-12) for RDF and in some occasions the extensions for Turtle from the RDF-star[[12]](#footnote-13) draft specification[[13]](#footnote-14).

Identity

Person details such as name, birth date and visa can be expressed with UCO, but there is no predicate available for gender. The following listing provides the suggested datatype to be introduced. It uses the UCO approach for open vocabularies. I.e., suggested/default values are given for normalization, but are open for extension at time of use.

vocabulary:GenderVocab
 a rdfs:Datatype ;
 rdfs:label "Gender Vocabulary"@en-US ;
 rdfs:comment "Defines an open-vocabulary for gender."@en-US ;
 owl:equivalentClass [
 a rdfs:Datatype ;
 owl:onDatatype xsd:string ;
 owl:oneOf (
 "Male"^^vocabulary:GenderVocab
 "Female"^^vocabulary:GenderVocab
 "Other"^^vocabulary:GenderVocab
 ) ;
 ] ;
 .

1. UCO extension for gender information

Location

Most location related predicates used for the CTD have a direct mapping to UCO such as street, locality, postal code as well as latitude / longitude coordinates. Also, the same structure of location objects with various ‘groups’ of location attributes (address and coordinate based) was used. However, a few properties had no mapping:

Country code

While UCO supports country names, it would for normalization purposes be ideal to add support of country codes e.g., based on the ISO 3166 series of standards.

Open vocabulary

Following the design principles of UCO this could be provides as an open vocabulary for country codes. A distinction could be made between two- and three-letter codes, however creating separate vocabularies would probably reduce usability while not substantially improving the rigor of the ontology. The vocabulary definition would be like the gender vocabulary and is not listed here for brevity.

External ontology

To relieve UCO from the burden of specifying country codes, even as an open vocabulary, it could also be an option to use an external ontology such as the Languages, Countries and Codes specification[[14]](#footnote-15) from the Object Management Group (OMG). This also allows users of information expressed with UCO to also use information on country entities under the knowledge graph adage *things, not strings*.

@prefix lcc-cr: <https://www.omg.org/spec/LCC/Countries/CountryRepresentation/> .

[

 sh:class core:Facet ;

 sh:or (

 lcc-cr:Alpha2Code,

 lcc-cr:Alpha3Code

 )

 sh:path observable:countryCode ;

]

1. Implementation in UCO would entail extending the SimpleAddressFacet shape with an extra property.

Instead of adding an addition predicate, the range of the existing country predicate can be extended to accept both strings as well as country (code) entities.

Opaque street address strings

UCO contains a street predicate; however, no predicates exist for full street address strings. Such opaque strings are undesirable from a data usage viewpoint, but sometimes e.g., when using online sources such information is not available in a structured form. Attempts may of course be performed to (automatically) decompose such strings, but this may fall short and often it is useful to retain the original information. Street address strings can be added to UCO in the same way as street names are expressed.

House number

UCO surprisingly does not seem to support expressing house numbers or related information. Adding such a predicate can be simply added to UCO as a predicate with a string range for the SimpleAdressFacet type. Inspiration can be taken from the Financial Industry Business Ontology (FIBO) Foundations Domain (FND) ontology for addresses[[15]](#footnote-16). Here this street address component is named *primary address number*.

Use of FIBO FND

More generally, the address related parts of UCO could borrow from the FIBO FND ontology. For all components of SimpleAddressFacet a mapping exists to FIBO FND except for addresType. However, such re-design would probably have a difficult migration path if possible.

Employment

UCO lacks support for expressing employment relationships between organizations and individuals. As for address information, UCO could be extended using the FIBO FND ontology for formal organizations[[16]](#footnote-17). This ontology has the required notions of employment, employee, and employer.

However, for the CTD use case in DEFRAUDify additional information needs to be expressed:

* Start and end dates of the employment.
* Position of the employee at the employer, e.g., software developer or CEO.

Social media and accounts/profiles

Information on social media accounts/profiles can be expressed for DEFRAUDify using the Profile and (Digital)Account types from UCO. Some additional elements are required though:

* Screen/display name of a profile. The name predicate from the UCO core part seems much too broad for this type of information.
* Biography/description for a profile. The UCO core description predicate can be used for this purpose, but again, the use for (social media) profiles has more specific meaning as it is provided through the account itself instead.

Risk scores

Based on the attack surface risk score tool developed in DEFRAUDify, risk scores can be reported for persons. UCO provides the Analysis type and shape which has no further constraints. The attack-surface risk scores could be expressed as such and would require extra predicates for the risk score and an associated timestamp.

The listing below provides a possible extension to UCO for this use case:

:AttackSurfaceRiskAnalysisResult

 a

 owl:Class ,

 sh:NodeShape

 ;

 rdfs:subClassOf core:UcoObject ;

 rdfs:label "AttackSurfaceRiskAnalysisResult"@en ;

 rdfs:comment "The result of an attack surface risk analysis action."@en ;

 sh:property

 [

 sh:datatype xsd:decimal ;

 sh:maxCount "1"^^xsd:integer ;

 sh:nodeKind sh:Literal ;

 sh:path :riskScore ;

 ]

 ;

 sh:targetClass :AttackSurfaceRiskAnalysisResult ;

 .

:riskScore

 a owl:DatatypeProperty ;

 rdfs:label "riskScore"@en ;

 rdfs:comment "The risk score value of an attack surface risk analysis."@en ;

 rdfs:range xsd:decimal ;

 .

1. Class and node shape for attack surface risk analysis.

Examples of how the risk analysis results were previously expressed within DEFRAUDify and using the suggested UCO extension are provided below.

@prefix p: <http://itea4.org/defraudify/ontology/v1/persons#> .

@prefix src: <http://itea4.org/defraudify/ontology/v1/sources#> .

@prefix attack-surface: <http://itea4.org/defraudify/ontology/v1/attack-surface#> .

:person a p:Person .

:person attack-surface:riskScore "0.13"^^xsd:decimal {|

 src:source [ src:timestamp "2023-08-01T11:06"^^xs:dateTime ]

|}

1. Attack surface risk score example using the original DEFRAUDify ontology.

The listing above expresses a person instance with an associated attack-surface risk score. The timestamp of the risk score is associated to the risk score through the RDF-star annotation syntax[[17]](#footnote-18).

The listing below expresses the same information using UCO with the extensions described above. Note that blank nodes are used only for brevity.

@prefix action: <https://ontology.unifiedcyberontology.org/uco/action/> .

@prefix analysis: <https://ontology.unifiedcyberontology.org/uco/analysis/> .

@prefix core: <https://ontology.unifiedcyberontology.org/uco/core/> .

@prefix identity: <https://ontology.unifiedcyberontology.org/uco/identity/> .

:person a identity:Person .

[

 a analysis:AnalyticResult ;

 analysis:originatingAnalysis [

 a analysis:Analysis ;

 action:object :person ;

 action:startTime "2023-08-01T11:06"^^xsd:dateTime

 ] ;

 analysis:resultContent [

 a :AttackSurfaceRiskAnalysisResult ;

 :riskScore "0.13"^^xsd:decimal

 ]

] .

1. Attack surface risk score example using UCO with DEFRAUDify extensions.

Honey tokens

One of the tools developed in the DEFRAUDify project manages so called honey tokens and registers them being triggered. E.g., honey tokens can be (hidden) URLs that (when accessed) may indicate fraudulous or otherwise malicious intent. UCO has no direct support for such concepts. Extension of UCO using the DEFRAUDify ontology can be achieved by introducing honey tokens as ObservableObject and the trigger events as Action.

The following listing provides these as classes and node shapes following the UCO ontology:

@prefix core: <https://ontology.unifiedcyberontology.org/uco/core/> .

@prefix observable: <https://ontology.unifiedcyberontology.org/uco/observable/> .

@prefix xsd: <http://www.w3.org/2001/XMLSchema#> .

@prefix : <tag:example> .

:HoneyToken

 a

 owl:Class ,

 sh:NodeShape

 ;

 rdfs:subClassOf observable:ObservableObject ;

 rdfs:label "HoneyToken"@en ;

 rdfs:comment "A honey token is fake data that looks attractive to cybercriminals

 and contains a marker for tracking."@en ;

 sh:targetClass :HoneyToken ;

.

:Trigger

 a

 owl:Class ,

 sh:NodeShape

 ;

 rdfs:subClassOf observable:ObservableObject ;

 rdfs:label "Trigger"@en ;

 rdfs:comment "A trigger is an event of (the marker of) a honey token being

 accessed."@en ;

 sh:targetClass :Trigger ;

.

:TriggerFacet

 a

 owl:Class ,

 sh:NodeShape

 ;

 rdfs:subClassOf core:Facet ;

 rdfs:label "TriggerFacet"@en ;

 rdfs:comment "A trigger facet is a grouping of characteristics unique to a the

 triggering of a honey token"@en ;

 sh:property

 [

 sh:datatype xsd:dateTime ;

 sh:maxCount "1"^^xsd:integer ;

 sh:nodeKind sh:Literal ;

 sh:path core:startTime ;

 ]

 ;

 sh:targetClass :TriggerFacet ;

.

1. UCO extension for honey tokens based on the DEFRAUDify ontology.

The listing below provides an example instantiation using the original DEFRAUDify ontology for a profile for which a honey token is deployed and a single trigger with an IPv4 address and an associated location.

@prefix accounts: <http://itea4.org/defraudify/ontology/v1/accounts#> .

@prefix honey-tokens: <http://itea4.org/defraudify/ontology/v1/honey-tokens#> .

@prefix locations: <http://itea4.org/defraudify/ontology/v1/locations#> .

@prefix wgs84: <http://www.w3.org/2003/01/geo/wgs84\_pos#> .

:profile-1

 a accounts:Account .

:token-1

 a honey-tokens:HoneyToken ;

 honey-tokens:tracks :profile-1 .

:token-1-event-1

 a honey-tokens:TriggerEvent ;

 core:hasFacet [

 a :TriggerFacet ;

 honey-tokens:triggered :token-1 ;

 honey-tokens:timestamp "2023-08-01T11:06"^^xsd:dateTime ;

 honey-tokens:sourceAddress <urn:ipv4:172.16.17.18> ;

 locations:location [

 locations:coordinate [ wgs84:lat "53.3602"^^xsd:float ;

 wgs84:long "5.2143"^^xsd:float ]

 ]

 ]

.

1. Example expression of honey token information in the DEFRAUDify ontology.

The following listing conveys the same information as the listing above but mapped to UCO with extensions derived from the DEFRAUDify ontology.

@prefix core: <https://ontology.unifiedcyberontology.org/uco/core/> .
@prefix location: <https://ontology.unifiedcyberontology.org/uco/location/> .
@prefix observable: <https://ontology.unifiedcyberontology.org/uco/observable/> .

:profile-1 a observable:Profile .

:token-1 a :HoneyToken ;

[
 a core:Relationship ;
 core:source :token-1 ;
 core:target :profile-1 ;
 core:kindOfRelationship "tracks"
] .

1. Example expression of honey token information using UCO with extensions based on the DEFRAUDify ontology.

An example of the trigger information for a honey token is provided in the following listing.

:token-1-event-1
 a :TriggerEvent ;
 core:startTime "2023-08-01T11:06"^^xsd:dateTime .

:token-1-event-1-ip-address
 a observable:IPv4Address ;
 core:hasFacet [ a observable:IPv4AddressFacet ;
 observable:addressValue "172.16.17.18" ] .

[
 a core:Relationship ;
 core:source :token-1-event-1 ;
 core:target :token-1-event-1-ip-address ;
 core:kindOfRelationship "triggered-by"
] .

:token-1-event-1-location
 a location:Location ;
 core:hasFacet [ a location:LatLongCoordinatesFacet ;
 location:latitude "53.3602"^^xsd:float ;
 location:longitude "5.2143"^^xsd:float ] .

[
 a core:Relationship ;
 core:source :token-1-event-1 ;
 core:target :token-1-event-1-location ;
 core:kindOfRelationship "triggered-from"
] .

1. Example expression of honey token information using UCO with extensions based on the DEFRAUDify ontology.

Possibly, a more compact mapping to UCO is possible by not using relationship objects to relate the trigger event with the IP address and location. An example of this is provided in the listing below. However, based on the description of at least the ipAddress predicate[[18]](#footnote-19), it is debatable whether this is intended usage.

:token-1-event-1
 a :TriggerEvent ;

 core:hasFacet [

 core:startTime "2023-08-01T11:06"^^xsd:dateTime ;

 observable:ipAddress [

 a observable:IPv4Address ;
 core:hasFacet [

 a observable:IPv4AddressFacet ;
 observable:addressValue "172.16.17.18"

 ]

 ] ;

 observable:location [

 a location:Location ;
 core:hasFacet [

 a location:LatLongCoordinatesFacet ;
 location:latitude "53.3602"^^xsd:float ;
 location:longitude "5.2143"^^xsd:float

 ]

 ]

 ] .

1. Example expression of honey token information using UCO with extensions based on the DEFRAUDify ontology without use of Relationship objects.
	1. Unified Cybersecurity Ontology usability in DEFRAUDify

Information domain coverage

As can be seen from the previous sections, UCO requires only a few extensions to support the DEFRAUDify Cyber Threat Dashboard use case.

Composability

While re-use of existing ontologies is considered a best practise for linked data[[19]](#footnote-20), this is not always easily achieved. UCO does not use other ontologies to express (sub) domains that have support in other domains such as addresses and locations while there are strong ontologies available in this area. The composability of UCO itself and UCO compatible tools with other ontologies may also be hindered by the particular design principles maintained by UCO[[20]](#footnote-21).

Provenance

Provenance in the sense of tracking evidence, and other metadata in (cyber) criminal investigations is covered by the CASE ontology. Tracking derivations of data with UCO can be expressed using the analysis (sub)ontology.

Time dimension

Time metadata in UCO can be expressed through the core startTime and endTime predicates for various objects. Also, relationships in UCO can be given a (bi-)temporal dimension using the aforementioned predicates.

General usability observations

During the mapping of the DEFRAUDify ontology and Cyber Threat Dashboard UCO some usability issues where identified.

Complexity of facets

The facet concept creates an extra indirection in RDF between domain objects and things stated about these objects. Within the context of DEFRAUDify the advantages of this indirection haven’t been experienced yet.

Open vocabularies versus ‘things’

The notion of open vocabularies within UCO allows for run-time extensibility. This suits tools that don’t deal directly with an RDF representation. However, in terms of data quality and model usability, it can be beneficial to support extensibility through custom *instances*. This allows users to create new instances and share them across investigations, applications, etc. and add information for these instances themselves. *Things, not strings* can be seen as one of the main drivers between the development of technologies such as RDF and related standards.

String relationship types versus predicates / types

In a similar vein as for open vocabularies, the use of a string typed predicate to indicate the kind of the relationship seems brittle. While the need for the ability to further describe relationships beyond the participating nodes, the current design may lead to confusion with its users. RDF-star could address this issue, although it must be considered that this specification is not yet completed.

Relying on usage of opaque strings allows for extensibility, but it is also a possible pitfall for interoperability and data quality. A possible alternative design strategy could be to allow for user-defined instances as also suggested in the previous paragraph.

Followup and conclusion

Elements from existing standards (mainly UCO) were taken as a basis for implementations in DEFRAUDify. Completely new standardization proposals were not created, but based on experiences, some small improvements can be suggested towards UCO. Project partner TNO is involved in the CASE/UCO working group and will consider to suggest these proposals.

1. <https://oasis-open.github.io/cti-documentation/stix/intro.html> [↑](#footnote-ref-2)
2. <https://cybox.mitre.org/about/> [↑](#footnote-ref-3)
3. <https://oasis-open.github.io/cti-documentation/taxii/intro.html> [↑](#footnote-ref-4)
4. <https://www.misp-project.org> [↑](#footnote-ref-5)
5. <https://www.misp-project.org/taxonomies.html#_dark_web> [↑](#footnote-ref-6)
6. <https://www.europol.europa.eu/media-press/newsroom/news/eu-forensic-experts-call-for-action-new-cyber-investigation-standard> [↑](#footnote-ref-7)
7. <https://www.anita-project.eu> [↑](#footnote-ref-8)
8. Inference and ontologies; Ulicny, Moskal, Kokar, Abe, Smith; 2014 [↑](#footnote-ref-9)
9. <https://github.com/oasis-tcs/tac-ontology> [↑](#footnote-ref-10)
10. <https://interpol-innovation-centre.github.io/DW-VA-Taxonomy/> [↑](#footnote-ref-11)
11. <https://www.w3.org/TR/turtle/> [↑](#footnote-ref-12)
12. RDF-star is an extension for RDF being standardized by the W3C to allow for expressing RDF statements with triples as subject and/or object. [↑](#footnote-ref-13)
13. <https://w3c.github.io/rdf-star/cg-spec/editors_draft.html#turtle-star> [↑](#footnote-ref-14)
14. <https://www.omg.org/spec/LCC/> [↑](#footnote-ref-15)
15. <https://spec.edmcouncil.org/fibo/ontology/FND/Places/Addresses/> [↑](#footnote-ref-16)
16. <https://spec.edmcouncil.org/fibo/ontology/FND/Organizations/FormalOrganizations/> [↑](#footnote-ref-17)
17. <https://w3c.github.io/rdf-star/cg-spec/editors_draft.html#dfn-annotation-syntax> [↑](#footnote-ref-18)
18. *Specifies the corresponding ip address for a whois entry. Usually corresponds to a name server lookup.* taken from <https://github.com/ucoProject/UCO/blob/master/ontology/uco/observable/observable.ttl>. [↑](#footnote-ref-19)
19. <https://www.w3.org/TR/ld-bp/> [↑](#footnote-ref-20)
20. <https://unifiedcyberontology.org/resources/uco_design_document.html> [↑](#footnote-ref-21)