COM DIG 10. Classificazioni, tassonomie, ontologie nella cybersecurity

Elisabetta Zuanelli
Università degli Studi «Tor Vergata»

Le attività CSIRT (Nagashima et alii, «CSIRT Activities at NTT» , NTT technical Review, 2010, pp.1-7)

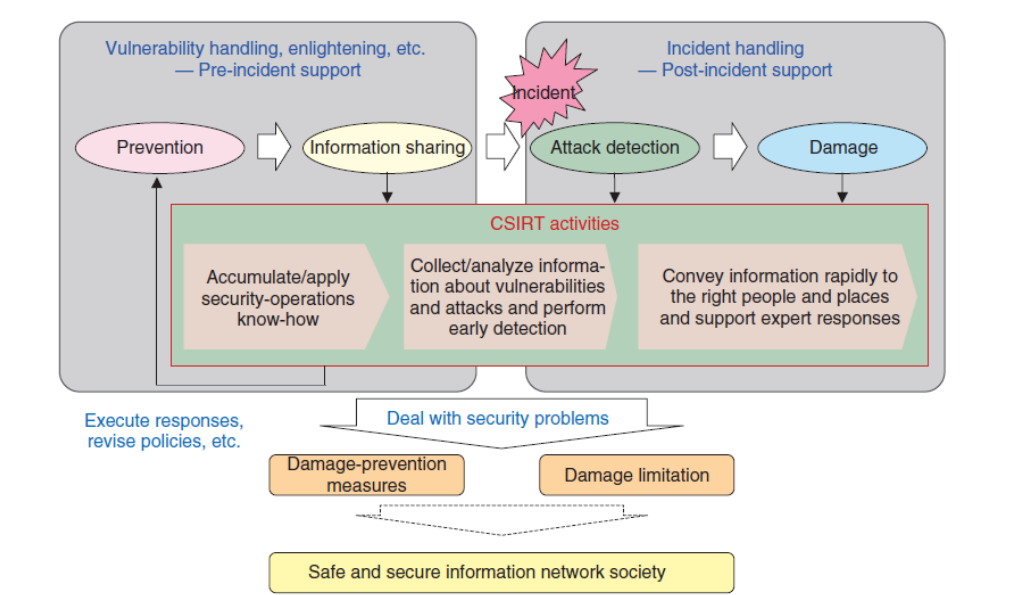
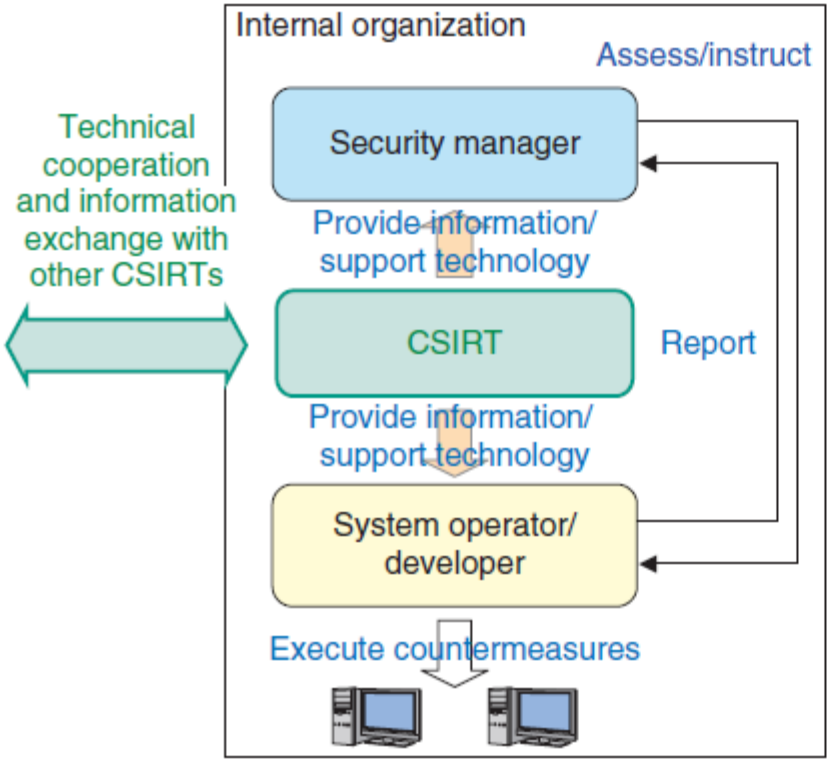


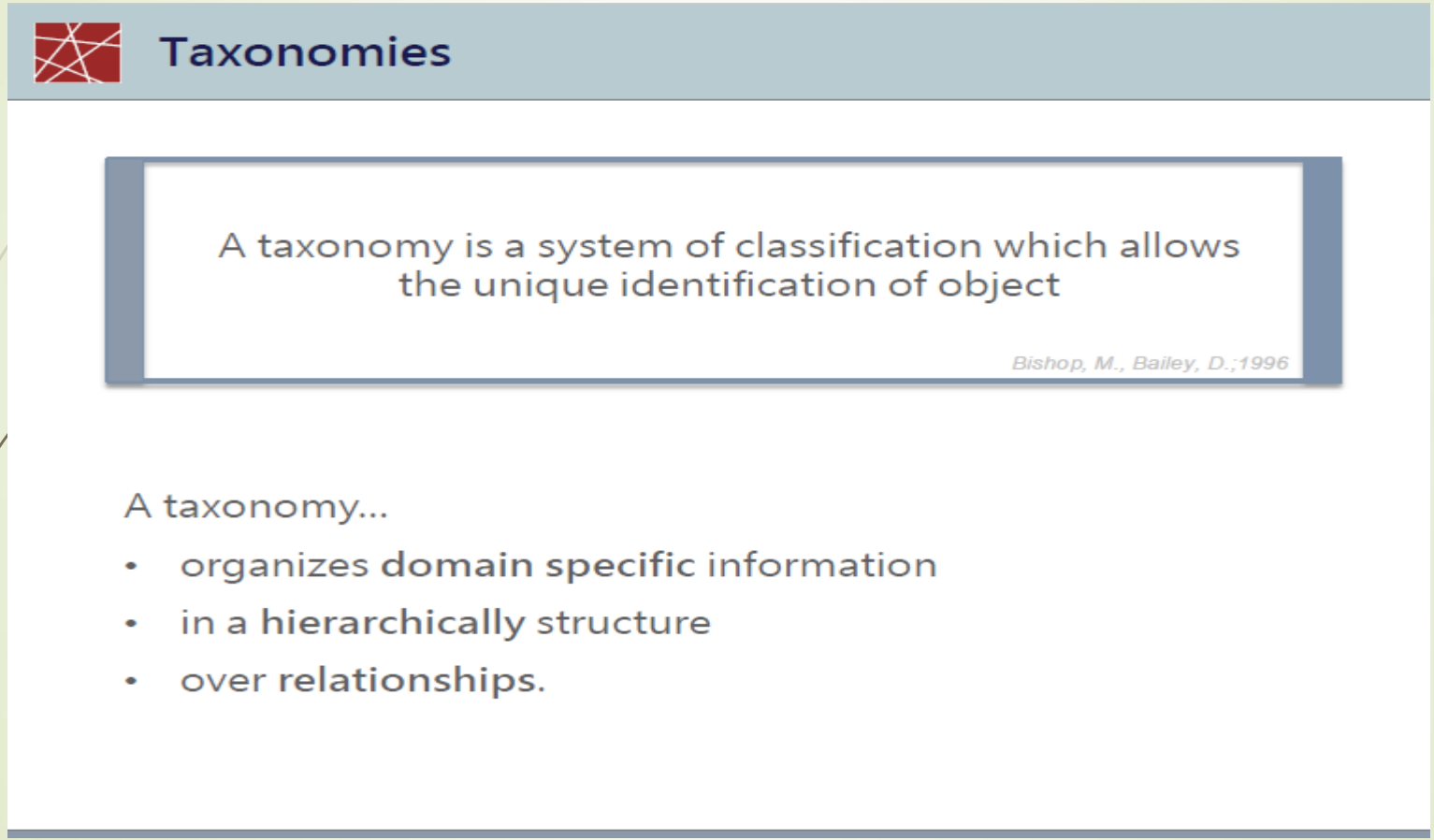
Fig. 1. CSIRT activities.



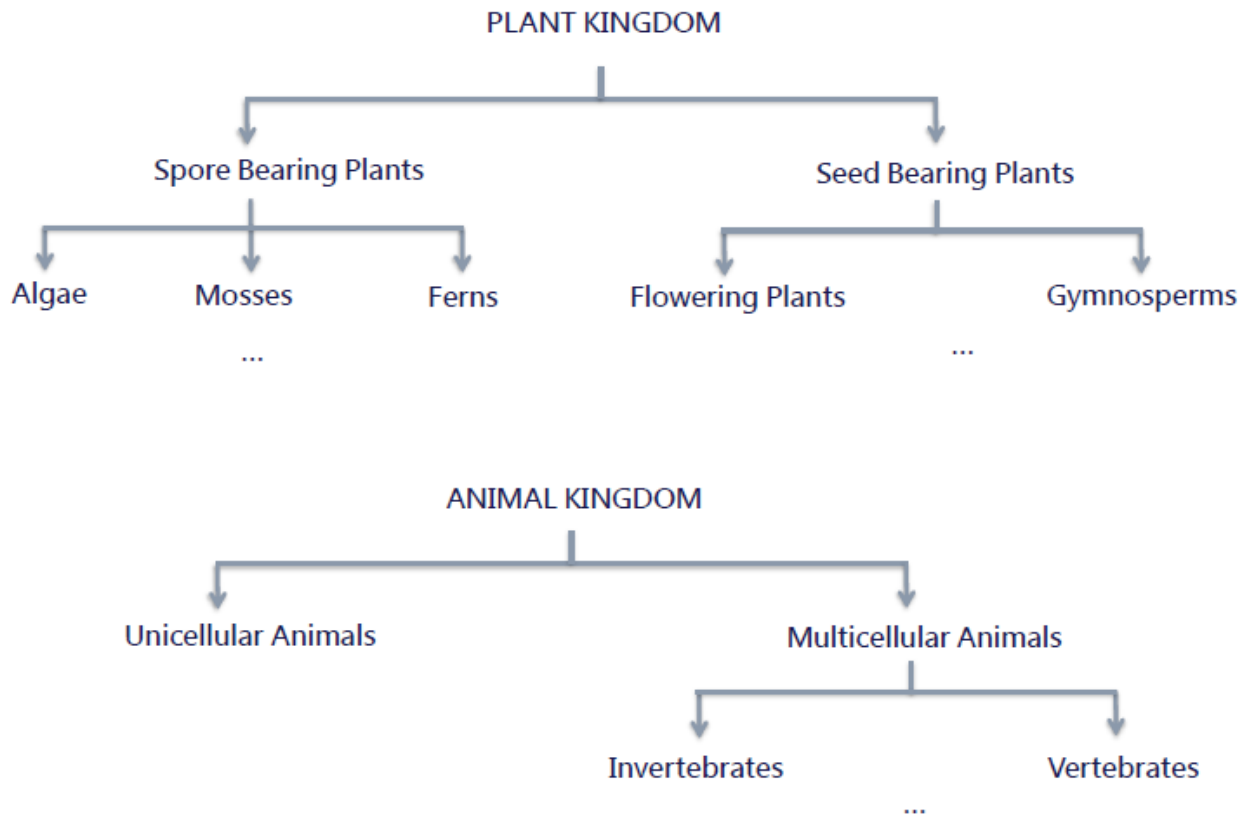
Function	Main roles
Security manager	Assess/instruct responses, manage countermeasures
CSIRT	Coordinate responses, support technology, collect/analyze/provide information
System operator/developer	Execute countermeasures (apply patches, upgrade versions, modify settings, etc.)

Fig. 2. CSIRT roles.

Natascha Abrek 2014 (Lehrstuhl Netzarchitekturen und Netzdienste Institut für InformatikTechnische Universität München)

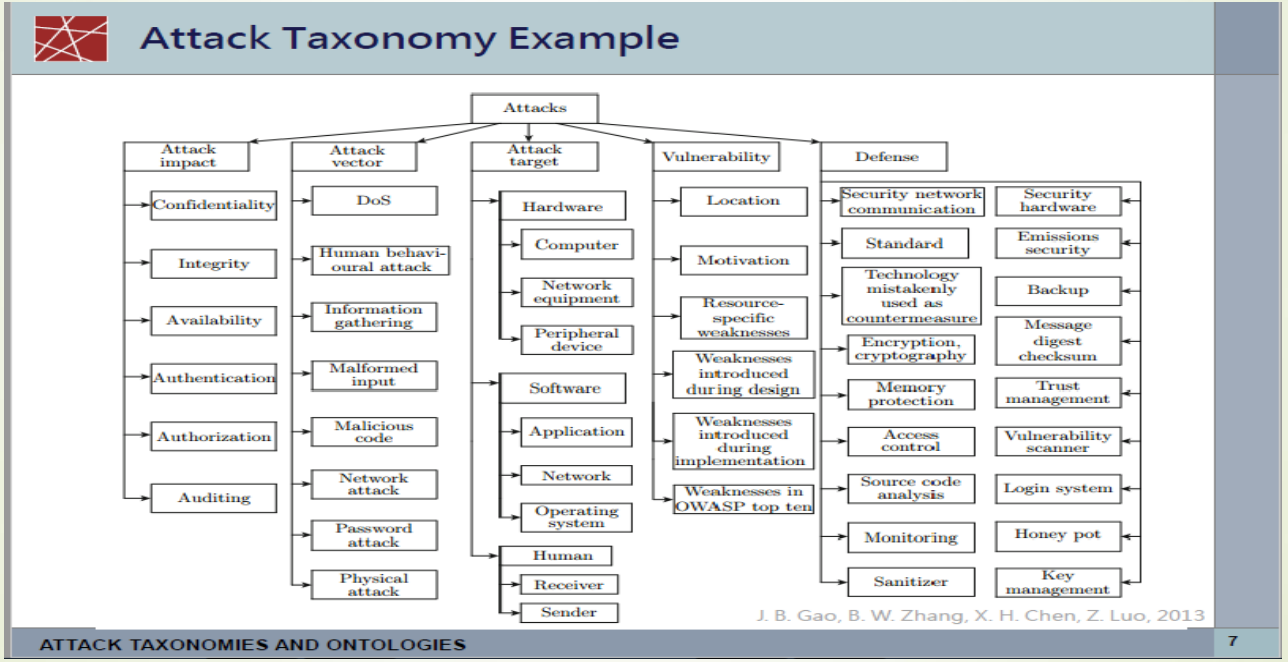


Well-known Taxonomies

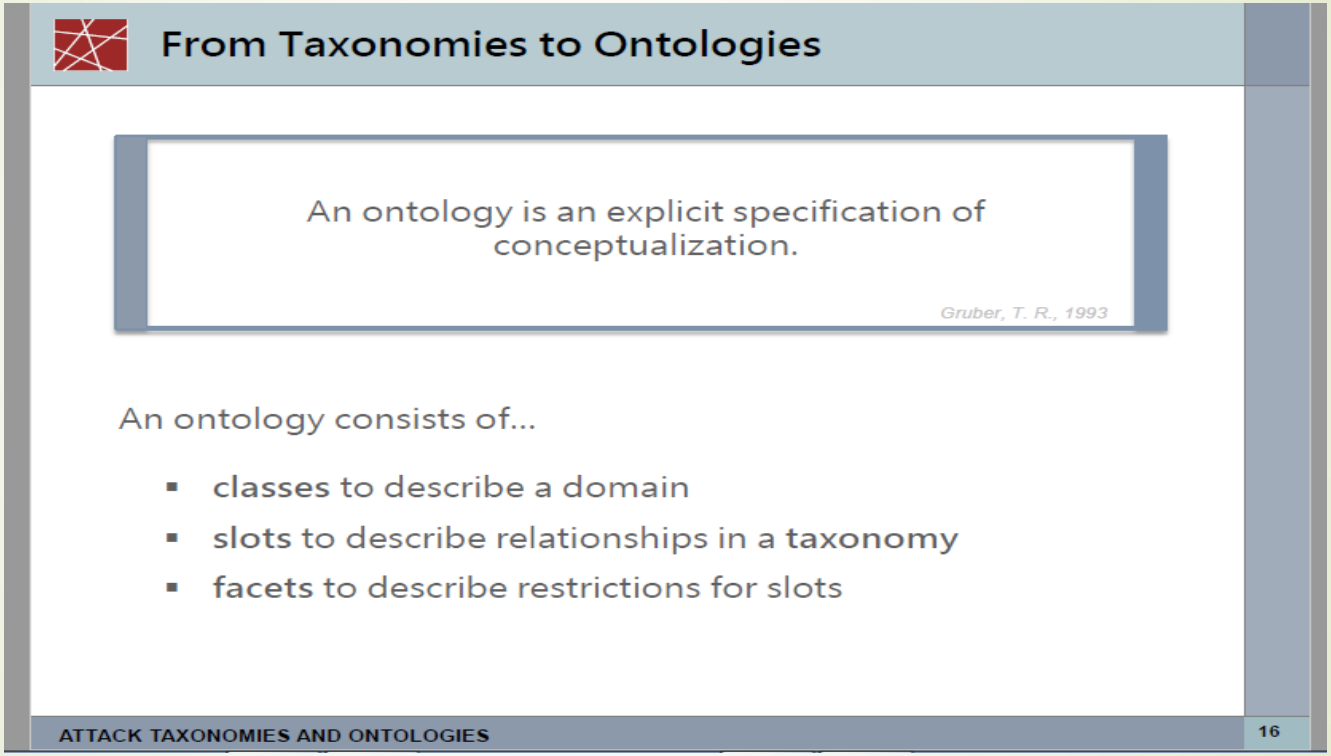


Tassonomie e menù

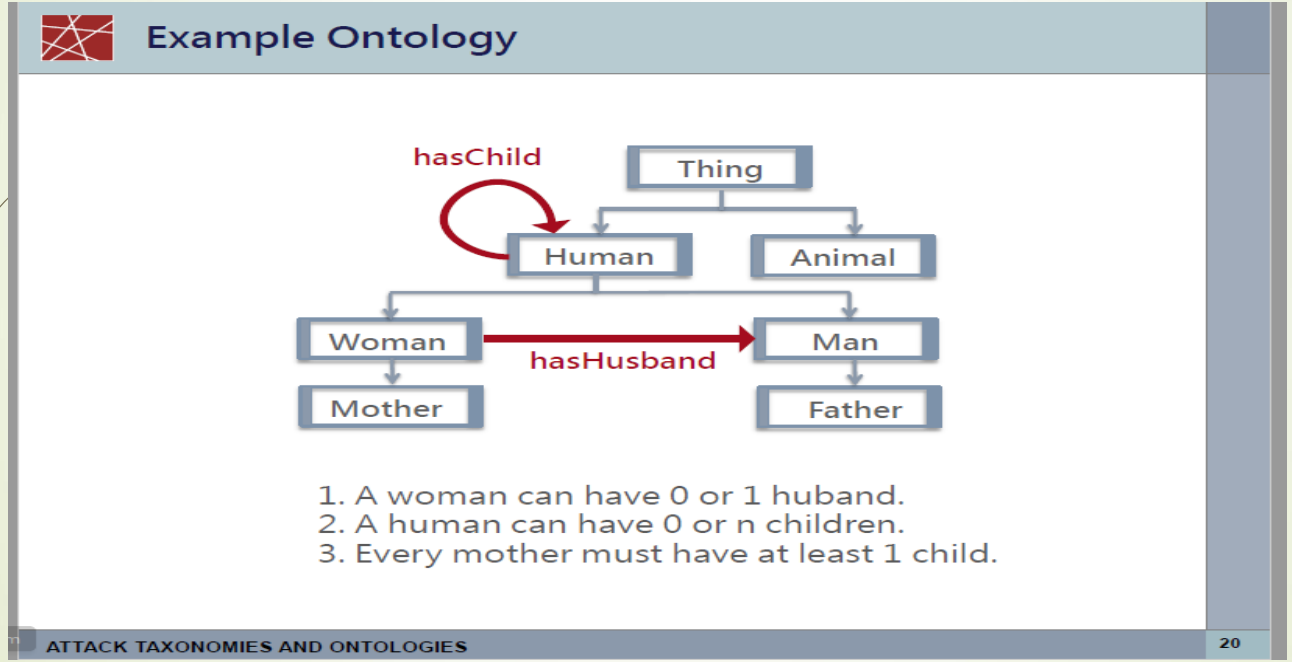
Esempio di tassonomia degli attacchi



Ontologie

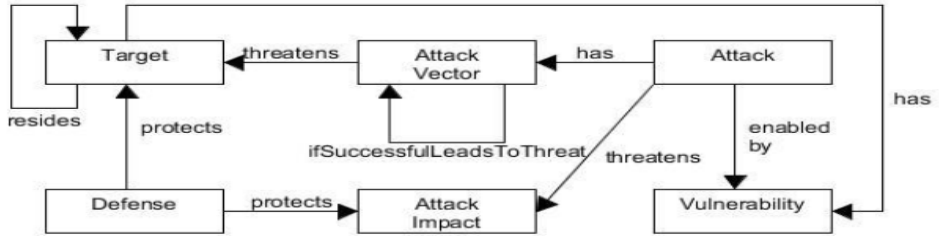

From Taxonomies to Ontologies
An ontology is an explicit specification of conceptualization.
Gruber, T. R., 1993
An ontology consists of...
classes to describe a domain
slots to describe relationships in a taxonomy
facets to describe restrictions for slots
ATTACK TAXONOMIES AND ONTOLOGIES
16

Esempio di ontologia



Esempio di ontologia dell'attacco

Attack Ontology Example



The SQL Slammer **is** a computer worm and **has** the attack vectors buffer overflow and denial of service. The attack **is enabled** by the vulnerabilities due to implementation flaws. **Threatened** targets are networks. If a Slammer attack **succeeds** he can cause further DoS attacks.

J. B. Gao, B. W. Zhang, X. H. Chen, Z. Luo, 2013

ENISA, A good practice guide of using taxonomies in incident prevention and detection, dicembre 2016

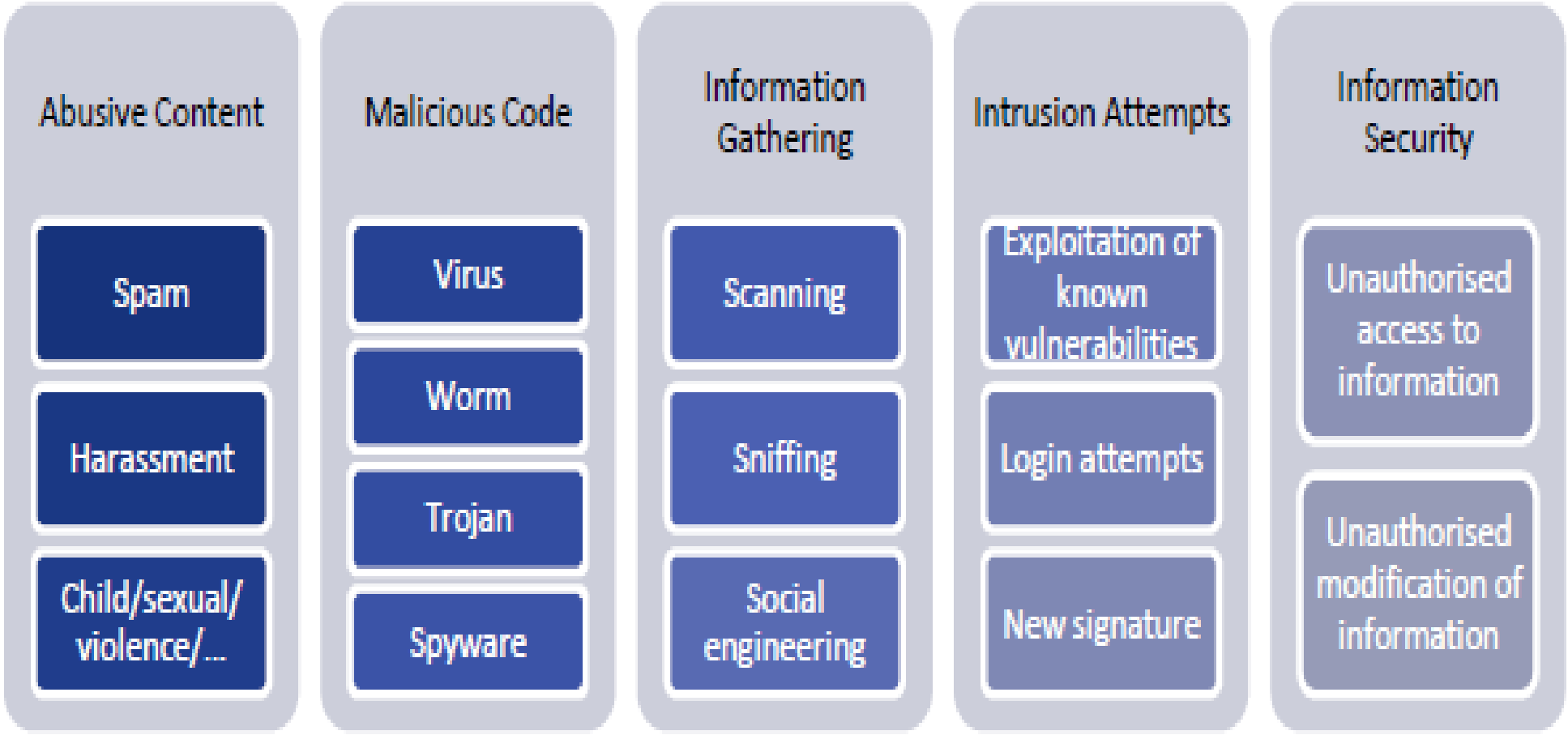
- **Taxonomy**
- a form of **classification scheme** to group related things together and to define the relationship these things have to each other
- a **semantic vocabulary** describes knowledge and information assets
- a **knowledge map** gives users an immediately grasp of the overall structure of the knowledge domain covered by the taxonomy, which should be comprehensive, predictable and easy to navigate.

Enisa 2016

- **inheritance**: different kinds of objects often have a certain amount in common with each other

- **differentiation**: characteristics of an object, which allows to differentiate an object from another object. For example, in case of malware you can have the category “malware” with the related characteristics “downloaders”, “rootkits”.
  - https://www.avast.com/it-it/c-rootkit

Example of a comparison between different versions



ENISA 2016

CERT.PT	CERT.BE	CESNET CERT	ECSIRT.NET MKII
Malware	Spam	Spam	Spam
Botnet Drone	Abusive Content	Bounce	Harassment
Ransomware	Malware	Virus	Child/Sexual/Violence/...
Malware Configuration	Scan	Malware	Virus
C&C	System/Account Compromised	Trojan	Trojan
DDoS	(D)DoS	Malware	Spyware
Scanner	Phishing	Probe	Dialler
Exploit	Vulnerability Report	Crack	Rootkit
Brute-force	Other	Botnet	Scanning
IDS alert		Dos	Sniffing
Defacement		Copyright	Social Engineering
Compromised		Scam	Exploiting of known Vulnerabilities
Backdoor		Phishing	Login attempts
Drop zone		Pharming	New attack signature
Phishing		Other	Privileged Account Compromise
SPAM		Unknown	Unprivileged Account Compromise
Vulnerability			Application Compromise
Service			Bot
Other			DoS

Confronto terminologico

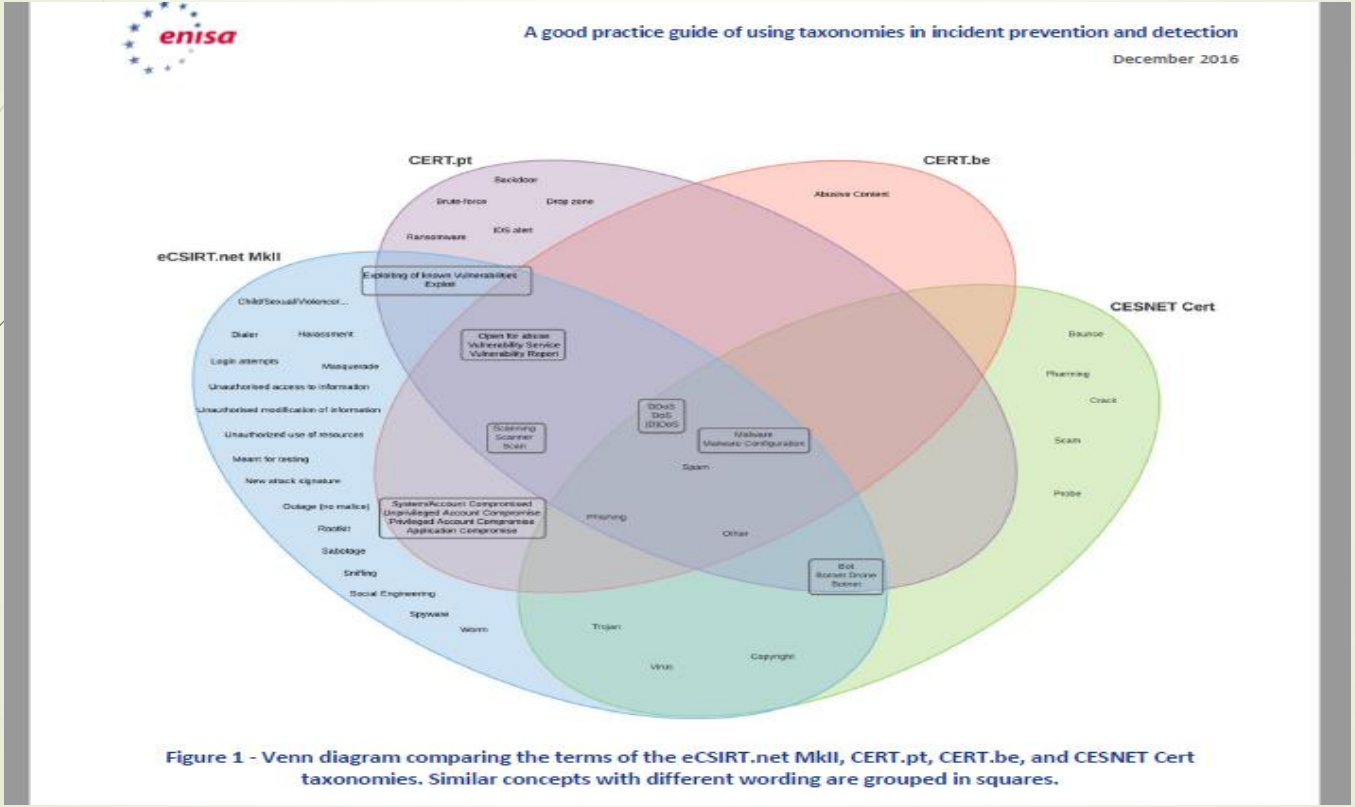


Figure 1 - Venn diagram comparing the terms of the eCSIRT.net MkII, CERT.pt, CERT.be, and CESNET Cert taxonomies. Similar concepts with different wording are grouped in squares.

Enisa 2016 Conclusioni

- There is currently **no consensus on concepts and definitions** related to taxonomies. Clear definitions reflecting the operational interpretation of the CSIRTs should be considered as **a key success factor** towards increasing cooperation between EU Member States.
- **Taxonomies currently lack terms to properly handle the impact of an incident,** incidents with no malice intended, explicit fields for ransomware, whether the incident is confirmed, and the differentiation between intrusion attempts and intrusions.
- The identified areas for potential improvement of existing taxonomies are based on the **complexity, contextual information, mutual exclusivity or ambiguity, performance measurement, impact, sensitivity, confidentiality, and purpose of taxonomies.**

Human interventions/organization

- (CERTs, CSIRTs, CIRTs, SIEM, SOC)
- Legislation
- Education and training: awareness
- R&D
- Public private partnerships

- Cybersecurity diplomacy
- Cybersecurity by design

Cybersecurity as a service

- Big data analytics
- AI applications: ontologies, taxonomies, data architectures
- Knowledge representation and info-sharing
- Resilience
- Technological solutions (detection, removal, alarm, etc.): prevention and prediction

The cybersecurity ecosystem and knowledge representation

- Conceptual definitions and analyses of the cybersecurity domain and sub-domains: **prospective standards for cybersecurity digital knowledge representation** and **related tools**
- Applications needed in risk assessment and evaluation: ISO, COBIT, NIST framework, etc.
- Quality/quantity metrics for risk evaluation
- Standards and tools for cyber security analytics and applications in defense and resilience:

  taxonomies /ontologies
- vulnerabilities/threats
- semantic web metalanguages/logical semantic modeling

Tipologia di impatto logico

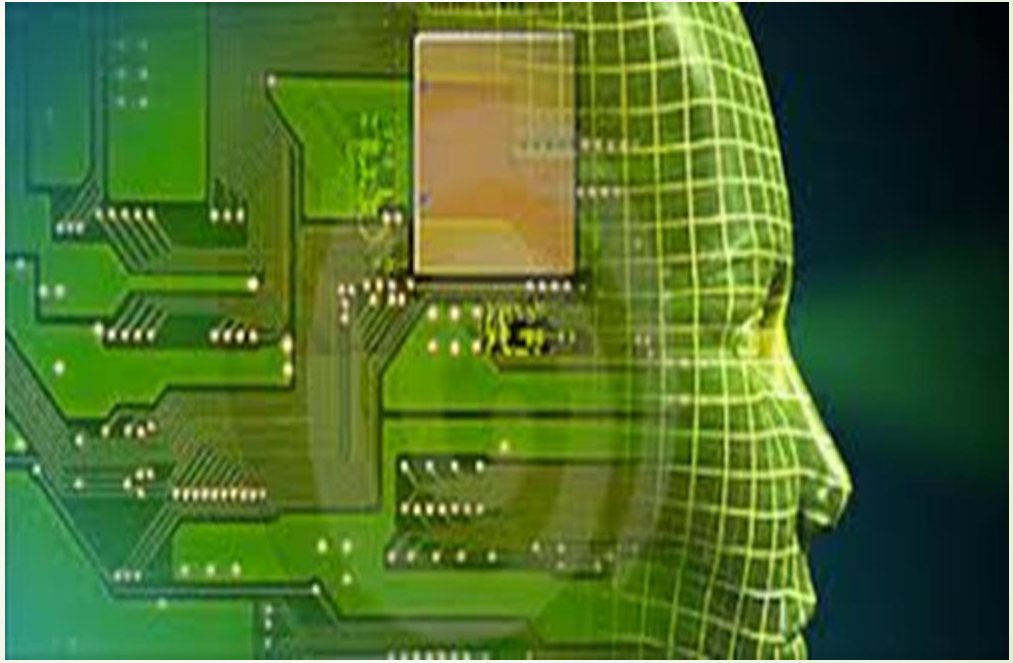
- Espionage (political, institutional, industrial, commercial, etc.)
- Data exfiltration
- Data destruction
- Data manipulation
- Denial of service
- Data encryption

N. Guarino (ed.), Formal Ontology in Information Systems, IOS Press, Amsterdam, 1998

- Some twenty years ago Guarino postulated the increasing relevance of ontology in the fields of **Artificial Intelligence, Computational Linguistics and Database Theory** and mentioned specific research fields such as **knowledge engineering, knowledge representation, qualitative modelling, language engineering, database design, information modelling and integration, object oriented analysis, information retrieval and extraction, knowledge management and organization, agent-based systems design**.

- At the methodological level he stressed the main peculiarity of an **ontology as its being a highly interdisciplinary approach where philosophy and linguistics play a fundamental role**

The digital mind, artificial intelligence and big data architecture



Ontologie e tassonomie

- Ontologies: logical semantic systems of entities and relationships based on a high level definition as applied to the cybersecurity domain
- Best definitions are contextualized entities and relations

- Taxonomies: mainly hierarchical classes with single decontextualized entities

Artificial intelligence e dati

- Modeling of data and of logical semantic relationships
- Design and development of the model: data cluster, univocal definition of terminology, search functions
- Technological translation into the platform and data insertion
- Metadata languages
- Metadata applications
- Data representation formats

Ontologies and taxonomies: tools and standards

- Definition and approaches
- Top level, middle level, domain ontology, pragmatic ontology
- Conceptual specifications: metalanguages for technological interoperability and logical semantic relationships
- Domains and subdomains

ENISA NIS 2017: the modus operandi

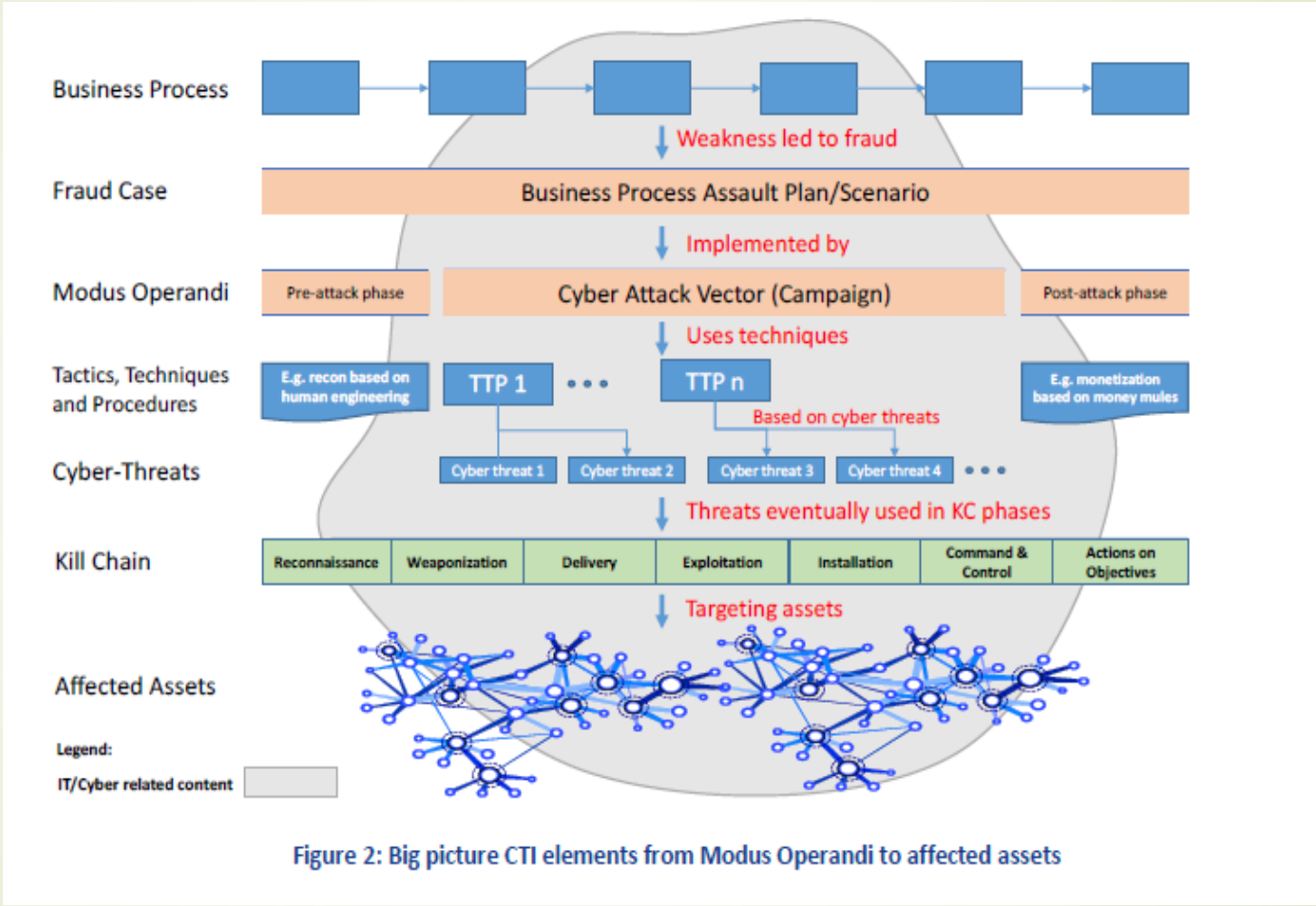


Figure 2: Big picture CTI elements from Modus Operandi to affected assets

Artificial intelligence and data

- Modeling of data and of logical semantic relationships
- Design and development of the model: data cluster, univocal definition of terminology, search functions
- Technological translation into the platform and data insertion
- Metadata languages
- Metadata applications
- Data representation formats

The Babel conceptualization: typology of analysis

- General vs domain and subdomain ontologies
- Ontologies and taxonomies relationships
- Vocabulary standards
- Goals of description

General and domain subdomain ontologies

- Oltramari et alii 2014: **ontology of cybersecurity**/Dolce/ Secco/Osco
- Syed et alii 2016: UCO **a unified cybersecurity ontology** (semantic web languages and UCO)
- Pragmema/Zuanelli 2017: the Poc **ontology platform** / 3level and pragmatic domain ontology)
- Mavroeidis and Bromander 2017 : **cyber threat intelligence** comparison and model

Domain/sub-domain ontologies

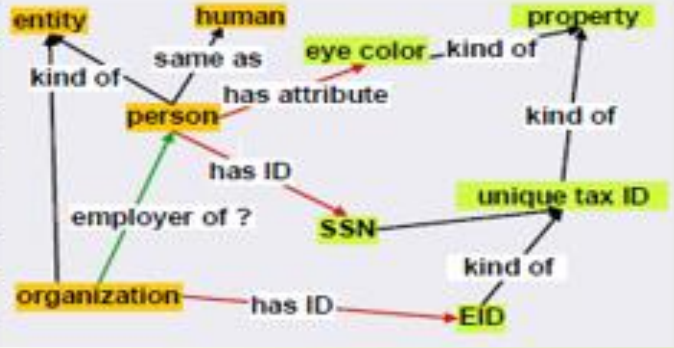
- Enisa 2011: Ontology and taxonomy of **resilience**
- Bromander et alii 2017: **Semantic threat modeling** (threat agent/threat scenario)
- Mavroeidis and Bromander 2017: **Cyber threat intelligence** model/Taxonomies, ontologies in cyberthreat intelligence
- Nistir 2016: **Vulnerability ontology**
- Silva and Rodriguez 2017: **Network ontology/Cyber threat intelligence** comparison and model

- **Attack taxonomies**
- Van Heeerden et alii 2015: attack taxonomy

- **Taxonomies in incident prevention and detection**
- Enisa 2016

The NIST/Mitre corporaton
Machine processable data

Ontologies, Controlled Vocabularies and Semantic Interoperability

	Controlled Vocabulary		Ontology
Definition	A controlled vocabulary (CV) is a set of lexical expressions that are vetted according to some criteria, such as their accepted usage in a community. • CVs are structured by one or more ordering relations, such as "narrower-than," "broader-than," or "related-to." • Structure is machine processable and semantics are human interpretable.		An ontology specifies the meaning of a controlled vocabulary in the form of a conceptual model. • Ontologies can be independent of any given controlled vocabulary. • Structure is machine processable and semantics are machine interpretable.
Example	**Terms**	**Relation**	
	entity	broader-than person broader-than organiz.	
	> person	narrower-than entity	
	>> eye color	related-to person	
	>> SSN	related-to person	
	>> employer	related-to person	
	> organization	narrower-than entity	
	>> EID	related-to organization	

Controlled Vocabularies for Standards: contents and representation NIST/MITRE

- – CEE: Common Event Expression
- – CPE: Common Platform Enumeration
- – CRE: Common Remediation Enumeration
- – CVE: Common Vulnerability Enumeration
- – CWE: Common Weakness Enumeration
- – MAEC: Malware Attribute Enumeration and Characterization
- – OVAL: Open Vulnerability and Assessment Language
- – XCCDF: Extensible Configuration Checklist Description Format

- ■ Both MITRE and NIST maintain public repositories and Web sites for
- the various standards: http://nvd.nist.gov/ http://oval.mitre.org/repository/ http://measurablesecurity.mitre.org/

CVE (SR-13/03/2018)/MITRE

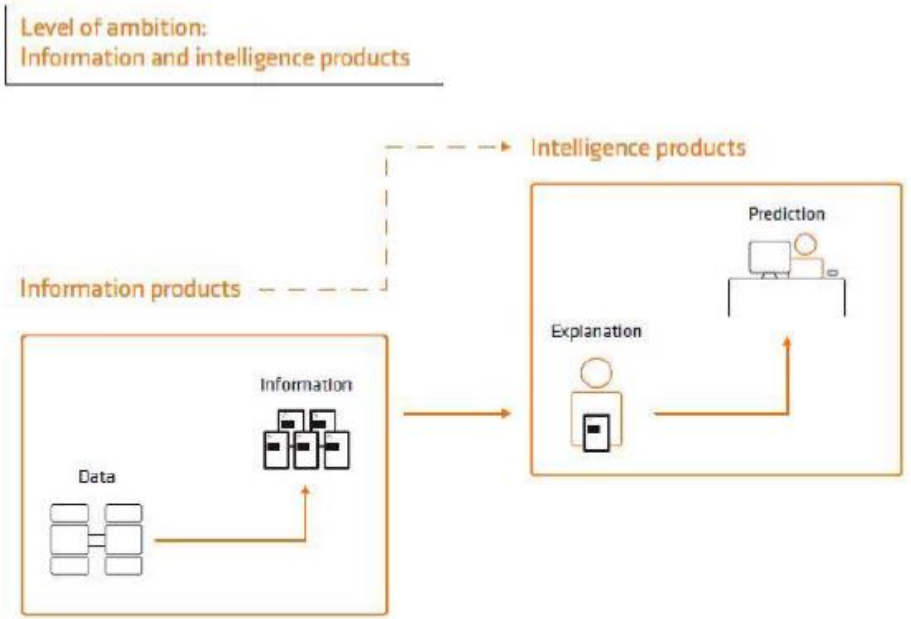
Incident	TXT	HTML	XML
CVE-2018-7580	Name: CVE-2018-7580 Status: Candidate URL: http://cve.mitre.org/cgi-bin/cvename.cgi?name=CVE-2018-7580 Phase: Assigned (20180301) Category: ** RESERVED ** This candidate has been reserved by an organization or individual that will use it when announcing a new security problem. When the candidate has been publicized, the details for this candidate will be provided. Current Votes: None (candidate not yet proposed)	Name: CVE-2018-7580<p> <p>Description: ** RESERVED ** This candidate has been reserved by an organization or individual that will use it when announcing a new security problem. When the candidate has been publicized, the details for this candidate will be provided. <p>Status: Candidate Phase: Assigned (20180301) <p>Votes: <pre></pre>	<item seq="2018-7580" name="CVE-2018-7580" type="CAN"><status>Candidate</status><phase date="20180301">Assigned</phase><desc>** RESERVED ** This candidate has been reserved by an organization or individual that will use it when announcing a new security problem. When the candidate has been publicized, the details for this candidate will be provided.</desc><refs></refs><votes> </votes><comments> </comments></item>
CVE-2018-7581	Name: CVE-2018-7581 Status: Candidate URL: http://cve.mitre.org/cgi-bin/cvename.cgi?name=CVE-2018-7581 Phase: Assigned (20180301) Category: ** RESERVED ** This candidate has been reserved by an organization or individual that will use it when announcing a new security problem. When the candidate has been publicized, the details for this candidate will be provided. Current Votes: None (candidate not yet proposed)	Name: CVE-2018-7581<p> <p>Description: ** RESERVED ** This candidate has been reserved by an organization or individual that will use it when announcing a new security problem. When the candidate has been publicized, the details for this candidate will be provided. <p>Status: Candidate Phase: Assigned (20180301) <p> Votes: <pre></pre>	<item seq="2018-7581" name="CVE-2018-7581" type="CAN"><status>Candidate</status><phase date="20180301">Assigned</phase><desc>\ProgramData\WebLog Expert\WebServer\WebServer.cfg in WebLog Expert Web Server Enterprise 9.4 has weak permissions (BUILTIN\Users:(ID)C), which allows local users to set a cleartext password and login as admin.</desc><refs><ref url="https://www.exploit-db.com/exploits/44270/" source="EXPLOIT-DB">44270</ref><ref url="http://hyp3rlinx.altervista.org/advisories/WEBLOG-EXPERT-WEB-SERVER-ENTERPRISE-v9.4-AUTHENTICATION-BYPASS.txt" source="MISC">http://hyp3rlinx.altervista.org/advisories/WEBLOG-EXPERT-WEB-SERVER-ENTERPRISE-v9.4-AUTHENTICATION-BYPASS.txt</ref><ref url="http://packetstormsecurity.com/files/146697/WebLog-Expert-Web-Server-Enterprise-9.4-Weak-Permissions.html" source="MISC">http://packetstormsecurity.com/files/146697/WebLog-Expert-Web-Server-Enterprise-9.4-Weak-Permissions.html</ref></refs><votes> </votes><comments> </comments></item>

ACT, TOCSA and Oslo Analytics (2017)

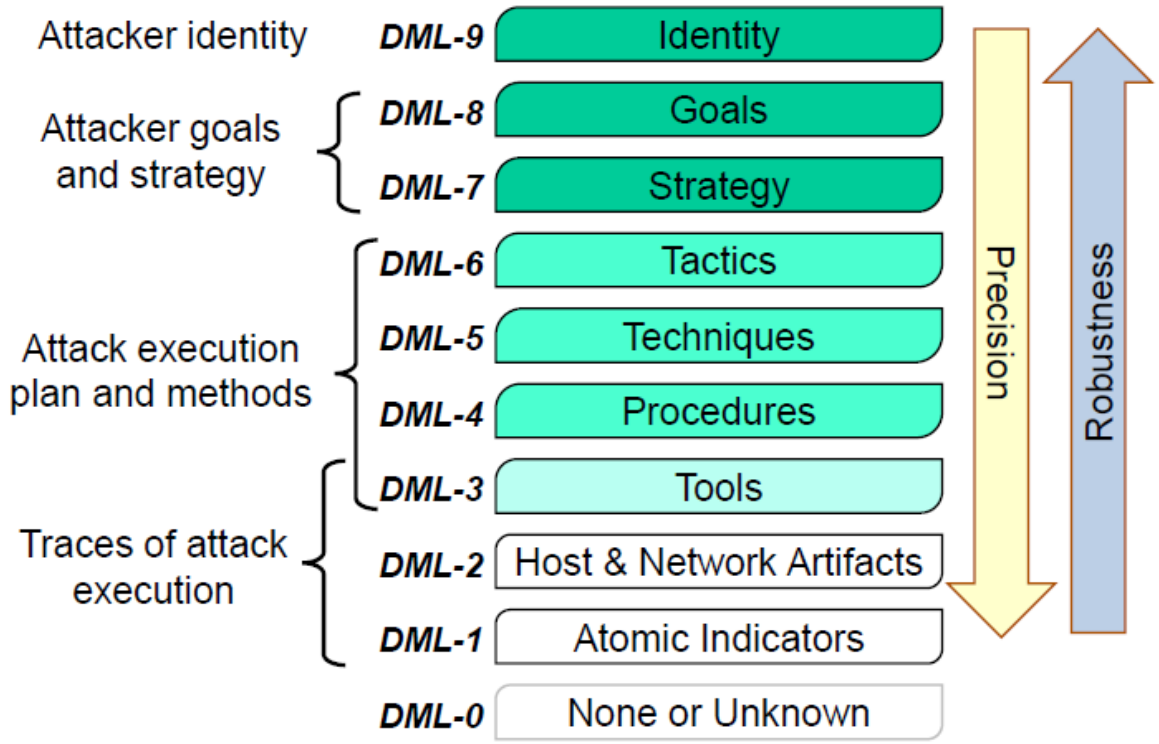
•Semi-Automated Cyber Threat Intelligence (ACT)
-Open Source Threat Intelligence Platform
-https://www.mnemonic.no/research-and-development/semi-automated-cyber-threat-intelligence/
•Threat Ontologies for Cyber Security Analytics (TOCSA)
-Ontologies
-PhD Project
-https://www.mnemonic.no/no/research-and-development/threat-ontologies-for-cybersecurity-analytics/
-http://www.mn.uio.no/ifi/english/research/projects/tocsa/
•Operable Subjective Logic Analysis Technology for Intelligence in Cybersecurity (Oslo Analytics)
-Analytics
-Subjective Logic (quantifying uncertainty)
-Trust Networks
-Academic
-http://www.mn.uio.no/ifi/english/research/projects/oslo-analytics/

The approach

- Threat Information vs Threat Intelligence

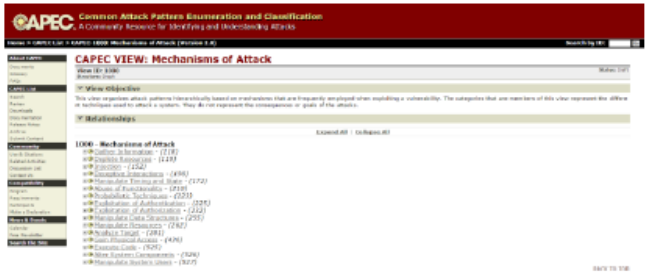


The Detection Maturity Level (DML) Model



Semantic Feature Extraction

- Formal definitions of
  - -Goals
  - -Strategy
  - -Tactics
  - -Techniques
  - -Procedures
- Relevant initiatives
  - -MITRE CAPEC
    - https://capec.mitre.org
  - -MITRE ATT&CK
    - https://attack.mitre.org
  - -MITRE CAR
    - https://car.mitre.org



mnemonic

Network resilience ontology Enisa 2011

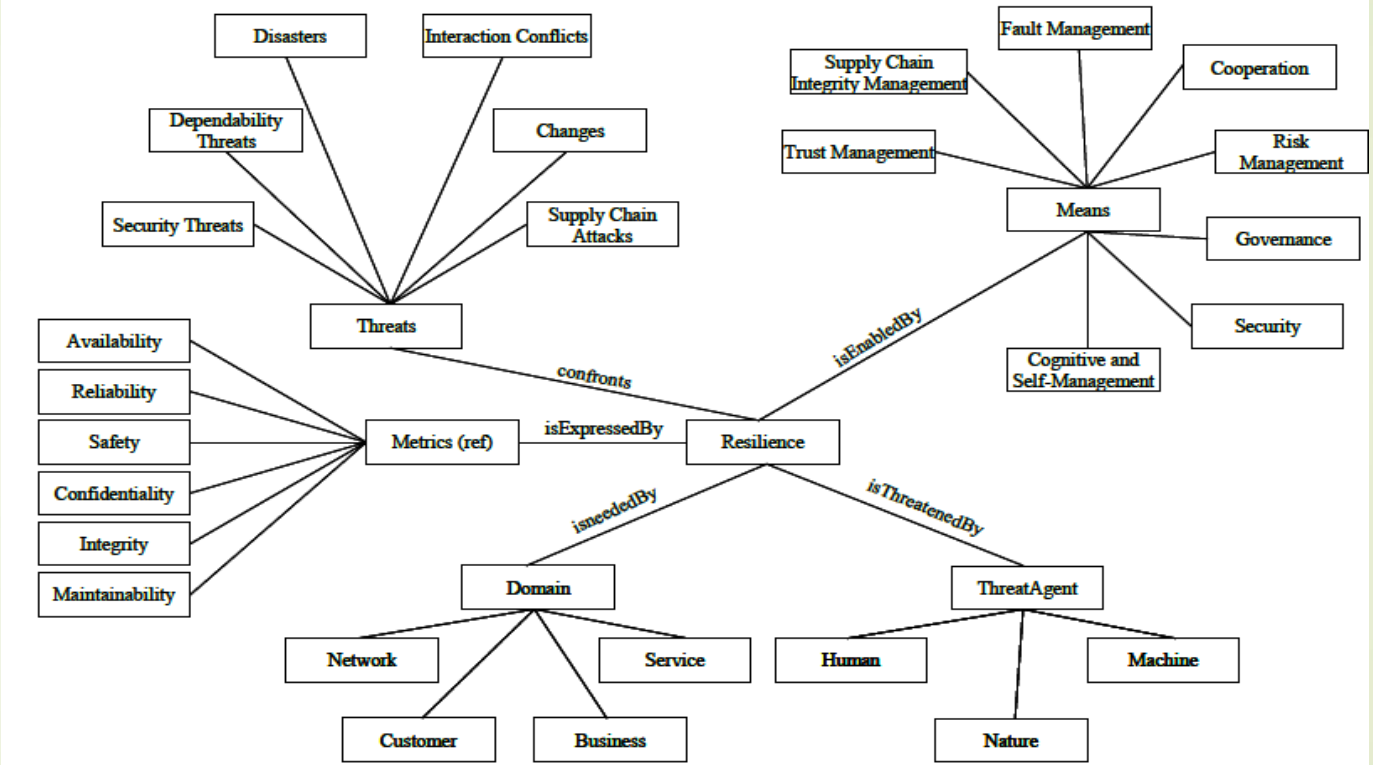


Figure 11: High level overview of resilience ontology

Business ontology (sub-domain)

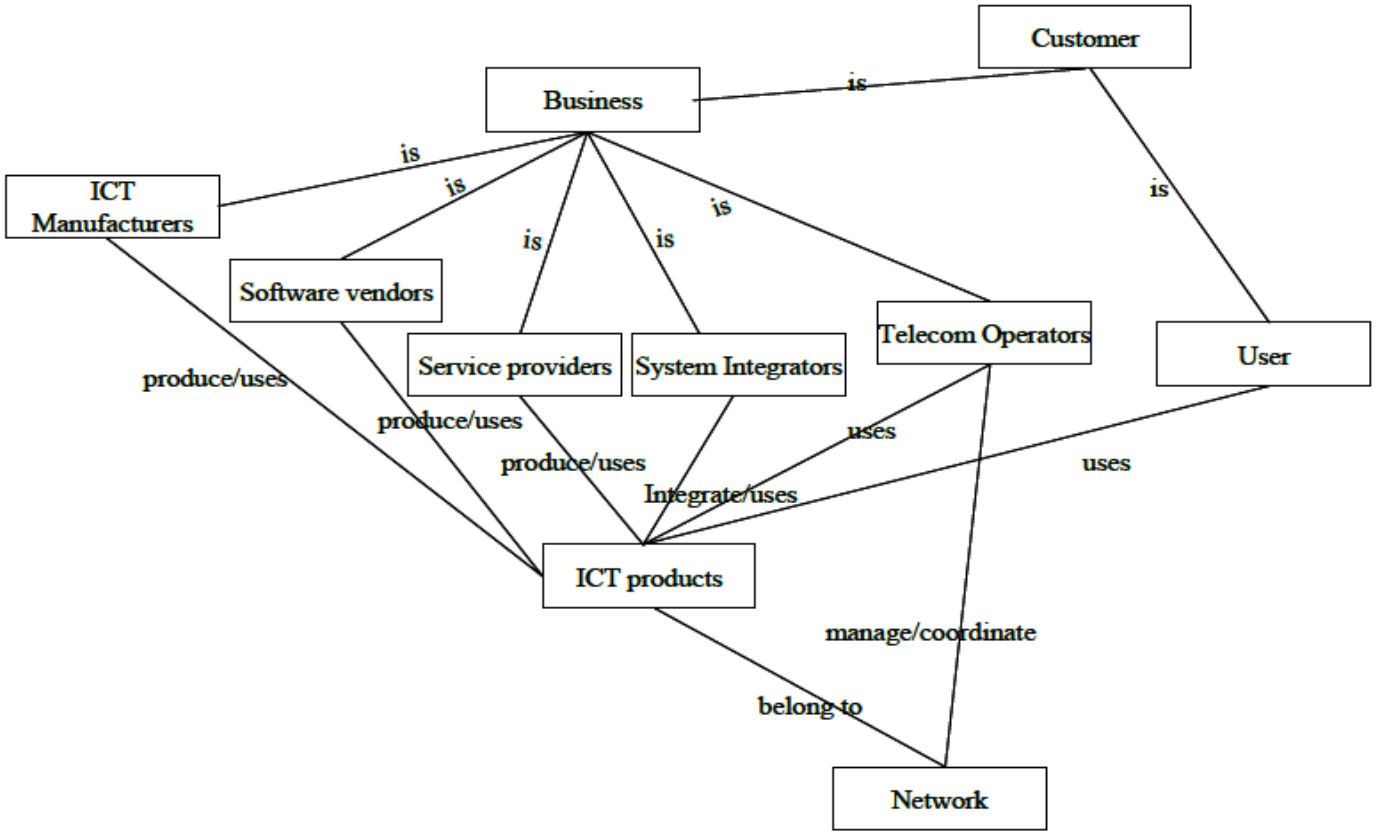


Figure 19: Business domain

Network ontologies

- Network security ontologies: aspects/ comparison (V. Silva and G. Rodriguez 2017 in https://arxiv.org/pdf/1704.02441)

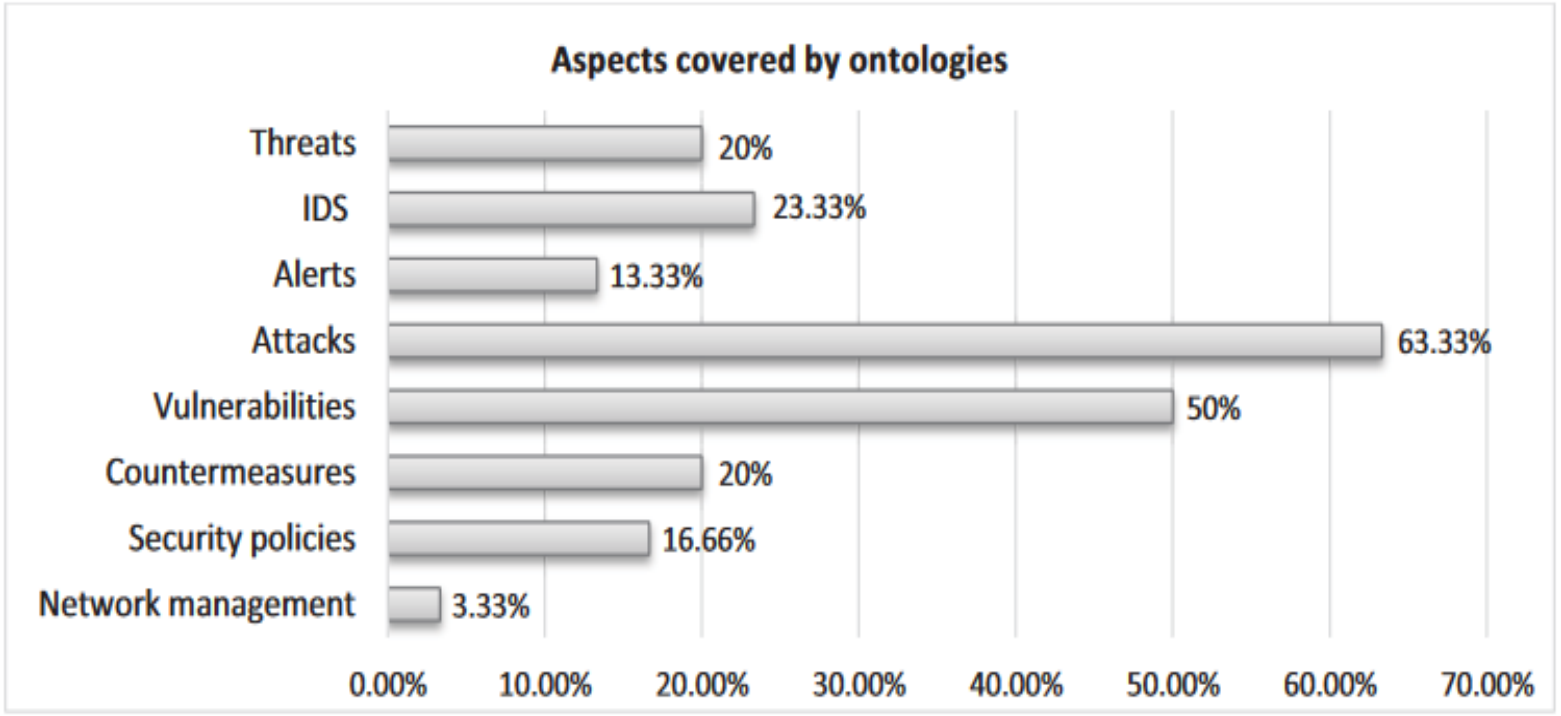


Figure 1: Aspects covered by ontologies

Comparison features (Silva & Rodriguez 2017)

- 63.33% of the ontologies make reference to attacks and their taxonomical structure. Their focus is mainly on the network layer **missing attacks at the application layer**.

- 80% of the papers reviewed **do not present the results obtained from test scenarios**, and therefore it is unachievable to evaluate the ontology and determin if it adapts to the requirements or to measure its effectiveness.

- Only 13.33% of the papers validate their proposals, trying to identify the **correct use of the language, the accuracy of the taxonomic structure, the validity of the vocabulary, and the adequacy of the requirements** for the purpose of documenting the process of development to verify if the proposal complies with the terms specified …

.

...Comparison features

- One of the challenges that constitutes a potentially interesting area arises when data is collected from **different safety equipment** (IDS, Intrusion prevention system, firewall, antivirus system, system security audit, honeynet,etc.).

- The **safety equipment is distributed in different domains in the network**, which is required to develop an ontology that can integrate real-time data from this safety equipment and allows the captured data to be properly administered

The proposal: neither ontology nor taxonomy (Silva and Rodriguez)

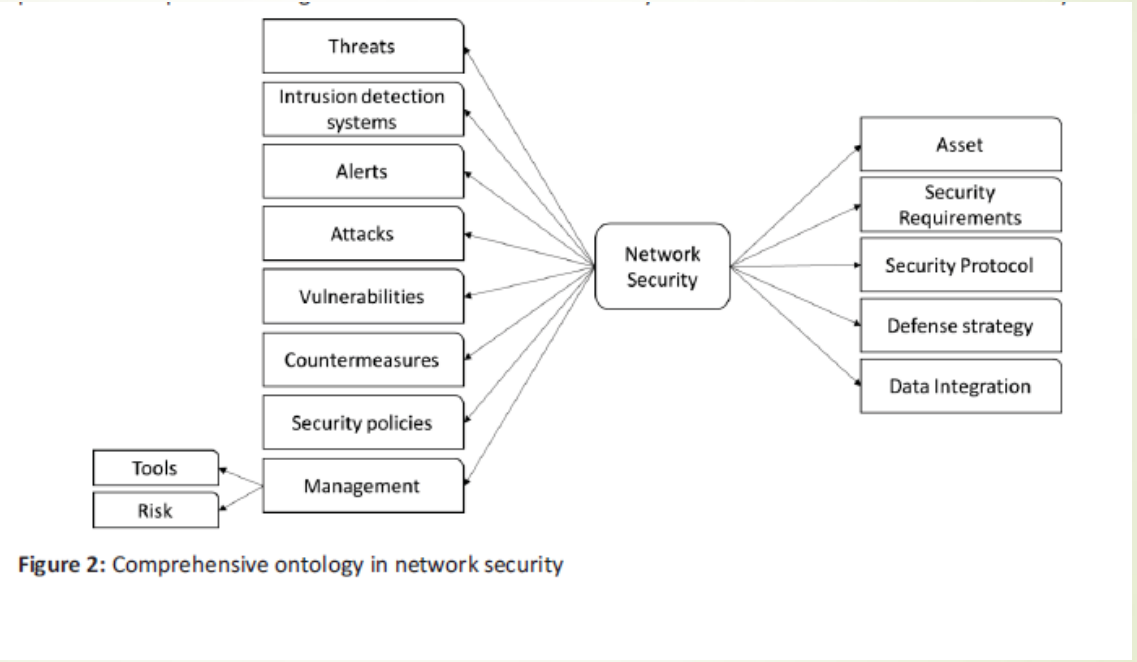


Figure 2: Comprehensive ontology in network security

Table 1: Syntax and Semantics of Description Logic constructors

Name	Syntax	Semantics	Symbol
Top	$\top$	$\Delta^{\mathcal{I}}$	$\mathcal{AL}$
Bottom	$\bot$	ϕ	$\mathcal{AL}$
Intersection	$C \sqcap D$	$C^{\mathcal{I}} \cap D^{\mathcal{I}}$	$\mathcal{AL}$
Union	$C \sqcup D$	$C^{\mathcal{I}} \cup D^{\mathcal{I}}$	$\mathcal{U}$
Negation	$\neg C$	$\Delta^{\mathcal{I}} \setminus D^{\mathcal{I}}$	$\mathcal{C}$
Value restriction	$\forall R.C$	$\{a \in \Delta^{\mathcal{I}} \mid \forall b.\ (a,b) \in R^{\mathcal{I}} \rightarrow b \in C^{\mathcal{I}}\}$	$\mathcal{AL}$
Existential quant.	$\exists R.C$	$\{a \in \Delta^{\mathcal{I}} \mid \forall b.\ (a,b) \in R^{\mathcal{I}} \wedge b \in C^{\mathcal{I}}\}$	$\mathcal{E}$
Nominal	I	$I^{\mathcal{I}} \subseteq \Delta^{\mathcal{I}}$ with $\lvert I^{\mathcal{I}}\rvert = 1$	$\mathcal{O}$
Qualified Number restriction (less than)	$\leq nR.C$	$\{a \in \Delta^{\mathcal{I}} \mid \lvert\{\forall b \in \Delta^{\mathcal{I}} \mid (a,b) \in \mathcal{R}^{\mathcal{I}} \wedge b \in C^{\mathcal{I}}\}\rvert \leq n\}$	$\mathcal{Q}$
Qualified Number restriction (equal than)	$= nR.C$	$\{a \in \Delta^{\mathcal{I}} \mid \lvert\{\forall b \in \Delta^{\mathcal{I}} \mid (a,b) \in \mathcal{R}^{\mathcal{I}} \wedge b \in C^{\mathcal{I}}\}\rvert = n\}$	$\mathcal{Q}$
Qualified Number restriction (greater than)	$\geq nR.C$	$\{a \in \Delta^{\mathcal{I}} \mid \lvert\{\forall b \in \Delta^{\mathcal{I}} \mid (a,b) \in \mathcal{R}^{\mathcal{I}} \wedge b \in C^{\mathcal{I}}\}\rvert \geq n\}$	$\mathcal{Q}$
Role Hierarchy	$R_1 \sqsubseteq R_2$	$\{(a, b) \in \Delta^{\mathcal{I}} \times \Delta^{\mathcal{I}} \mid (a, b) \in R_1^{\mathcal{I}} \rightarrow (a, b) \in R_2^{\mathcal{I}}\}$	$\mathcal{H}$
Role Inverse	R^{-}	$\{(b, a) \in \Delta^{\mathcal{I}} \times \Delta^{\mathcal{I}} \mid (a, b) \in R^{\mathcal{I}}\}$	$\mathcal{I}$
Role Composition	$R_1 \circ R_2$	$\{(a, c) \mid \exists b.\ (a, b) \in R_1^{\mathcal{I}} \wedge (b, c) \in R_2^{\mathcal{I}}\}$	$\mathcal{R}$

UCO conceptual relationships

- In addition to mapping to STIX, UCO has also been extended with a number of **relevant cybersecurity standards, vocabularies and ontologies** such as CVE4, CCE5, CVSS6, CAPEC7, CYBOX8, KillChain9 and STUCCO10

- To support diverse use cases,UCO ontology has been mapped to general **world knowledge** available through Google's knowledge graph, Dbpedia knowledge base (Auer et al. 2007), Yago knowledge base (Suchanek, Kasneci, and Weikum 2008) etc.

- Linking to these knowledge sources provides **access to large number of datasets for different domains** (e.g. geonames) as well as terms in different languages (e.g. Russian

UCO's 'important' classes present in UCO ontology

- 1. **Means**: This class describes various **methods of executing an attack** and consists of sub-classes like BufferOver-Flow, SynFlood, LogicExploit, TcpPortScan etc., which can further consist of their own sub-classes. The Means class maps to TTP field in STIX which characterizes specific details of observed or potential attacker Tactics, Techniques and Procedures.
- 2. **Consequences:** This class describes the **possible outcomes of an attack**. It consists of sub-classes like DenialOfService, LossOfConfiguration, PrivilegeEscalation,UnauthUser, etc. It maps to Observables in STIX.
- 3. **Attack**: This class characterizes a **cyber threat attack** and is mapped to Incident in STIX.
- 4. **Attacker**: This class represents **identification or characterization of the adversary** and is mapped to ThreatActor in STIX.

UCO classes

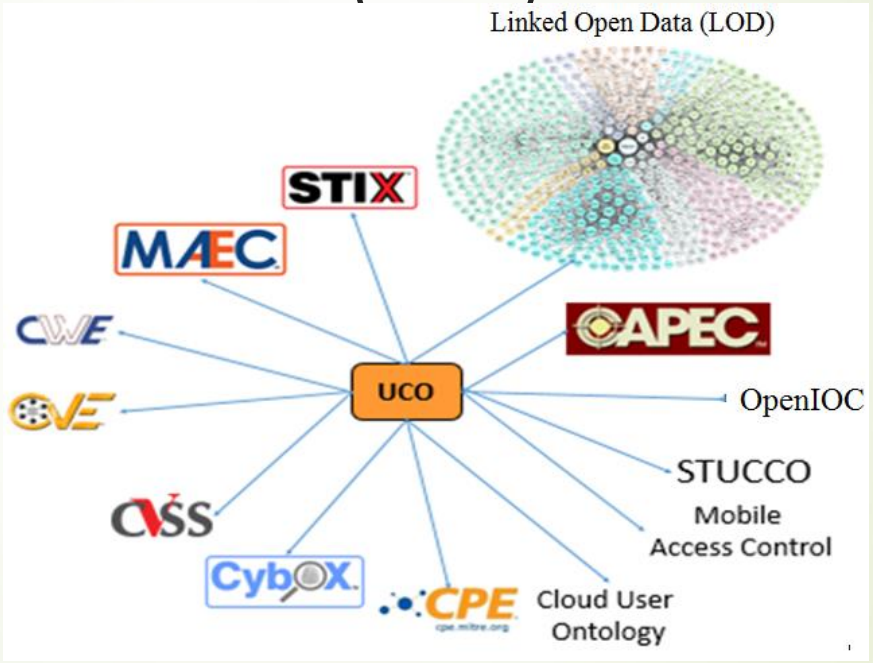
- 5. **Attack Pattern**: Attack Patterns are **descriptions of common methods for exploiting software** providing the attackers perspective and guidance on ways to mitigate their effect. An example of attack pattern is Phishing.

- 6. **Exploit:** This class characterizes **description of an individual exploit and maps** to ExploitType in STIX schema.

- 7. **Exploit Target**: **Exploit Targets are vulnerabilities or weaknesses** in software, systems, networks or configurations that are targeted for exploitation by the TTP (cyber threat adversary Tactic, Technique or Procedure).

- 8. **Indicator**: A cyber threat indicator is made up of **a pattern identifying certain observable conditions as well as contextual information** about the patterns meaning, how and when it should be acted on, etc. This class is mapped to IndicatorType in STIX schema and Indicator class in CAPEC ontology.

UCO ontology serves as the core for cybersecurity
Linked Open Data (LOD) cloud



Limitations of approach

- **Difference of conceptual relational descriptors** in a metadata language such as OWL as opposed to logical semantic entities as defined by (fuzzy) logic criteria in terminology

- UCO classes are **loose entities definition**: no logical semantic definition

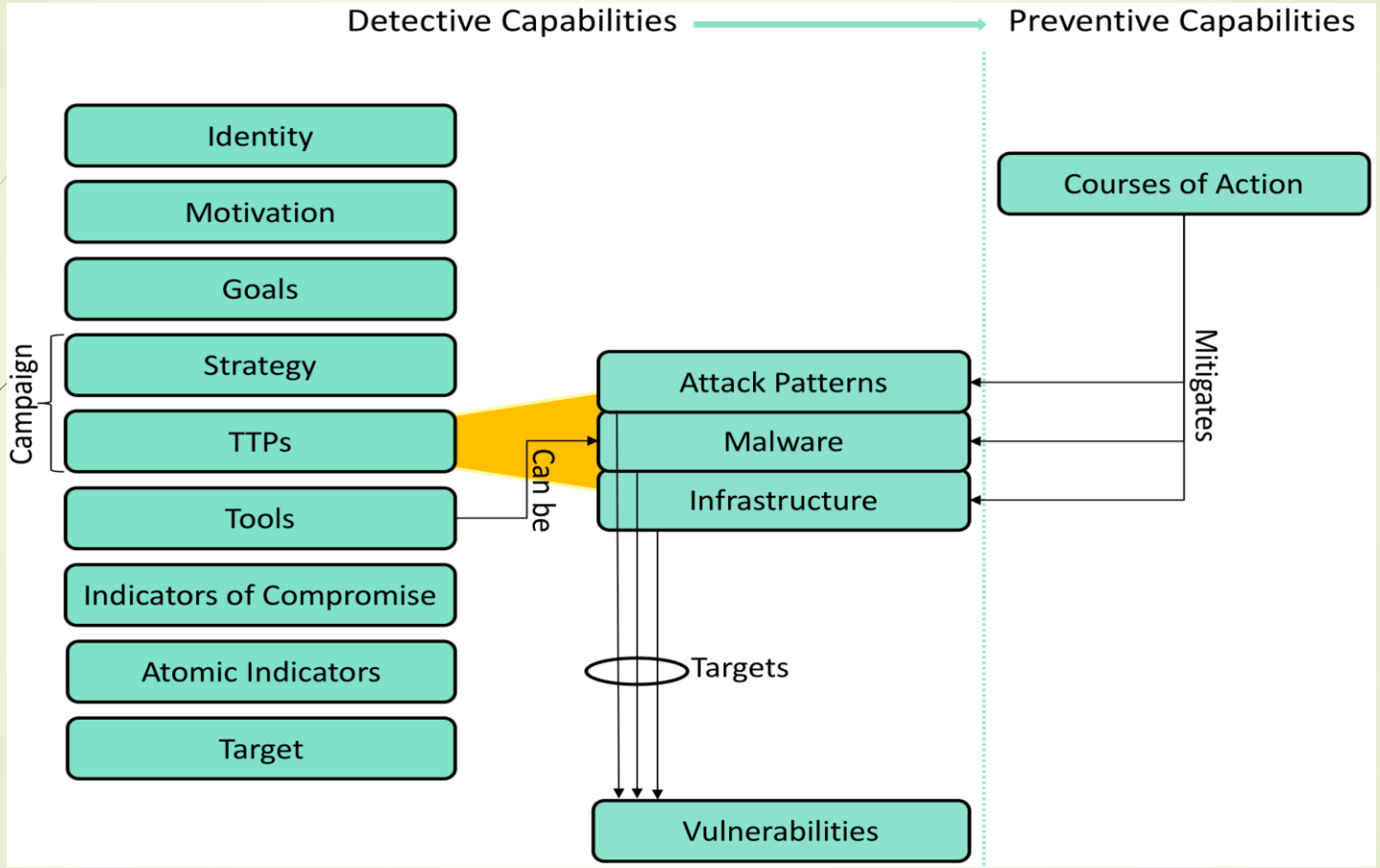
- **Useful linked open data**

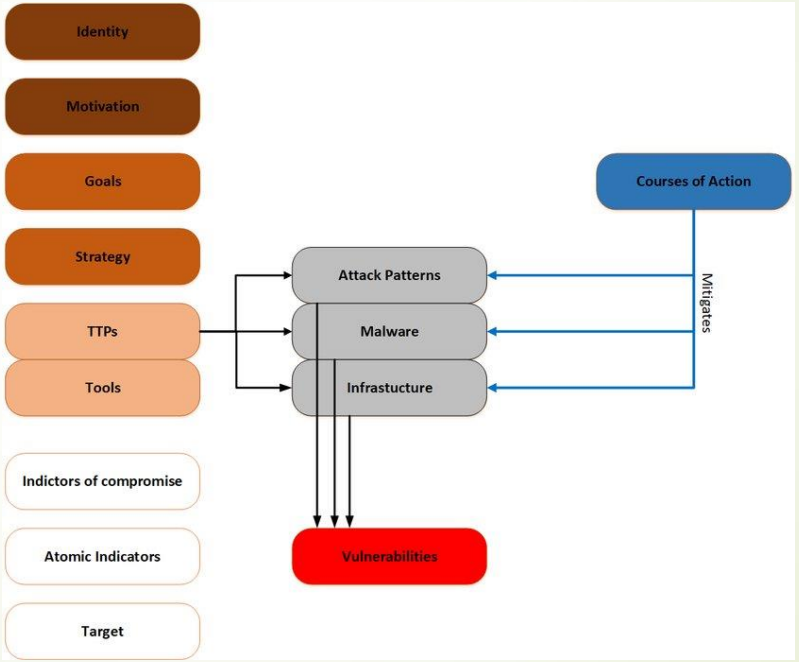
Cyber threat intelligence model: taxonomies, sharing standards and ontologies

TABLE I
CTI EVALUATION: TAXONOMIES, SHARING STANDARDS, AND ONTOLOGIES

		Identity	Motivation	Goal	Strategy	TTP	Tool	IOC	Atomic Indicator	Target	COA
Taxonomies	TAL [8]	*									
	Threat Agent Motivation [5]	*	*								
	CVE [9]								*		
	NVD [10]								*		
	CPE [11]								*		
	CWE [12]					*			*		*
	CAPEC [13]					*		*			*
	ATT&CK [14]	*				*	*				
	CVSS [15]								*		
	CWSS [16]								*		
Sharing Standards	STIX 1 [18]	*	*	* (Intended Effect:taxonomy)	*	*	*	*	*	*	*
	STIX 2 [36]	*	*	* (Objectives:string)	*	*	*	*	*	*	*
	MAEC [19]							*			
	OpenIOC [37]					*	*	*	*		
Ontologies	Fenz & Ekelhat (2009) [21]								*		
	Wang & Guo (2009) - OVM [22]					*			*		*
	Orbst et al. (2012) [23]	*					*		*	*	
	More et al. (2012) [26]					*			*		
	Oltramari et al. (2014) - CRATELO [27]	*				*			*	*	
	Gregio et al. (2014) [28]						* (malware)		*		
	Salem & Wacek (2015) - ICAS [29]					*			*		
	Iannacone et al. (2015) - STUCCO [30]	*				*	*		*		
	Gregio et al. (2016) - MBO [31]						* (malware)	*	* (it may provide)		
	Fusun et al. (2015) - ASR [32]					*			*	*	
	Pendelton et al. (2016) - Security Metrics Ontology [34]					*					*
	Syed et al. (2016) - UCO [35]	*	*	*	*	*	*	*	*		*
	Unified Cyber Ontology (2016) - UCO [38]	*	*	*	*	*	*	*	*	*	*

Cyber threat intelligence model




Identity
Motivation
Goals
Strategy
TTPs
Tools
Indictors of compromise
Atomic Indicators
Target
Courses of Action
Attack Patterns
Malware
Infrastucture
Mitigates
Vulnerabilities

Le premesse metodologiche

- Cybersecurity, *data analytics*, AI
- Le tassonomie, le classificazioni e le ontologie della *cybersecurity*
- La soluzione POC: *tool*/piattaforma automatica di intelligenza artificiale *cybersecurity*/protezione dati e servizi

Quale cybersecurity: una delimitazione del campo

- La sicurezza informatica è la condizione in cui il **ciberspazio è protetto**, rispetto a **eventi volontari** o accidentali consistenti **nell'acquisizione ed esfiltrazione di dati, nella loro modifica o distruzione illegale o il blocco di sistemi informatici**, grazie ad appropriati sistemi di sicurezza.
- Queste misure includono **verifiche di sicurezza, gestione degli aggiornamenti o correzioni, procedure di autenticazione, gestione degli accessi, analisi dei rischi, individuazione e risposta a incidenti/attacchi, mitigazione degli impatti, recupero delle componenti soggette ad attacco, addestramento ed educazione del personale**, verifica e incremento della sicurezza fisica dei locali dove sono situati i sistemi di informazione e comunicazione.

La funzione dei dati: masse di dati del contesto/corpora di dati d'attacco, minacce, vulnerabilità (eventi e incidenti)

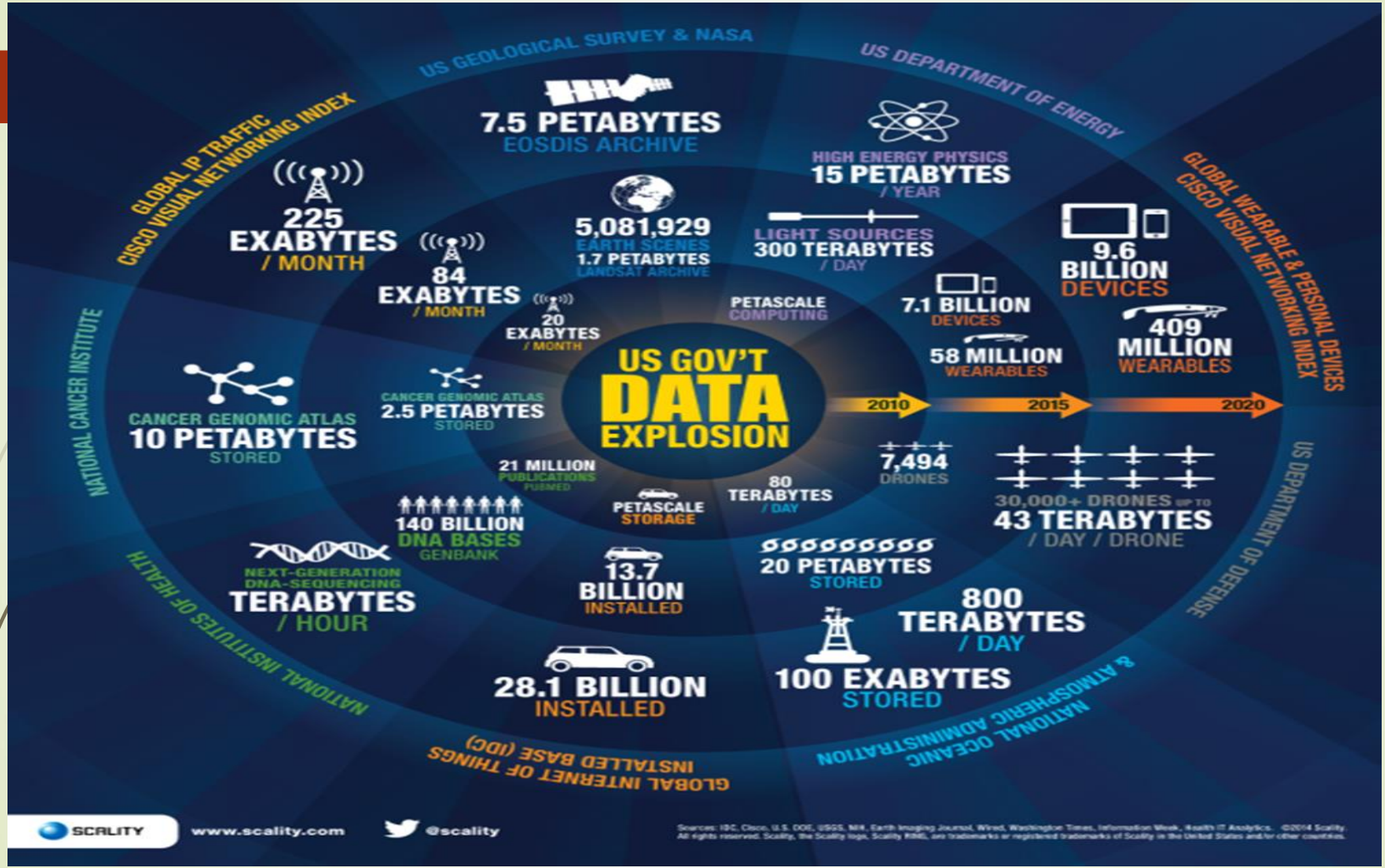
Finalità: pronti interventi, mitigazione, resilienza, prevenzione, predittività

- Reportistica
- Metriche dei danni
- Statistiche
- Standard
- Certificazioni di sicurezza dei prodotti e dei servizi

Percorsi d'azione:

- L'acquisizione dei dati (detentori dei dati)
- Le tassonomie dei dati
- Le classificazioni dei dati

Cybersecurity, dati, big data analytics


US GEOLOGICAL SURVEY & NASA
US DEPARTMENT OF ENERGY
GLOBAL IP TRAFFIC CISCO VISUAL NETWORKING INDEX
GLOBAL WEARABLE & PERSONAL DEVICES CISCO VISUAL NETWORKING INDEX
NATIONAL CANCER INSTITUTE
US DEPARTMENT OF DEFENSE
NATIONAL INSTITUTES OF HEALTH
NATIONAL OCEANIC & ATMOSPHERIC ADMINISTRATION
GLOBAL INTERNET OF THINGS INSTALLED BASE (IDC)
7.5 PETABYTES
EOSDIS ARCHIVE
HIGH ENERGY PHYSICS
15 PETABYTES
/ YEAR
225
EXABYTES
/ MONTH
84
EXABYTES
/ MONTH
5,081,929
EARTH SCENES
1.7 PETABYTES
LANDSAT ARCHIVE
LIGHT SOURCES
300 TERABYTES
/ DAY
9.6
BILLION
DEVICES
7.1 BILLION
DEVICES
20
EXABYTES
/ MONTH
PETASCALE
COMPUTING
409
MILLION
WEARABLES
58 MILLION
WEARABLES
US GOV'T
DATA
EXPLOSION
2010
2015
2020
CANCER GENOMIC ATLAS
10 PETABYTES
STORED
CANCER GENOMIC ATLAS
2.5 PETABYTES
STORED
7,494
DRONES
30,000+ DRONES UP TO
43 TERABYTES
/ DAY / DRONE
21 MILLION
PUBLICATIONS
PUBMED
80
TERABYTES
/ DAY
PETASCALE
STORAGE
140 BILLION
DNA BASES
GENBANK
NEXT-GENERATION
DNA-SEQUENCING
TERABYTES
/ HOUR
13.7
BILLION
INSTALLED
20 PETABYTES
STORED
800
TERABYTES
/ DAY
28.1 BILLION
INSTALLED
100 EXABYTES
STORED
SCALITY
www.scality.com
@scality
Sources: IDC, Cisco, U.S. DOE, USGS, NIH, Earth Imaging Journal, Wired, Washington Times, Information Week, Health IT Analytics. ©2014 Scality.
All rights reserved. Scality, the Scality logo, Scality RING, are trademarks or registered trademarks of Scality in the United States and/or other countries.

Big data analytics: masse astrutturate (e strutturate)di dati da analizzare e strutturare (ri-strutturare)

- Obiettivi: analisi e classificazione per scopi rimediali e preventivi nei diversi settori: militari, industriali, istituzionali, economico-finanziari, sanitari, ecc.
- Tecnologie disponibili per l'analisi: *libraries*, *big data software*, nuove logiche di *storage* ed elaborazione
- Limiti: carenza di modellizzazioni e applicazioni: tassonomie, classificazioni, ontologia
- Ambiguità semantiche: dati e metadati (XML, OWL)
- Utilità: destinazioni tecniche (CERT, CSIRT, SIEM, ecc.) e laiche
- Formalizzazione e gestione automatica dei dati

Soluzioni

- Modellistica ontologica di utilità generale
- Definizione, correlazione e valore univoco dei dati
- Sviluppo tecnologico interoperabile di piattaforme di analisi, *assessment* e valutazione del rischio
- Implementazione additiva dell'architettura

L’intelligenza artificiale: knowledge, informazioni, dati

Stima del mercato di AI

- Crescita stimata del mercato AI al 2022: 16,06 miliardi di dollari. Top investitori Amazon e Google

L'intelligenza artificiale: knowledge, informazioni, dati

- Immissione di informazioni sensoriali multiple dal mondo esterno
- Utilizzo di diverse tipologie di input sensoriale/modale
- Canalizzazioni sensoriali: specificità dei dati
- Interpretazione e modellizzazione dei dati
- Linguaggi ibridi
- Linguaggi traduttivi

Ontologie e tassonomie: la modellistica

Un riferimento: N. Guarino (ed.), Formal Ontology in Information Systems, IOS Press, Amsterdam, 1998

- Some twenty years ago Guarino postulated the increasing relevance of ontology in the fields of **Artificial Intelligence, Computational Linguistics and Database Theory** and mentioned specific research fields such as **knowledge engineering, knowledge representation, qualitative modelling, language engineering, database design, information modelling and integration, object oriented analysis, information retrieval and extraction, knowledge management and organization, agent-based systems design**.

- At the methodological level he stressed the main peculiarity of an **ontology as its being a highly interdisciplinary approach where philosophy and linguistics play a fundamental role.**

La piattaforma ontologica per la cybersecurity POC Pragmema

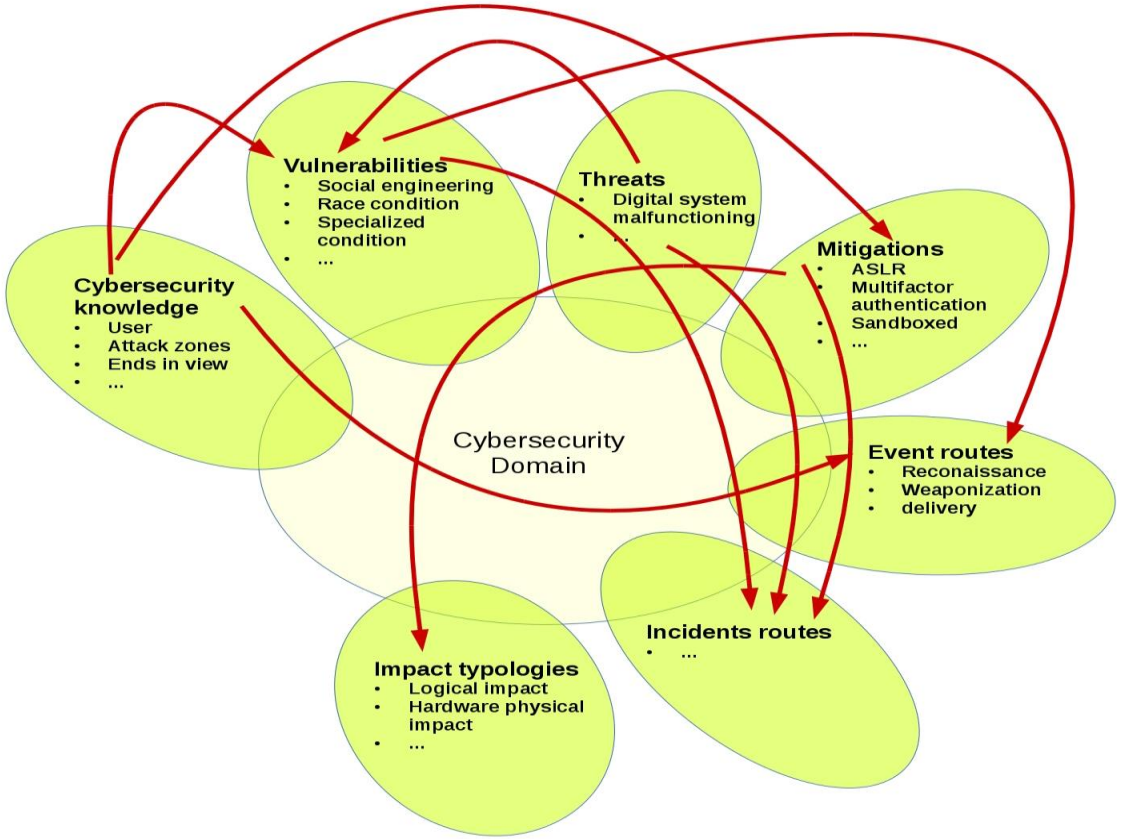
- La piattaforma POC su infrastruttura Liferay, è la prima nel contesto europeo e internazionale
- POC integra la *Vulnerability ontology* del NIST, classificazioni ENISA, analisi e proposte ontologiche e classificatorie specifiche (TOCSA) riformulandone la valenza e l'impianto strutturale e ampliandone la portata conoscitiva e applicativa
- Necessaria per l'analisi e la classificazione dei dati e la descrizione univoca degli eventi e incidenti

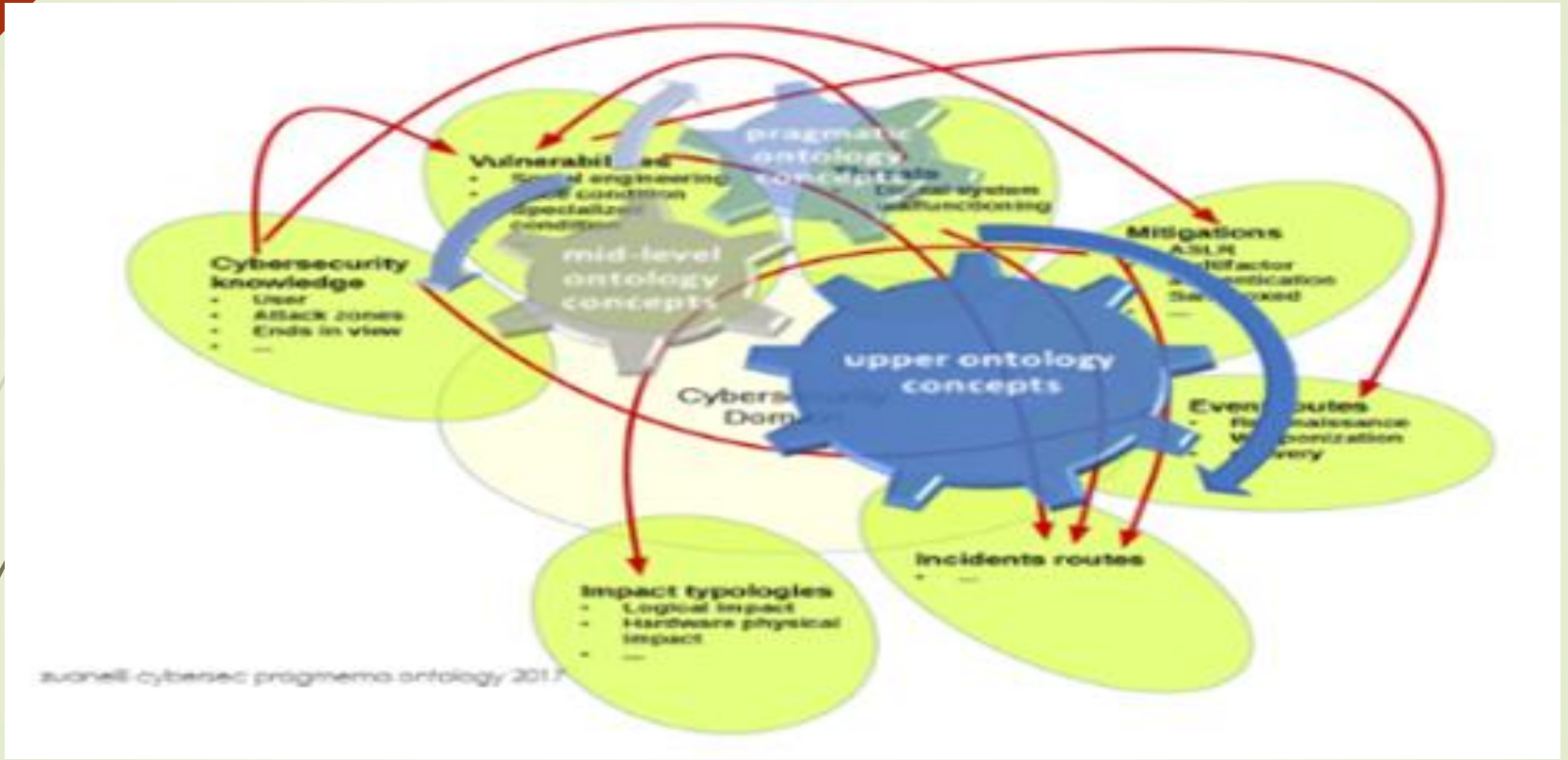
Criteri definitori ontologici

- Entità
- Attributi
- Proprietà
- Relazioni logico-semantiche (sintattiche-testuali)
- Vocabolari controllati
- Traduzione tecnologica

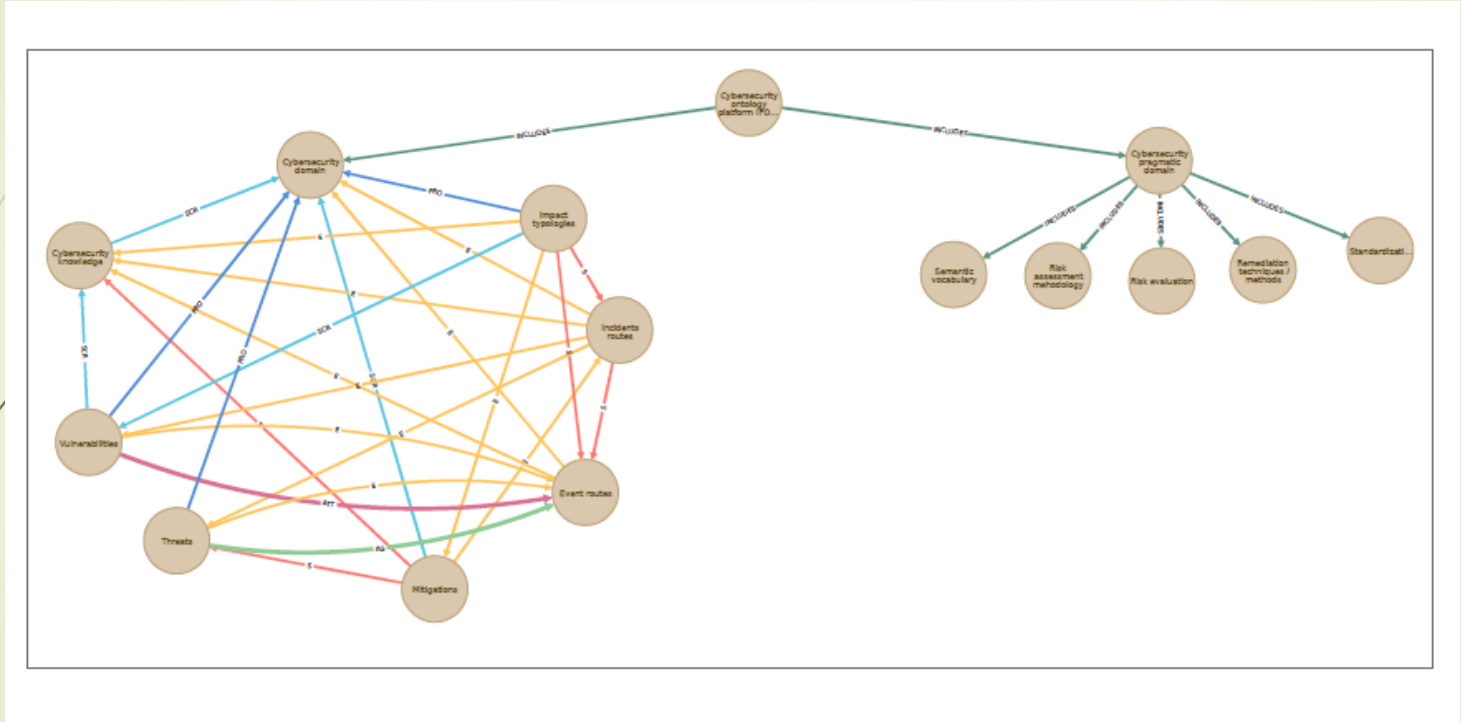
- Di immediato interesse applicativo nel contesto normativo di ***compliance*** per la sicurezza nella Direttiva NIS e nel GDPR
- Customizzabile per utenti diversi nel pubblico e nel privato
- Con funzioni di interoperabilità

La rete di relazioni su entità di primo livello

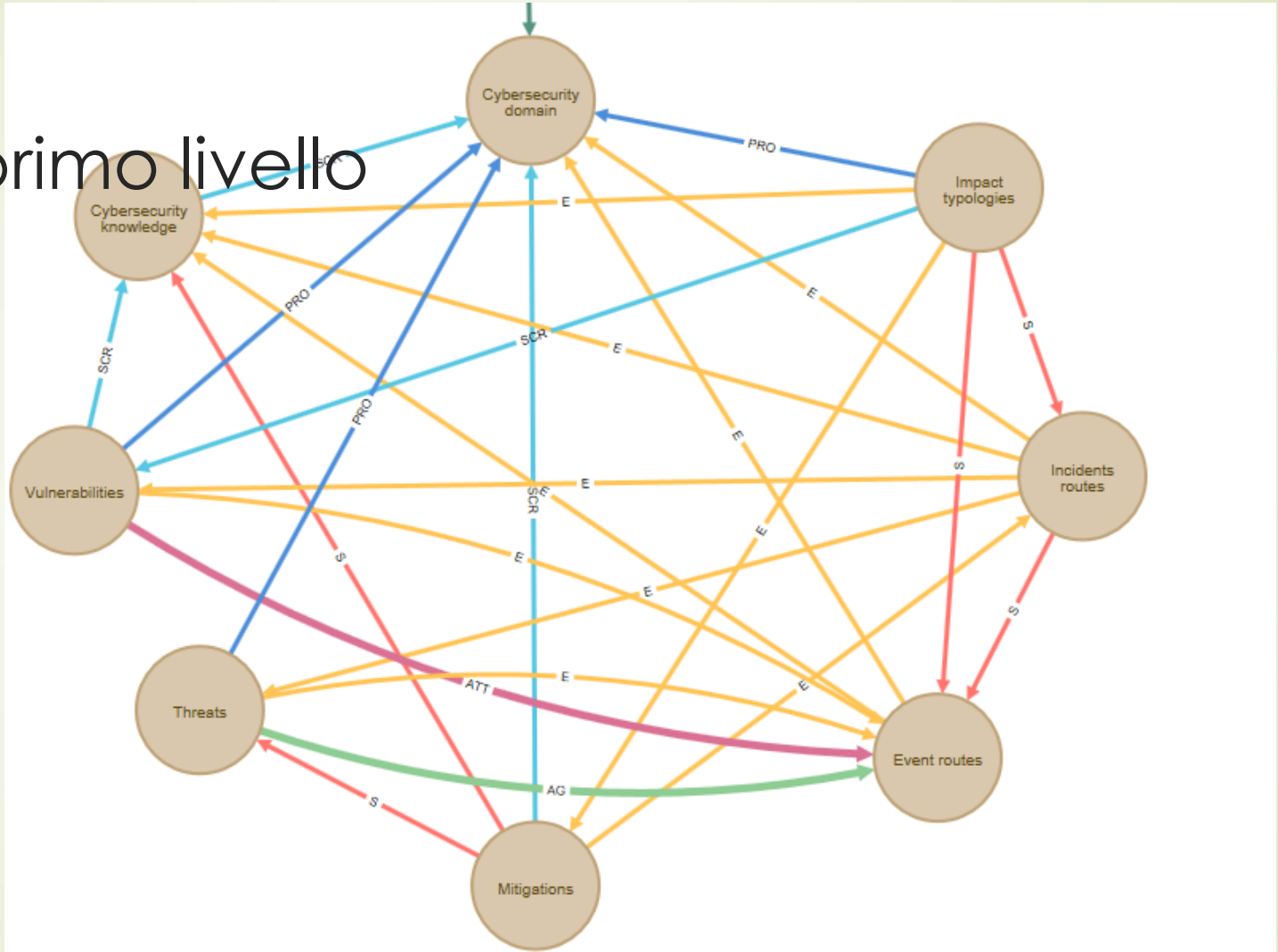



Cybersecurity knowledge
Mitigations
mid-level ontology concepts
upper ontology concepts
Incidents routes
Impact typologies
Logical impact
Hardware physical impact

Il modello a grafi



Reti di primo livello



- POC dispone di due livelli corrispondenti a due oggetti ontologici/tecnologici: l'**ontologia di dominio cybersecurity** e l'**ontologia pragmatica di servizi** alla cybersecurity

- I due oggetti sono **commerciabili separatamente o in combinazione**. Possono essere **integrati ai dispositivi aziendali cybersecurity già installati o da installare**, costituendo un livello di automazione e *data analytics* unico

Le funzioni dell'ontologia di dominio

Oltre 600 entità del dominio *cybersecurity* e relativi nodi relazionali di tipo logico-semantico

La rappresentazione ontologica della *cybersecurity knowledge* e delle vulnerabilità, *threat*, impatto, resilienza, ecc. consente:

- la possibilità di **prevenire e individuare** eventi e incidenti cibernetici
- la definizione di **metodi e tecnologie per il *risk assessment* e la *risk evaluation***
- l'applicazione di **sistemi rimediali**, tecnologici e comportamentali
- l'impiego di **standard** per l'automazione della sicurezza
- la conoscenza, la definizione e la rappresentazione degli **elementi costitutivi di eventi e di incidenti *cybersecurity***, necessarie per la reportistica dei dati
- l'**analisi automatica delle variabili tipologiche** che definiscono gli eventi e gli incidenti
- modalità di **rappresentazione dei dati**

Le metriche

- L'ontologia ricomprende la rappresentazione strutturata di masse di dati
- Correlazione tra *asset tecnologici* dell'ente/azienda, *cybersecurity implementation, eventi e incidenti*
- Tipologia di registrazione del valore di rischio e *assessment* degli incidenti

La Direttiva NIS e il GDPR (DPIA): correlazione e automazione compliance

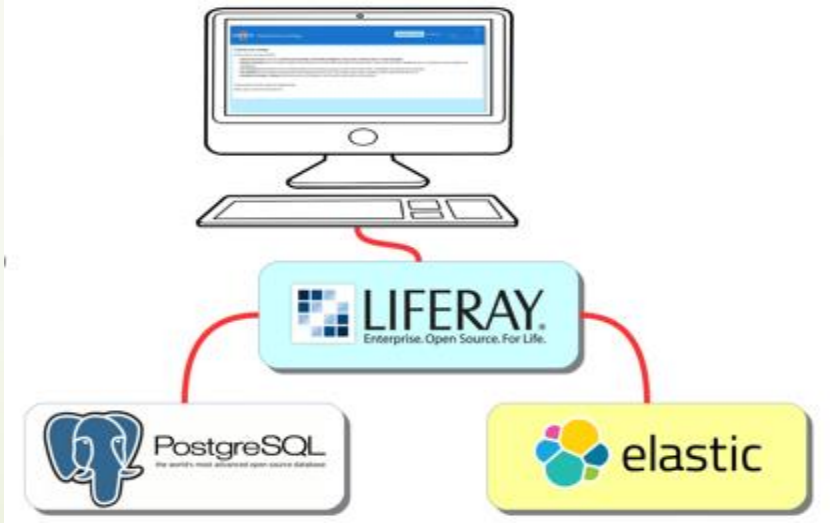
- Operative da maggio 2018
- *Risk assessment* e *risk evaluation*
- Adozione standard tecnologici sicurezza (NIS)per aziende di servizi essenziali e aziende di servizi digitali (*search engine*, *e commerce*, *cloud*, DNS, IXP e Registrar)

Stakeholder privilegiati

- **CERT, CSIRT, SIEM, SOC e aziende di infrastrutture critiche e servizi essenziali** come da Direttiva NIS, GDPR e standard sicurezza internazionali
- I dati aziendali inseriti nel database customizzato consentono di analizzare le evoluzioni del sistema di sicurezza nella logica della ***big data analytics e AI***. La correlazione dei dati consente di pervenire a una interpretazione preventiva e, in prospettiva, predittiva degli attacchi
- La piattaforma ontologica può essere trasferita su tool tecnologici di AI e big data: varie *libraries,* ecc.

Architettura e infrastruttura tecnologica

- POC è sviuppato su Liferay, portale orizzontale scritto in Java;
- Post-gres,database relazionale e ricerca elastic; motore di ricerca scritto in Java



La forza dell'architettura POC: il parametro strutturale

- Il **server applicazioni**: Liferay è costruito sui microservizi Java e progettato per essere un server a *clustered persistence*
- Il **database relazionale** Postgres è progettato per essere replicato
- Il **server di indicizzazione e ricerca** Elastic search è progettato per gestire ricerche distribuite

I servizi tecnologici alla cybersecurity della piattaforma POC

I servizi della piattaforma/ontologia pragmatica POC Pragmema

Tutti i servizi sono retti dall'ontologia di dominio *cybersecurity*:

•il **vocabolario semantico controllato** della *cybersecurity* (VSC) (e i vocabolari generici) è il servizio di definizione delle informazioni e dei dati utili per gestire univocamente le attività di cybersecurity

•***risk assessment***: il servizio consente di verificare lo stato dell'arte preliminare e in progress del sistema e nel tempo l'andamento ovvero il *check* di sistema di sicurezza inerente dati e tecnologie delle aziende e delle istituzioni. Su richiesta del cliente il modello *check* può essere personalizzato e interrelato ai sistemi tecnologici di *cybersecurity* aziendali/istituzionali

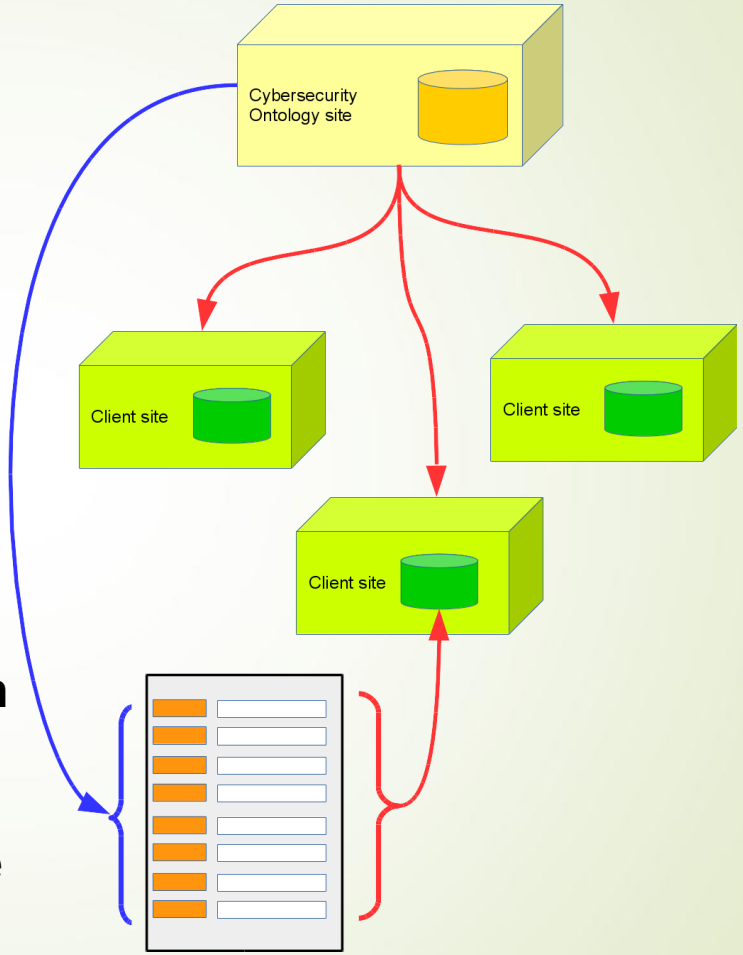
•***risk evaluation***: il servizio consente di valutare il rischio economico dell'azienda in caso di attacchi *cyber*. Può essere customizzato …

I servizi alla cybersecurity

- ***remediation techniques***: il servizio consente di definire le attività di resilienza in caso di attacchi. Può essere customizzato
- ***standard references***: il servizio offre e correla la gamma di standard internazionali richiesti per la *cybersecurity*. E' connesso ai **form/*application tools***
- gli **application tools** contengono form/schemi da compilare e memorizzare sul *data base* dell'azienda o dell'istituzione correlandosi ai servizi di sicurezza presenti nel sistema informativo informatico ed effettuare le seguenti attività: reportistica di eventi ed incidenti, statistiche, valutazione del rischio, ***assessment***, ecc. Consente **l'automazione del servizio** e dei **segnali di rischio**

Features of all services

- Every client has **his/her own service site**
  - Only subscribed services are accessible
  - **Client data remain in the client site**
- Each service displays:
  - Structured forms to **gather field data**
  - Each field data has a **contextual explanation/implementation** based on the Cybersecurity Ontology
  - Forms data **are stored in the client site and represent a time series**



Risk assessment service

- Acquisition forms for:
  - Asset analysis
  - Reporting
  - Incidents identifiers
  - Event description
  - Metrics
  - Statistics
- Search and correlation of the acquired data
- Access to the related knowledge base in the Cybersecurity Pragmatic Domain

Risk evaluation service

- Acquisition forms for:
  - Metrics
  - Statistics
- Search and correlation of the acquired data
- Access to the related knowledge base in the Cybersecurity Pragmatic Domain

Remediation techniques service

- Acquisition forms for:
  - Preparation
  - Identification
  - Containment
  - Eradication
  - Recovery
  - Post-incident activity
- Search and correlation of the acquired data
- Access to the related knowledge base in the Cybersecurity Pragmatic Domain

Risk evaluation service

- Acquisition forms for:
  - Metrics
  - Statistics
- Search and correlation of the acquired data
- Access to the related knowledge base in the Cybersecurity Pragmatic Domain

Standards reference service

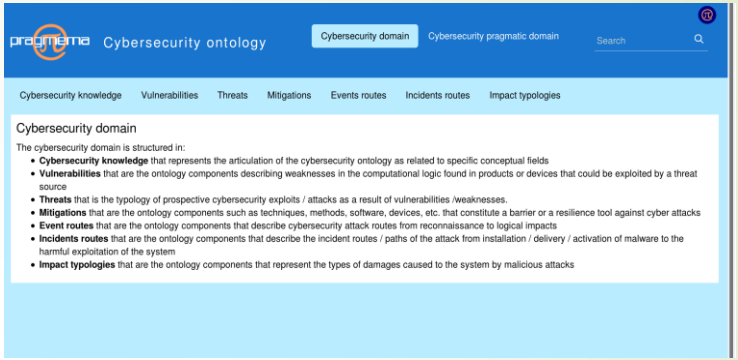
- Access to the knowledge base in the Cybersecurity Pragmatic Domain
  - Access to relevant standards with annotated focal points
  - Standard documents correlation with Cybersecurity Ontology structure
  - Standard documents are organized following the Cybersecurity Ontology structure

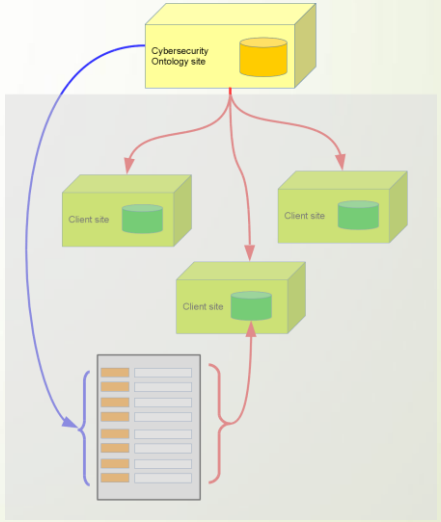
Controlled Semantic Vocabulary service

- Access to the knowledge base in the Cybersecurity Domain
- Multiple vocabularies with search functions
  - Cybersecurity Ontology Domain Vocabulary
  - Common attack Pattern Enumeration and Classification (CAPEC™)
  - Others

Pragmema Cybersecurity Ontology Domain

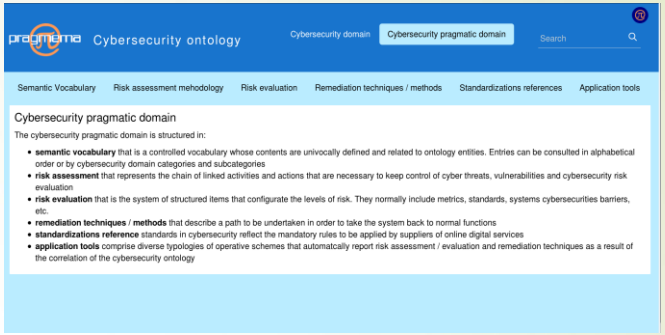
- Cybersecurity domain is structured in:
  - Cybersecurity knowledge
  - Vulnerabilities
  - Threats
  - Mitigations
  - Event routes
  - Incidents routes
  - Impact typologies

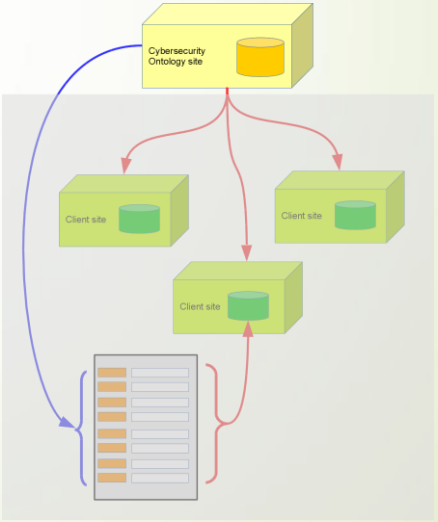




Pragmema Cybersecurity Pragmatic Domain

- Cybersecurity pragmatic domain is structured in:
  - semantic vocabulary
  - risk assessment
  - risk evaluation
  - remediation techniques / methods
  - standardizations reference





Gli sviluppi

- Machine learning
- Predittività
- Resilienza automatica