Cybersecurity: tra big data analytics e intelligenza artificiale
SOIEL 2017

Elisabetta Zuanelli

Presidente CReSEC/Università degli Studi di Roma «Tor Vergata»

Coordinatore del Piano di formazione nazionale in cybersecurity, cyberthreat, privacy

Gli argomenti

- La rilevanza per chi e per cosa
- A che punto siamo
- Quali le prospettive dell'occupazione e del mercato

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 e trasferimento 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)

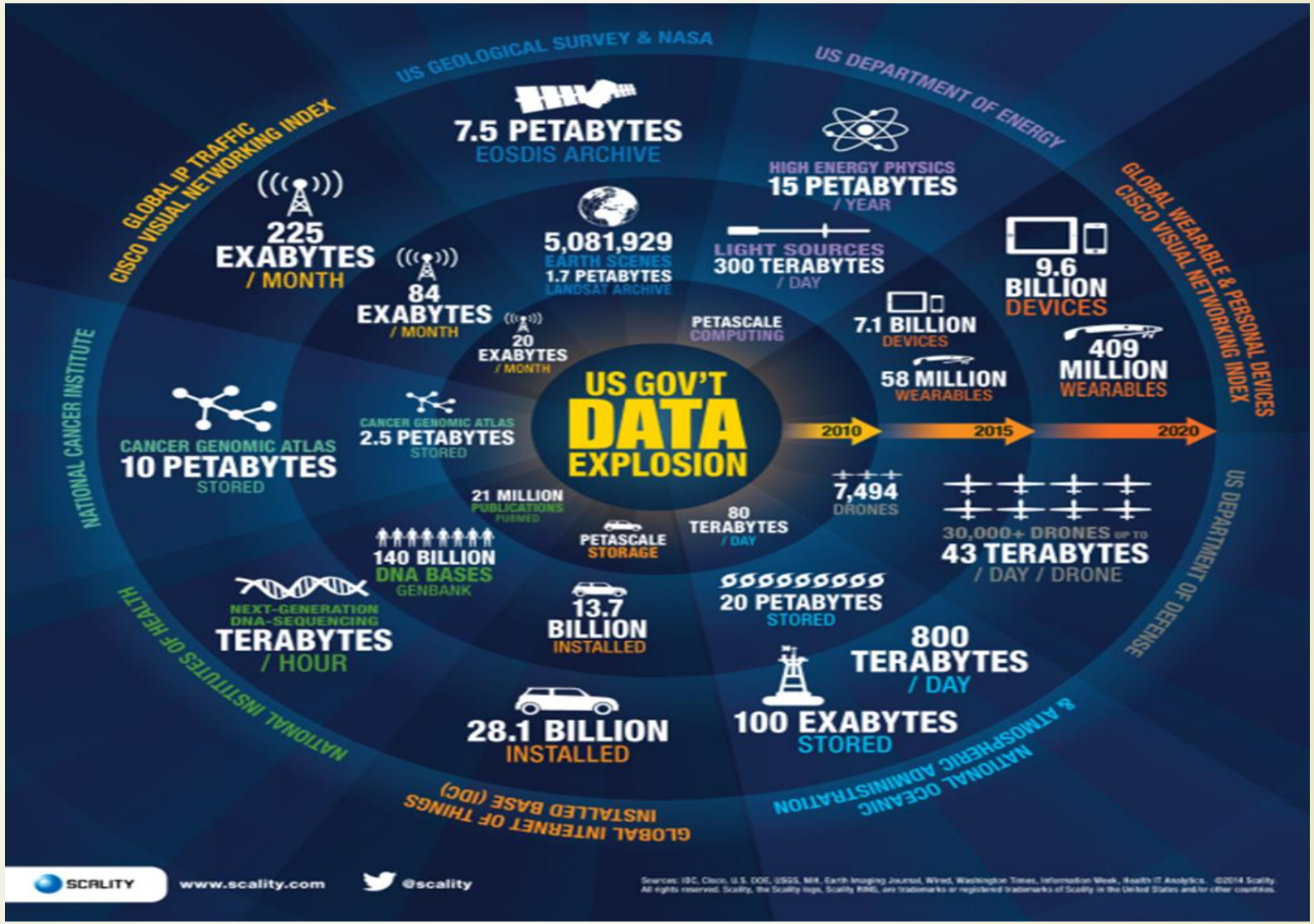
Finalità: pronti interventi, mitigazione, resilienza, prevenzione, predittività

- 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

L'esplosione dei dati



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
- Utilità: destinazioni tecniche (CERT, CSIRT, SIEM, ecc.) e laiche
- Formalizzazione e gestione automatica dei dati

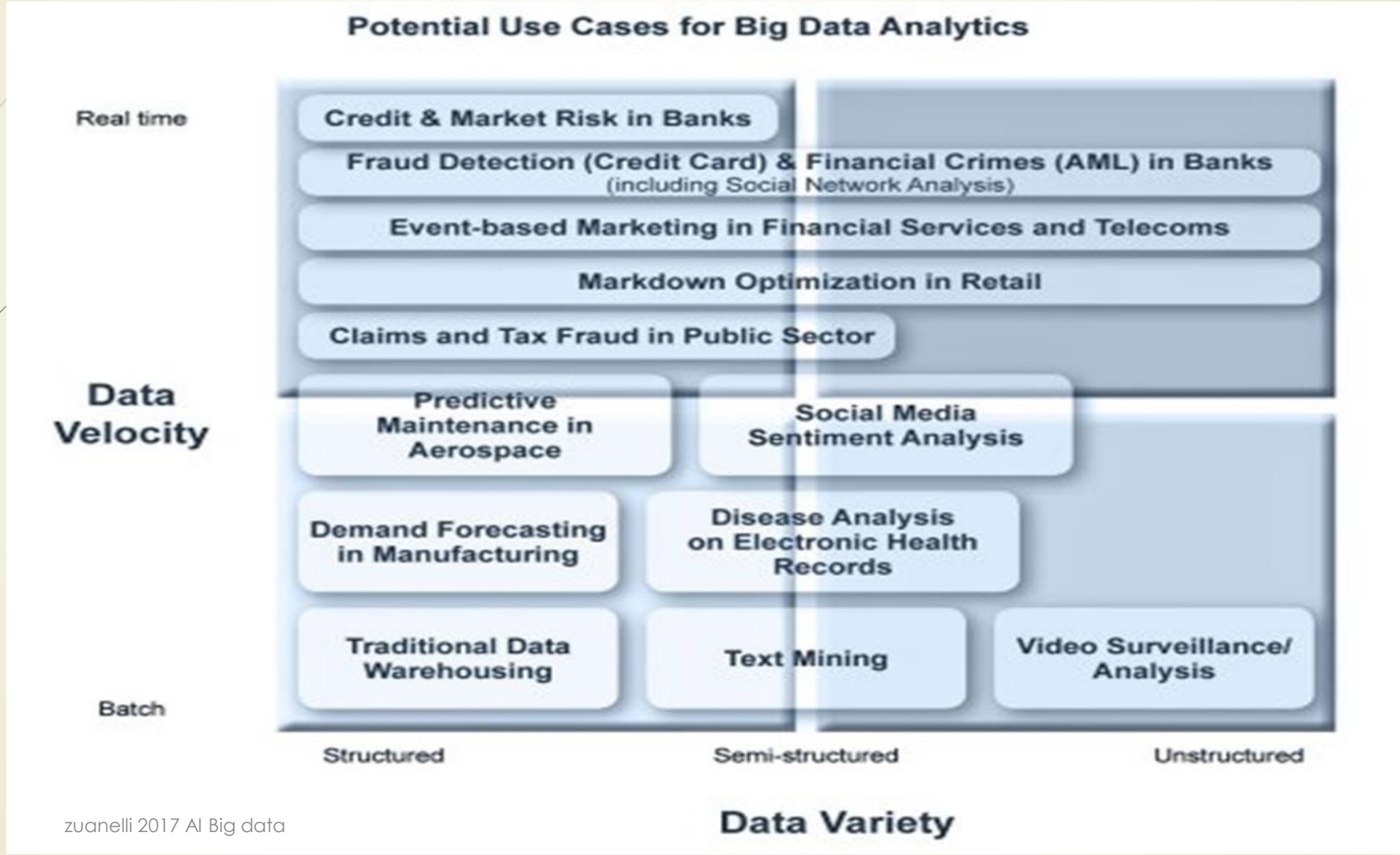
ENISA 2016
Analisi comparativa tassonomie cybersecurity

"**There is currently no consensus on concepts and definitions related to taxonomies**"

attualmente le tassonomie esistenti

"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".

Tipologia e qualità 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

Collecting

Sorting
Consolidating

Asset exposure

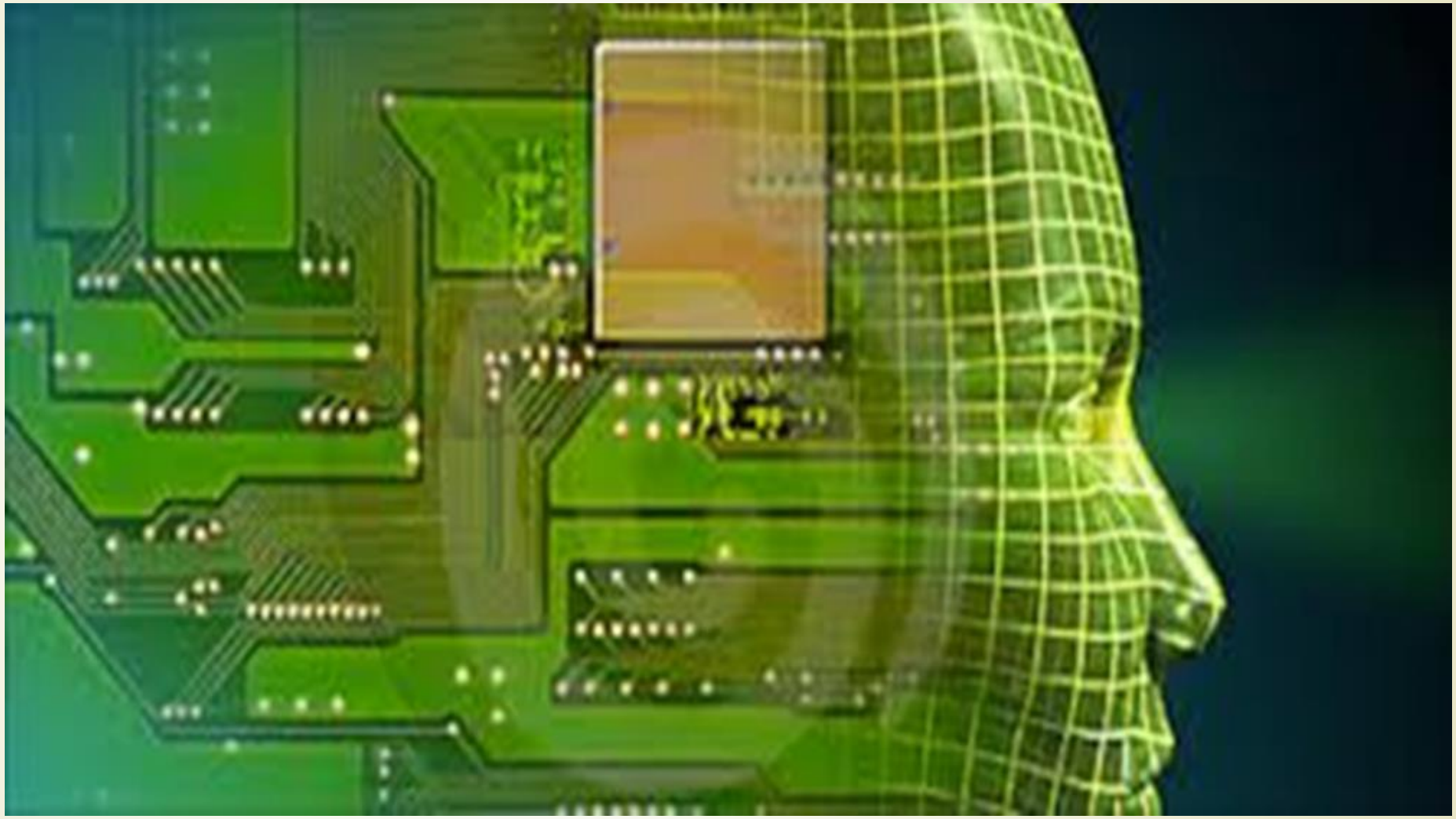
Vulnerability exploitation

Figure 1: ENISA Threat Taxonomy and its use-cases

AI: emulazione del funzionamento della mente umana

- 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

La mente digitale



L'intelligenza artificiale e i dati

- Modellizzazione dei dati e delle relazioni logico-semantiche
- Definizione e sviluppo del modello
- Definizione dei metadati
- Linguaggi di metadati
- Formati di rappresentazione dei dati

Le metriche

- Tipologia di registrazione del valore di rischio e assessment degli incidenti
- Necessitano di un'ontologia
- L'ontologia deve ricomprendere la rappresentazione strutturata di masse di dati

La conoscenza della cybersecurity: l'ontologia della vulnerabilità (NISTIR 8138 Draft 2016, et alii

- Elementi di conoscenza costitutivi della rappresentazione della **vulnerabilità** intesa come «debolezza nella logica computazionale dei prodotti e servizi che può essere sfruttata da una fonte di minaccia»
- Teatri remoti esterni: Internet (DNS services, web browsers, siti pubblici)
- Remoti limitati: cellulari, wireless, bluetooth, ecc.
- Bugs degli applicativi interni (sistemi operativi, programmi installati)
- Localizzazione degli attacchi esterni: piattaforme ecommerce, social, cloud, WoT, ecc.
- Tipologia malware e finalità degli attacchi: per esfiltrazione dati, corruzione, controllo di sistema, ecc.
- Logiche di percorso: es. Kill Chain

Vocabolari controllati per gli standard

- Gli standard richiedono vocabolari controllati CV a livello di contenuti e rappresentazione– Standard principali (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

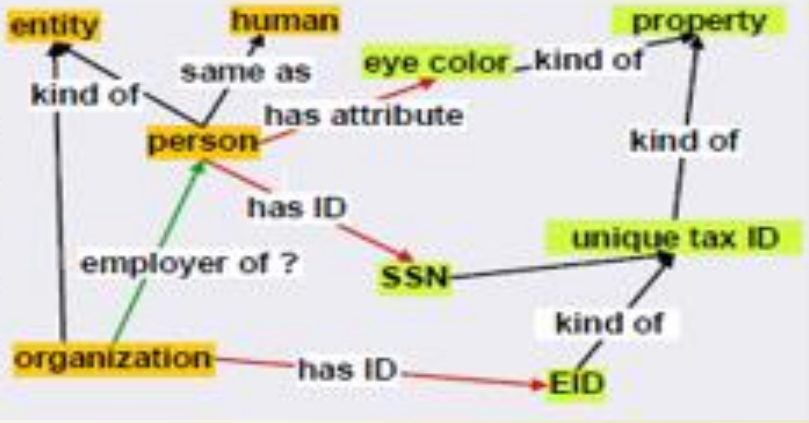
- ■ Both MITRE and NIST maintain public repositories and Web sites for the various standards: http://nvd.nist.gov/ ,
- http://oval.mitre.org/repository/ http://measurablesecurity.mitre.org/

Criteri definitori ontologici

- Entità
- Attributi
- Proprietà
- Relazioni logico-semantiche
- Vocabolari controllati
- Traduzione tecnologica

Esempio di ontologia, CV e interoperabilità semantica

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	

Le nuove professionalità

- Business Intelligence Developer, Web Developer, Software Developer, Business Developer, Analyst Developer, Applications Developer, Database Developer and Front-End Developer

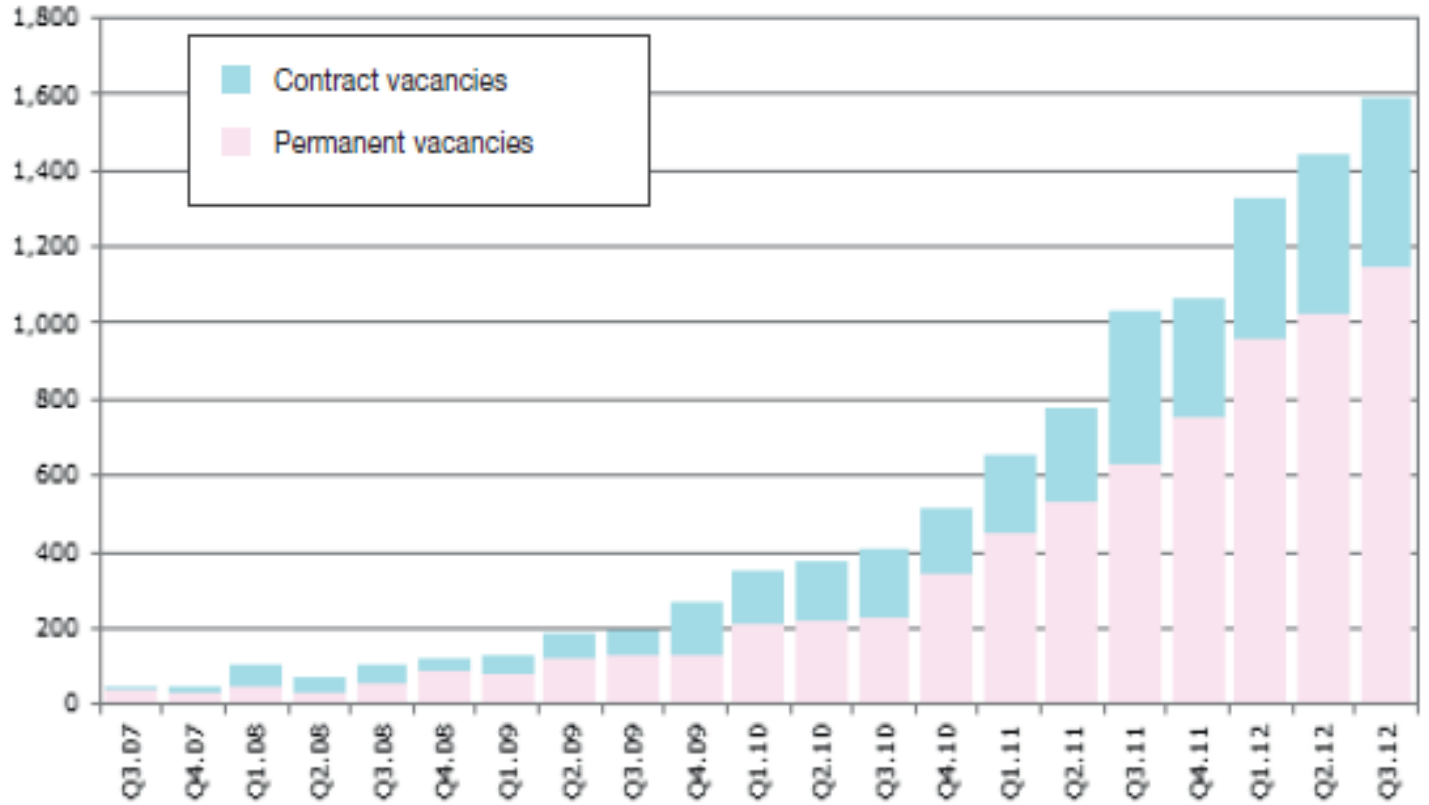
- The top three skills required of big data developers are

NoSQL, Java and SQL

Domanda di sviluppatori di Big data nell'UK

- Stima dell'aumento di domanda 2007-2012 del 673% per anno

Figure 5: Demand for Developers from big data recruiters 2007–2012



Average number of vacancies per quarter	
2007	60
2008	100
2009	200
2010	410
2011	880
2012*	1,460

* Average for Q1-Q3 only

Source: e-skills UK analysis of data provided by IT Jobs Watch

Stima del mercato di AI

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

Partenariato PPP

- Piano di Formazione Nazionale in materia di Cyber Security, Cyber Threat, Privacy
- La Direttiva UE e il Regolamento privacy del 2016 richiederanno per il 2018 l'attivazione di nuove competenze e figure professionali in chiave multidisciplinare finalizzate allo sviluppo di specifiche competenze in materia di Cyber Security, Cyber Threat e Privacy per rispondere ai rischi crescenti.
- L'Università degli Studi di Roma "Tor Vergata", attraverso il CReSEC (Centro di Ricerca e Sviluppo sull'EContent), ha promosso un Partenariato per la realizzazione di un Piano di Formazione Nazionale e di sviluppo tecnologico in cybersecurity, cyberthreat, privacy.

Le collaborazioni sono gradite.
Grazie