

Alma Mater Studiorum Università di Bologna

Corso di Laurea Magistrale in Ingegneria e Scienze Informatiche
Campus di Cesena

**Una proposta di ontologia compatibile agli
standard CIDOC-CRM del tracciato
ICCD - OA**

Relatore:
Prof. Antonella Carbonaro

Presentata da:
Antonio Leo Folliero
Mat:0000702696

SESSIONE III
Anno Accademico 2015-2016.

*Alla mia famiglia,
Mamma, Papà, Nonni,
ad Alessandro, Alessia e
Deborah e a tutte le
persone che mi hanno
sostenuto.*

INDICE:

1.0 - INTRODUZIONE	6
1.1 - EVOLUZIONE DEL WEB	8
1.1.1 - Web 1.0	10
1.1.2 - Web 2.0	11
1.1.3 - Web 3.0	13
1.2 - STACK DEL WEB SEMANTICO	16
1.2.1 - Strato di base URI (Uniform Resource Identifiers)	16
1.2.2 - Strato di mark-up XML (eXtensible Markup Language)	17
1.2.3 - Strato dei modelli RDF (Resource Description Framework) e RDF Schema	18
1.2.4 - Strato ontologico	20
1.2.5 - Strato logic e proof	21
1.2.6 - Strato Trust	22
1.3 - LOD - Linked Open Data	23
1.3.1 - LOD e gli Archivi Culturali	24
2.0 - IL PATRIMONIO CULTURALE	27
2.1 - Il patrimonio culturale e il Web	28
2.2 - STRATEGIE DI INTEGRAZIONE SU PORTALI WEB	30
2.2.1 - Ricerca federata	31

2.2.2 - Data warehousing	32
2.3 - PUBBLICAZIONI COLLABORATIVE	32
2.3.1 - Standard catalografico italiano - ICCD	33
2.3.2 - CIDOC Conceptual Reference Model	35
2.3.3 - Caso di studio - Europeana	37
2.3.4 - Caso di studio - ZERI FOUNDATION - Progetto ZERI & LODE	38
2.3.5 - Caso di studio - Amsterdam Museum	41
2.3.5 - Caso di studio - Semantic Web Services	42
3.0 - INTRODUZIONE AL PROGETTO DI TESI	51
3.1 - ONTOLOGIE DI RIFERIMENTO	52
3.2 - CREAZIONE DEL MODELLO - Protège	56
3.3 - MAPPING - CICDO - Central Institute for Cataloguing and Documentation Ontology	57
3.3.1 - Mapping CD - Codici	59
3.3.2 - Mapping OG - Oggetto	63
3.3.3 - Mapping LC - Localizzazione geografico-amministrativa	64
3.3.5 - Mapping DT - Cronologia	66
3.3.6 - Mapping MT - Dati tecnici	68
3.3.7 - Mapping TU - Condizione giuridica e vincoli	71
3.3.8 - Mapping DO - Fonti e documenti di riferimento	73

3.3.9 - Mapping AD - Accesso ai dati	77
3.3.10 - Mapping CM - Compilazione	79
3.3.11 - Grafo ontologia CICDO	83
4.0 - MAPPING TRACCIATO OA - CICDO-OA ontology	84
4.1 - SPECIALIZZAZIONI DEL MODELLO CICDO-OA	87
4.2 - MAPPING CICDO-OA	91
4.2.1 - Mapping AC - Altri codici	92
4.2.2 - Mapping LDC - Collocazione Specifica	93
4.2.3 - Mapping RE - Modalità di reperimento Oggetto D'arte	96
4.2.4 - Mapping CO - Conservazione	99
4.2.5 - Mapping DO - Documentazione fotografica	101
4.2.6 - Grafo ontologia CICDO-OA	105
5 - REASONER HeremiT	106
6 - CONCLUSIONI	110
BIBLIOGRAFIA	111

1.0 - INTRODUZIONE

Il Web semantico si prefigge di dare la possibilità ai documenti pubblicati sul Web di essere associati ad informazioni e dati che ne specificano il contesto semantico, in questo modo si dà la possibilità ai calcolatori di elaborare e interpretare il contenuto dei documenti in maniera del tutto automatica.

Le informazioni anzichè essere indicizzate tramite le pagine che le contengono vengono indicizzate tramite i metadati che le descrivono.

Essendo questa evoluzione sostanziale in quanto le informazioni iniziano ad essere indicizzate attraverso i metadati che le descrivono e non più rispetto alle pagine che le contengono si inizia a parlare del Web semantico come Web di dati (Web 3.0) e non più di Web di documenti ipertestuali (Web 1.0).

Con il Web di dati la descrizione di un oggetto diventa essa stessa un'informazione riutilizzabile nel Web, di conseguenza si vuole una grado di relazioni tra informazioni standardizzato e attendibile, in modo da fornire una base di conoscenza leggibile ed utilizzabile da macchine e persone.

Il mondo degli enti catalogatori e il mondo di Internet sono entrambi interessati all'integrazione delle informazioni sulla Rete; il primo per garantire la visibilità e l'usabilità dei dati, il secondo per sfruttare al meglio il proprio ruolo di generatori di informazione di qualità per il Web of data.

Nei paragrafi successivi si ripercorrerà tutta l'evoluzione del Web fino ad arrivare al Web semantico e al suo obiettivo principale cioè quello di estendere il Web come un bene comune di dati interconnessi tra loro. Inoltre ci specializzeremo nell'ambito dei Beni Culturali da sempre dati fortemente strutturati e affidabili e quindi con informazioni riutilizzabili in un'ottica Linked Open Data. Il bene culturale, soggetto della tesi, per il quale si vuole proporre un ontologia di riferimento è l'Opera e oggetto D'arte, documentato attraverso la normativa OA[1] (definita dall' Istituto Centrale per il Catalogo e la

Documentazione[2]). La tesi dapprima fornirà un'ontologia comune a più beni culturali, nello specifico Opera e oggetto d'arte, Fotografia, Reperto Storico, denominata CICDO Ontology, in seguito specializzerà il modello ontologico ottenuto, aggiungendo la descrizione in termini ontologici delle informazioni relative solo al tracciato OA ottenendo un'ontologia dedicata definita CICDO-OA Ontology. In entrambi i casi si descriverà la mappatura delle informazioni catalografiche in statement RDF[3] compatibili allo standard CIDOC-CRM[4].

Andiamo ora a descrivere la struttura di questo lavoro di tesi.

Nel primo capitolo si descrive l'evoluzione del web dalla prima versione contenente documenti ipertestuali ossia Web 1.0, al Web semantico ossia Web 3.0, dopodichè si andrà ad analizzare nello specifico la struttura del web semantico e si introdurrà l'argomento di Linked Open Data.

Il secondo capitolo descrive il patrimonio culturale e il modo in cui i dati relativi ai beni culturali vengono esposti nel Web. Si analizzeranno varie tecniche di integrazione di questi dati su portali web e si vedranno alcuni casi reali.

Il terzo capitolo descrive come si è proceduto a definire una prima versione dell'ontologia CICDO-OA, proponendo l'ontologia CICDO ontologia comune a più tracciati ICCD, e descrivendone peculiarmente la concettualizzazione OWL.

Il quarto capitolo descrive l'estensione e la specializzazione dell' ontologia CICDO in CICDO-OA ossia una proposta di ontologia compatibile agli standard CIDOC che descrive i concetti espressi nel modello ICCD OA 'Opera e oggetto d'arte', nonchè l'ontologia soggetto della tesi.

Il quinto capitolo illustrerà alcuni ragionamenti inferiti dal reasoner HeremiT sull'ontologia CICDO-OA, il quale è stato utilizzato anche per garantire la coerenza dell'ontologia costruita.

1.1 - EVOLUZIONE DEL WEB

Il 12 marzo del 1989, Tim Berners-Lee, informatico britannico allora capo ricercatore del CERN (The European Organization for Nuclear Research[5]), suggerì la creazione di uno spazio di ipertesto globale in cui ogni informazione fosse accessibile tramite la rete indicandola attraverso un Universal Resource Identifier (URI[6]).

Nel Novembre 1990 – Nei laboratori del Cern di Ginevra, Tim Berners-Lee creò il primo indirizzo Web (<http://nxoc01.cern.ch>), ossia un protocollo (HTTP[7]) che permise di raggiungere una stessa pagina HTML da postazioni/calcolatori differenti. La finalità principale di questo progetto derivava dalla necessità di condividere files all'interno di uno spazio comune. Per visualizzare tale pagina sullo schermo, Berners-Lee creò un'interfaccia, il browser World Wide Web. La pagina, raggiungibile solo attraverso i calcolatori del Cern, si presentava come un documento ipertestuale, cioè una pagina di collegamento che permette di raggiungere tutti i documenti allora presenti online. Nacque così il World Wide Web o Web 1.0[8], un grande ipertesto volto ad agevolare la condivisione dei documenti.

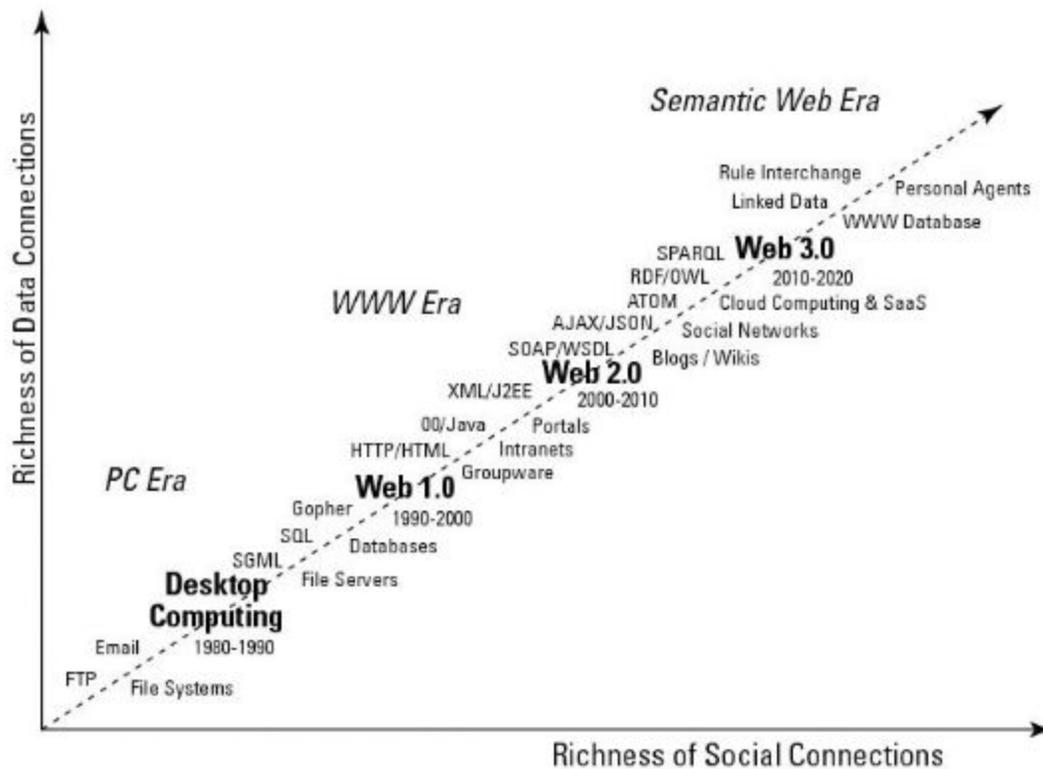


Figura 1.0: Evoluzione del Web

Nel Web 1.0, o prima versione del Web, i siti Web erano semplici testi statici condivisi in uno spazio comune accessibile da più postazioni. Queste pagine avevano lo scopo di fornire consulenza e informazioni senza permettere l'interazione fra utente e contenuto. I siti erano dunque formati da pagine ricche di ipertesti, pagine contenenti collegamenti ad altre pagine che formavano un enorme rete di cui i documenti sono i nodi. L'impossibilità di questa rete di interagire con i lettori spinse i ricercatori a trovare nuove soluzioni o se non altro un'evoluzione di questo primo step evolutivo per rendere dinamico e più interattivo il Web. La trasformazione ebbe inizio con la possibilità di inserire dei commenti all'interno delle pagine Web, creando così i primi forum e i primi blog. Da qui iniziarono a nascere le prime community fino ad arrivare all'avvento dei social network e dei wiki (dove gli utenti possono reperire informazioni, modificarle e aggiungerne altre, es: Wikipedia) fin ad arrivare a quello che

attualmente è il Web 2.0 (termine coniato da Tim O'Reilly[9]) il Web dinamico. In questo step evolutivo si è data grande importanza all' usabilità e al modo di condividere i contenuti, ora gli utenti possono condividere e creare un'enorme quantità di informazioni si pensi alle comunità di migliaia di utenti (Es.Facebook, LinkedIn, Twitter, Youtube), che ogni giorno acquisiscono e forniscono nuovi contenuti. Il Web non ha comunque smesso di evolversi e ad oggi di fatto stiamo entrando nel Web 3.0[10]: il Web della semantica e delle cose, Web of Things. Si parla di un unico enorme database, il WebDatabase, dove tutte le informazioni di internet confluiranno per velocizzare ricerche e semplificare la gestione dei dati, tutto sarà legato alle parole chiave legate ai documenti o alle informazioni ad essi associate e tutte le ricerche saranno legate a queste parole. Si parla anche di intelligenze artificiali che grazie ad algoritmi sempre più sofisticati permetteranno una ricerca migliore di contenuti in una rete sempre più affollata.

Dopo questa introduzione vediamo ora nello specifico i vari step evolutivi.

1.1.1 - Web 1.0

Web 1.0 è la prima generazione di Web durato dal 1989 al 2005, è stato progettato come una rete di connessioni tra informazioni. Tim Berners-Lee considera il Web 1.0 come "Read-only" Web, infatti esso fornisce molto poca interazione ma solo fruizione di informazioni.

Il Web 1.0 è stato sostanzialmente definito come "uno spazio di informazione in cui gli elementi di interesse tra cui le risorse sono identificate da un identificatore globale chiamato Uniform Resource Identifiers (URI) ".

Il Web di prima generazione aveva pagine e contenuti statici al solo scopo di consegna, con molta poca interazione con l'utente in termini di contribuzione di contenuti. Le principali caratteristiche sono:

- Le tecnologie incluse nel Web 1.0 sono HTML, HTTP e URI.

- L'obiettivo principale del Web 1.0 è rendere le informazioni disponibili a chiunque in qualsiasi momento.
- Le pagine web sono statiche.
- La principale limitazione del Web 1.0 è che le pagine web sono comprese solo da esseri umani e non da macchine, inoltre l'holder del sito web è l'unico responsabile e creatore di informazione all'interno di esso.

1.1.2 - Web 2.0

Il Web 2.0 rappresenterebbe, rispetto al Web 1.0, una versione nuova e sostanzialmente differente dal precedente al punto di meritarsi una nuova definizione:

“Web 2.0 is the business revolution in the computer industry caused by the move to the internet as platform, and an attempt to understand the rules for success on that new platform. Chief among those rules is this: Build applications that harness network effects to get better the more people use them. (This is what I've elsewhere called "harnessing collective intelligence.")” [T. O'Reilly, 2004].

I principi sul quale è stato definito il Web 2.0 sono:

- Il Web 2.0 è inteso come piattaforma:

In genere si distingue tra piattaforma di elaborazione e piattaforma di sviluppo.

Con il Web 2.0 si arriva ad applicazioni e servizi software, accessibili e utilizzabili direttamente on line, il processo si sposta dal proprio computer alla piattaforma di elaborazione Web, che mette a disposizione degli utenti le applicazioni in modalità di servizio. Il proprio computer diventa un

terminale di accesso al quale si chiede solo un collegamento internet e un browser (Thin client). Oltre al processo di elaborazione di applicazioni, vi è il processo di costruzione di nuove applicazioni, in questa situazione il web si configura come un ambiente che mette a disposizione tutti gli elementi necessari per creare nuove applicazioni (Web as a service).

Oltre a ciò il Web è inteso anche come piattaforma di condivisione e fruizione di informazione dove ogni utente può creare e diffondere contenuti.

- Il Web 2.0 è partecipativo:

Il Web 2.0 è volto a favorire l'aumento delle interazione degli utenti, sia tra di loro che rispetto alla Rete stessa. Gli utenti sono i principali creatori e diffusori di contenuti, la Rete diventa portatrice di un'intelligenza collettiva in continua crescita che si sviluppa attraverso le associazioni mediante hyperlink di informazioni fatte dagli utenti stessi. Ogni qual volta un contenuto viene pubblicato on-line qualsiasi utente può leggerlo, commentarlo e più importante di tutti citarlo o referenziarlo, in questo modo la rete cresce organicamente come risultato di un'attività collettiva dove i principali realizzatori sono gli utenti stessi.

Ci sono tante applicazioni e servizi che si propongono di sviluppare forme concrete di intelligenza collettiva, tra questi Google basato sulla struttura dei link e il PageRank o Wikipedia che qualsiasi utente può consultare liberamente contribuendo a migliorare o a integrare descrizione in un'ottica di partecipazione attiva.

- Il vero valore del Web 2.0:

Elemento centrale del Web 2.0 è rappresentato dal fatto che i dati prodotti dagli utenti siano il valore aggiunto del Web infatti essi contribuiscono costantemente ad arricchire i database della rete di informazioni sempre più numerose e dettagliate.

Secondo O'Reilly "la gestione dei database è una competenza centrale delle società Web 2.0 al punto che spesso abbiamo definito queste applicazioni come "infoware" piuttosto che semplicemente software". Un esempio di successo è Amazon che ha sfruttato i dati prodotti dai propri utenti nel campo dell'editoria. Amazon ha sistematicamente arricchito i dati in suo possesso, aggiungendo quelli forniti dagli editori, come le immagini delle copertine, le tavole dei contenuti, gli indici e materiali campioni diventando così una delle fonti primarie dei dati bibliografici sui libri.

1.1.3 - Web 3.0

John Markoff nel 2006 creò il termine Web 3.0 per identificare una terza generazione del Web.

L'idea di base del Web 3.0 è quello di definire dati strutturati e collegarli per rendere più efficace la ricerca, l'integrazione e il riutilizzo attraverso varie applicazioni.

Tim Berners-Lee e l'intera comunità W3C individuarono una sorta di errore di fondo nello sviluppo della Rete a partire dalla considerazione che tutta l'informazione all'interno del Web fosse disegnata per essere fruibile dall'uomo, la Rete invece, fu disegnata come uno spazio di informazioni, con l'obiettivo di essere non solo utile per la comunicazione uomo-uomo ma, anche macchina-macchina.

Peter Mika [2006] parla di knowledge gap, ovvero della differenza di conoscenza da parte di un calcolatore rispetto ad un utente per eseguire una computazione, come ad esempio una ricerca. Il knowledge gap si origina a causa della mancanza di quel background di conoscenza che gli umani posseggono, ed è spesso non deducibile dal contesto delle pagine Web, ciò fa sì che i calcolatori non possano lavorare con efficienza. A partire da questo errore evolutivo cioè lo sviluppo di un Web fatto di documenti, il Web Semantico ha come obiettivo quello di dare alle macchine la possibilità di intervenire attivamente e consapevolmente nel processo di gestione e elaborazione delle informazioni, in una Rete machine readable.

La visione di Berners-Lee riguardante il Web Semantico può essere letta in questa citazione:

“I have a dream for the Web [in which computers] become capable of analyzing all the data on the Web – the content, links, and transactions between people and computers. A “Semantic Web”, which makes this possible, has yet to emerge, but when it does, the day-to-day mechanisms of trade, bureaucracy and our daily lives will be handled by machines talking to machines. The “intelligent agents” people have touted for ages will finally materialize. “ [1999] .

Con il Web Semantico i documenti pubblicati (pagine HTML, file, immagini ...) sono associati ad informazioni e dati (metadati) che ne specificano il contesto semantico in un formato adatto all'interrogazione e all'interpretazione (es. tramite motori di ricerca) e, più in generale, all'elaborazione automatica.

Lo scopo dei metadati consiste nello specificare il contesto semantico in formati adatti all'interpretazione e all'elaborazione automatica. L'idea del Web machine readable non va pensato come un totale cambiamento bensì come l'aggiunta di uno strato di informazione, nel Web già presente, visibile solo alle macchine.

Per rendere i dati comprensibili alle macchine, anzichè indicizzare le pagine Web, si indicizzano i metadati che le descrivono, il Web Semantico è quindi un Web di dati e non più di documenti. Uno dei presupposti essenziali è costituito dall'utilizzo di linguaggi come ad esempio XML (eXtensible Markup Language[11]), che hanno introdotto la possibilità di strutturare i dati e sulla scia di esso RDF (Resource Description Framework) che attualmente rappresenta lo standard proposto da W3C per la codifica, lo scambio e riutilizzo dei metadati strutturati, esso infatti inquadra le strutture di dati secondo tassonomie gerarchiche rendendo possibile l'esecuzione di procedure inferenziali, inoltre si crea un valore aggiunto per via dell'opportunità di mescolare e integrare tra loro i dati. Il formato di serializzazione RDF con la sua sintassi XML è un formato funzionale e leggibile dalle macchine può esprimere le stesse informazioni di un database relazionale.

RDF non descrive la semantica ma bensì fornisce un modello per descrivere le risorse e le proprietà. L' RDF definisce quindi degli statement RDF basandosi su triple di oggetti risorsa-proprietà-predicato, definendo una risorsa come un qualsiasi oggetto che sia identificabile univocamente tramite un Uniform Resource Identifier (URI) al quale si associano delle caratteristiche con un certo valore.

1.2 - STACK DEL WEB SEMANTICO

Tim Berners-Lee ha proposto un'architettura a strati per il web semantico[12] spesso rappresentata con un diagramma. La figura seguente fornisce una rappresentazione di questo diagramma.

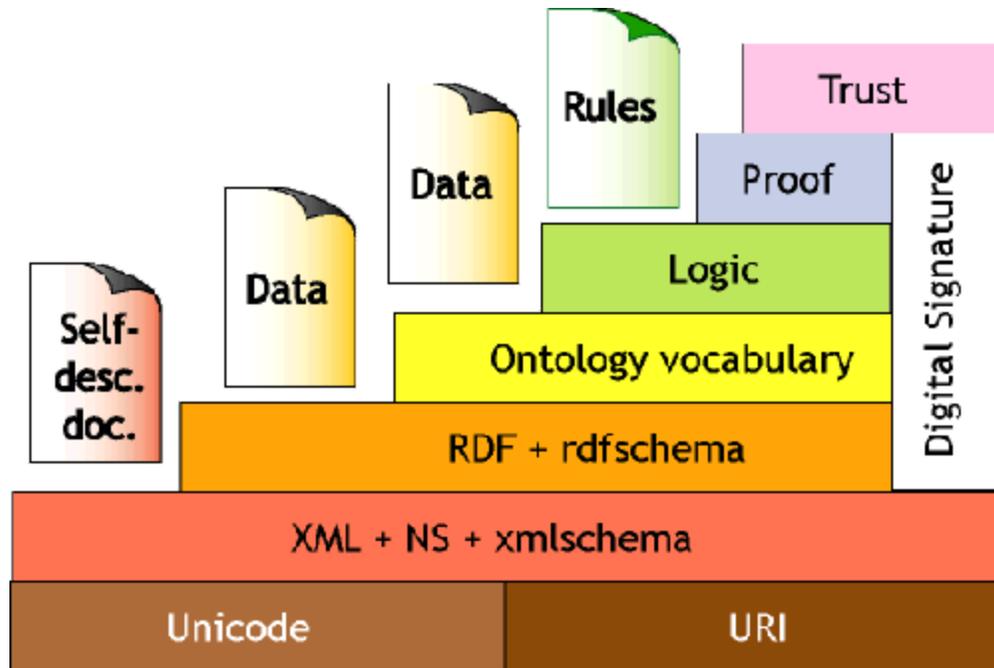


Figura 1.2: Stack del Web Semantico

1.2.1 - Strato di base URI (Uniform Resource Identifiers)

Il Web Semantico si basa sullo standard URI (Uniform Resource Identifiers), un sistema di identificatori unificato. Il nome deriva dal fatto che ogni elemento identificato viene considerato una risorsa. Gli URI sono utilizzati da RDF per codificare l'informazione in un documento. Gli URI costituiscono la tecnologia di base ideale con la quale costruire un Web globale. Possiamo definire un URI per un qualsiasi oggetto, e qualsiasi cosa che ha un URI può essere considerata sul

Web. Gli URI sono il fondamento del Web, mentre ogni parte del Web stesso può essere rimpiazzata, gli URI no. Anche per identificare le pagine sul Web utilizziamo identificatori: sono i tipi più comuni di URI, gli indirizzi URL (Uniform Resource Locator).

1.2.2 - Strato di mark-up XML (eXtensible Markup Language)

Al livello superiore si trova XML (eXtensible Markup Language), XML è il primo passo per assegnare una semantica ai tag e supportare le transazioni sul Web, permettendo lo scambio di informazioni tra database diversi. XML è un metalinguaggio per la definizione di linguaggi di markup, ovvero un linguaggio marcatore basato su un meccanismo sintattico che consente di definire e controllare il significato degli elementi contenuti in un documento o in un testo. Nel definire un documento XML e i relativi tag bisogna, rispettare determinati parametri (regole), in modo che il documento risulti corretto (well formed) per un parser che effettui la verifica su di esso. Tali insiemi di parametri vengono chiamati grammatiche formali, proprio per indicare il fatto che ogni documento XML è considerato well formed se e solo se rispetta le regole previste dalla grammatica a cui fa riferimento. Le grammatiche XML (come per es. DTD o XML schema) una volta definite diventano normative per tutti i documenti che vi fanno riferimento. Nella visione del Web Semantico l'XML permette di aggiungere contenuto machine readable e interpretabile in maniera differente a seconda di come verrebbe istruito il sistema che lo interpreta. Per far ciò è sufficiente dichiarare l'esistenza di blocchi semantici minimi (tag) e definire una grammatica con delle regole sintattiche volta ad assicurare, attraverso i tag, la correttezza formale nell'uso degli oggetti costruiti. Uno dei vantaggi dell'utilizzo dell'XML come linguaggio di markup universale risiede nella possibilità di accedere a informazioni provenienti da più fonti, un documento XML infatti può utilizzare anche più di una grammatica. Tale caratteristica rappresenta

sicuramente un grande vantaggio, ma ha anche una serie di inconvenienti tra cui il problema della disambiguazione del significato o polisemia. I documenti strutturati, nel momento in cui vengono confrontati o integrati, creano inevitabilmente conflitti tra termini, tag, etc. Ciò dipende essenzialmente dal fatto di essere sviluppati in maniera indipendente rispetto alla grammatica. Il problema è risolto dai namespace, essi permettono di etichettare ogni tag o attributo generato in un documento XML mediante degli URI che li identificano i tag in maniera univoca sul Web, indicandone la grammatica. Questa libertà lo rende poco adatto, però, a definire completamente la struttura e l'interscambio di informazioni tra diverse realtà, quindi è stata favorita la creazione di un nuovo linguaggio RDF.

1.2.3 - Strato dei modelli RDF (Resource Description Framework) e RDF Schema.

Le possibilità offerte dai namespace e in generale dagli strumenti di XML si sono dimostrate insufficienti per un'adeguata “semantizzazione” del Web. Il W3C infatti, considerando l'XML poco adatto a definire in maniera esaustiva la struttura delle informazioni, ha ritenuto opportuno sviluppare un nuovo linguaggio: RDF. Il Resource Description Framework (RDF “struttura per la descrizione di risorse”), sulla scia dell'XML, è linguaggio assertivo appositamente studiato per esprimere proposizioni utilizzando degli specifici vocabolari formalizzati e per consentire l'accesso all'informazione e l'utilizzo dei servizi del Web Semantico. Sia l'XML che l'RDF prevedono, oltre che l'utilizzo dei namespace, la distinzione dei documenti dagli schemi; questi ultimi, anche detti vocabolari. RDF è lo strumento base per la codifica, lo scambio e il riutilizzo di metadati strutturati, e consente l'interoperabilità tra applicazioni che si scambiano sul Web informazioni *machine-understandable*. RDF, quindi, non descrive la semantica, ma fornisce una base comune per poterla esprimere,

permettendo di definire la semantica dei tag XML. RDF è costituito da due componenti:

- *RDF Model and Syntax*: definisce il modello di dati RDF e la sua codifica XML, senza definire alcun livello di gerarchia o di relazione.
- *RDF Schema*: permette di definire specifici vocabolari per i metadati e creare nessi tra oggetti.

Quindi da una parte abbiamo i dati e dall'altra uno schema che definisce come i dati si strutturano e come sono in relazione fra loro. Qualunque cosa descritta dall'RDF è detta "risorsa". Principalmente una risorsa è reperibile sul Web, ma l'RDF può descrivere anche risorse che non si trovano direttamente sul Web. Ogni risorsa è identificata da un URI (Uniform Resource Identifier). Il modello dei dati dell'RDF è formato da risorse, proprietà e valori. Le proprietà sono delle relazioni che legano tra loro risorse e valori. Un valore è una risorsa ad un dato semplice quale, ad esempio, un numero. Questo modello è basato sull'idea di fare "statements" (dichiarazioni, asserzioni) sulle risorse (o sulla rappresentazione delle informazioni) nella forma di "subject - predicate - object" (soggetto - predicato - oggetto), queste espressioni sono conosciute con il nome di triple. Il modello dei dati RDF permette quindi di definire un modello semplice per descrivere le relazioni tra le risorse, in termini di proprietà identificate da un nome e dai relativi valori. Questo modello ha diversi formati di file che lo serializzano, quindi il meccanismo con il quale una risorsa o una tripla viene codificata varia da formato a formato alcuni di essi sono ad esempio: RDF/XML[13], Turtle[14].

Il modello dei dati, tuttavia, non fornisce nessun meccanismo per definire le relazioni tra queste proprietà ed altre risorse: questo compito è realizzato dall'RDF Schema. L'RDF Schema permette di definire dei vocabolari, quindi l'insieme delle proprietà semantiche individuato da una particolare comunità.

RDF Schema permette di definire significato, caratteristiche e relazioni di un insieme di proprietà, compresi eventuali vincoli sul dominio e sui valori delle singole proprietà. Inoltre, implementando il concetto (transitivo) di classe e sottoclasse, consente di definire gerarchie di classi. L' "RDF Schema" (RDFS o RDF-S), è un linguaggio estensibile per la rappresentazione della conoscenza, fornisce gli elementi di base per la descrizione dei vocabolari RDF, progettati per strutturare le risorse RDF. Molti dei componenti dell'RDF Schema sono inclusi nel linguaggio OWL Web Ontology Language[15] che permette più espressività.

1.2.4 - Strato ontologico

Lo strato superiore comprende le ontologie e i linguaggi relativi, in particolare OWL "Web Ontology Language", è quello che permette di concettualizzare le informazioni strutturate, dotandole di relazioni. OWL rappresenta lo standard sviluppato dal W3C per l'iniziativa del Web Semantico. Se XML e RDF si occupano di garantire un ordine, una struttura per le singole informazioni, le ontologie agiscono invece a un livello più alto. Secondo la definizione ufficiale un'ontologia è "una specificazione di una concettualizzazione" [Gruber, 1993 [16]]. In altre parole, un'ontologia permette di specificare i concetti e le relazioni che caratterizzano un certo dominio/base di conoscenza. Una base di conoscenza è un insieme coeso e coerente di dati singolarmente strutturati rispetto ad un concetto. OWL a differenza degli RDF Schema, che estende, aumenta le inferenze che un software è in grado di compiere.

OWL è diviso in tre sottolinguaggi a potenza espressiva crescente, ognuno dei quali è designato per un uso specifico.

- OWL Lite: serve a supportare la rappresentazione di classificazioni gerarchiche e vincoli semplici. Esso fornisce anche una migrazione veloce di tassonomie.

- OWL DL: supporta la massima espressività, ma conserva la completezza computazionale (cioè tutte le asserzioni sono garantite essere computabili) e la decidibilità (cioè tutte le elaborazioni terminano in un tempo finito). Esso comprende tutti i costrutti del linguaggio OWL ma, devono essere usati sotto certe condizioni. L'aggettivo "DL" è dovuto alla corrispondenza del linguaggio con la "Description Logic" (logica descrittiva) che è un campo della ricerca che ha studiato le logiche che formano la base formale dell'OWL.

- OWL Full: sviluppato per usi che necessitano la massima espressività e libertà sintattica dell'RDF: questo però non dà garanzie computazionali. L'OWL Full permette alle ontologie di aumentare il significato di vocabolari predefiniti, appartenenti sia all'RDF che all'OWL.

L'OWL Full può essere visto come un'estensione dell'RDF, mentre l'OWL Lite e l'OWL DL possono essere visti come estensioni di una visione ristretta dell'RDF. Ogni documento OWL è un documento RDF ed ogni documento RDF è un documento di OWL Full, ma solo certi documenti RDF saranno un documento valido per l'OWL Lite o per l'OWL DL. A causa di questo è necessaria, quindi, molta attenzione quando si vuole trasferire un documento da RDF a OWL.

1.2.5 - Strato logic e proof

Fino al livello delle ontologie non vi è inferenza, ma solo rappresentazione della conoscenza, questo strato comprende l'uso della logica per rispondere a interrogazioni al fine di poter estrarre informazioni utili e derivare nuova conoscenza mediante regole di inferenza e altre tecniche logiche. Lo strato della Logica rende quello che era un linguaggio dichiarativo, con limitate capacità espressive, un linguaggio logico con inferenze e funzioni, consentendo ad applicazioni RDF diverse di connettersi.

I sistemi deduttivi non sono normalmente interoperabili, per cui invece di progettare un unico sistema onnicomprensivo per supportare il ragionamento, si è pensato di definire un linguaggio universale per rappresentare le dimostrazioni. I sistemi potrebbero quindi autenticare con la firma digitale queste dimostrazioni ed esportarle ad altri sistemi che le potrebbero incorporare nel Web Semantico.

1.2.6 - Strato Trust

L'ultimo strato riguarda avere una garanzia di qualità delle informazioni sul web e un grado di fiducia nella risorsa che fornisce tali informazioni. Firmando digitalmente le istruzioni RDF, chi le incontrerà potrà essere sicuro della loro autenticità. Ogni utente fisserà il suo personale livello di fiducia e sarà il computer a decidere a cosa (e quanto) credere. È però difficile avere fiducia in un gran numero di persone differenti e questo potrebbe limitare l'utilizzo del Web. Si vuole quindi costruire il *Web of Trust*: lo scenario che si vuole realizzare è che un utente comunica al computer che ha fiducia in una persona, a sua volta questa persona ha fiducia in altre persone e queste ultime in altre (Fiducia come una proprietà transitiva). Tutte queste relazioni di fiducia si aprono a ventaglio e formano il *Web of Trust*. Ognuna delle relazioni ha un grado di fiducia (o di sfiducia) associata con esso. Sarà poi l'utente ad istruire il software del proprio computer su quali firme digitali fidarsi oppure no.

1.3 - LOD (Linked Open Data)

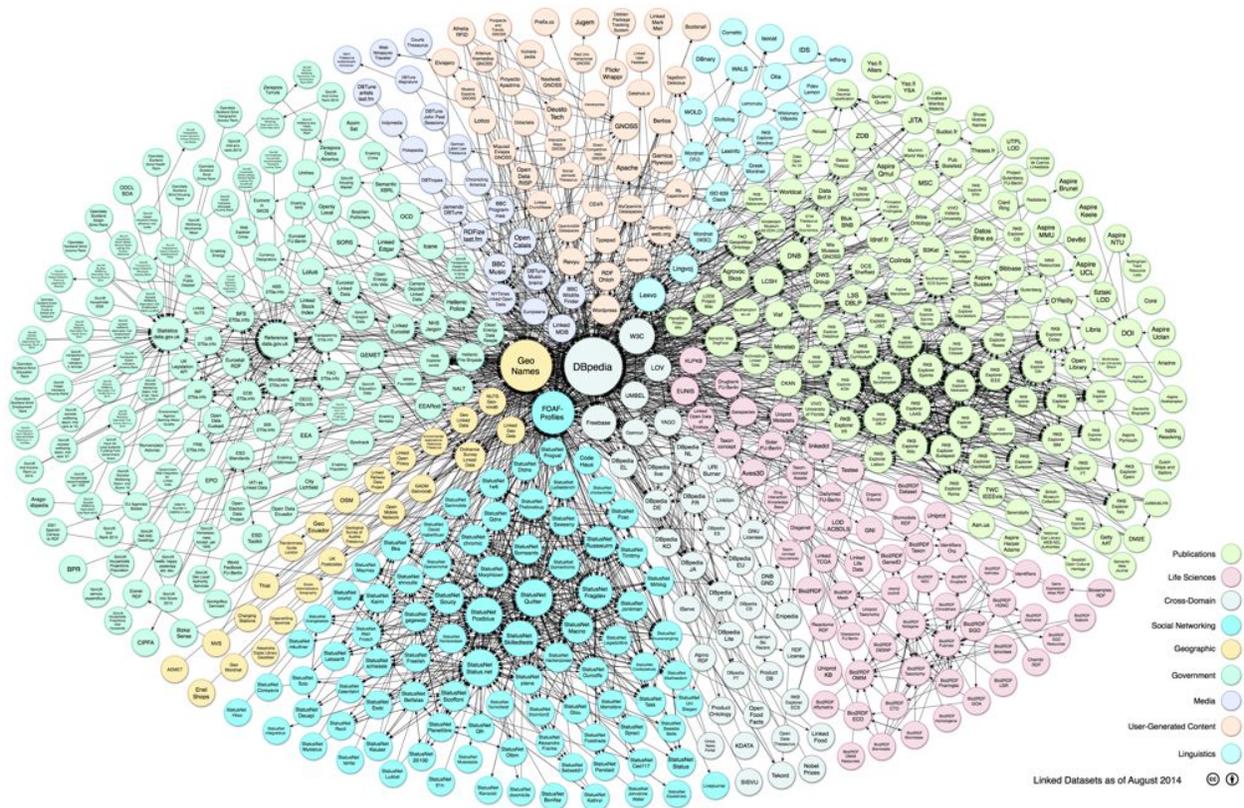


Figura 1.3: Grafo dei Linked Open Data

Con Linked Open Data, ci si riferisce a dati pubblicati sul web in una modalità leggibile e interpretabile da una macchina, il cui significato sia esplicitamente definito tramite una stringa costituita da parole e marcatori. I dati collegati tra loro (Linked Data[17]) costituiscono un dominio che definisce il contesto di partenza, essi vengono a loro volta collegati ad altri set di dati esterni, in un contesto di relazioni sempre più estese. Il concetto di Linked Data è strettamente connesso al Web Semantico in quanto riguardano lo stesso ambito applicativo, eppure il Web Semantico non si limita solo ai Linked Data, ma richiede inoltre il rispetto di alcune importanti regole finalizzate alla creazione di uno strato di contenuti accessibili a processi automatizzati, i Linked Data sono solo la

tecnologia adoperata per la realizzazione del web semantico o web di dati. Dal sito del W3C si può leggere: Il modello dati del web semantico è direttamente connesso col modello dei database relazionali[18]. Un database relazionale è costituito da tabelle, realizzate da righe o record. Ogni record è costituito da una serie di campi. Il record non è altro che il contenuto dei suoi campi, proprio come un nodo RDF non è altro che i suoi collegamenti: i valori delle proprietà. La mappatura è diretta: un record è un nodo RDF; il nome del campo (colonna) è il tipo di proprietà RDF; il campo (la singola cella) è il valore.

Un punto di forza principale del web semantico è sempre stata l'espressione, sul web, della grande quantità di informazioni dei database relazionali formulate in una modalità processabile da una macchina. Il formato di serializzazione RDF, con la sua sintassi XML, è un formato funzionale a esprimere le informazioni di un database relazionale.

Il web di dati è, dunque, un network globale di asserzioni collegate tramite link qualificati e autoparlanti che diventano una collezione di conoscenza leggibile e utilizzabile da una macchina, prima che da una persona.

1.3.1 - LOD e gli Archivi Culturali

Le biblioteche hanno sempre prodotto una grande quantità di dati di qualità in record bibliografici e di autorità fortemente strutturati, rispondenti a regole condivise e diffuse. Il mondo delle biblioteche e il mondo di Internet sono entrambi interessati all'integrazione in rete; il primo per garantire la visibilità e l'usabilità dei dati, il secondo per sfruttare informazioni e creare reticoli sempre più ampi e significativi. La quantità e qualità delle informazioni che viaggiano in rete sono due aspetti spesso inversamente proporzionali, l'aumento dell'informazione in rete, causa strumenti di pubblicazione sempre più utilizzati come i social network non è sempre sinonimo di qualità.

La crescita e l'uso esponenziale dell'informazione disponibile in rete non coincide con la crescita di fiducia nelle notizie, in quanto il grado di affidabilità è basso. Gli utenti a seguito di una ricerca, per trovare un'informazione significativa devono fruire molte informazioni con basso contenuto informativo prima di arrivare a una notizia attendibile, a questo punto quale criterio di ricerca potrebbe aiutare a discriminare notizie attendibili? L'autorevolezza della fonte diventa l'elemento discriminante, così facendo la selezione avviene a monte, preferendo scegliere la risorsa sulla base dell'autorevolezza di chi l'ha creata, anziché a valle, scegliendo acriticamente sulla base del ranking delle notizie che appaiono sulla pagina. La qualità della fonte, la certezza della provenienza diventano, dunque, elementi determinanti nel percorso esplorativo del ricercatore.

Il ruolo delle biblioteche, degli archivi e dei musei diviene pertanto rilevante, in quanto enti storicamente produttori di informazioni di alta qualità, dunque queste istituzioni assumono il ruolo di generatrici di informazione di qualità per la rete ed è per questo che i loro dati sono ambiti, ma devono essere resi disponibili al web. Di fatto molte di queste raccolte dati e prodotti dalle biblioteche, cataloghi, la cui redazione ha richiesto elaborazione di normative, competenze professionali e finanziamenti, non sono sul Web, ma isolati dal Web, non sono quindi interrogabili, pur essendo il Web il luogo in cui la maggior parte degli utenti lavora e ricerca informazioni. Un passaggio fondamentale da affrontare sarà quindi la trasformazione di questi dati inaccessibili attraverso la rete in Linked Open Data.

Un buon riferimento metodologico per pubblicare questi dati sotto forma di Linked Data è fornito da Boris Villazón-Terrazas in "Methodological guidelines for publishing Linked Data"[19], che riproduce il ciclo di vita per la produzione di Linked Data in sette passi:

- Identificazione della fonte dati;

- Modellizzazione del vocabolario, con l'adozione di ontologie esistenti, espresse in OWL, Web Ontology Language, o RDF(S) o con la creazione (più complessa) di nuove ontologie;
- Generazione dei dati in formato RDF, tramite diversi linguaggi di mappatura disponibili, anche in relazione al formato di origine del dato. In questa fase l'operazione più delicata è la creazione di URI, poiché essi sono la chiave per allineare risorse eterogenee provenienti da fonti differenti;
- Pubblicazione dei dati in RDF;
- Bonifica dei dati prodotti, per individuare eventuali e possibili errori di conversione e rendere il dato qualitativamente usabile;
- Creazione di collegamenti tra dataset differenti, con l'identificazione di dataset di interesse che possano diventare linking target, identificando relazioni tra i singoli dati, validando le relazioni individuate;
- Rendere concreto l'utilizzo dei dati, con differenti passi, tra cui la pubblicazione del dataset ottenuto dal processo sul CKAN Registry (Comprehensive Knowledge Archive Network), un registro per la pubblicazione di dati e pacchetti open, che rende possibile la loro scoperta, la condivisione e il riutilizzo.

Il dataset ottenuto con i sette passi suggeriti da Boris Villazón-Terrazas può essere poi valutato tramite un sistema di rating definito da Tim Berners-Lee per assegnare un punteggio ai siti che espongono dati sul web, definito le 5 stelle per gli Open Linked Data[20]:

- Make your stuff available on the web (whatever format);
- Make it available as structured data (e.g. excel instead of image scan of a table);
- Non-proprietary format (e.g. csv instead of excel);
- Use URLs to identify things, so that people can point at your stuff;

- Link your data to other people's data to provide context.

La valutazione sugli Open Linked Data prodotti dev'essere realizzata considerando dunque cinque punti fondamentali:

- 1) I propri dati siano disponibili sul web (in qualsiasi formato);
- 2) Il materiale messo sul web sia disponibile come dato strutturato (per esempio, in excel anziché come scansione dell'immagine di una tabella);
- 3) Siano stati scelti formati non proprietari (per esempio, in csv invece che excel);
- 4) Siano stati utilizzati URL per identificare gli oggetti, in modo che gli utenti possano puntare a questi oggetti;
- 5) I propri dati siano stati collegati a dati prodotti da altri in modo da definire un contesto.

2.0 - IL PATRIMONIO CULTURALE

La gestione dei dati per il settore dei beni culturali ha bisogno di ospitare un assortimento di diversi tipi di informazioni rilevanti per l'identificazione, la descrizione, interpretazione, estetica, interventi tecnici, valutazione delle condizioni, e lo sfondo storico di oggetti d'arte, monumenti e siti storici.

I dati catalogatori di queste opere, sono prodotti ed elaborati da una vasta gamma di scienziati, come ingegneri, archeologi, storici, restauratori, bibliotecari nazionali ed internazionali che impiegano una varietà di metodi e tecniche di catalogazione atti a conservare e descrivere nel miglior modo possibile i beni.

La scienza della conservazione, può far luce su diversi aspetti di un' opere d'arte, rivelando la sua struttura, la quale, a sua volta, fornisce informazioni importanti relative ai materiali e alla tecnica utilizzata per la creazione dell'oggetto. L'identificazione dei materiali porta ad una comprensione migliore della fase di

creazione del manufatto, mentre i materiali sono fortemente connessi con l'artista, il periodo di creazione e le tecniche di creazione utilizzate. Una corretta documentazione tuttavia, può essere ostacolata dal deterioramento o alterazione dei materiali dovute al passare del tempo, e dal non rispetto delle norme di catalogazione vigenti le quali potrebbero compromettere totalmente la qualità del documento di catalogazione.

Gli enti pubblici catalogatori come biblioteche, archivi e musei sono sempre più interessati all'integrazione sul Web, in quanto hanno sempre prodotto dati di qualità in record fortemente strutturati, rispondenti a regole condivise e diffuse secondo standard nazionali e internazionali. La gestione della pubblicazione di tale ricchezza e varietà di contenuti sul Web, e allo stesso tempo il mantenimento di essi pone sfide che ad oggi sono molto attuali ed importanti.

L'applicazione dei principi e delle tecnologie riguardanti i Linked Data e il Semantic Web è un promettente approccio per affrontare questo sviluppo. Ad oggi l'evoluzione della catalogazione dei beni storici e culturali sta portando alla creazione di grandi portali di condivisione nazionali e internazionali, come Europeana, e di grandi archivi di dati aperti, come il linked Open Data Cloud, con massicce pubblicazioni di dati collegati negli Stati Uniti, Europa e Asia. Questo fenomeno di catalogazione e pubblicazione di Beni culturali è diventato uno dei campi di applicazione di maggior successo riguardante i Linked data e tecnologie del Semantic Web.

2.1 - Il patrimonio culturale e il Web

Il Web è diventato un mezzo sempre più importante per la pubblicazione di contenuti riguardanti i beni culturali di diverso tipo. Molte biblioteche e archivi storici hanno digitalizzato le loro collezioni dando la possibilità di fruire i propri beni attraverso un qualsiasi browser. Ci sono grandi progetti nazionali e internazionali che hanno come obiettivo quello di pubblicare online contenuti

affidabili e ben documentati provenienti da diverse fonti attraverso servizi centralizzati. Questi sistemi forniscono un unico punto di accesso a collezioni eterogenee e estremamente vaste da fonti di informazioni autorevoli. In contrasto con le mostre fisiche tradizionali, questi tipi di servizi Web sono aperti tutto il tempo e il numero di opere non è limitata dallo spazio fisico disponibile, inoltre si dà la possibilità al cliente di fruire le opere nel modo più consono alle sue esigenze. Naturalmente, il Web non può sostituire l'esperienza fisica di visitare un museo o una mostra reale ma fornisce un'alternativa complementare per accedere ai dati di raccolta in qualsiasi tempo da qualsiasi luogo. Per i ricercatori in scienze umane, la disponibilità dei dati relativi ai beni culturali in maniera digitale ha aperto un nuovo paradigma di ricerca denominato Digital Humanities.

Essenzialmente ci sono due tipi di pubblicazioni relativi ai beni culturali sul Web. In primo luogo, vi è una grande varietà di sistemi ad hoc specifici per certi domini, molto personalizzati chiamati a tema chiuso. L'altro tipo di sistema sono grandi banche dati, aperte a chiunque voglia partecipare a contribuire a rifornire la base di conoscenza il cui contenuto non è tematicamente focalizzato, spesso vi partecipano più istituzioni che acconsentono a fornire i propri dati formattati secondo gli standard della piattaforma offerente il servizio. Questi prendono il nome di portali federati e possono essere trovati in molti paesi a livello internazionale. Un'applicazione di punta è Europeana, la sua base di conoscenza comprende milioni di oggetti da collezione provenienti da diverse organizzazioni e istituti europei. All'interno della piattaforma si possono effettuare ricerche di oggetti e raffinare ulteriormente la ricerca selezionando filtri supplementari. Un'altra piattaforma simile è la "United University Library of Congress", che mette a disposizione, gratuitamente, significativi materiali quali ad esempio manoscritti, mappe, libri rari, spartiti musicali, registrazioni, film, stampe, fotografie. Tali portali facilitano lo scambio di conoscenze principalmente fra tutti gli enti catalogatori, e per le organizzazioni che

contribuiscono, tali sistemi sono l'occasione per entrare in contatto con un pubblico più ampio sul Web con nuove modalità di interazione, e spingono inoltre la collaborazione con altre organizzazioni. Da un punto di vista sociale, questi portali migliorano la comunicazione fra enti spingendoli ad unirsi a formare un'unica base di conoscenza fruibile ed interrogabile da tutti, coerente e affidabile.

2.2 - STRATEGIE DI INTEGRAZIONE SU PORTALI WEB

Ogni organizzazione pubblica in genere espone i propri contenuti catalogafici sui propri portali customizzando la ricerca sul proprio dominio, l'utente finale per poter usufruire di tutte le basi di conoscenza dovrebbe imparare a utilizzare diversi sistemi di ricerca e interrogare e raccogliere dati più volte con interfacce Web differenti, evidentemente questo crea un ostacolo alla ricerca. L'utente finale piuttosto ha bisogno di un unico servizio globale per l'accesso ai dati situati in diversi database, per questo i contenuti locali devono essere aggregati e combinati in qualche modo, e poi utilizzati tramite un'interfaccia utente armonizzata.

Un portale Web combina dati provenienti da fonti diverse e ne fornisce una visione unificata. L'integrazione dei dati può avvenire in diverse fasi, ci sono due strategie principali:

- Integrazione dei dati in modo dinamico durante l'elaborazione di query.
- Integrare i dati in una fase di pre-elaborazione.

2.2.1 - Ricerca federata

Ricerca federata consiste in una query di ricerca inviata ai database locali distribuiti, i risultati vengono poi combinati insieme ed esposti all'utente finale. Questo tipo di ricerca è anche noto come multi-ricerca o meta-ricerca.

La ricerca federata è costituita dai seguenti passaggi.

1. Trasformare la query di interrogazione per ogni database esposto e trasmetterlo.
2. Raccogliere i risultati dei servizi e fonderli in un unico set di risultati.
3. Presentare il set di risultati in un formato conciso con la minima duplicazione di informazione.
4. Fornire all'utente finale un interfaccia per ispezionare il set di risultati derivanti dalla fusione.

Un notevole vantaggio della ricerca federata è che richiede solo che le organizzazioni partecipanti devono essere d'accordo sul protocollo di interrogazione, mentre i database dei singoli rimangono indipendenti. Grazie alla sua semplicità, la ricerca federata è stata applicata ampiamente nei portali di beni culturali, e ci sono molti sistemi di raccolta culturali nazionali sul Web basati su di esso, come ad esempio es: Europeana. Questo è particolarmente utile quando la ricerca si basa su metadati simili, come i libri nelle biblioteche. Una limitazione fondamentale della ricerca federata è che l'elaborazione della query è indipendente per ogni database locale, le dipendenze a livello globale, vale a dire, le associazioni tra oggetti attraverso diverse collezioni sono difficili da trovare. Dal momento che trovare collegamenti semantici tra elementi della raccolta è uno dei principali obiettivi dei linked data, un portale di beni culturali basato sulla ricerca federata rischia di sacrificare gran parte del potenziale del Semantic Web.

2.2.2 - Data warehousing

L'approccio alternativo alla ricerca federata è quello di consolidare prima le collezioni locali in un repository globale, un data warehouse, e quindi cercare nel database globale. Modelli concettuali reciprocamente condivisi, ontologie, possono essere utilizzati per arricchire i contenuti e per agevolare l'interoperabilità dei dati, nel risultato finale gli elementi di raccolta possono essere collegati tra loro attraverso associazioni semantiche.

In questo caso è importante il processo di creazione di contenuti, che dovrà essere separato e incentrato a consolidare il repository globale sulla base di banche dati locali, inoltre sarà necessario programmare anche il suo mantenimento-allineamento con i database partecipanti in costante evoluzione.

In questo caso è necessaria una maggiore collaborazione e coordinamento tra i fornitori di contenuti e il portale.

2.3 - PUBBLICAZIONI COLLABORATIVE

Il patrimonio culturale nonché i beni culturali sono eterogenei, ma allo stesso tempo semanticamente interconnessi.

Ad esempio, la biografia narrativa di una persona, opere artistiche create, luoghi di interesse, articoli di Wikipedia o romanzi, parenti, ed eventi storici in cui il soggetto è coinvolto, sono contenuti multi-prospettiva e rischiano di essere creati da diverse organizzazioni in modo indipendente l'uno dall'altro, utilizzando diversi schemi di metadati, vocabolari, convenzioni di catalogazione, e lingue.

Il collegamento in questo ambiente dei contenuti distribuiti è impegnativo se si vuole offrire agli utenti finali un dato uniforme e coerente.

Il Web Semantico e i Linked Data agevolano l'interoperabilità dei dati e la creazione di contenuti distribuiti allo stesso tempo, una rete semantica condivisa,

un'infrastruttura un'ontologia, composta da ontologie allineate può porsi come infrastruttura per descrivere tutti i contenuti del sistema distribuito. I fornitori di contenuti relativi ai beni culturali utilizzando gli stessi concetti, vengono automaticamente linkati l'uno con l'altro costituendo un'unica base di conoscenza. Esempio: un dipinto creato da Picasso proveniente da un museo d'arte, può essere arricchito (linked) con le biografie di Wikipedia e da altre fonti, foto scattate, libri di una libreria che descrive le sue opere d'arte e così via, così facendo i metadati del dipinto sono semanticamente arricchiti, allo stesso tempo, il contenuto di ogni organizzazione sul portale relativo a Picasso è arricchito dai metadati del nuovo lavoro entrato nel sistema.

Questo è chiaramente una situazione che avvantaggia tutti i partecipanti al sistema. La collaborazione avvantaggia tutti ma in Europa il patrimonio culturale viene distribuito a diverse istituti nazionali che definiscono i propri standard indipendentemente vincolando questa visione comune di condivisione. Ad oggi si sta collaborando con progetti internazionali per far sì che almeno virtualmente attraverso il Web semantico e i linked data queste informazioni siano interconnesse tra loro e formino un enorme base dati comune nonché una visione di patrimonio culturale mondiale allineata fra i vari enti nazionali.

2.3.1 - Standard catalografico italiano - ICCD

L'ICCD Istituto Centrale per il Catalogo e la Documentazione gestisce il Catalogo generale del patrimonio archeologico, architettonico, storico artistico e etnoantropologico italiano a livello nazionale, coordina la ricerca per la definizione degli standard di catalogazione per le diverse tipologie di beni culturali che afferiscono agli ambiti di tutela del MiBAC[21], il cui obiettivo principale è quello di creare un catalogo nazionale centralizzato del patrimonio culturale italiano.

Compito dell'Istituto è definire, in accordo con le regioni, metodologie e procedure standard per la catalogazione secondo criteri omogenei, con lo scopo di promuovere l'incremento del catalogo nazionale del patrimonio archeologico, architettonico, storico artistico, etnoantropologico, scientifico-tecnologico e dei beni naturalistici, nelle sue articolazioni territoriali.

Gli standard catalografici sono costituiti dalle normative, da specifici strumenti terminologici e da un insieme di indirizzi di metodo, definiti per attuare la catalogazione secondo criteri omogenei e condivisi a livello nazionale, in modo funzionale alla gestione informatizzata.

La relazione tra le autorità locali e ICCD è completamente orientata verso la conoscenza, la tutela e la valorizzazione del patrimonio culturale.

In questo contesto ICCD fornisce:

- Norme, metodologie e guide per la gestione tecnologica del catalogo/tracciato; le procedure di catalogazione sono monitorate e valutate attraverso l'ICCD Observatory for Cataloguing, un comitato interno preposto alla gestione delle attività di catalogazione.
- Strumenti per la gestione dei dati, e soprattutto la SIGECweb (Information System General Catalogue), un'applicazione software creata con l'obiettivo di unificare e semplificare i processi relativi alle attività di catalogazione del patrimonio culturale, e garantire la qualità dei dati prodotti e la loro conformità con le norme nazionali.

ICCD offre un documento di catalogazione standard per ogni tipo di tracciato o tipologia di bene da catalogare, che è costituito da un insieme di regole e linee guida che illustrano i metodi da seguire per l'acquisizione e la produzione di documentazione ufficiale, comprensivo di :

- Schema catalogo: modello, moduli e vincoli per la raccolta di informazioni in modo strutturato.

- Linguaggio comune e condiviso per ogni campo, essenziale per un corretto riutilizzo delle informazioni.

Il formato ICCD in cui si salvano i dati, è una struttura XML interoperabile all'interno di un'applicazione, sarebbe molto utile anche l'interoperabilità tra applicazioni, in quanto con il web of data si vuole puntare a condivisione totale di conoscenza.

L'adozione di XML facilita la gestione di collezioni di documenti, e costituisce un supporto fondamentale per la pubblicazione di informazioni a livello internazionale.

L'approccio catalografico adottato per le schede di catalogo relative ai beni culturali dall'ICCD differisce profondamente da qualsiasi struttura rappresentabile in RDF e quindi elaborabile tramite le tecnologie del Web dei Dati. Il sistema di catalogazione tende infatti a schematizzare in campi e sottocampi i vari attributi e le relazioni di una entità, nascondendo le relazioni semantiche.

Al fine di portare i repository xml in una forma consona ai linked data si ha bisogno di trasformarlo in un triple store e descriverlo con un'ontologia. L'ontologia che si presta di più a livello internazionale per i beni culturali è il CIDOC CRM.

2.3.2 - CIDOC Conceptual Reference Model

Con il CIDOC CRM si è voluto definire una ontologia standard comune ai quali tutti possono riferire al fine di ottenere una base di conoscenza affidabile e interoperabile comune dal quale dedurre inferenze.

CIDOC-CRM riconosciuto come standard ISO(21127:2006)[22] è l'ontologia più ampiamente utilizzata nel dominio dei beni culturali e rappresenta un modello

Anche le proprietà sono organizzate in gerarchia e sottoproprietà come in RDFS. CIDOC CRM è in uso in diverse organizzazioni ed è attivamente mantenuto da una community modulata dalle regole imposte dal CIDOC ICOM, International Council of Museums[23].

2.3.3 - CASO DI STUDIO - Europeana

Inaugurata il 20 novembre 2008, Europeana è un'iniziativa dell'Unione Europea, un portale che finora ha raccolto metadati su milioni di oggetti, rappresenta più di 3.000 organizzazioni culturali, allo scopo di mostrare come sia possibile ottenere una vera interoperabilità che non solo attraversi i confini nazionali per confluire in un patrimonio europeo unificato, ma soprattutto abbatta le barriere tra i diversi domini per arrivare ad un contesto semantico condiviso.

Il portale Europeana è modellato secondo tre principi di funzionamento:

- Dati facilmente usufruibili.
- Scambio reciproco di informazioni.
- Dati digitali sempre autentici, affidabili e robusti, che siano facili da creare e da cui i partner della rete Europeana possano trarre beneficio nella condivisione.

Europeana dà accesso a diversi tipi di contenuti da diversi tipi di istituzioni del patrimonio culturale. Gli oggetti digitali che gli utenti possono trovare in Europeana non vengono memorizzati su un computer centrale, ma rimangono con l'istituzione culturale e sono ospitati sui loro portali. Europeana raccoglie informazioni contestuali e metadati sugli elementi, gli utenti cercano le informazioni contestuali e una volta trovato quello che stanno cercando, accedono al contenuto completo della voce attraverso il sito originale che contiene il contenuto. Come spiegato al paragrafo 2.2.2 Europeana si basa sulla ricerca federata, infatti ogni organizzazione organizza i propri contenuti in modi

diversi e per diversi standard e per rendere l'informazione ricercabile sul portale di Europeana quest'ultima offre uno standard comune, noto come Europeana Semantic Elements (ESE). Quest'ultimo standard di metadati avendo un basso approccio di interoperabilità di risorse in quanto i metadati erano in formato testuale e non URI, venne nel 2010 sostituito con uno standard di metadati più ricco basato su RDF, l' European Data Model o EDM[24]. I dataset in questa versione contengono milioni di testi, immagini, video e suoni raccolti da Europeana, che possono essere liberamente fruiti in RDF e formato XML. ESE è un sottoinsieme di EDM e può essere mappato direttamente su EDM, infatti nonostante la poca riusabilità del formato ESE, per motivi storici ancora molte istituzioni hanno molti dati in questo formato e quindi hanno la possibilità di mapparli in EDM. European Data Model è un framework basato sul Web semantico fa una distinzione semantica tra le descrizioni intellettuali e tecniche degli oggetti e le rappresentazioni digitali dell'oggetto, inoltre distingue tra un oggetto e i metadati che lo descrivono e prevede il supporto per le risorse contestuali, in modo da poter connettere direttamente quest'ultime all'oggetto di interesse (si pensi alla fondamentale importanza delle informazioni geografiche relative ad un reperto archeologico). EDM non è uno schema fisso che determina il modo di rappresentare i dati, ma piuttosto un quadro concettuale o un'ontologia in cui più modelli specifici possono essere collegati tra loro migliorando l'interoperabilità.

2.3.4 - CASO DI STUDIO - ZERI FOUNDATION - Progetto ZERI & LODE

La Fondazione Federico Zeri è un centro di ricerca e di formazione specialistica nel campo della Storia dell'arte. È stata istituita nel 1999 dall'Università di Bologna con lo scopo di tutelare e divulgare l'opera e la figura di Federico Zeri, conservare e valorizzare lo straordinario lascito dello studioso: la Biblioteca

d'arte (46.000 volumi, 37.000 cataloghi d'asta) e la Fototeca (290.000 fotografie di opere d'arte).

Per rendere disponibile alla comunità scientifica questo patrimonio, la Fondazione ha intrapreso un innovativo progetto per la catalogazione informatizzata e la messa online dei nuclei più significativi della Fototeca.

Questo ha portato alla creazione di banche dati digitali ricchissime con immagini di grande qualità, che rendono accessibile a tutti la preziosa documentazione raccolta da Federico Zeri.

Il progetto ZERI & LODE è il frutto della collaborazione tra la Fondazione Federico Zeri e un team di informatici ed esperti in digital humanities dell'Università di Bologna.

L'obiettivo è trasformare i dati del catalogo Fototeca Zeri in Linked Open Data, affinché siano accessibili, rintracciabili e riusabili da utenti e altre applicazioni, secondo le esigenze del web semantico.

La necessità di esporre i dati sotto forma di Linked Open Data si inserisce nel contesto del progetto PHAROS(International Consortium of Photo Archives)[25], a cui la Fondazione aderisce. Il Consorzio riunisce 14 archivi fotografici di storia dell'arte di importanza internazionale, in Europa e negli Stati Uniti, le cui risorse assommano a 32 milioni di immagini. Il Consorzio promuove la collaborazione attiva tra le istituzioni partner in vista della creazione di una piattaforma comune per la ricerca su immagini e metadati relativi alle opere d'arte.

I LOD sono stati scelti come il veicolo per la condivisione dei dati. Per garantire la più ampia interoperabilità semantica, i dati sono stati mappati prevalentemente sullo standard internazionale CIDOC-CRM.

Un aspetto importante del progetto ZERI & LODE è stata la ricerca nel campo dello sviluppo delle ontologie.

Prima di tutto i dati sono stati strutturati secondo un modello ontologico che rispecchia l'analiticità delle schede ministeriali ICCD OA e F, utilizzate per la descrizione delle fotografie della Fototeca Zeri, dopodiché sono stati mappati in

CIDOC CRM e integrati con altre ontologie consolidate nell'ambito per sopperire ad alcune lacune di esso, ciò ha valorizzato e rafforzato la ricchezza semantica del catalogo della Fototeca Zeri.

Il dataset RDF così ottenuto comprende attualmente dati relativi alla pittura italiana del XVI secolo, e sono già disponibili sul sito web della Fondazione Zeri attraverso l'apposita interfaccia.

Attraverso i LOD è stato possibile arricchire il catalogo online della Fototeca Zeri, integrandolo con link ad authorities e dataset disponibili in Linked Open Data: link a Viaf, DBpedia, Wikidata e pagine Wikipedia sono disponibili per artisti e fotografi; link al Getty Art and Architecture Thesaurus (AAT) sono previsti per i termini relativi agli oggetti artistici e alle fotografie; infine, link a record Geonames sono forniti per i nomi di luoghi geografici.

Possiamo riassumere i principali passi affrontati dal progetto ZERI & LOD in:

- Catalogazione dei beni in XML seguendo la normativa ICCD vigente;
- Mappatura dei file XML ottenuti in CIDOC-CRM;
- Creazione del triple store RDF.
- Arricchimento delle triple tramite link ai LOD principali.
- Creazione di un'interfaccia per la fruizione dati del cliente.

Attraverso la condivisione di queste collezioni in un formato leggibile dalla macchina si è riuscito a rendere queste opere d'arte e fotografie più rintracciabili, e a promuovere un riuso creativo dei dati nella ricerca e altre applicazioni.

La prima versione del set di dati Zeri RDF, rilasciati nell'Aprile 2016 è disponibile presso il portale ZERI ed è individuabile anche attraverso il portale Europea.

I dati per lo più riguardano opere d'arte moderna (15 ° 16 ° secolo): circa 19.000 opere d'arte e più di 30.000 fotografie che ritraggono tali opere, queste vengono accuratamente descritte per mezzo di 11 milioni di dichiarazioni RDF.

2.3.5 - CASO DI STUDIO - Amsterdam Museum

Europeana ritiene che un fiorente e sano dominio pubblico è fondamentale e sostiene, dunque, che le rappresentazioni digitali di opere di dominio pubblico dovrebbero essere liberamente accessibili. Tuttavia, questa decisione non è facile da attuare per le istituzioni, soprattutto quando traggono profitto dalla vendita di queste immagini.

Nel 2011, il Rijksmuseum nei Paesi Bassi ha iniziato a rilasciare le immagini di pubblico dominio on-line, e nel 2013, è stato messo a disposizione l'intero catalogo nella più alta risoluzione possibile, senza alcun diritto d'autore o restrizione. Il Rijksmuseum è il museo nazionale olandese dedicato all'arte e alla storia ed è stato fondato nel 1800. Da allora, ha raccolto oltre 1.000.000 di oggetti fisici. Fra il 2003 e il 2013, l'edificio simbolo di Amsterdam era in fase di ristrutturazione e, pertanto, chiuso per la parte più grande. Nel 2013 si è aumentato di 22.000 metri quadrati lo spazio espositivo dell'edificio, ma anche in questo spazio più grande, solo circa 8.000 oggetti sono stati esposti.

Per mostrare l'intera raccolta, il Rijksmuseum ha impiegato un enorme sforzo nella creazione di un catalogo digitale disponibile tuttora online. Il museo fornisce punti di accesso multipli a queste immagini, tra cui una API (Application Programming Interface) e un sito web dedicato chiamato Rijksstudio[25], dove possono essere facilmente scaricate. Poiché queste immagini sono libere, possono anche essere trovate in altri luoghi, come Wikimedia Commons, Kennisnet, ARTstor e vari altri siti web.

Prima di raggiungere la decisione di rilasciare tutto il materiale senza limitazioni, ci sono state molte discussioni interne ed esterne, ma una volta approvato tutto ciò ha creato un grande ritorno di immagine al Museo infatti le immagini provenienti da una fonte attendibile, quindi le buone copie digitali fornite dal museo sono state rapidamente adottate da grandi piattaforme di condivisione

della conoscenza come Wikipedia, mettendo in risalto la fonte e rendendo le immagini non autoritarie impopolari.

Tutto ciò è dovuto anche alla collaborazione con Europeana che nel 2010 ha iniziato a lavorare con Rijksmuseum e che ha sempre supportato la causa di mettere a disposizione di essa più materiale possibile autoritario e libero da qualsiasi vincolo.

La collezione completa del Museo di Amsterdam è disponibile come Linked Open Data (LOD) dal 2011 e questo rende i dati del Museo i primi del loro genere ad essere ufficialmente convertiti e resi disponibili in questo formato.

I dati sono stati recuperati da un Web API XML Adlib del museo e convertiti in RDF conforme al modello di dati Europeana (EDM).

I risultati sono più di 5 milioni di triple RDF che descrivono più di 70.000 oggetti del patrimonio culturale legati alla città di Amsterdam.

Per la conversione in EDM è stato fatto un processo in due fasi, utilizzando gli script di conversione XML/RDF. Nella prima fase, l'XML viene convertito in 'RDF base', ogni nodo nella struttura XML dà luogo a un nodo RDF. Nella seconda fase, l'RDF base viene convertito in EDM-RDF, utilizzando le regole di riscrittura XML/RDF. In questa fase, i record vengono convertiti in EDM, ed oltre a ciò si arricchiscono tramite link ai LOD principali.

2.3.5 Caso di studio - Semantic Web Services

Un campo di particolare interesse nel Web Semantico riguarda l'integrazione con i Web Services, il Semantic Web Service. Vediamo prima una piccola introduzione al funzionamento dei web service senza l'aggiunta di espressività della semantica dei servizi.

Un Web Service, secondo W3C, è un sistema software progettato per supportare l'interoperabilità tra diversi elaboratori in un contesto distribuito.

Tale caratteristica si ottiene associando all'applicazione un'interfaccia software (descritta attraverso il Web Services Description Language - WSDL) che espone all'interno del contesto distribuito il servizio associato ed il modo di interagire con esso. Altri sistemi possono interagire con l'applicazione eseguendo le operazioni descritte nell'interfaccia, attivando servizi o richieste di procedure remote.

I messaggi di richiesta comunicano tramite protocollo HTTP e sono formattati secondo lo standard XML. Grazie all'utilizzo di standard basati su XML, tramite un'architettura basata sui Web Service (Service oriented Architecture, SOA) applicazioni software scritte in diversi linguaggi di programmazione e implementate su diverse piattaforme hardware possono quindi essere utilizzate, tramite le interfacce che queste espongono.

L'interazione che avviene tra processi e principalmente composta da quattro componenti:

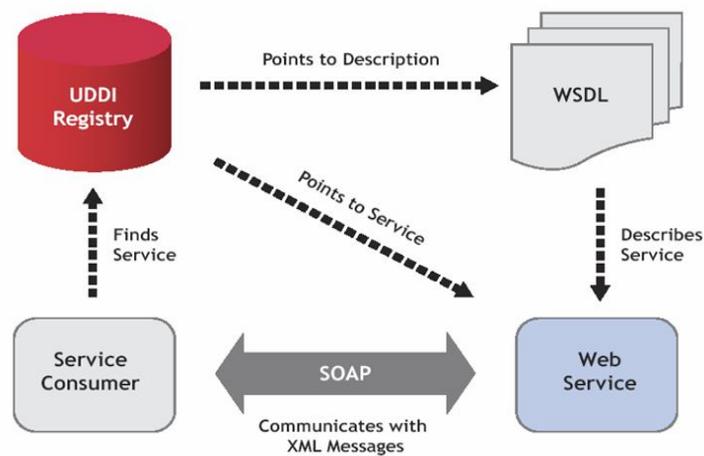


Figura 2.3.5: architettura di un web service

- Trasporto dei messaggi tra le applicazioni in rete, include protocolli quali HTTP, SMTP, FTP.
- XML Messaging: tutti i dati scambiati sono formattati mediante "tag" XML in modo che gli stessi possano essere utilizzati ad entrambi i capi delle

connessioni; il messaggio può essere codificato conformemente allo standard SOAP o REST principalmente.

- Descrizione del servizio: l'interfaccia pubblica di un Web Service viene descritta tramite WSDL (Web Services Description Language), il WSDL document è descritto in XML ed è usato per definire la service description, definisce le funzionalità offerte dal servizio e il comportamento di esso
- Elencazione dei servizi disponibili in rete: si centralizzano i WSDL dei servizi disponibili in un registro UDDI (Universal Description Discovery and Integration), sarà poi il richiedente del servizio a scegliere quale dei servizi descritti dai WSDL soddisfa al meglio la propria richiesta.

Il WSDL ricopre un ruolo importante nell'architettura dei web service perchè descrive il contratto tra il Service Consumer (l'entità che richiede il servizio) ed il Service Provider (l'entità che fornisce il servizio) in modalità indipendente dal linguaggio di sviluppo e dalla piattaforma utilizzata. Partendo dal documento WSDL è possibile generare le classi stub da usare lato client utilizzando l'opportuno tool WSDL-to- che permette di semplificare notevolmente lo sviluppo di un client riducendo la quantità di codice da scrivere e senza fare riferimento diretto alle API di basso livello di gestione dati/protocollo. Combinando un Design SOA con la tecnologia Web Services possiamo abbattere le barriere tecnologiche tra le varie piattaforme grazie ad una esposizione dei servizi omogenea e interoperabile. Le ricerche UDDI ad oggi devono prevedere un intervento umano per decidere effettivamente quale servizio utilizzare nel proprio business process visto che generalmente le ricerche avvengono mediante criteri di identità (nome del servizio o una parola chiave). La situazione dunque è analoga a una ricerca sintattica mediante un motore di ricerca Web dove una volta effettuata una ricerca e ottenuto il risultato, l'informazione deve essere interpretata scorrendo una lunga quantità di elenchi, dunque è necessaria una forte "human interaction" per potere fare interagire tra di loro le applicazioni. La problematica

dell'annettere, ai dati di un servizio, informazioni semantiche e poterle dunque acquisirle tramite una metodologia "machine understandable" è una problematica del tutto attuale. Ad oggi è prioritario avere a disposizione non solo la descrizione sintattica dei dati (WSDL) ma anche la relativa semantica. Nel WSDL però si descrive il servizio in modo prettamente sintattico, ma non si esplicita alcuna semantica degli stessi. WSDL e le ontologie condividono lo stesso linguaggio (XML) e gli stessi servizi (XML Schema, Namespace etc.), ma si posizionano a livelli decisamente differenti risultando, di fatto, complementari. Lo scopo della Semantic Integration è aggiungere informazioni semantiche ai servizi (Semantici Web Services) mediante un meccanismo di tagging e l'introduzione di Ontologie, correlare quindi mediante un'opportuna annotazione semantica i dati (la cui descrizione sintattica risiede nel WSDL) al loro significato (descritto nella relativa ontologia). UDDI, come discusso in precedenza non offre molte possibilità per la ricerca di servizi web. I servizi web possono essere ricercati per tipi attraverso i loro tModel. In particolare è possibile ricercare tutti i servizi che hanno una certa descrizione WSDL oppure che hanno certi valori associati con i tModel. Dal momento che la ricerca in UDDI è per parole chiave non c'è alcuna possibilità di inferenza semantica. Unendo le tecnologie interoperabili dei Web Services alla potenza semantica tipiche delle Ontologie si aprono nuove strade all'orizzonte nel mondo dell'integrazione. Non esiste ad oggi uno standard unico riconosciuto per associare contenuto semantico ai servizi applicativi ed in generale ai Web Services. Negli ultimi anni sono stati sviluppati molti tool e framework che supportano la pubblicazione, la scoperta e la composizione di Semantic Web Services. Queste iniziative includono OWL-S, WSMO, SWSF e WSDL-S.

- OWL-S (Ontology Web Language for Services): Owl-s è un'ontologia di servizi definita in Owl, permettere una ricerca semantica di Web Services e l'invocazione di essi da parte di software agents.

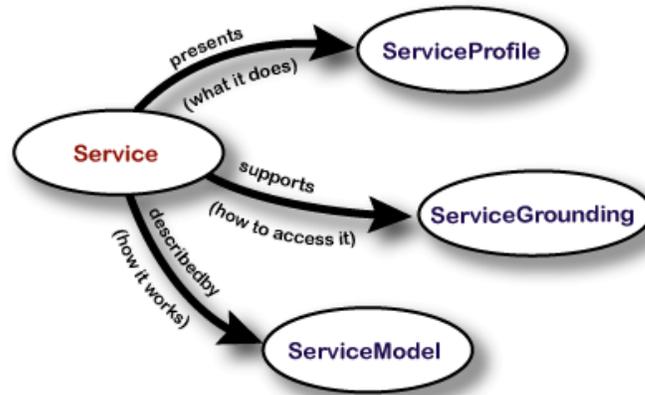


Figura 2.3.5.1: struttura OWL-S

La struttura di Owl-s può essere suddivisa in tre parti principali:

- Service Profile : fornisce informazioni utili, che agevolano la ricerca e scelta di un servizio o meno, descrive principalmente cosa fa il servizio e cosa fornisce. La descrizione di un servizio è costituita da proprietà funzionali, come ad esempio gli input, gli output, precondition e effect, e non funzionali come il nome del servizio e parametri per definire i metadati riguardanti il servizio stesso.
- Process Model: Descrive come è composto il servizio in termini dei loro processi costituenti. Ciò è utilizzato sia per verificare se un modello è eseguibile in un dato contesto, sia per controllare l'invocazione di un servizio.
- Grounding di servizio: definisce l'interoperabilità di un dato servizio, specificando i dettagli dei protocolli e i formati dei messaggi supportati da esso, quale protocollo di trasporto utilizza e le informazioni sul suo indirizzamento.

La combinazione del modello di processo OWL-S con WSDL completa e amplia la descrizione del servizio in una maniera machine readable, i motori di ricerca avranno a disposizione registri che contengono descrizioni OWL-S di

servizi web e saranno capaci di effettuare ricerche semantiche. Ad oggi questa metodologia è ancora in fase di sviluppo in quanto il problema principale è che ci dovrebbero essere ontologie dei domini comuni e condivise per ogni fornitore di servizio e per ogni tipologia di servizio (cosa attualmente in sviluppo).

- WSMO(Web Service Modeling Ontology): fornisce le specifiche ontologiche per gli elementi fondamentali di SWS. WSMO si basa sui seguenti principi di progettazione:

Web Compliance - WSMO eredita il concetto di URI (Universal Resource Identifier) per l'identificazione univoca delle risorse, Inoltre adotta il concetto di namespace e utilizza XML.

Ontology Based - Le ontologie sono utilizzati come modello di dati in tutto WSMO, il che significa che tutte le descrizioni delle risorse, nonché tutti i dati scambiati durante l'utilizzo del servizio si basano su ontologie. L'ampio utilizzo di ontologie permette maggiori deduzioni semanticamente delle informazioni.

Strict Decoupling - Le risorse WSMO sono definite in modo isolato, il che significa che ogni risorsa è specificata in modo indipendente, senza alcun accenno all'utilizzo o interazioni con altre risorse.

Centrality of Mediation - Come principio di progettazione complementare alla Strict Decoupling la mediazione si rivolge alla gestione delle eterogeneità che si presentano naturalmente in ambienti aperti e distribuiti. L'eterogeneità può verificarsi in termini di dati e ontologie di riferimento, protocollo o processo.

Descrizione contro Attuazione - WSMO distingue tra le descrizioni di elementi dei SWS (descrizione) e le tecnologie utilizzabili (Attuazione).

Un servizio Web è un'entità di calcolo che è in grado, una volta invocato di perseguire un'obiettivo utente. Il valore effettivo di un servizio è misurato in base

al soddisfacimento della richiesta dell'utente. WSMO fornisce i mezzi per descrivere i servizi Web che forniscono l'accesso (la ricerca, l'acquisizione, ecc) ai servizi. WSMO è progettato come un mezzo per descrivere il servizio web e non per sostituire le funzionalità di quest'ultimo.

- WSMO è composto da quattro elementi principali:

- **Ontologies:** forniscono la terminologia utilizzata da altri elementi WSMO per descrivere gli aspetti rilevanti dei domini.
- **Goals:** definiscono i problemi che un Web Services dovrebbe risolvere, rappresentano la richiesta utente. Le ontologie possono essere utilizzate per descrivere gli aspetti rilevanti. I goals modellano la visione dell'utente nel processo di utilizzo del servizio Web.
- **Descrizioni dei Web Services:** definiscono i vari aspetti di un web service, queste descrizioni comprendono le capacità, interfacce e funzionamento interno del web service. Tutti questi aspetti di un Web Service sono descritti utilizzando ontologie.
- **I Mediators** che bypassano i problemi riguardanti l'interoperabilità. I mediatori sono il concetto di base per risolvere le incompatibilità a livello di dati, processi e di protocollo, risolvono gli squilibri tra diverse terminologie (livello dati) e la comunicazione tra i servizi web (livello di protocollo).

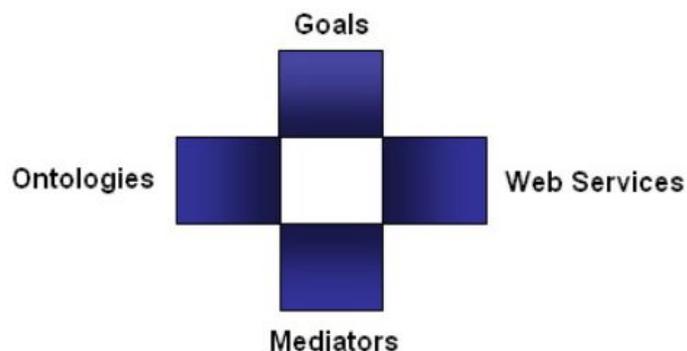


Figura 2.3.5.2: componenti WSMO

- SWSF (Semantic Web Services Framework): si compone di due parti principali: SWSL, SWSO.

Semantic Web Services Language (SWSL) è utilizzato per descrivere le caratteristiche formali dei concetti di un WS e per descrivere i singoli servizi.

Semantic Web Services Ontology (SWSO), presenta un modello concettuale attraverso il quale i WS possono essere descritti. Più specificamente, SWSL è un linguaggio logico general-purpose, con alcune feature che lo rendono utilizzabile con i linguaggi base e con l'infrastruttura del Web attuale. Queste feature comprendono le URI, l'integrazione dei tipi incorporati da XML, i namespace e i meccanismi di import compatibili con XML.

- WSDL-S (Web Service Semantics): ha la caratteristica principale di essere facilmente utilizzabile in quanto ha un approccio più diretto per aggiungere semantica ai Web Services. WSDL-S punta tutto sull'aggiungere semantica ai WSDL, i concetti sono simili a quelli di OWL-S, ma risulta più facilmente utilizzabile in quanto gli utenti possono descrivere sia la semantica che i dettagli relativi alle operazioni del servizio nel WSDL stesso. Le ontologie sono esterne al WSDL e si possono aggiungere annotazioni con un linguaggio ontologico a scelta, questo permette di descrivere le operazioni e i dati corredandoli di informazioni contenute in ontologie. WSDL-S basandosi su WSDL è compatibile ai tool relativi al trattamento di WSDL e questo comporta un enorme vantaggio in relazione all'integrazione. I registri UDDI di norma non supporterebbero la pubblicazione e la scoperta di semantic web services, ma per abilitare il supporto alla semantica è stata definita un'infrastruttura per mappare WSDL-S su UDDI, così facendo un servizio WSDL-S è incapsulato utilizzando la Business Service entity di UDDI, mentre i portType e le operation sono incorporate nel Technical Model (tModel). Il framework di WSDL-S inoltre fornisce un tool grafico denominato Lumina, il quale è stato sviluppato per facilitare il discovery di Semantic Web Services.

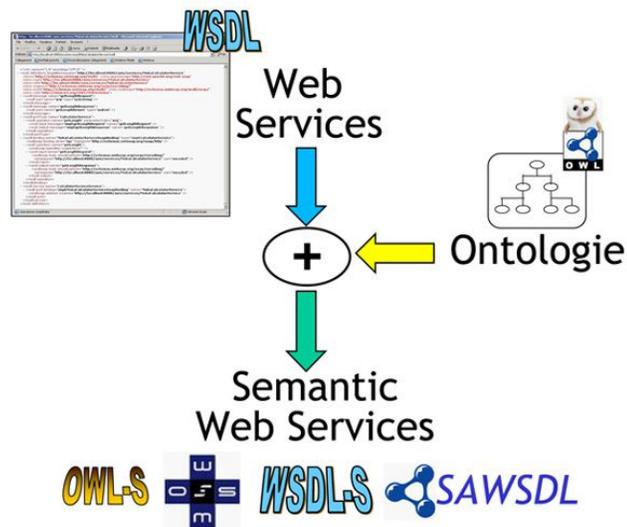


Figura 2.3.5.3: Semantic Web Services

Dopo quest'analisi dobbiamo precisare che ancora nessun tool o framework per il Semantic Web fornisce tutto ciò che è richiesto ad una piattaforma di modellazione generale di Web Services. Tutti questi standard sono ancora incompleti e potrebbero non adempiere alle future richieste delle industrie come una maggiore complessità, scalabilità, affidabilità, solo per citarne alcune. Inoltre, le informazioni semantiche di un Web Service devono essere sufficientemente generali per permettere il supporto di interazioni automatizzate tra Web Services e software agent. Idealmente, il linguaggio dei Semantic Web Services dovrebbe permettere il dinamismo in tutti i tipi di utilizzo di un Web Service, come ad esempio la selezione, la scoperta, la composizione, l'invocazione, la negoziazione e il ripristino dopo un guasto. Avendo a disposizione servizi la cui descrizione non è ambigua ed è comprensibile anche a una macchina si potranno effettuare una serie di operazioni, ricerche e correlazioni che ad oggi è possibile effettuare solo con intervento umano. Innanzitutto i servizi potranno essere categorizzati in base alle loro informazioni semantiche. Questo agevolerà le operazioni di discovery e renderanno possibili

operazioni di matchmaking tra richiesta e offerta basate sulle descrizioni semantiche dei servizi. Formalizzando le relazioni tra i WS e le ontologie si creerà uno "strato" semantico che potrebbe permettere l'utilizzo di "Agenti Software" in grado di analizzare i contenuti semantici ed automatizzare le ricerche dei servizi cosa che ad oggi è possibile effettuare solo con intervento umano.

3.0 - INTRODUZIONE AL PROGETTO DI TESI

Il contesto in cui si pone il progetto è quello del Web semantico e dei linked data, nel progetto si è potuto disporre di un repository di documenti XML descrittivi numerose categorie di beni culturali identificati come tracciati secondo la normativa ICCD, nello specifico i documenti descrivono Opere D'arte (TRACCIATO OA), Reperti Storici (TRACCIATO RA) e Fotografie (TRACCIATO F).

L'XML come detto in precedenza consente di descrivere le diverse parti di un documento secondo uno specifico dettaglio, in questo caso gli elementi XML che descrivono il documento sono gli elementi presenti in una scheda ICCD atti a documentare le informazioni di un'opera d'arte. Un documento così formato può essere elaborato per usi diversi; ad esempio per estrazione di informazioni, per l'adattamento ad altri formati o visualizzazione delle informazioni in base alla capacità del terminale. I documenti XML così formati non consentono però una definizione semantica adeguata in quanto sono poco adatti al Web attuale che per sua natura è distribuito e decentralizzato. Con XML non si definisce alcun meccanismo esplicito per qualificare le relazioni tra i concetti definiti nei documenti, inoltre non si possono organizzare in tassonomie le classi e le relazioni (Properties) del dominio a seconda della loro generalità, questo rende difficile inferire conoscenza semantica ai vari contenuti dei documenti e impoverisce molto il contenuto informativo della collezione di documenti messa a disposizione, in una visione di Information Retrieval. I documenti così costruiti rimangono dunque poveri di semantica e si prestano poco al riutilizzo delle

informazioni da essi specificate. L'obiettivo di questo progetto di tesi è quello di fornire un'ontologia comune a più beni culturali, nello specifico Opera e oggetto d'arte, Fotografia, Reperto Storico, denominata CICDO Ontology, in seguito specializzare il modello ontologico ottenuto, aggiungendo la descrizione in termini ontologici delle informazioni relative solo al tracciato OA ottenendo un'ontologia dedicata definita CICDO-OA Ontology, allo scopo di fornire una concettualizzazione semantica delle informazioni espresse nei documenti XML forniti. I modelli ontologici di dominio, che si andranno a creare, descriveranno le informazioni più significative relative alle conoscenze espresse nei documenti, tali concetti saranno espressi tramite una serie di triple RDF compatibili allo standard CIDOC-CRM, inoltre con l'obiettivo di perseguire l'interoperabilità semantica attraverso l'approccio Linked Open Data, saranno utilizzati i principali vocabolari utilizzati nel contesto dei beni culturali per modellare i dati.

Il progetto dunque sarà così suddiviso:

- Identificazione dei principali vocabolari utili a modellare il dominio.
- Mappatura concetti Iccd dei documenti XML, secondo le entità dello schema ontologico CIDOC-CRM.
- Creazione triple RDF descrittive le relazioni presenti fra i concetti delle schede.
- Arricchimento delle triple tramite link ai LOD principali.
- Validazione modello ontologico.

3.1 - ONTOLOGIE DI RIFERIMENTO

Ogni asserzione, espressa in forma di tripla riguardante i campi dei tracciati, diventa generatrice di una nuova informazione; più gli ambiti di appartenenza di queste asserzioni crescono e si intersecano, più la rete semantica presente e

disponibile sul web si arricchisce e diventa informazione classificata, in altre parole tante più caratteristiche o vincoli strutturali del dominio possiamo esprimere con il linguaggio ontologico, tanto più potente potrà essere il ragionamento che potremo applicare ai dati.

Le ontologie consentono di rappresentare le entità tramite la descrizione delle loro caratteristiche sono utilizzate soprattutto per esprimere categorizzazioni ed ampliare i ragionamenti deduttivi. Collegando le nostre entità a delle ontologie di dominio già esistenti e associate in termini di qualità della concettualizzazione, non faremo altro che contribuire ad aggiungere informazioni di dominio aumentando la capacità informativa del modello e l'interoperabilità dei dati.

Le ontologie individuate in pre-analisi utili al nostro modello sono state:

- Erlangen CRM / OWL: un'ontologia che descrive in modo più fedele alle specifiche il modello CIDOC-CRM.
- FOAF (Friend Of A friend): un'ontologia utilizzata per descrivere le persone, le loro attività, le loro relazioni con altre persone o cose, utile nel nostro caso a descrivere le persone come i responsabili o i redattori dei beni catalogati;
- HICO (Historical context): HiCo è un'ontologia utilizzata per descrivere contesto storico di oggetti del patrimonio culturale. Il contesto storico coinvolge principalmente gli aspetti esplicitamente descritti in un oggetto di interesse come una descrizione di un evento in un documento, ma anche impliciti come il seguire di un particolare movimento artistico o di una situazione, precedente non menzionato in un documento, di un autore. Questi sono tutti elementi utili al fine di comprendere chiaramente il contenuto dell'oggetto di interesse.
- CITO (Citation Typing Ontology): Cito è una ontologia per consentire caratterizzazione e descrizione di tipo o di natura di citazioni, sia fatte

materialmente sia retoricamente. Le citazioni caratterizzate possono essere sia dirette (come il riferimento a un articolo di giornale) o indirette (una citazione di un documento più recente, dallo stesso gruppo di ricerca sullo stesso argomento). Nel nostro caso CITO è utilizzata per esprimere una motivazione relativa ad una certa datazione di un reperto rispetto ad un'altra.

- PROV Ontology: fornisce un insieme di classi, proprietà e limitazioni che possono essere utilizzati per rappresentare e scambiare informazioni di provenienza generate in sistemi diversi e in contesti diversi.
- FRBR: è un modello di metadati approvato da IFLA che definisce in modo chiaro ed univoco i metadati utilizzati da una biblioteca digitale. L'obiettivo di FRBR è quello di correlare i dati contenuti nei record bibliografici ai requisiti degli utenti definendo un livello minimo di funzionalità per i record bibliografici creati dalle diverse agenzie nazionali es ICCD. I metadati FRBR vengono utilizzati per descrivere gli attributi e le proprietà delle risorse delle biblioteche digitali e possono essere facilmente inclusi in RDF. FRBR utilizza il modello Entità Relazione per la rappresentazione del modello concettuale delle basi di dati. Per entità si intende un qualsiasi elemento della realtà che può essere chiaramente individuato, FRBR comprende diversi gruppi di entità:
Gruppo 1: rappresentano i diversi aspetti di interesse per un utente in un'opera, essi vengono classificati in:

- WORK: opera intellettuale o artistica, si tratta di un entità astratta che non rappresenta nessun oggetto fisico specifico.
- EXPRESSION: è la realizzazione del WORK sotto forma di testo, musica, suono, immagine o una qualunque loro combinazione, non specifica la forma fisica della realizzazione.

- MANIFESTATION: definisce la rappresentazione fisica di una specifica espressione.
- ITEM: singolo esemplare di manifestazione, ci possono essere più item di una stessa manifestazione.

Gruppo 2: rappresentano i responsabili della custodia intellettuale o artistica delle entità del gruppo 1 e la produzione fisica e/o la distribuzione delle manifestazioni del gruppo 1.

Gruppo 3: includono un concetto, oggetto, evento o luogo. Queste entità fungono di solito da soggetti delle opere. Possono essere correlate con una sola opera o una moltitudine di opere e ciascuna opera può includere varie entità del gruppo 3.

I vantaggi di questo modello sono quindi essenzialmente due:

- 1) Sul piano teorico, normativo e didattico, esso offre uno schema di analisi logico e coerente, di applicabilità generale;
- 2) Sul piano pratico, esso permette di sviluppare, in grandi basi dati bibliografiche, percorsi di ricerca e consultazione più semplici e nello stesso tempo più potenti.

- FaBio FRBR - (Aligned Bibliographic Ontology) è un'ontologia basata su FRBR originariamente sviluppata per registrare descrizioni di entità bibliografiche pubblicate o potenzialmente pubblicabilcate. Le entità in FaBio sono principalmente pubblicazioni testuali, ma anche pagine web, dataset, metadati di documenti, specifiche formali e vocabolari, cataloghi e collezioni. Le classi FaBio sono strutturate secondo lo schema FRBR di opere, espressioni, manifestazioni ed elementi. Ulteriori oggetti sono stati aggiunti estendendo il modello di dati FRBR collegando: Opere e Manifestazioni (fabio: hasManifestation e Fabio:

isManifestationOf); Opere e Articoli (fabio: hasPortrayal e Fabio: isPortrayedBy); Espressioni e Articoli (fabio: hasRepresentation e Fabio: isRepresentedBy). L'opera d'arte è un oggetto fisico potenzialmente sottoposto a diversi eventi di trasformazione (deterioramenti, restauri ...); perciò l'essenza dell'opera d'arte è rappresentata dal livello Work di FRBR, il risultato di ogni trasformazione dal livello Manifestation e le sue caratteristiche fisiche dal livello Item. Tali caratterizzazioni sono il risultato di un'attività di produzione che coinvolge degli agenti in uno specifico momento nel tempo e sono utili per disambiguare e specificare i vari stadi di un'opera artistica.

3.2 - CREAZIONE DEL MODELLO - Protégè

Si è provveduto a creare le ontologie attraverso Protégé, un editor di ontologie open-source. Protégé permette di costruire un modello ontologico per intero, dalla definizione della semantica, alla conoscenza (RDF / RDFS → OWL).

Con RDF/RDFS definiremo una sequenza di triple, basate sull'idea di fare dichiarazioni circa le risorse sotto forma di espressioni soggetto-predicato-oggetto.

Protégè inoltre ci permette di definire:

- rdfs: Class: definisce un gruppo di individui che appartengono ad un insieme perché condividono alcune proprietà, nel nostro caso ogni campo del documento ICCD corrisponderà ad una classe.
- rdfs: subclassOf: organizza le classi in una gerarchia di specializzazione utilizzando subclassOf, nel nostro caso di studio ogni sottocampo del documento ICCD sarà sottoclasse del campo padre.

- `rdf: Property`: definisce i rapporti tra individui o da soggetti a valori di dati, le property ci serviranno a relazionare le informazioni delle varie classi che identificano i campi.
- `rdfs: subPropertyOf`: organizza le proprietà in modo gerarchico definendo ad esempio che una proprietà è una sottoproprietà di una o più proprietà, nel nostro caso sfrutteremo le proprietà dell'ontologia Erlanger che descrive in modo accurato le relazioni fra i campi CIDOC-CRM sui quali avremo modellato le classi del nostro caso di studio.
- `rdfs: dominio`: definisce il dominio di una proprietà limitando gli individui a cui la proprietà è applicata, sarà utile a modellare meglio le proprietà del nostro caso di studio.
- `rdfs: range`: limita gli individui che possono essere referenziati a una specifica proprietà, utile anch'esso a modellare meglio le proprietà del nostro caso di studio.

3.3 - MAPPING IN COMUNE - CICDO - Central Institute for Cataloguing and Documentation Ontology

Dopo un'attenta analisi abbiamo stabilito una prima fase di mappatura che comprende la scrittura in triple RDF delle informazioni ICCD presenti nel repository di documenti XML messo a disposizione. Il repository era comprensivo di diversi tracciati (OA,F,RA), quindi di diversi documenti che esprimono diverse categorie di beni culturali. Lo schema di catalogazione in tutti e tre i casi era modellato con dei metadati in comune, per questo motivo abbiamo voluto creare un modello ontologico comune "CICDO" che facesse da modello base per le tre ontologie specifiche di ogni tracciato. Il risultato ottenuto è essenzialmente una concettualizzazione delle informazioni comuni espresse dai diversi tracciati descritte tramite triple RDF composte da soggetto (range) - predicato - oggetto (domain) del livello inventariale (campi minimi e obbligatori da compilare per

far sì che la scheda fosse consistente) presente nelle schede. I campi in comune tra le tre schede prese in esame sono (Rif. documentazione ufficiale ICCD[26]):

- **CD** - Codici: In questo paragrafo vengono forniti i dati che permettono di individuare la scheda negli archivi costituiti dalle Istituzioni preposte alla catalogazione e di risalire all'Ente che l'ha prodotta ed all'Ente che ha la competenza sul bene schedato. In particolare, il codice univoco serve da 'chiave' per identificare univocamente un bene a livello nazionale.
- **OG** - Oggetto: Il paragrafo contiene le informazioni essenziali per un'immediata identificazione del bene catalogato.
- **LC** - Localizzazione geografico-amministrativa: In questo paragrafo vengono registrati i dati relativi alla localizzazione geografico-amministrativa del bene nel momento in cui viene compilata la scheda.
- **DT** - Cronologia: Informazioni sugli estremi cronologici del bene catalogato.
- **MT** - Dati tecnici: Informazioni relative agli aspetti materici e tecnici del bene in esame.
- **TU** - Condizione giuridica e vincoli: Indicazioni relative alla proprietà del bene.
- **DO** - Fonti e documenti di riferimento: Informazioni sulla documentazione che si ritiene utile allegare alla scheda o semplicemente citare.
- **AD** - Accesso ai dati: In questo paragrafo vengono registrate le indicazioni relative all'accesso ai dati, secondo le indicazioni dell'Ente fornitore, per l'utenza generica.
- **CM** - Compilazione: In questo paragrafo vengono registrate le informazioni relative all'elaborazione e alla validazione scientifica della scheda. Le informazioni riguardano le date in cui tali operazioni sono state svolte e le persone intervenute e/o responsabili.

Di seguito verrà descritta la mappatura dei campi e sottocampi ICCD in classi e sottoclassi CIDOC-CRM inoltre verranno relazionati i vari concetti-classi con delle proprietà rispettando i vincoli di domain e di range imposti da CIDOC. Oltre a ciò le classi sono state collegate a delle ontologie esterne per referenziare al meglio i concetti. In questa fase per cercare di mantenere il modello ontologico il più comune possibile si è cercato di generalizzare al meglio i concetti in modo da comprendere il più possibile aspetti simili alle tre schede.

Premessa: partendo dal punto di vista CIDOC CRM il modello della Scheda OA-F-RA riguarda la descrizione del bene culturale in oggetto, dal punto di vista CIDOC-CRM sono oggetti fisici creati dall'uomo, quindi la classe usata per la rappresentazione è 'E22_man_made_object', ad esso sono legati altri documenti che descrivono le varie caratteristiche della scheda come ad esempio DO, oltre alle informazioni relative all'oggetto in se relazionate direttamente all'entità E22, come ad esempio i campi ICCD CD,OG,LC.

3.3.1 - Mapping CD - Codici:

Il paragrafo cd contiene a sua volta altri sottocampi, segue una lista di essi con relativa descrizione ICCD:

- TSK Tipo scheda: Indica la sigla che contraddistingue il modello di scheda previsto per il bene catalogato. Vocabolario chiuso.
- LIR Livello ricerca: Indica la sigla che individua il livello di indagine effettuato in sede di catalogazione. Vocabolario chiuso.
- NCT CODICE UNIVOCO: Indica, in forma univoca, il codice di collegamento tra la scheda e il bene. Il codice univoco si genera dalla concatenazione dei due sottocampi Codice Regione (NCTR) e Numero catalogo generale (NCTN), che viene assegnato dall'ICCD.
- NCTR Codice Regione: Indica il numero di codice che individua la Regione in cui ha sede l'Ente competente sul bene catalogato. Vocabolario chiuso.

- NCTN Numero catalogo generale: Indica il numero, composto da otto cifre, assegnato dall'ICCD a ciascuna scheda di catalogo, secondo l'ordine progressivo relativo ad una determinata Regione.
- ESC Ente schedatore: Indica in codice l'Ente che ha curato la compilazione della scheda. Vocabolario aperto.
- ECP Ente competente: indica in codice l'Ente sotto la cui competenza ricade la tutela o la delega alla tutela del bene catalogato. Vocabolario aperto.

- Descrizione in triple RDF dei campi relativi alla sezione CD:

Nome Campo ICCD	Codice Campo ICCD	Entità/Range CIDOC	Proprietà CIDOC	Domain CIDOC	Voc.	Rip.
Codici	CD	E42 Identifier	P48 has pref id	E22Man-Made bject		
Tipo scheda	TSK	xsd:string	P48 has pref id	E22Man-Made bject	C	
Livello ricerca	LIR	xsd:string	P48 has pref id	E22Man-Made bject	C	
Codice univoco	NCT	E42 Identifier	P48 has pref id	E22Man-Made bject		
Codice regione	NCTR	E42 Identifier	P48 has pref id	E22Man-Made bject	C	
Numero-catalogo-gen	NCTN	E42 Identifier	P48 has pref id	E22Man-Made bject		
Ente schedatore	ESC	E42 Identifier	P48 has pref id	E22Man-Made bject		
Ente competente	ECP	E42 Identifier	P48 has pref id	E22Man-Made bject		

- Descrizione dei concetti (Classi) del campo CD:

Classe	Vocabolario	Descrizione	Superclasse
'Cultural Heritage(Physical object)'	Frbr- Manifestation	Classe degli oggetti fisici soggetti alla catalogazione	'E22 Man-MadeObject'
'CD (Codes)'		Classe che rappresenta l'insieme dei codici	'E42 Identifier'
'(TSK) Scheda Type'		Classe dei vari tipi di scheda	'CD (Codes)'
'(LIR) Research Level'		Classe dei i vari tipi di ricerca	'CD (Codes)'
'(NCT) Unique Code'		Classe dei codici univoci assegnati ad ogni scheda	'CD (Codes)'
'(NCTN) General Catalog Number'		Classe relativa ai codici di catalogazione	'CD (Codes)'
'(NCTR) Region Code'		Classe dei codici regione in cui risiede l'ente	'CD (Codes)'
'(ECP) Competent Entity'		Classe degli enti schedatori	'CD (Codes)'
'(ESC) Cataloguer Entity'		Classe degli enti catalogatori	'CD (Codes)'

- Definizione delle proprietà utilizzate in CD Object Property:

Proprietà	Dominio	Range	Super-proprietà
'has identifier'	'Cultural Heritage (Physical Object)'	'(CD) Codes'	'P48 has preferred identifier'

Tutte le classi create in cd sono sottoclassi di 'CD Codes', quindi di un'entità 'E42 Identifier' sono tutte modellate dalla stessa proprietà 'has Identifier' che ha come range il cultural heritage (E22-Man-Made Object) e domain 'CD Codes'. Essendo tutte le classi che modellano i campi di CD sottoclassi della classe 'CODE CD' anch'esse saranno relazionate all'individuo 'E22-Man-Made Object' tramite la proprietà 'has Identifier'.

- Definizione delle Data property degli individui del campo CD:

Proprietà	Dominio	Range	Super-proprietà
'(TSK) has scheda type'	'(TSK) Scheda Type'	'xsd:string'	'owl:topDataProperty'
'(LIR) has search level'	'(LIR) Research Level'	'xsd:string'	'owl:topDataProperty'

3.3.2 - Mapping OG - Oggetto:

Il paragrafo OG contiene a sua volta altri sottocampi, segue una lista di essi con relativa descrizione ICCD:

- OGT Oggetto: il campo contiene indicazioni che consentono la corretta e precisa individuazione, sia tipologica che terminologica, del bene catalogato.
- OGTD Definizione : Nome o locuzione che individua e determina l'oggetto in base alla connotazione funzionale e morfologica espressa secondo la tradizione degli studi storico artistici.

Di seguito sono state descritte in triple RDF solo le informazioni ritenute importanti al modello:

Nome Campo ICCD	Codice Campo ICCD	Entità/Range CIDOC	Proprietà CIDOC	Domain CIDOC	Voc.	Rip.
Oggetto	OG	E55 Type	P2 has type	E22 Man-Made Object		
	OGTD	E55 Type	P2 has type	E22 Man-Made Object	A	

- Descrizione dei concetti (Classi) del campo OG:

Classe	Vocabolario	Descrizione	Superclasse
'(OG) Object'		Classe relativa alla descrizione dell'oggetto catalogato	'E55 Type'
'(OGTD) Object Definition'		Classe delle definizioni dell'oggetto catalogato	'(OG) Object'

- Definizione delle proprietà utilizzate in OG Object Property:

Proprietà	Dominio	Range	Super-proprietà
'(OG) is type of'	'Cultural Heritage (Physical Object)'	'(OG) Object'	'P2 is type of'

3.3.3 - Mapping LC - Localizzazione geografico-amministrativa:

Il paragrafo LC contiene a sua volta altri sottocampi, segue una lista di essi con relativa descrizione ICCD:

- PVC Localizzazione geografico-amministrativa attuale: Informazioni sulla localizzazione geografico-amministrativa del bene al momento in cui viene compilata la scheda, relativa al territorio italiano oppure ad organizzazioni amministrativo-territoriali di Paesi esteri (nel caso, ad esempio, di beni presenti nelle aree di pertinenza delle ambasciate italiane all'estero).
- PVCS Stato: Indica il nome dello stato nella quale si trova il bene catalogato.
- PVCR Regione: Indica il nome della regione nella quale si trova il bene catalogato, secondo le denominazioni ufficiali delle regioni italiane. Qualora il bene non si trovi in Italia, al momento della catalogazione valorizzare il campo con '00' (bene situato all'estero). Vocabolario chiuso.
- PVCP Provincia : Indica la sigla corrispondente alla provincia nella quale si trova il bene catalogato. Qualora il bene non si trovi in Italia, al momento della catalogazione valorizzare il campo con '00' (bene situato all'estero).Vocabolario chiuso.
- PVCC Comune: Indicare il comune nel quale si trova il bene catalogato, senza abbreviazioni, secondo le denominazioni ISTAT dei comuni italiani. Qualora il bene non si trovi in Italia, al momento della catalogazione valorizzare il campo con '00' (bene situato all'estero).

Di seguito si è mappato in triple RDF le risorse espresse nei sottocampi.

Nome Campo ICCD	Codice Campo ICCD	Entità/Range CIDOC	Proprietà CIDOC	Domain CIDOC	Voc.	Rip.
Localizzazione geografico amministrativa	LC	E53 Place	P55 has current location	E22 Man-Made Object		
	PVC					
Stato	PVCS	E53 Place	P89 falls within	E53 Place	C	
Regione	PVCR	E53 Place	P89 falls within	E53 Place	C	
Provincia	PVCP	E53 Place	P89 falls within	E53 Place	C	
Comune	PVCC	E53 Place	P89 falls within	E53 Place	C	

- Descrizione dei concetti (Classi) del campo LC:

Classe	Vocabolario	Descrizione	Superclasse
'(LC) Geographic Administrative Localization'		Classe relativa alla descrizione dell'oggetto catalogato	'E53 Place'
'(PVC) Actual Localization'		Classe che contiene l'indirizzo completo	'(LC) Geographic Localization'
'(PVCC) Comune'		Classe dei Comuni	'(PVC)'
'(PVCP) Province'		Classe delle Province	'(PVC)'
'(PVCR) Region'		Classe delle regioni	'(PVC)'
'(PVCS) State'		Classe degli stati	'(PVC)'

- Definizione delle proprietà utilizzate in LC Object Property:

Proprietà	Dominio	Range	Super-proprietà
'(LC) has current location'	'Cultural Heritage (Physical Object)'	'(LC) Geographic Administrative Localization'	'P55 has current location'
'P89 falls within'	'E53 Place'	'E53 Place'	'P89 falls within'

3.3.5 - Mapping DT - Cronologia:

- DTZ Cronologia generica: Indicazione della fascia cronologica di riferimento. Questo campo può essere ripetitivo nei casi in cui un oggetto abbia subito interventi successivi validi per il suo insieme (es.: una statua del sec. XV rilavorata nel XVIII).
- DTZG Secolo: Indicazione dei secoli in numeri romani. Due secoli sono separati da una “/” seguita da uno spazio. Es.: sec. XVI secc. XIII/ XIV.
- DTS Cronologia specifica: Datazione specifica in anni, eventualmente anche a cavallo di secoli. Se esiste solo una datazione generica, va riportato il relativo intervallo in anni
- DTSI Da : Indicazione in anni della data iniziale dell'esecuzione dell'opera.
- DTSF A: Indicazione in anni della data finale dell'esecuzione dell'opera.
- DTM Motivazione cronologia: Indicazioni sulle motivazioni della datazione proposta. Il campo è ripetitivo.

Di seguito la descrizione dei campi di DT in triple RDF:

Nome Campo ICCD	Codice Campo ICCD	Entità/Range CIDOC	Proprietà CIDOC	Domain CIDOC	Voc.	Rip.
Cronologia	DT					SI
Cronologia generica Secolo	DTZG	E52 Time-Span	P4 has time span	crm:E65 Creation		
Cronologia specifica Da:	DTSI	E61 Time Primitive	ti:hasInt.StartDate	E52 Time-Span		
Cronologia specifica A:	DTSF	E61 Time Primitive	ti:hasInt.EndDate	E52 Time-Span		
Motivazione cronologia	DTMM	hico:Int.Criterion	hico:hasI.Criterio.	hico:Activity	C	Si

- Descrizione dei concetti (Classi) del campo DT:

Classe	Vocabolario	Descrizione	Superclasse
'(DTZG) Century Span'	Prov-o	Classe contenente il secolo in formato generico	'E52 Time-Span'
'(DTMM) Date Attribution'	Hico	Classe contenente la motivazione relativa ad una certa datazione	'hico:Activity'

- Definizione delle proprietà utilizzate in DT Object Property:

Proprietà	Dominio	Range	Super-proprietà
'(DTZG) has century span'	'E2 Temporal Entity'	'(DTZG) Century Span'	'P4 has time-span'

3.3.6 - Mapping MT - Dati tecnici:

- MTC Materia e tecnica: Indica la materia e la tecnica di esecuzione del bene catalogato, nella forma. Vocabolario aperto
- MIS MISURE: Indicazioni relative alle dimensioni massime e/o al peso del bene catalogato (nel caso di manufatti metallici di piccole dimensioni quali medaglie e monete). E' obbligatorio compilare almeno uno dei sottocampi che indicano il tipo di misura, oltre al sottocampo MISU nel quale è specificata l'unità di misura. Il campo è ripetitivo.
- MISA Altezza: Indicare l'altezza del bene catalogato.
- MISL Larghezza: Indicare la larghezza del bene catalogato.
- MISP Profondità: Indicare la profondità del bene catalogato.
- MISD Diametro: Diametro massimo del bene catalogato.
- MISN Lunghezza: Indicare la lunghezza del bene catalogato.
- MISS Spessore: Indicare lo spessore del bene catalogato.
- MISG Peso: Indicare il peso in grammi, con precisione ai due decimali.
- MISV Varie: Indicare altre misure utili, specificando sia il tipo, sia la parte presa in esame, sia il valore. Il sottocampo è ripetitivo.
- MISR Mancanza: Indicare, con la sigla MNR (Misure Non Rilevate) l'impossibilità di rilevare le misure del bene catalogato. Vocabolario chiuso.

Di seguito la descrizione delle informazioni di MT in triple RDF:

Nome Campo ICCD	Codice Campo ICCD	Entità/Range CIDOC	Proprietà CIDOC	Domain CIDOC	Voc.	Rip.
Dati tecnici	MT					
Indicazione di colore	MTX	E26 Physical Feature	P56 bears feature	E22 Man-Made Object		
Materia e tecnica	MTC	E55 Type	P2 has type	E22 Man-Made Object	A	Si
Misure	MIS					
Unità di misura	MISO	E16 Measurement	P91 has unit	E55 Type		
Altezza	MISA	xsd:string	P90 has value	E54 Dimension		
Larghezza	MISL	xsd:string	P90 has value	E54 Dimension		
Spessore	MISS	xsd:string	P90 has value	E54 Dimension		
Diametro	MISD	xsd:string	P90 has value	E54 Dimension		

- Descrizione dei concetti (Classi) del campo MT:

Classe	Vocabolario	Descrizione	Superclasse
'(MTX)'		Classe relativa ai colori	'E26 Physical Feature'

'(MTC) Technique'		Classe relativa alla tecnica di realizzazione	'E55 Type'
'(MISO) Measurement Type'		Classe dei tipi di unità di misura	'E16 Measurement'
'(MISD, 3.0) Diameter'		Classe indicativa del diametro di un oggetto	'E54 Dimension'
'(MISS, 3.0) Height'		Classe indicativa dell'altezza di un oggetto	'E54 Dimension'
'(MISS, 3.0) Thickness'		Classe indicativa dello spessore di un oggetto	'E54 Dimension'
'(MISS, 3.0) Width'		Classe indicativa della larghezza di un oggetto	'E54 Dimension'

- Definizione delle proprietà utilizzate in MT Object Property:

Proprietà	Dominio	Range	Super-proprietà
'(MTX) has colour'	'Cultural Heritage (Physical Object)'	'(MTX) Colour'	'P56 bears feature'
'(MTC) has technique'	'Cultural Heritage (Physical Object)'	'(MTC) Technique'	'P2 has type'
'(MISO, 3.0) is measurement type of'	'(MISO, 3.0) Measurement Type'	'E16 Measurement'	'P2 is type of'

- Definizione delle Data property degli individui del campo MIS:

Proprietà	Dominio	Range	Super-proprietà
'P90 has value'	'E54 Dimension'	'xsd:string'	'owl:topDataProperty'

Sulle classi indicanti le misure di altezza, spessore e larghezza, è stato posto un vincolo OWL per modellare il fatto che ogni tipo di misurazione è memorizzata come String, in modo tale da avere tutte le misurazioni in stringhe e identificarle poi con il 'Measurement type' corrispondente al tipo di unità di misura. Il vincolo posto sulle classi indicanti le misure (MISS) è:

sub class of:

'E54 Dimension'

'P90 has value' only xsd:string'

3.3.7 - Mapping TU - Condizione giuridica e vincoli:

- CDG Condizione giuridica: Informazioni relative all'attuale proprietà o detenzione del bene catalogato.
- CDGG Indicazione generica: Indica la personalità giuridica della proprietà o, qualora essa non sia accertabile, quella del detentore. Vocabolario chiuso.
- CDGS Indicazione specifica: Indica l'esatta denominazione dell'Amministrazione, dell'Ente, del privato che hanno la proprietà del bene. Qualora questi non siano noti, va indicata la denominazione del detentore o del possessore. Il sottocampo è ripetitivo.

Di seguito i sottocampi di TU espressi in triple RDF:

Nome Campo ICCD	Codice Campo ICCD	Entità/Range CIDOC	Proprietà CIDOC	Domain CIDOC	Voc.	Rip.
Condizione giuridica e vincoli	TU					
Condizione giuridica	CDG					
Indicazione generica	CDGG	E55 Type	P2 has type	E8 Acquisition	C	
Indicazione specifica	CDGS	E39 Actor	P52 has cur.owner	E18 Physical Thing		Si

- Descrizione dei concetti (Classi) del campo TU:

Classe	Vocabolario	Descrizione	Superclasse
'(CDGG) Acquisition Type'		Classe indicativa del tipo di acquisizione	'E55 Type'
'(CDGS) S. Acquisition Type'		Classe indicativa del tipo di acquisizione	'E39 Actor'

- Definizione delle proprietà utilizzate in TU Object Property:

Proprietà	Dominio	Range	Super-proprietà
'(CDGG) has acquisition type'	'Cultural Heritage (Physical Object)'	'(CDGG) Acquisition Type'	'P2 has type'

3.3.8 - Mapping DO - fonti e documenti di riferimento:

Informazioni sulla documentazione che si ritiene utile allegare alla scheda o semplicemente citare. Come descritto dalla documentazione ufficiale ICCD questa è un documento in aggiunta alla scheda ma legata comunque univocamente ad essa, esprimiamo questa relazione con la classe CIDOC CRM 'P70 Documents' che si lega all'entità effettiva 'E31 document'.

(E31→P70 Documents→E22 onceptual Object).

Andremo ora a relazionare le altre entità-concetti a questa prima classe, descrivendo tutti i sottocampi di DO.

- FTAX Genere: Indica se si tratta di documentazione allegata alla scheda di catalogo o di altra documentazione nota relativa al bene in esame; Trattandosi di un vocabolario chiuso di tipo stringa, abbiamo interpretato il campo Genere come una caratteristica legata al documento relazionandola attraverso la proprietà CIDOC 'P3_has_type' e legandola all'entità 'E55 type'. (E31→P2→E55).

- FTAP Tipo: Indica il tipo di documentazione allegata e/o esistente; Il tipo di scheda è stato interpretato come un'entità CIDOC 'E55 Type' legata al documento tramite la proprietà 'P2_has_type'. (E31→P2→E55).

- FTAN Codice identificativo: Indica il codice identificativo della documentazione allegata o esistente; Come si percepisce dalla documentazione ufficiale questa è una caratteristica che si può esprimere con l'entità CIDOC 'E42 Identifier' che si relaziona al Documento E31 tramite la proprietà 'P48 has preferred id'. (E31→P48→E42).

- FTAT Note: La intendiamo come una nota testuale esprimibile quindi attraverso la proprietà CIDOC 'P3_has_note' e relazionandola all'entità 'E62 String'. (E31 → P3 → E62).

Di seguito le triple RDF appena descritte:

Nome Campo ICCD	Codice Campo ICCD	Entità/Range CIDOC	Proprietà CIDOC	Domain CIDOC	Voc.	Rip.
Fonti e documenti di riferimento	DO	E22 Man Made Object	P70 documents	E31 Document		
Genere	FTAX	E55 Type	P2 has type	E31 Document	C	
Tipo	FTAP	E55 Type	P2 has type	E31 Document	A	
Codice identificativo	FTAN	E42 Identifier	P48 has preferred identifier	E31 Document		
Note	FTAT	E62 String	P3 has note	E31 Document		

- Descrizione dei concetti (Classi) del campo DO:

Classe	Vocabolario	Descrizione	Superclasse
'Reference Documentation'		Classe contenente informazioni documento di riferimento	'E32 Authority Document' sub of 'E31 Document'
'(FTAX) Documentation Genre'		Classe dei generi	'E55 Type'
'(FTAP) Documentation Type'		Classe dei tipi di documento	'E55 Type'
'(FTAN) Documentation Identifier Code'		Classe identificativi documento	'E42 Identifier'

- Definizione delle proprietà utilizzate in DO Object Property:

Proprietà	Dominio	Range	Super-proprietà
'documents'	'E31 Document'	'Cultural Heritage (Physical Object)'	'P70 documents'
'(FTAX) has genre'	'Reference Documentation'	'(FTAX) Documentation Genre'	'P2 has type'
'(FTAP) has documentation type'	'Reference Documentation'	'(FTAP) Documentation Type'	'P2 has type'
'(FTAN) has identifier code'	'Reference Documentation'	'(FTAN) Documentation Identifier Code'	'P48 has preferred identifier'

- Definizione delle Data property del campo FTAT di DO:

Proprietà	Dominio	Range	Super-proprietà
'has note'	'Reference Documentation'	'xsd:string'	'P3 has note'

- Di seguito si illustra il grafo risultante in Protégè del campo DO rappresentato in RDF/RDFS to OWL schema, con le proprietà appena descritte.

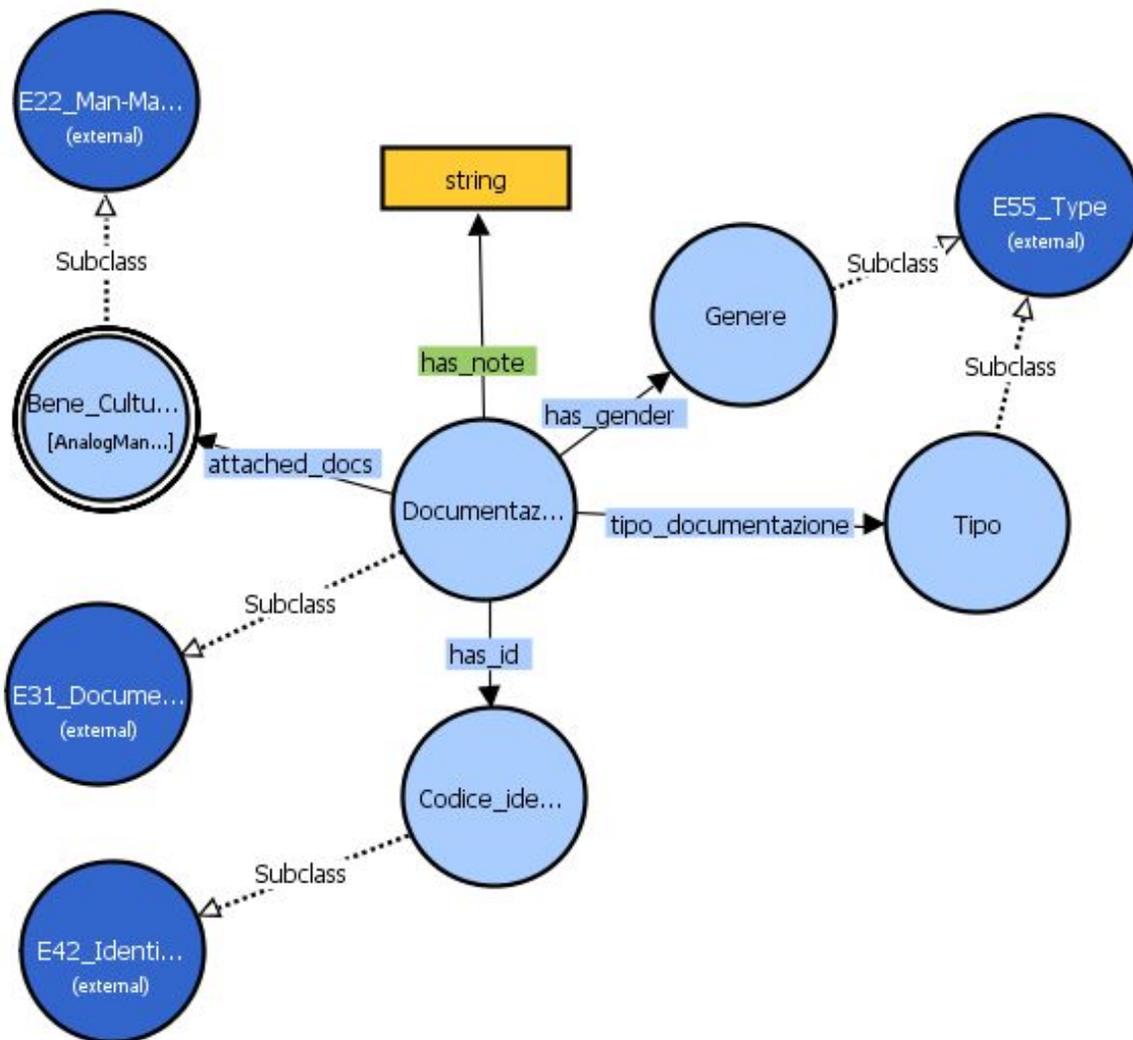


Figura 3.3.8: grafo OWL del campo DO

3.3.9 - Mapping AD - Accesso ai dati:

In questo paragrafo vengono registrate le indicazioni relative all'accesso ai dati, secondo le indicazioni dell'Ente fornitore, per l'utenza generica.

Anche in questo caso trattandosi di una documentazione la esprimiamo con l'entità 'E31 Document' e la relazioniamo all'entità 'E22 Conceptual Object' con la proprietà 'P70 Documents'. (E31 → P70 Documents → E22 Conceptual Object)

- ADSP Profilo di accesso: i profili da documentazione ICCD sono dei numeri ai quali si riferisce la definizione scelta, li rappresentiamo attraverso l'entità 'E42 Identifier' e la relazioniamo alla documentazione 'E31 Document' attraverso la proprietà 'P48 has preferred identifier'. (E31 → P48 → E42).

- ADSM Motivazione: La intendiamo come una nota testuale esprimibile attraverso 'E62 String', questa nota è relativa all'identifier in quanto esprime il motivo dell'assegnazione di un certo profilo d'accesso ad un documento, sarà quindi relazionata all'entità 'E42 Identifier' attraverso la proprietà CIDOC 'P3 has note'. (E42 → P3 → E62).

Di seguito le triple RDF che descrivono i campi:

Nome Campo ICCD	Codice Campo ICCD	Entità/Range CIDOC	Proprietà CIDOC	Domain CIDOC	Voc.	Rip.
Accesso ai dati	AD	E22 Man Made Object	P70 documents	E31 Document		
	ADS					
Profilo d'accesso	ADSP	E42 Identifier	P48 preferred identifier has	E31 Document	C	

Motivazione	ADSM	E62 String	P3 has note	E42 Identifier	C	
-------------	------	------------	-------------	----------------	---	--

- Descrizione dei concetti (Classi) del campo AD:

Classe	Vocabolario	Descrizione	Superclasse
'(ADSP) Access Profile'		Classe dei profili di accesso	'E42 Identifier'

- Definizione delle proprietà utilizzate in AD Object Property:

Proprietà	Dominio	Range	Super-proprietà
'documents'	'E31 Document'	'Cultural Heritage (Physical Object)'	'P70 documents'
'(ADSP) has access profile'	'Scheda'	'(ADSP) Access Profile'	'P48 has preferred identifier'

- Definizione delle Data property del campo ADSM di AD:

Proprietà	Dominio	Range	Super-proprietà
'(ADSP) has motivation'	'(ADSP) Access Profile'	'xsd:string'	'P3 has note'

- Di seguito si illustra il grafo risultante in Protégè del campo AD rappresentato in RDF/RDFS to OWL schema, con le proprietà appena descritte.

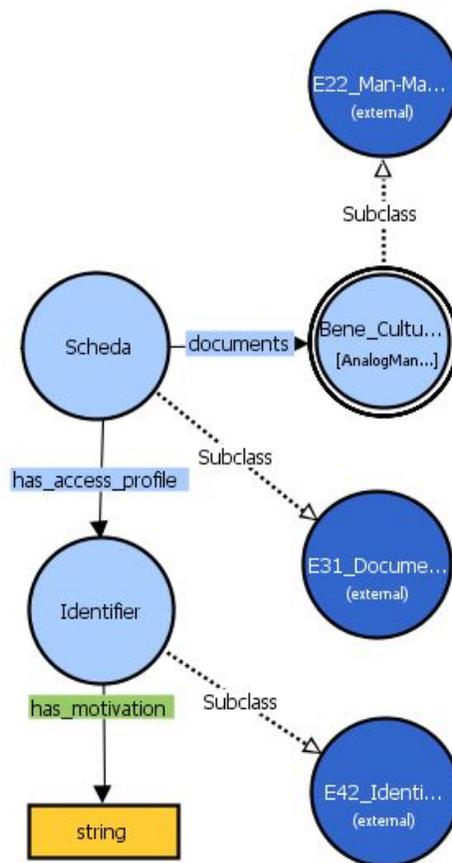


Figura 3.3.9: grafo OWL del campo AD

3.3.10 - Mapping CM - Compilazione:

In questo paragrafo vengono registrate le informazioni relative all'elaborazione e alla validazione scientifica della scheda. Le informazioni riguardano le date in cui tali operazioni sono state svolte e le persone intervenute e/o responsabili

intervenuti al momento della creazione della scheda, anch'esso è interpretato come documento allegato, in quanto ad ogni oggetto fisico E22 è associata una scheda di catalogo, Le informazioni di creazione del documento saranno rappresentate con l'entità E65 Creation di CIDOC alla quale verranno relazionate le altre informazioni contenute nei sottocampi di CM. (E31 Document→ P70 documents→ E22 Man Made Object→ P94 was created by → E65 Creation)

- CMPD Data: Anno di redazione della scheda, relazioniamo la data di creazione del documento tramite la proprietà 'P4_has_time_span'. (E65 Creation → P4_has_time_span → E52 Time Span)

- CMPN Nome: Indica il nome del compilatore della scheda nella forma 'cognome, nome'. (E65 Creation → P14 carried out by → E21 Person)

- FUR Funzionario responsabile: Il campo indica il funzionario responsabile della campagna di catalogazione. (E65 Creation → P14 carried out by→ E21 Person)

Di seguito le triple RDF che descrivono i campi:

Nome Campo ICCD	Codice Campo ICCD	Entità/Range CIDOC	Proprietà CIDOC	Domain CIDOC	Voc.	Rip.
Compilazione	CM	E65 Creation	P94 was created by	E31 Document		
	CMP					
Data	CMPD	E52 Time Span	P4 has time span	E65 Creation		Si
Nome	CMPN	E21 Person	P14 carried out by	E65 Creation		Si
Funzionario Responsabile	FUR	E21 Person	P14 carried out by	E65 Creation		Si

- Descrizione dei concetti (Classi) del campo CM:

Classe	Vocabolario	Descrizione	Superclasse
'(CM) Compilation'		Classe contenente i dati di compilazione	'E65 Creation'
'(CMPD) Drafting Date'		Classe delle date relative alla scrittura della scheda	'E52 Time-Span'
'(CMPN) Editor'	Foaf	Classe degli edotri della scheda	'E21 Person'
'(FUR) Supervisor'	Foaf	Classe dei supervisori della produzione scheda	'E21 Person'

- Definizione delle proprietà utilizzate in CM Object Property:

Proprietà	Dominio	Range	Super-proprietà
documents	'E31 Document'	'Cultural Heritage (Physical Object)'	'P70 documents'
'created_by'	'Scheda'	'(CM) Compilation'	'P70 documents'
'(CMPD) has draft date'	'(CM) Compilation'	'(CMPD) Drafting Date'	'P4 has time-span'
'(CMPN) compiled by'	'(CM) Compilation'	'(CMPN) Editor'	'P14 carried out by'
'(FUR) supervised by'	'(CM) Compilation'	'(FUR) Supervisor'	'P14 carried out by'

- Di seguito si illustra il grafo risultante in Protégè del campo CM rappresentato in RDF/RDFS to OWL schema, con le proprietà appena descritte.

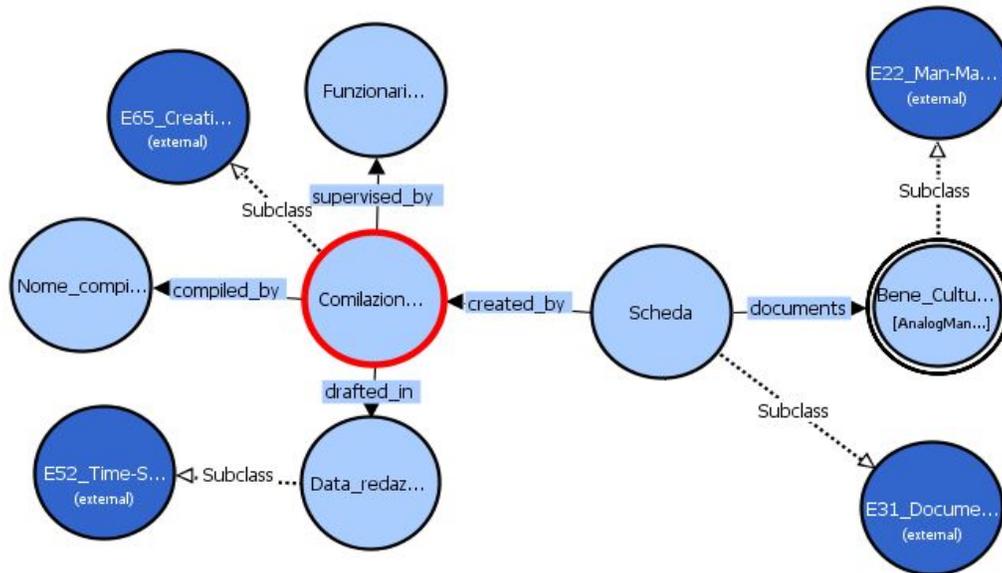


Figura 3.3.10: grafo OWL del campo CM

Grafo risultante DO-AD-CM :

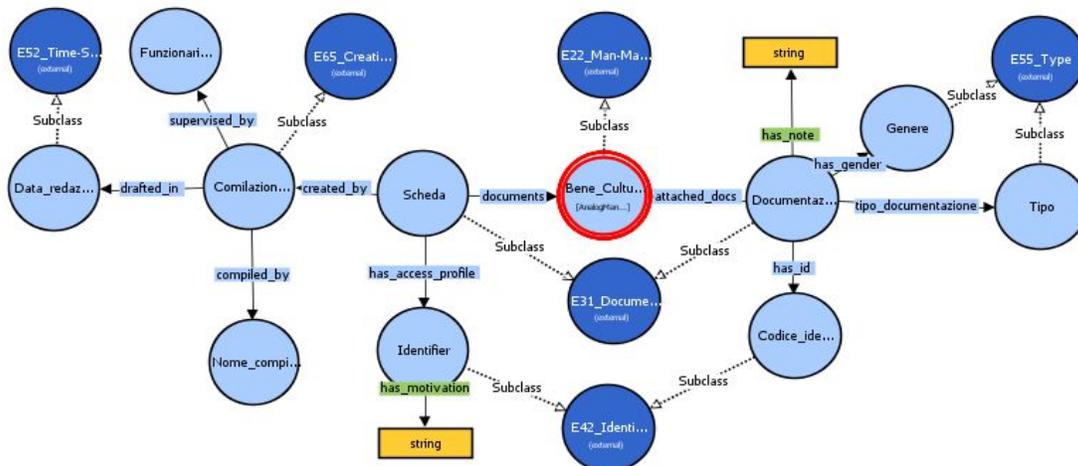


Figura 3.3.8: grafo OWL dei campi DO-AD-CM

4.0 - Mapping Tracciato OA - CICDO-OA ontology.

Una volta implementato i campi in comune non ci resta altro che integrare le informazioni aggiuntive relative al mapping specifico del tracciato OA ossia le schede catalografiche delle Opere D'arte. Le informazioni aggiuntive si allegheranno all'ontologia CICDO per formare una nuova appunto CICDO-OA comprensiva delle informazioni relative solo al tracciato OA.

Anche in questa nuova ontologia un ruolo fondamentale lo occupano le varie rappresentazioni che un'opera d'arte può avere, questa disambiguazione di concetti e specificazione di informazione viene effettuata tramite le classi FaBio, tramite esse possiamo classificare l'opera d'arte come un oggetto fisico potenzialmente sottoposto a diversi eventi di trasformazione (deterioramenti, restauri ...); l'essenza dell'opera d'arte è stata rappresentata dal livello Work di FRBR, mentre il risultato di ogni sua trasformazione dal livello Manifestation, in CIDOC-CRM esse sono associate corrispettivamente ad 'E28 Conceptual Work' - FaBio Work, mentre 'E22 Physical Object' - Fabio Manifestation.

Di seguito si propone il modello FaBio da documentazione ufficiale, e il corrispettivo applicato all'ontologia CICDO-OA.

Fabio Model:

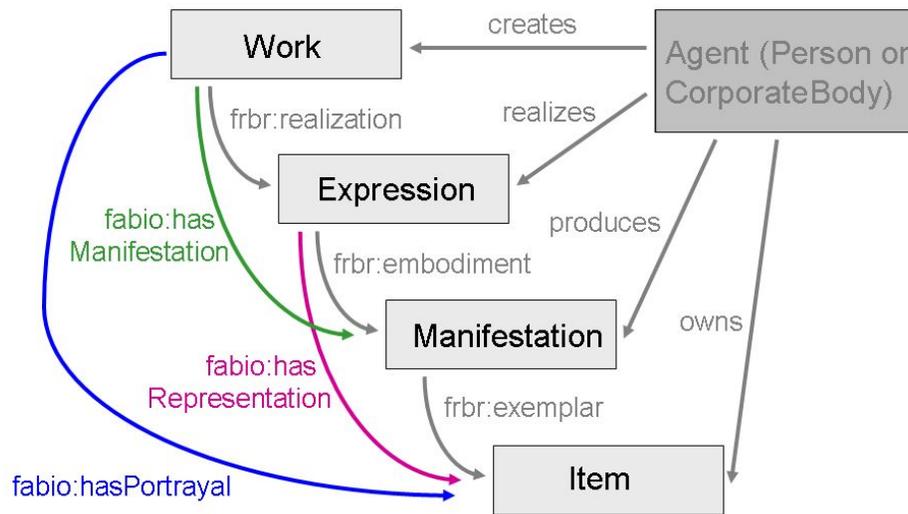


Figura 4.0.0: Fabio Model

CICDO-OA model:

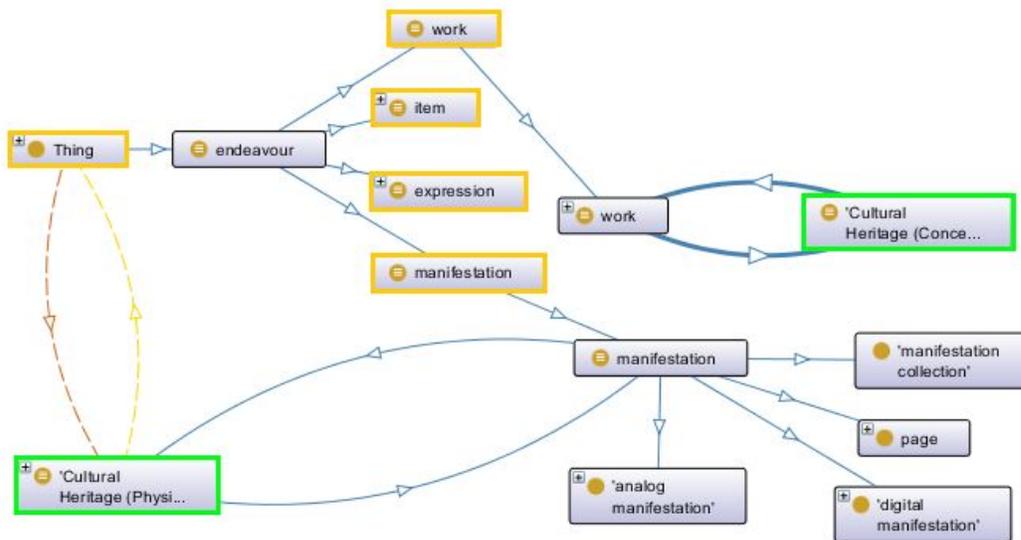


Figura 4.0.1: Fabio Model utilizzato in CICDO-OA

Come si può notare è stato importato il modello FaBio, e sono state relazionate le entità Cultural Heritage (Conceptual Object) al Work di Fabio e l'entità Cultural Heritage (Physical Object) alla manifestation di Fabio, corrispettivamente per indicare appunto la concettualizzazione di un'opera d'arte e la manifestazione fisica di essa, inoltre si eredita la proprietà Fabio 'has_manifestation' che definisce il Cultural Heritage (Physical Object) come una manifestazione del Cultural Heritage (Conceptual Object).

Le proprietà CIDOC-CRM ci hanno permesso di definire inoltre che un Cultural Heritage (Conceptual Object) è identificato con la proprietà CIDOC 'E28 Conceptual Object', e che l'entità corrispettiva al Cultural Heritage (Physical Object) è 'E22 Man-Made Object'. Il modello Fabio è quindi stato utile ad esprimere una relazione tra oggetto d'arte fisico e concettuale che l'ontologia CIDOC non avrebbe potuto esprimere, in quanto l'entità 'E28' ed 'E22' non sono relazionate tra loro nel modello CIDOC.

Le relazioni espresse nel modello sono state quindi :

Classe	Superclasse	Classe equivalente
'Cultural Heritage (Physical Object)'	'E22 Man-Made Object'	'manifestation'
'Cultural Heritage (Conceptual Object)'	'E28 Conceptual Object'	'work'
'work'	FRBR:endeavour	'Cultural Heritage (Conceptual Object)'
'manifestation'	FRBR:endeavour	'Cultural Heritage (Physical Object)'

Endeavour descrive i diversi aspetti di interesse degli utenti nelle opere artistiche quindi è la superclasse definita da FRBR che raccoglie tutti e 4 gli stadi in cui può essere categorizzato un oggetto e cioè: Work, Expression, Manifestation e Item.

L'object property che viene ereditata dalle due entità definite in CIDOC 'E22 Man-Made Object' e 'E28 Conceptual Object' è 'has manifestation' che esprime come un Work - Cultural Heritage (Conceptual Object) si manifesta in una manifestation - Cultural Heritage (Physical Object).

Vediamo ora l'object property implementata:

Proprietà	Dominio	Range	Super-proprietà
'has manifestation'	'Cultural Heritage (Conceptual Object)';work;	'Cultural Heritage (Physical Object)';manifestation;	'has related endeavour'

4.1 SPECIALIZZAZIONI DEL MODELLO CIDOC-OA:

Mentre nel modello comune a tutte le schede avevamo definito i concetti di scheda e di bene culturale in modo più generale possibile (rispettivamente come 'E22 Man made object', e 'E31 Documento' generale), nel nuovo modello CIDOC-OA possiamo ora categorizzare meglio le entità esprimendo l'opera d'arte soggetta alla documentazione come un'entità 'Opera D'arte' sottoclasse di Cultural Heritage (Physical Object) espressa dunque tramite 'E22 Man Made Object' che eredita tutte le proprietà definite per la sua superclasse. Inoltre i documenti allegati all'opera d'arte sono la scheda OA e le fotografie documentali. In questo modello dunque i due documenti allegati ,che documentavano un oggetto, corrispondenti nel modello CIDOC alle entità Scheda e Reference Documentation, vengono, nel nuovo modello definiti in modo più approfondito, esprimendo al meglio le informazioni della scheda OA, di seguito si elencano le nuove entità costruite e le loro relazioni.

Classe	Superclasse	Descrizione	Classe equivalente
'Opera D'Arte (Physical Object)'	'Cultural Heritage (Physical Object)'	Entità che descrive un'opera d'arte fisica	'manifestation'
'Photograph (Physical Object)'	'Cultural Heritage (Physical Object)'	Entità che identifica il bene culturale foto	'manifestation'
'OA Entry '	'E31 Document'; 'Scheda'	Identifica un documento che contiene metadati relativi a un'opera d'arte	
'F Entry '	'E31 Document'; 'Scheda'	Identifica un documento che contiene metadati relativi a una foto	

Avendo costruito nuove entità si son dovute definire nuove relazioni, segue un elenco con descrizione delle Object Property:

Proprietà	Dominio	Range	Super-proprietà
'has_attached_photo'	'Cultural Heritage (Physical Object)'	'Reference Documentation'	'P70 is documented in'
'photo_documents'	'Reference Documentation'	'Photograph (Physical Object)'	'P70 documents'
'photo_is_documented_in'	'Photograph (Physical Object)'	'F_Entry'	'P70 is documented in'
'is documented in'	'Opera D'Arte (Physical Object)'	'OA_Entry'	'P70 is documented in'

'documents'	'Scheda'	'Cultural Heritage (Physical Object)'	'P70 documents'
-------------	----------	------------------------------------------	-----------------

In questa modellazione si intende l'entità 'Reference Documentation', come le foto allegate ad un bene culturale, in quanto da specifica i documenti allegati alla scheda OA sono fotografie. Quello che si è voluto modellare è anche che una fotografia è in sé un bene culturale quindi anch'esso rappresentato da un'entità 'E22 Man Made Obj' con annesso una scheda ICCD relativa, la scheda F; per questo motivo si è costruito sia l'entità OA Entry (rappresentativa della scheda OA), sia la F Entry (rappresentativa della scheda F) e le abbiamo rispettivamente relazionate con le proprietà 'photo_is_documented_in' e 'is documented in' sopra descritte. Si può riassumere come il modello descriva che un'opera d'arte è un E22 Man Made Object ossia un bene culturale descritto da una scheda OA, al quale vengono allegate fotografie documentali che a loro volta sono beni culturali descritti da schede F.

Il grafo risultante la concettualizzazione di queste informazioni è il seguente:

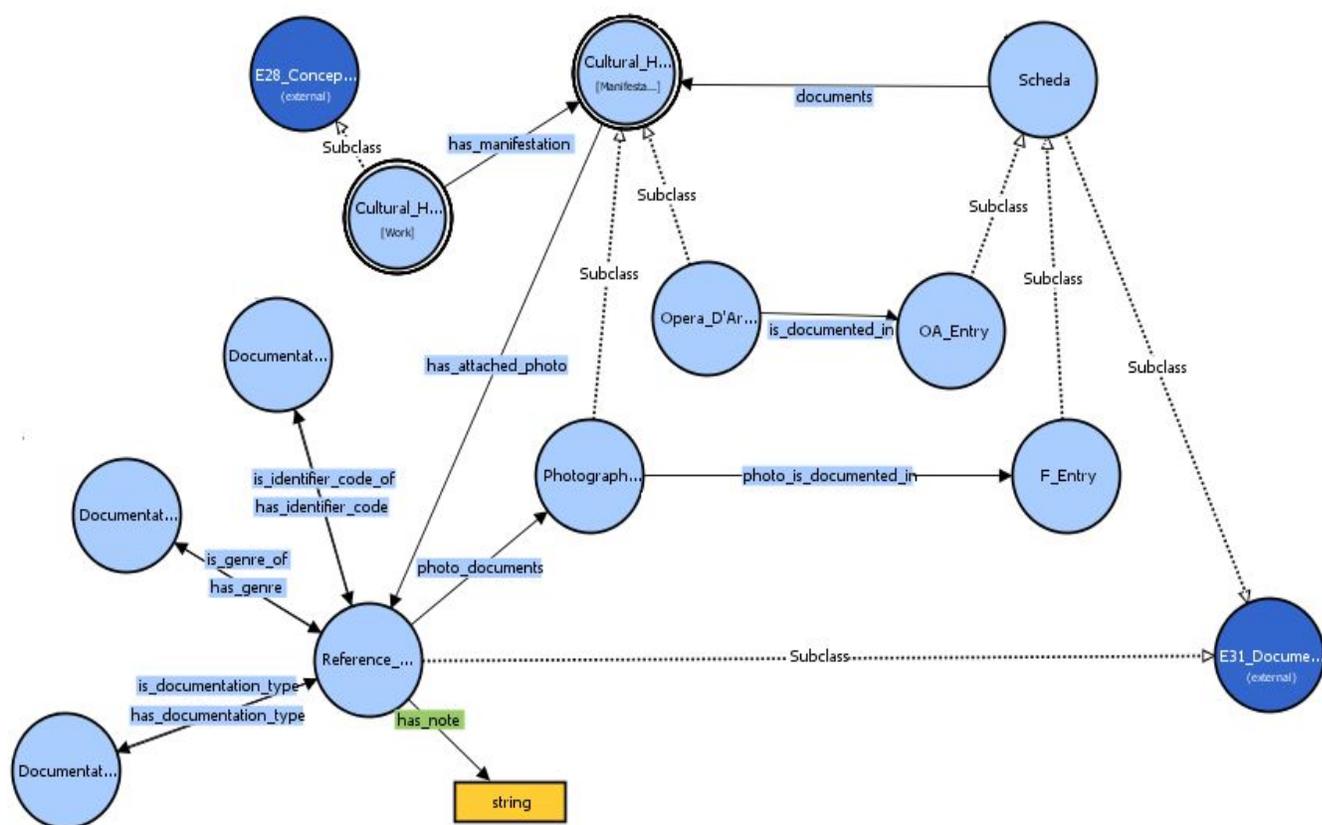


Figura 4.1: Parte del grafo riguardante l'ontologia CICDO-OA

Ricapitolando dal punto di vista CIDOC CRM il modello della Scheda OA riguarda la descrizione dell'opera d'arte intesa come bene culturale, dal punto di vista CIDOC-CRM sono oggetti fisici creati dall'uomo, quindi la classe usata per la rappresentazione è "E22_man_made_object" così come nel modello in comune CICDO. Le modifiche aggiunte per esprimere meglio le informazioni relative al tracciato OA sono legate ai documenti allegati all'opera d'arte, che in questo caso

sono state referenziati ad ontologie esterne, nello specifico l'entità 'OA Entry' è stata referenziata alla 'OA Entry Ontology' in modo da specificare che la scheda in questione sia la scheda OA e allegare tutte le informazioni provenienti da questa ontologia in ambito di classificazione della scheda OA. L'altro documento allegato esprime le informazioni del metadato DO della scheda OA, nello specifico caso del tracciato OA sono documenti fotografici quindi si è voluto esprimere questa informazione relazionando il documento espresso dall'entità CIDOC 'E31 Document' - 'Reference Documentation' con la nuova classe 'Photograph (Physical Object)' così da esprimere il concetto che un'opera d'arte è documentata da una fotografia che a sua volta è un bene culturale.

4.2 - MAPPING CIDOC-OA:

Avendo adeguato le classi e le object property appena descritte, si è potuto inoltre beneficiare di tutto il restante mapping dei campi in comune del modello ontologico CIDOC, dunque i metadati del tracciato OA ereditati dal modello comune ed espressi in triple RDF sono:

- **CD** - Codici competenza sul bene schedato.
- **OG** - Oggetto: contenente le informazioni essenziali per un'immediata identificazione del bene catalogato.
- **LC** - Localizzazione geografico-amministrativa.
- **DT** - Cronologia: Informazioni sugli estremi cronologici del bene catalogato.
- **MT** - Dati tecnici: Informazioni relative agli aspetti materici e tecnici del bene in esame.
- **TU** - Condizione giuridica e vincoli: Indicazioni relative alla proprietà del bene.
- **AD** - Accesso ai dati: contenente le indicazioni relative all'accesso ai dati.
- **CM** - Compilazione: contenente le informazioni relative all'elaborazione e alla validazione scientifica della scheda.

Oltre a questi si è mappato i campi obbligatori del tracciato OA per completare il modello ontologico CICDO-OA, e sono i seguenti:

- **AC** - Altri codici
- **LDC** - Collocazione Specifica
- **RE** - Modalità di reperimento
- **CO** - Conservazione
- **DO** - Fonti e documenti di riferimento

4.2.1 - Mapping AC - Altri codici

Il campo ICCD AC esprime un metadato di una scheda OA relativo ad altri codici utilizzati per identificare l'opera d'arte, nel modello ontologico CICDO-OA verrà rappresentato dall'entità CIDOC 'E42 Identifier' e relazionato all'oggetto fisico dell'opera d'arte identificato con 'E22 Man-Made Object'.

Di seguito segue la descrizione in triple RDF dei campi relativi alla sezione AC:

Nome Campo ICCD	Codice Campo ICCD	Entità/Range CIDOC	Proprietà CIDOC	Domain CIDOC	Voc.	Rip.
Altri Codici	AC					
Altro codice bene	ACC	'E42 Identifier'	'P48 has pref id'	'E22 Man-Made Object'		Si

- Descrizione dei concetti (Classi) del campo AC implementati:

Classe	Vocabolario	Descrizione	Superclasse
'(ACC) Alternative Code'		Classe relativa ai codici secondari alternativi alla classe TSK specifici per il tracciato OA	'E42 Identifier'

- Definizione delle proprietà utilizzate in AC Object Property:

Proprietà	Dominio	Range	Super-proprietà
'has_alternative_code'	'Opera D'Arte (Physical Object)'	'(ACC) Alternative Code'	'P48 has preferred identifier'

4.2.2 - Mapping LDC - Collocazione Specifica

LDC è un campo della scheda OA che esprime informazioni specifiche sulla collocazione del bene catalogato.

Il paragrafo LDC contiene a sua volta altri sottocampi, segue una lista di essi con relativa descrizione ICCD:

- LDCN: è un sottocampo di LDC che esprime la denominazione della struttura contenente l'opera d'arte. Per la denominazione degli edifici si fa riferimento, ove a fonti ufficiali, come ad esempio gli Annuari Diocesani per gli edifici ecclesiastici.
- LDCC: sottocampo di LDC che esprime il complesso monumentale a cui appartiene l'edificio in cui è collocato il bene.
- LDCU: sottocampo che esprime i dati riguardanti gli spazi viabilistici, pubblici o di uso pubblico, che consentono di individuare la collocazione di

un bene in un esterno o all'interno di una casa che non ha una denominazione nota o ancora in una collezione privata. Il campo contiene il nome della via o della piazza e il relativo numero civico, o l'indicazione della progressione chilometrica, nel caso della viabilità extraurbana. Ove non siano disponibili precedenti informazioni possono essere indicate in questa voce dati più generici riguardanti elementi ambientali (geografici, topografici, architettonici, naturalistici, ecc.).

- LDCS: sottocampo contenente le informazioni peculiari sulla localizzazione dell'oggetto all'interno del contenitore. La descrizione deve procedere dal generale al particolare, indicando, prioritariamente, la denominazione.

Di seguito segue la descrizione in triple RDF dei campi relativi alla sezione LDC:

Nome Campo ICCD	Codice Campo ICCD	Entità/Range CIDOC	Proprietà CIDOC	Domain CIDOC	Voc.	Rip.
Collocazione Specifica	LDC					
Denominazione	LDCN	E44 Place_appellation	P87_is_identifier_by	E53 Place		
Complesso di appartenenza	LDCC	E48 Place name	P87_is_identifier_by	E53 Place		
Denominazione spazio viabilistico	LDCU	E45 Address	P87_is_identifier_by	E53 Place		
Specifiche	LDCS	E53_Place	P55 has_current_loc	E22 Man-Made-Obj j		

- Descrizione dei concetti (Classi) del campo LDC implementati:

Classe	Vocabolario	Descrizione	Superclasse
'(LDCS) Specific Information'		Classe contenente la descrizione del contenitore specifico dell'oggetto d'arte	'(LC) Geographic Administrative Localization'; '(LDC) Specific Location'
'(LDCN) Denomination'		Classe delle denominazioni della struttura contenente l'opera d'arte	'E41 Appellation'; Equivalent to: 'E44 Place Appellation'
'(LDCC) Belonging Complex'		Classe dei nomi dei complessi di appartenenza dell'edificio in cui risiede l'oggetto d'arte	'E44 Place Appellation'
'(LDCU) Address'		Classe degli indirizzi specifici	'E44 Place Appellation'

- Definizione delle proprietà utilizzate in LDC Object Property:

Proprietà	Dominio	Range	Super-proprietà
'loc_identified_by'	'(LDC) Specific Location'	'(LDCN) Denomination'	'P87 is identified by'

4.2.3 - Mapping RE - Modalità di reperimento Oggetto D'arte.

Nel tracciato OA il campo RE è il campo che specifica come un oggetto d'arte sia stato recuperato, tramite questo campo si cataloga l'attività che hanno portato al ritrovamento dell'oggetto catalogato.

Rispettivamente:

- Il campo RCG esprime l'attività con il quale si è entrato in possesso dell'oggetto d'arte, è stato identificato con la classe CIDOC 'E7_Activity' e relazionato all'entità 'E8_acquisition' esprime la classe delle acquisizioni e con il quale è stato mappato il campo RE, tramite la proprietà 'P117_occurs_during'.
- Il campo NCUN contiene il riferimento univoco assegnato da ICCD, relativo all'attività di acquisizione dell'oggetto documentato nella scheda, è stato mappato con l'entità 'E42 Identifier' e viene relazionato attraverso la proprietà 'P1_Is_identified_by' all'entità 'E7_Activity'.
- RCGA indica il responsabile scientifico sovrintendente dei lavori di acquisizione. È mappato nel modello ontologico tramite l'entità CIDOC 'E39_Actor' e relazionato all'entità Activity attraverso la proprietà 'P14_carried_out_by'.
- RCGE indica il motivo per cui è stata avviata l'attività di acquisizione. Rappresenta una nota testuale quindi è rappresentata nel modello attraverso una stringa relazionata all'entità Activity attraverso la proprietà P3_has_note.
- RCGM indica il metodo con cui è stato reperito l'oggetto. Questa informazione si esprime attraverso l'entità 'E55_type' la quale è relazionata all'attività attraverso la proprietà CIDOC 'P32_used_general_technique'.
- RCGD indica la data dell'attività di acquisizione. Come le altre informazioni che esprimono una data presenti nel modello, è rappresentata attraverso

l'entità 'E52 Time-Span' e relazionata all'entità 'E7_Activity' tramite la proprietà 'P4_has_time_span'.

Di seguito viene data la descrizione in triple RDF dei campi relativi alla sezione RE:

Nome Campo ICCD	Codice Campo ICCD	Entità/Range CIDOC	Proprietà CIDOC	Domain CIDOC	Voc.	Rip.
Modalità di reperimento	RE	E8_acquisition	P24_i_chenge_own_through	E22 Man-Made-Obj		
Ricognizioni	RCG	E7_Activity	P117_occurs_during	E8_acquisition		
Codice univoco ICCD	NCUN	E42_Identifier	P1_Is_identified_by	E7_Activity		
Responsabile scientifico	RCGA	E39_Actor	P14_carried_out_by	E7_Activity		
Motivo	RCGE	xsd:string	P3 has note	E7_Activity		
Metodo	RCGM	E55_type	P32_used_general_technique	E7_Activity		
Data	RCGD	E52 Time-Span	P4_has_time_span	E7_Activity		

- Descrizione dei concetti (Classi) del campo RE implementati:

Classe	Vocabolario	Descrizione	Superclasse
'(RE) Acquisition OA'		Classe delle acquisizioni	'E8 Acquisition'
'(RCG) Inspection'		Classe delle ispezioni	'E7 Activity'
'(NCUN) Code'		Classe dei codici identificativi dell'attività	'E42 Identifier'

'(RCGA) Scientific_Sup'		Classe dei responsabili	'E21 Person'
'(RCGM) Method'		Classe dei metodi di acquisizione	'E55 Type'
'(RCGD) Acquisition Date'		Classe delle date	'E52 Time-Span'

- Definizione delle proprietà utilizzate in RE Object Property:

Proprietà	Dominio	Range	Super-proprietà
'acquisition_through'	'Opera D'Arte (Physical Object)'	'(RE) Acquisition OA'	'P24 changed ownership through'
'acquisition_during'	'(RE) Acquisition OA'	'(RCG) Inspection'	'P117 occurs during'
'(NCUN) has_activity_ID'	'(RCG) Inspection'	'(NCUN) Code'	'P48 has preferred identifier'
'(RCGA) in charge by'	'(RCG) Inspection'	'(RCGA) Scientific_Supervisor'	'P14 carried out by'
'(RCGM) used method'	'(RCG) Inspection'	'(RCGM) Metod'	'P32 used general technique'
'(RCGD) has date'	'(RCG) Inspection'	'(RCGD) Acquisition Date'	'P4 has time-span'

- Definizione delle Data property del campo RCGE:

Proprietà	Dominio	Range	Super-proprietà
'(RCGE) has note'	'(RCG) Inspection'	'xsd:string'	'P3 has note'

4.2.4 - Mapping CO - Conservazione

CO contiene le informazioni relative allo stato di conservazione di un'opera d'arte, contiene diversi sottocampi e sono:

- STC Stato di conservazione: è il campo che esprime la qualificazione dello stato di integrità dell'opera d'arte in rapporto alla condizione originaria.
- STCC Stato di conservazione: esprime un giudizio sintetico sullo stato dell'opera. Il vocabolario di questo campo è di tipo chiuso e può essere: buono cattivo discreto mediocre.
- STCS: è il campo che contiene indicazioni più specifiche riguardo al cambiamento di stato dell'opera d'arte rispetto allo stato originario, nella forma e nella consistenza dell'opera.

Nome Campo ICCD	Codice Campo ICCD	Entità/Range CIDOC	Proprietà CIDOC	Domain CIDOC	Voc.	Rip.
Conservazione	CO	E14_Condition_Assessment	P34i_was_assessed_by	E22 Man-Made Object		
Stato di conservazione	STC	E3_Condition_State	P35_has_identified	E14_Condition_Ass.		
Stato di conservazione	STCC	E55 Type:xsd:string	P2 has type	E3_Condition_State	C	
Indicazioni specifiche	STCS	E55 Type:xsd:string	P2 has type	E3_Condition_State		

- Descrizione dei concetti (Classi) del campo CO implementati:

Classe	Vocabolario	Descrizione	Superclasse
'(CO) OA Condition'		Classe complessiva di tutte le valutazioni delle condizioni	'E14 Condition Assessment'
'(STC) OA Condition State'		Classe delle condizioni dello stato di un'opera d'arte	'E3 Condition State'

- Definizione delle proprietà utilizzate in CO Object Property:

Proprietà	Dominio	Range	Super-proprietà
'(CO) has_condition'	'Opera D'Arte (Physical Object)'	'(CO) OA Condition'	'P34 was assessed by'
'(STC) has_state'	'(CO) OA Condition'	'(STC) OA Condition State'	'P35 has identified'

- Definizione delle Data property del campo STCC-STCS:

Proprietà	Dominio	Range	Super-proprietà
'(STCS) OA specification'	'(STC) OA Condition State'	xsd:string	'P3 has note'
'(STCC)_OA_cons_state'	'(STC) OA Condition State'	xsd:string	'P3 has note'

- Di seguito si illustra il grafo risultante in Protégè dei campi RE e CO rappresentato in RDF/RDFS to OWL schema, con le proprietà appena descritte:

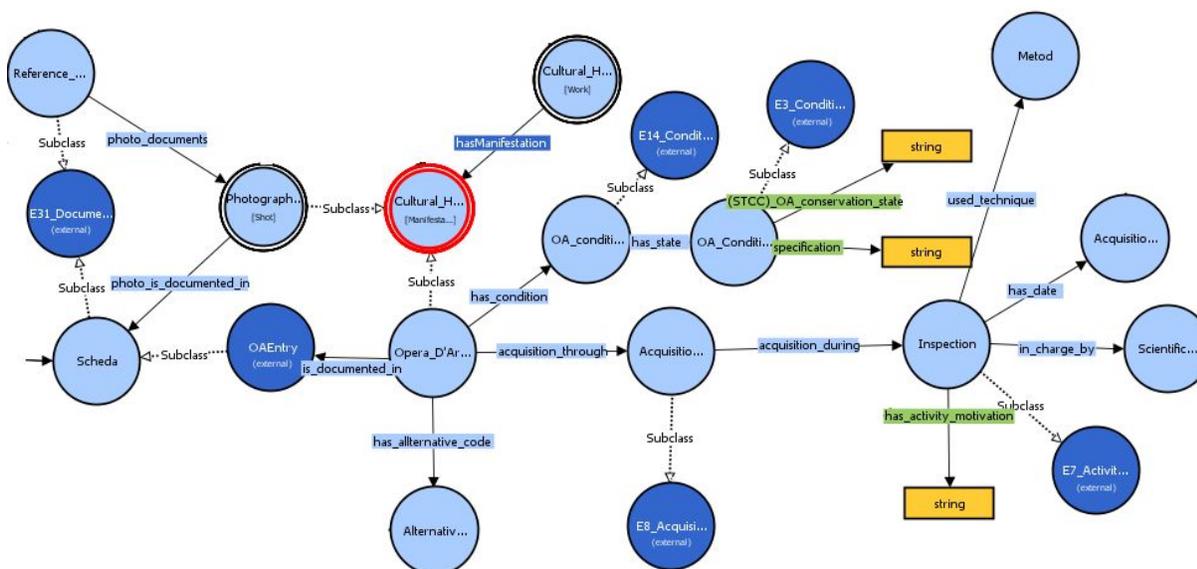


Figura 4.2.4: grafo OWL dei campi RE, CO dell'ontologia CICDO-OA

4.2.5 - Mapping DO - Documentazione fotografica

Contiene informazioni sulla documentazione fotografica che si ritiene utile allegare alla scheda OA. Come descritto dalla documentazione ufficiale ICCD questa è un documento in aggiunta alla scheda, ma legata comunque univocamente ad essa, esprimiamo questa relazione con la classe CIDOC CRM ‘

P70 Documents ‘ che si lega all’entità ‘E31 document’ equivalente nel caso della scheda OA all’entità 'Reference Documentation'. Parte delle relazioni e concetti relativi a questa classe sono già stati definiti nel mapping CICDO e quindi li andremo semplicemente ad ereditare ed espandere con gli altri campi specifici della scheda OA, di seguito si analizzano i sottocampi di DO e si aggiungerà il mapping in triple RDF di quelli non specificati in CICDO. Campi mappati in triple RDF ereditati dal modello comune CICDO:

- FTAX Genere: Indica se si tratta di documentazione allegata alla scheda di catalogo o di altra documentazione nota relativa al bene in esame; Trattandosi di un vocabolario chiuso di tipo stringa, abbiamo interpretato il campo Genere come una caratteristica legata al documento relazionandola attraverso la proprietà CIDOC ‘P3_has_type’ e legandola all’entità ‘E55 type’. ('Reference Documentation' = E31→P2→E55).
- FTAP Tipo: Indica il tipo di documentazione allegata e/o esistente; Il tipo di scheda è stato interpretato come un’entità CIDOC ‘E55 Type’ legata al documento tramite la proprietà ‘P2_has_type’. ('Reference Documentation' = E31→P2→E55).
- FTAN Codice identificativo: Indica il codice identificativo della documentazione allegata o esistente, questa è una caratteristica che si può esprimere con l’entità CIDOC ‘E42 Identifier’ che si relaziona al Documento E31 tramite la proprietà ‘P48 has preferred id’. ('Reference Documentation' = E31→ P48→ E42).
- FTAT: Note; La intendiamo come una nota testuale esprimibile quindi attraverso la proprietà CIDOC ‘P3_has_note’ e relazionandola all’entità ‘E62 String’. ('Reference Documentation' = E31 → P3 → E62).

Campi mappati in triple RDF specifici del tracciato OA:

- FTAE Ente proprietario; indica il proprietario ed è legato al singolo esemplare della manifestazione della fotografia, per cui il documento reale

è inteso come un 'Photograph (Physical Object)' relazionato attraverso la proprietà 'P51 has former owner' con l'entità 'E39 Actor' che esprime l'ente proprietario. (Photograph (Physical Object) → P51 → E39).

- FTAC Collocazione attuale; Allo stesso modo del campo precedente interpretiamo questa informazione come una caratteristica legata alla manifestazione del documento fotografico 'Photograph (Physical Object)' con il quale relazioniamo tramite la proprietà 'P55 has current location' l'entità 'E53 Place' (Photograph (Physical Object) → P55 → E53).

Di seguito le triple RDF non ereditate dal modello CICDO e aggiunte al modello CICDO-OA:

Nome Campo ICCD	Codice Campo ICCD	Entità/Range CIDOC	Proprietà CIDOC	Domain CIDOC	Voc.	Rip.
Fonti e documenti di riferimento	DO					
FTAE		'E39 Actor'	'P51 has former owner'	'Reference Documentation'		
FTAC		'E53 Place'	'P55 has current location'	'Reference Documentation'		

- Descrizione dei concetti (Classi) del campo DO:

Classe	Vocabolario	Descrizione	Superclasse
'(FTAE) photo owner'		Classe contenente i proprietari delle foto	'E21 Person'
'(FTAC) Photo_Actual_Place'		Classe delle locazioni	'E53 Place'

- Definizione delle proprietà utilizzate in DO Object Property:

Proprietà	Dominio	Range	Super-proprietà
'(FTAE) has_owner'	'Photograph (Physical Object)'	'(FTAE) photo owner'	'P52 has current owner'
'(FTAC) actual_location'	'Photograph (Physical Object)'	'(FTAC) Actual Location'	'P55 has current location'

5.0 - REASONER:

Il reasoner svolge un ruolo fondamentale nello sviluppo e utilizzo dell'ontologia. Il reasoner usato è stato HeremiT 1.3.8 disponibile in Protègè, HeremiT è un reasoner Open source capace di analizzare le ontologie scritte utilizzando l'Ontology Language Web (OWL). Dato un file OWL HeremiT è in grado di determinare se l'ontologia definita è consistente ossia coerente con le assunzioni fatte sulle classi e le relazioni fra esse.

Partendo dal modello di dati CICDO-OA espresso in OWL, questo agente software "intelligente" può usare queste meta-informazioni per fare dei semplici ragionamenti sui dati, ossia oltre a verificare tutte le asserzioni definite dall'ontologia, deduce delle inferenze logiche, le inferenze logiche sono regole logiche di dominio in aggiunta a quelle già definite nel modello. Questo processo è chiamato Reasoning ("ragionamento") è uno dei principi cardine del Semantic Web, in quanto consente di inferire nuova conoscenza ricavando affermazioni che non erano specificate esplicitamente nei dati iniziali, tante più sono le caratteristiche o vincoli strutturali del dominio espresse con linguaggio ontologico, tanto più potente potrà essere il ragionamento applicabile ai dati iniziali. Il reasoner dunque garantisce che il modello ontologico CICDO-OA è consistente ed inoltre inferisce nuova conoscenza ad esso, seguono alcuni esempi di inferenze offerte dal reasoner Heremit in Protègè.

Inferenze sull'entità Opera D'arte:

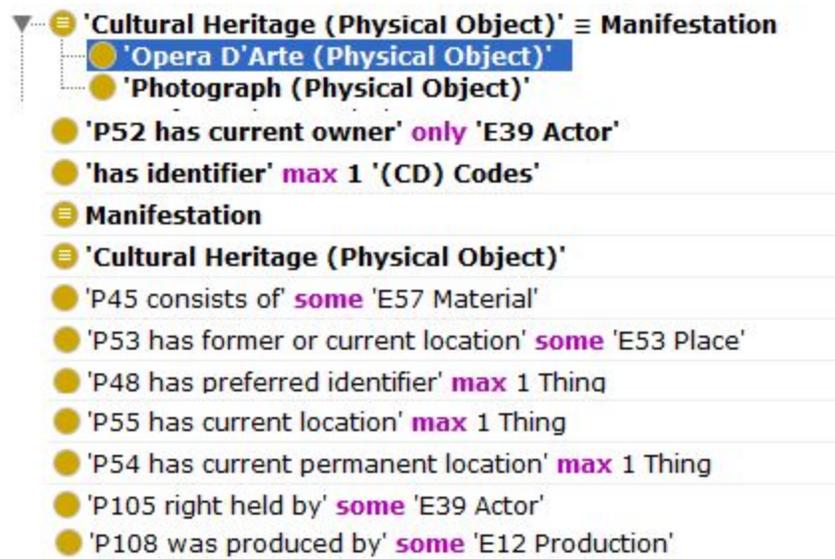


Figura 5.1: inferenze derivate sull'entità Opera D'arte

Le asserzioni dedotte dal reasoner riguardo la classe Opera d'arte sono ad esempio che un'opera d'arte è fatta di un certo materiale, ha al massimo un codice identificativo, ha al massimo una current location, ed è posseduto da qualche attore/ente, queste asserzioni rappresentano conoscenza aggiuntiva non definita ma bensì inferite dal reasoner.

Inferenze sull'entità Scheda OA:

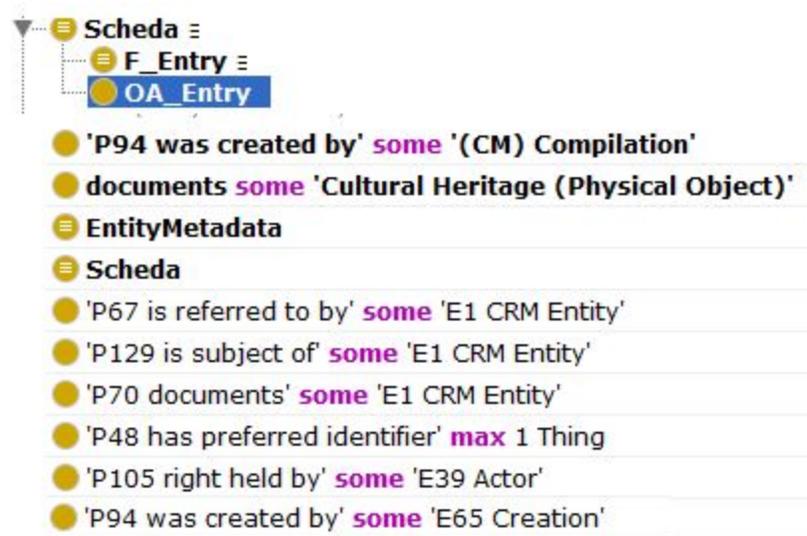


Figura 5.2: inferenze derivate sull'entità Scheda

La conoscenza inferita dal reasoner riguardo la scheda OA sono ad esempio che una scheda è creata attraverso una compilazione, documenta un bene culturale, è un'entità contenente metadati relativi ad un'opera d'arte.

Inferenze sulle entità Persona: 'scientific_Supervisor', 'Photo Owner', 'Supervisor', 'Editor':

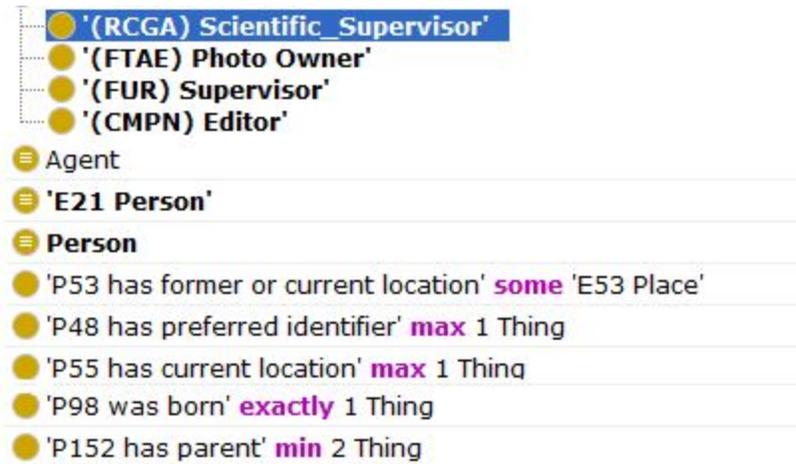


Figura 5.3: inferenze derivate sull'entità Persona

La conoscenza inferita dal reasoner riguardo le entità appena elencate sono ad esempio che sono tutte delle persone, e che sono nate esattamente una volta e che hanno un minimo di due parenti.

6.0 - CONCLUSIONI

Nel progetto dunque proponendo l'ontologia, CICDO-OA si vuole offrire un modello ontologico che rispecchia l'analiticità delle schede ministeriali OA.

L'ontologia definita costituisce un'infrastruttura per descrivere tutti i contenuti di più basi di dati distribuite, contenenti schede OA strutturate secondo lo standard ICCD. Dunque più risorse distribuite riguardanti lo stesso dominio possono essere accomunate e rese disponibili sul web come se fossero un'unica base di conoscenza. I fornitori di contenuti relativi ai beni culturali utilizzando gli stessi concetti, vengono automaticamente linkati l'uno con l'altro costituendo un'unica base di conoscenza affidabile e interoperabile comune dal quale dedurre inferenze. Le risorse così strutturate sono machine readable ed è possibile effettuare il reasoning che consente di inferire nuova conoscenza dalla base di conoscenza, ricavando affermazioni che non erano specificate esplicitamente nei dati iniziali, tante più risorse verranno espresse con le stesse ontologie tanto più saranno potenti le inferenze e le deduzioni logiche che si potranno fare su di essi. L'ontologia proposta rappresenta dunque uno strumento di conversione di dati strutturati secondo il modello ICCD OA, lo sviluppo futuro di questo progetto sarà dunque convertire le schede OA messe a disposizione per il progetto, attualmente descritte tramite XML, in RDF-S sfruttando il modello ontologico CICDO-OA, rendendo quindi i loro contenuti accessibili dal web in una modalità leggibile e interpretabile da una macchina e dunque introdurli al mondo del Web semantico e Linked Open Data.

BIBLIOGRAFIA

- [1] - <http://www.iccd.beniculturali.it/index.php?it/473/standard-catalografici/Standard/29>
- [2] - <http://www.iccd.beniculturali.it>
- [3] - <https://www.w3.org/RDF/>
- [4] - <http://www.cidoc-crm.org/>
- [5] - <https://home.cern/>
- [6] - <https://www.w3.org/Addressing/URL/uri-spec.html>
- [7] - https://it.wikipedia.org/wiki/Hypertext_Transfer_Protocol
- [8] - http://www.academia.edu//Web_1.0_to_Web_3.0_-_Evolution_of_the_Web
- [9] - <http://www.oreilly.com/pub/a/web2/archive/what-is-web-20.html>
- [10] - [https://www.w3.org/2009/Talks/0427-web30-tbl/#\(1\)](https://www.w3.org/2009/Talks/0427-web30-tbl/#(1))
- [11] - <https://www.w3.org/XML/>
- [12] - https://en.wikipedia.org/wiki/Semantic_Web_Stack
- [13] - <https://www.w3.org/TR/rdf-syntax-grammar/>
- [14] - <https://www.w3.org/TR/turtle/>
- [15] - https://it.wikipedia.org/wiki/Web_Ontology_Language
- [16] - <http://tomgruber.org/writing/ontologia-kaj-1993.pdf>
- [17] - https://it.wikipedia.org/wiki/Linked_data
- [18] - <https://www.w3.org/standards/semanticweb/data>
- [19] - https://www.lri.fr/~hamdi/datalift/tuto_inspire_2012/Suggestedreadings/egovld.pdf
- [20] - <http://5stardata.info/en/>
- [21] - <http://www.beniculturali.it/mibac/export/MiBAC/index.html>
- [22] - <http://www.cidoc-crm.org/>
- [23] - <http://icom.museum/the-committees/international-committees/international-committee/international-committee-for-documentation/>
- [24] - <http://pro.europeana.eu/page/edm-documentation>
- [25] - <http://pharosartresearch.org/>
- [26] - <http://www.iccd.beniculturali.it/>