

ALMA MATER STUDIORUM – UNIVERSITA' DI BOLOGNA
CAMPUS DI CESENA

SCUOLA DI INGEGNERIA E ARCHITETTURA
CORSO DI LAUREA MAGISTRALE IN INGEGNERIA E SCIENZE INFORMATICHE

**MODELLAZIONE E REALIZZAZIONE
DI UN'ONTOLOGIA FORMALE
PER LA RAPPRESENTAZIONE DI
INFORMAZIONI RELATIVE AI BENI
CULTURALI NEL WEB SEMANTICO**

Relazione finale in
Web Semantico

Relatore
Antonella Carbonaro

Presentata da
Andrea Pari

Sessione II
Anno Accademico 2016/17

ABSTRACT

La trattazione corrente si pone come obiettivo principale quello di fornire una descrizione esauriente del lavoro svolto relativo alla costruzione di un'ontologia per la rappresentazione di informazioni sui beni culturali nel Web semantico, la nuova estensione del Web che sta prendendo piede negli ultimi anni come modello standard per attribuire un significato alle informazioni contenute nei documenti della rete.

Il lavoro descritto deriva essenzialmente dalla necessità di definire un modello ontologico per la rappresentazione informatizzata dei dati relativi ai beni del patrimonio culturale, in contrasto con le modalità generali di rappresentazione nel Web attuale, estremamente limitanti e poco adeguate in un contesto per cui si dovrebbe invece favorire la diffusione della conoscenza a livello globale.

Il dominio applicativo da cui poi si è articolato l'intero lavoro è rappresentato dalle modalità con cui i dati culturali sono attualmente catalogati. In Italia, l'organo responsabile della catalogazione è l'*Istituto Centrale per il Catalogo e la Documentazione* (ICCD), che nel corso degli anni ha definito appositi standard normativi per la descrizione delle informazioni relative alle diverse tipologie di beni. Sebbene il lavoro faccia riferimento a normative di catalogazione utilizzate esclusivamente a livello nazionale, l'ontologia proposta rappresenta comunque un modello ampiamente valido anche per un suo riutilizzo a livello internazionale.

La descrizione del lavoro riportata nella trattazione corrente è strutturata secondo le tre fasi che hanno caratterizzato la realizzazione dell'ontologia in questione: analisi generale del dominio applicativo di riferimento; definizione dei concetti ontologici, rigorosamente allineati ai vincoli stabiliti dalle normative dell'ICCD, attraverso l'utilizzo di modelli concettuali e schemi già pubblicati nel Web semantico; costruzione dell'ontologia secondo i concetti ontologici definiti alla fase precedente.

RINGRAZIAMENTI

La stesura della tesi corrente è avvenuta durante un periodo particolarmente difficile e significativo a livello personale, che ha rappresentato per me un motivo di crescita e di cambiamento. Ritengo quindi doveroso ringraziare tutti coloro che mi hanno accompagnato, sostenuto e supportato in questi anni, in particolare negli ultimi mesi.

Ringrazio innanzitutto la professoressa Carbonaro per avermi proposto questo progetto di tesi, grazie al quale ho potuto concludere il mio percorso di studi. La ringrazio inoltre per la disponibilità, la precisione e la serietà dimostrata, che avevo già potuto riconoscere durante la stesura della tesi di laurea triennale.

In secondo luogo, ringrazio Valerio Capoccia di 3D Informatica per averci seguito con grande professionalità, attenzione e precisione durante tutte le fasi di progetto.

Ringrazio particolarmente i miei genitori Paola e Roberto per la pazienza e il supporto, grazie al quale ho potuto intraprendere la mia carriera universitaria e ottenere due lauree. Un ringraziamento speciale lo dedico ai miei zii paterni, su cui ho potuto sempre contare in questi anni, e a mia sorella che, seppur fisicamente lontana, ha sempre rappresentato per me un grande esempio e un punto di riferimento.

Ringrazio i colleghi di corso con cui ho partecipato alla realizzazione dei vari progetti universitari, in particolare Luca e Antonio con cui ho lavorato al progetto corrente.

Un ringraziamento speciale va a Riccardo, con cui ho condiviso praticamente tutti i momenti di studio degli ultimi anni, supportandoci e spronandoci a vicenda nel raggiungimento dei nostri obiettivi universitari.

In generale, ringrazio tutti i familiari, gli amici e le persone che hanno contribuito, anche solo con la loro presenza, a rendere più leggero e piacevole questo periodo.

Infine, ringrazio me stesso. Perché l'università mi ha insegnato che, come nella vita, per quanto abbiamo persone accanto a noi che ci supportano, è il nostro impegno e la nostra forza di volontà che ci permettono di raggiungere i traguardi che desideriamo.

INDICE

INTRODUZIONE	1
CAPITOLO 1. IL WEB SEMANTICO: UNA PANORAMICA GENERALE	5
1.1. Il World Wide Web.....	5
1.1.1. Nascita del Web.....	5
1.1.2. Evoluzione del Web	6
1.1.2.1. Il Web statico	7
1.1.2.2. Il Web dinamico.....	8
1.1.2.3. Verso il Web semantico.....	9
1.2. Problemi del Web attuale.....	10
1.2.1. Motori di ricerca	11
1.2.1.1. Information retrieval	12
1.2.1.2. Limiti dei motori di ricerca.....	13
1.2.2. Web Service, mash-up e accesso alle informazioni	14
1.2.3. Contestualizzazione di un'informazione	15
1.2.4. Collegamenti ipertestuali.....	16
1.2.5. Derivazione di conoscenza implicita.....	17
1.2.6. Problemi del Web attuale: un riepilogo.....	17
1.3. Stato dell'arte del Web semantico	18
1.3.1. Idee di base	18
1.3.2. La visione del Web semantico.....	20
1.3.3. Architettura del Web semantico	20
1.3.3.1. Lo strato sintattico	22
1.3.3.1.1. Uniform Resource Identifier e Unicode.....	22
1.3.3.1.2. eXtensible Markup Language (XML).....	22
1.3.3.2. Lo strato semantico	23

1.3.3.2.1. Resource Description Framework (RDF)	23
1.3.3.2.2. Web Ontology Language (OWL).....	26
1.3.3.2.3. SPARQL Protocol and RDF Query Language.....	26
1.3.4. Il progetto Linked Open Data.....	27

CAPITOLO 2. LINKED DATA RELATIVI AI BENI CULTURALI: PUBBLICAZIONE E UTILIZZO SUL WEB SEMANTICO 29

2.1. Il patrimonio culturale	30
2.2. Pubblicazione e utilizzo sul Web 2.0	31
2.2.1. Portali informativi culturali	32
2.2.2. Accesso alle informazioni	33
2.2.2.1. Federated search	33
2.2.2.2. Data warehousing	34
2.3. Pubblicazione e utilizzo sul Web semantico: un modello concettuale condiviso	35
2.3.1. Caratteristiche dei dati e interoperabilità.....	35
2.3.2. Modello concettuale per la descrizione dei dati	37
2.3.3. Portali informativi culturali semantici.....	38
2.3.3.1. Componenti di un portale semantico	38
2.3.3.2. Benefici per gli utilizzatori	39

CAPITOLO 3. UN'ONTOLOGIA PER RAPPRESENTARE I DATI DEI BENI CULTURALI NEL WEB SEMANTICO: INTRODUZIONE 43

3.1. Profilo dell'azienda committente.....	44
3.2. Motivazioni	45
3.3. Obiettivi e requisiti	45
3.4. Fasi di realizzazione del progetto	46

CAPITOLO 4. FASE 1: ANALISI DEL DOMINIO APPLICATIVO	49
4.1. L’Istituto Centrale per il Catalogo e la Documentazione (ICCD).....	50
4.2. Modalità di compilazione di una scheda.....	51
4.3. La normativa F: analisi del livello di inventario.....	54
4.3.1. Analisi dei paragrafi in comune	56
4.3.1.1. Il paragrafo DT	56
4.3.1.2. Il paragrafo MT.....	58
4.3.1.3. Il paragrafo TU	60
4.3.2. Analisi dei paragrafi esclusivi	61
4.3.2.1. Il paragrafo SG.....	61
4.3.2.2. Il paragrafo UB	62
4.3.2.3. Il paragrafo AU	63
4.3.3. Analisi dei campi esclusivi di paragrafi in comune.....	64
4.3.3.1. Il paragrafo OG	64
4.3.3.2. Il paragrafo LC.....	65
 CAPITOLO 5. FASE 2: MAPPING DEL LIVELLO INVENTARIALE	 67
5.1. Schemi e modelli di riferimento.....	68
5.1.1. CIDOC Conceptual Reference Model (CRM)	68
5.1.2. Dublin Core (DC)	70
5.1.3. FRBR-aligned Bibliographic Ontology (FaBiO)	71
5.1.4. Altri schemi e modelli secondari	74
5.2. Mapping del livello inventariale	75
5.2.1. Mapping dei paragrafi in comune.....	76
5.2.1.1. Il paragrafo DT	76
5.2.1.1.1. Il campo DTZ.....	77
5.2.1.1.2. Il campo DTS	78

5.2.1.1.3. Il campo DTM.....	78
5.2.1.2. Il paragrafo MT.....	79
5.2.1.2.1. Il campo MTX.....	80
5.2.1.2.2. Il campo MTC.....	80
5.2.1.2.3. Il campo MIS.....	80
5.2.1.3. Il paragrafo TU.....	82
5.2.1.4. Sintesi del mapping.....	83
5.2.1.5. Esempi di utilizzo.....	85
5.2.2. Mapping dei paragrafi esclusivi.....	88
5.2.2.1. Il paragrafo SG.....	88
5.2.2.1.1. Il campo SGT.....	88
5.2.2.1.2. Il campo SGL.....	88
5.2.2.2. Il paragrafo UB.....	89
5.2.2.3. Il paragrafo AU.....	91
5.2.2.4. Sintesi del mapping.....	91
5.2.2.5. Esempi di utilizzo.....	92
5.2.3. Mapping dei campi esclusivi di paragrafi in comune.....	94
5.2.3.1. Il paragrafo OG.....	94
5.2.3.1.1. Il campo OGT.....	94
5.2.3.1.2. Il campo QNT.....	95
5.2.3.2. Il paragrafo LC.....	95
5.2.3.3. Sintesi del mapping.....	97
5.2.3.4. Esempi di utilizzo.....	98

CAPITOLO 6. FASE 3: REALIZZAZIONE DELLE ONTOLOGIE CICDO E CICDO-F..... 101

6.1. Il software *Protégé*..... 102

6.2. Costruzione delle ontologie..... 103

6.2.1. Ontologia CICDO.....	104
6.2.1.1. Definizione delle classi.....	104
6.2.1.2. Definizione delle proprietà.....	106
6.2.1.2.1. Object property.....	106

6.2.1.2.2. Data property	111
6.2.2. Ontologia CICDO-F	111
6.2.2.1. Definizione delle classi	111
6.2.2.2. Definizione delle proprietà	113
6.2.2.2.1. Object property	113
6.2.2.2.2. Data property	117

CAPITOLO 7. UNA PROPOSTA DI APPLICAZIONE PER LA CONVERSIONE DI SCHEDE ICCD IN DOCUMENTI RDF..... 119

7.1. XSLT come tecnologia di conversione	120
7.1.1. Modello di trasformazione	120
7.1.2. Elementi sintattici di interesse	122
7.2. Architettura e strategie di sviluppo.....	123
7.2.1. Struttura del processo di conversione	123
7.2.2. Architettura generale	124
7.2.3. Il template XSLT	125
7.2.4. Un esempio di conversione	126
7.2.5. Osservazioni	128

CONCLUSIONI E SVILUPPI FUTURI 131

BIBLIOGRAFIA E SITOGRAFIA 135

INTRODUZIONE

Il termine *Web semantico* viene coniato per la prima volta agli inizi degli anni 2000 dal suo ideatore, Tim Berners-Lee, per indicare il nuovo modo di concepire il Web come un mondo in cui le informazioni contenute all'interno dei documenti hanno un significato associato ben preciso, che le macchine sono in grado di elaborare e "interpretare". Nel celebre articolo scientifico *The Semantic Web*, pubblicato dallo stesso Berners-Lee all'interno della rivista *Scientific American*, egli ipotizza e descrive uno scenario futuro ideale in cui "macchine e umani lavorano in cooperazione" per raggiungere determinati obiettivi. Con l'introduzione delle idee riportate in questo articolo cambia completamente il modo di vedere il Web, che ora non è più considerato un contenitore di soli documenti collegati fra loro, ma anche un enorme *repository* di informazioni collegate fra loro e con un significato ben definito. E' la nascita del cosiddetto *Web of data*.

Il Web semantico non rappresenta un nuovo tipo di Web, ma un'estensione del Web "tradizionale", che si propone essenzialmente di introdurre al suo interno una nuova componente ritenuta ormai essenziale: la *semantica* associata ai dati, esplicitata formalmente attraverso l'utilizzo di apposite "etichette", dette *metadati*.

In generale, le idee e le tecnologie finora introdotte per il Web semantico rappresentano una grande opportunità, che, se sfruttata adeguatamente, può favorire enormemente la diffusione della conoscenza a livello globale, facilitando la fruizione e l'accesso alle informazioni pubblicate. Per quanto questo scenario rappresenti di certo qualcosa di intrigante per il futuro del Web, sfortunatamente al momento l'avanzamento dei "lavori" è ancora a uno stato arretrato. Per questo motivo, spesso ci si riferisce alle idee del Web semantico con il termine di *visione*.

Il lavoro descritto all'interno della trattazione corrente nasce proprio dalle nuove opportunità emerse con la diffusione delle idee del Web semantico, e dalla necessità di sfruttare queste tecnologie per rappresentare in un nuovo formato le informazioni relative ai beni del patrimonio culturale. Diverse istituzioni della cultura, infatti, hanno

“colto la palla al balzo” con lo scopo di trarre beneficio da tali tecnologie per incrementare il valore dei propri dati. Sono nati così i cosiddetti portali informativi semantici, che raccolgono e presentano, in maniera centralizzata, le informazioni culturali sfruttando le tecnologie del Web semantico. E’ nel contesto di queste tecnologie applicate al dominio dei beni culturali che si sviluppa l’intero lavoro descritto nella trattazione corrente. L’ontologia proposta, infatti, si pone come obiettivo quello di rappresentare, in un nuovo formato, le informazioni contenute all’interno delle schede di catalogo normate dagli standard catalografici dell’ICCD, basandosi su schemi e ontologie già pubblicati nel Web semantico, al fine di garantire il riutilizzo e la fruizione di queste informazioni da parte di chiunque.

Prima di descrivere l’ontologia sviluppata e illustrare le fasi che hanno caratterizzato il processo di realizzazione di tale modello, si fornirà una panoramica sul mondo del Web semantico, analizzando dapprima l’evoluzione del Web dagli albori ai giorni nostri, al fine di comprendere le motivazioni e identificare i problemi del Web attuale da cui è nata la necessità di un “nuovo” tipo di Web. Si passerà successivamente a discutere lo stato dell’arte delle tecnologie sviluppate finora, illustrando l’architettura generale del Web semantico. Verrà infine descritto brevemente il progetto *Linked Open Data* proposto da Berners-Lee, e i principi di base per pubblicare e collegare fra loro le informazioni nel Web.

Conclusa questa panoramica, si passerà ad illustrare le caratteristiche dei dati relativi ai beni culturali, e a come questi sono pubblicati e utilizzati sia nel Web “tradizionale”, sia attraverso le tecnologie del Web semantico, descrivendo in questo ultimo caso il modello concettuale proposto per permettere alle istituzioni culturali e alle organizzazioni coinvolte nel mondo della cultura di pubblicare, in maniera standard, condivisa e interoperabile, le informazioni relative ai beni.

Una volta introdotti questi concetti e snocciolati i diversi aspetti, si passerà a descrivere in maniera dettagliata il lavoro di costruzione dell’ontologia proposta. Verrà fornita innanzitutto una breve panoramica del profilo dell’azienda che ha commissionato il lavoro, *3D Informatica*, per poi passare a illustrare gli obiettivi e le finalità che la stessa azienda si è prefissata con la realizzazione di questo progetto. Successivamente si

descrivono le tre fasi principali in cui si è articolato l'intero lavoro, riassunte nei punti seguenti:

1. *analisi del dominio*, per comprendere il sistema di compilazione delle schede di catalogo e analizzare le normative ICCD di interesse;
2. *definizione dei concetti ontologici*, per definire il nuovo formato di rappresentazione dei dati sulla base dei vincoli stabiliti dagli standard catalografici ICCD, attraverso l'utilizzo di schemi e modelli di riferimento già presenti nel Web semantico;
3. *costruzione dell'ontologia*, attraverso la definizione di classi di entità e delle loro relazioni, sulla base dei concetti ontologici modellati alla fase precedente.

Conclusa la trattazione del lavoro svolto, si descrivono brevemente le modalità di sviluppo che hanno caratterizzato il processo di realizzazione di un'apposita applicazione per la conversione di schede di catalogo in documenti con formato interpretabile dalle tecnologie del Web semantico, utilizzando l'ontologia costruita come punto di riferimento per la conversione.

CAPITOLO 1

IL WEB SEMANTICO: UNA PANORAMICA GENERALE

Per introdurre correttamente il mondo del *Web semantico* e comprenderne i principi alla base e le tecnologie utilizzate è necessario anzitutto fare un passo indietro, agli albori del *World Wide Web* (WWW). Vedremo quindi come questi è nato e si è evoluto nel corso del tempo, e come i limiti presenti nel Web dei giorni nostri abbiano portato alla necessità di introdurre un nuovo tipo di Web, il Web semantico per l'appunto. Successivamente verrà fornita una panoramica generale dello stato dell'arte del Web semantico, analizzandone gli obiettivi, la visione corrente e le tecnologie finora sviluppate.

1.1. Il World Wide Web

1.1.1. Nascita del Web

La data di nascita del World Wide Web è tradizionalmente riferita al *6 agosto 1991*, giorno in cui l'informatico britannico *Tim Berners-Lee* pubblicò in rete la prima pagina web. L'idea di creare un "contenitore virtuale" di informazioni multimediali e condivise venne però concepita due anni prima, nel 1989, presso il CERN¹ di Ginevra.

¹ *Conseil Européen pour la Recherche Nucléaire.*

Qui Berners-Lee presentò al proprio supervisore il documento “*Information Management: A Proposal*”, contenente alcune delle idee di base del progetto che avrebbero poi costituito le fondamenta del World Wide Web. Tra queste vi era la definizione del linguaggio di markup HTML² per la rappresentazione standard dei documenti, e la formalizzazione del protocollo di rete HTTP³/1.0 per la trasmissione di informazioni sul Web.

L’obiettivo principale del progetto di Berners-Lee, ideato insieme al collega e ricercatore belga *Robert Cailliau*, era quello di realizzare un software che permettesse la condivisione di documentazione scientifica in formato elettronico, indipendentemente dalla piattaforma utilizzata, al fine di facilitare la comunicazione e lo scambio di informazioni tra i diversi ricercatori del CERN.

Nei primi anni di vita, il Web fu utilizzato esclusivamente dalla comunità scientifica per tali scopi. In data 30 aprile 1993, il CERN decise di mettere il Web a disposizione del pubblico, rinunciando a qualsiasi diritto d’autore. In pochi anni, esso divenne la piattaforma più diffusa al mondo per l’invio e la ricezione di dati su Internet. Era l’inizio della cosiddetta *era del Web*.

1.1.2. Evoluzione del Web

Il World Wide Web rappresenta tuttora un mondo in continua evoluzione ed espansione. Questa crescita rapida, frenetica e talvolta incontrollata ha modificato profondamente nel corso degli anni le dinamiche socio-economiche e politiche che regolano la vita di tutti i giorni, sia per quanto riguarda le singole persone, sia per quanto concerne le aziende e le organizzazioni in generale. Si pensi ad esempio alla possibilità di condividere, anche in tempo reale, informazioni personali e momenti

² HyperText Markup Language.

³ HyperText Transfer Protocol.

della nostra vita sui canali di comunicazione social, o a come gli strumenti offerti dalla piattaforma Web siano ormai indispensabili per le aziende per pianificare e condurre le proprie strategie di mercato. La pervasività del Web nelle nostre vite è ormai un fattore che non possiamo più ignorare.

Evitando comunque di addentrarci in discorsi di tipo sociologico, l'evoluzione del Web può essere descritta dalla presenza di due tipologie di Web: il *Web statico* e il *Web dinamico*.

1.1.2.1. Il Web statico

Con i termini *Web statico* o *Web 1.0* si fa riferimento alle prime fasi di vita del Web. Formalmente, il Web statico rappresenta il paradigma di progettazione e pubblicazione sul Web ideato inizialmente da Berners-Lee e Cailliau e formalizzato nel loro progetto.

In origine, il Web era un semplice contenitore di documenti HTML collegati tra loro da collegamenti ipertestuali (i cosiddetti “link”) e pubblicati dagli autori dei siti. Gli utenti potevano soltanto consultare questo grande deposito di informazioni, senza avere alcuna possibilità di interagire con esse o di condividere contenuti personali. Da questo punto di vista, quindi, un utente era considerato un “lettore”, un semplice fruitore passivo dei contenuti.

Questo scenario relativo alle prime fasi di vita del Web deriva in realtà dalle limitazioni imposte dalle tecnologie presenti negli anni '90. I contenuti pubblicati dagli autori, siano essi testuali o audiovisivi, erano memorizzati su una macchina e trasmessi all'utente senza manipolazione alcuna da parte di applicazioni lato server. La tecnologia del tempo, infatti, rendeva poco conveniente – in termini di tempi di risposta del server – la presenza di software che potessero elaborare dinamicamente e “on the fly” i contenuti sulla base della richiesta dell'utente. Il termine “Web statico” si riferisce proprio a questo tipo di staticità dei siti web e dei relativi documenti HTML.

Ciò che a prima vista può sembrare una pesante limitazione se confrontata con gli attuali scenari del Web, in realtà la possibilità di trasmettere documenti HTML senza

alcuna elaborazione dei contenuti da parte del server ha rappresentato, agli albori del Web, un punto di forza per la sua rapida diffusione. Infatti:

- gli autori dei documenti non erano coinvolti in attività di programmazione di applicazioni, in quanto l'unica loro preoccupazione era quella di progettare e realizzare contenuti (nella forma di documenti HTML);
- ospitare un sito web statico richiedeva al server un insieme molto ristretto di risorse.

Per questo motivo, nonostante la diffusione di nuove tecnologie e l'evoluzione di quelle già presenti abbiano permesso di superare ampiamente tali limiti, i siti web statici vengono ancora in parte utilizzati per pubblicazioni di piccola entità o caratterizzate da modifiche poco frequenti dei contenuti e che non richiedano interazioni con l'utente.

1.1.2.2. Il Web dinamico

Con il termine *Web dinamico* si fa riferimento a un nuovo paradigma di pubblicazione e manipolazione dei contenuti, contrapposto al paradigma del Web statico.

Nel Web dinamico cambia radicalmente il modo con cui gli utenti si approcciano ai contenuti dei siti: non più una mera consultazione passiva delle informazioni contenute nei documenti, ma anche interazione attiva con essi e possibilità di contribuire alla pubblicazione di nuovi contenuti. Da questo punto di vista non vi è più distinzione tra "autore" che pubblica documenti e "lettore" che li consulta: il ruolo centrale e unico è ora quello dell'*utente*.

La fase di evoluzione corrispondente alla diffusione del Web dinamico è talvolta indicata con la locuzione *Web 2.0* per enfatizzare il cambiamento rispetto ai primi anni di vita del World Wide Web. Questo nuovo approccio verso i contenuti del Web è in realtà più di tipo filosofico che tecnologico, in quanto ora il focus è posto essenzialmente sulla possibilità di condividere informazioni con altri utenti, portando

la rete Internet ad avere un ruolo più “sociale”. E’ proprio durante questa fase che si diffondono le applicazioni caratteristiche del Web attuale, come ad esempio:

- i blog, i forum e le chat, utilizzati per condividere e scambiarsi informazioni e opinioni di qualsiasi tipo;
- le *wiki*, utilizzate per la diffusione delle informazioni e della conoscenza in generale, e all’interno delle quali ogni utente può dare il suo contributo;
- le piattaforme per la condivisione e la fruizione di elementi multimediali;
- i *Content Management System*, per la gestione semplificata dei contenuti dei siti web;
- i social network.

In generale, la caratteristica chiave delle applicazioni del Web dinamico è il modo con cui i contenuti sono presentati all’utente. A differenza del Web statico, in cui i documenti HTML vengono trasmessi all’utente così come sono, nel Web 2.0 i contenuti sono generati dinamicamente dalle applicazioni lato server. Nello specifico, al momento di una richiesta al server da parte dell’utente, le informazioni vengono estratte da una sorgente dati - tipicamente un database relazionale – in base al tipo di richiesta effettuata, e successivamente sono inserite all’interno di un documento HTML. Terminata l’elaborazione, il documento viene trasmesso all’utente.

Generalmente parlando, il Web dinamico ha cambiato profondamente il modo con cui le persone si avvicinano alla rete e condividono informazioni con altri utenti. Con la sua diffusione, il Web è diventato un vero e proprio strumento di tipo partecipativo, una piattaforma in cui tutti possono pubblicare e condividere liberamente informazioni e dati.

1.1.2.3. Verso il Web semantico

Con il termine *Web semantico* si fa riferimento alla nuova tipologia di Web che sta prendendo piede negli ultimi anni come “estensione” del Web attuale.

Nel Web semantico cambia nuovamente il modo con cui il Web è concepito: mentre nel Web dinamico il focus è posto sull'utente e sulle nuove modalità di pubblicazione e condivisione dei contenuti da parte di esso rispetto al Web statico, nel Web semantico l'attenzione si sposta sul concetto di *informazione* presente all'interno dei documenti del Web e sulla possibilità di collegare *semanticamente* tra loro le informazioni, in maniera più o meno analoga a come già avviene per i documenti.

Tutti gli aspetti e i principi relativi al Web semantico saranno approfonditi a breve. Prima di tutto, però, ci si soffermerà ad illustrare i problemi e i limiti del Web attuale, al fine di comprendere le motivazioni che hanno portato alla necessità di introdurre questo nuovo tipo di Web.

1.2. Problemi del Web attuale

Abbiamo visto come la diffusione del “modello” del Web dinamico abbia permesso di superare i limiti presenti in un Web di tipo statico, tuttavia, con il passare del tempo, anche nel Web attuale sono emersi diversi problemi, riconducibili nell'insieme a un'unica, grande causa: la pubblicazione e la produzione esponenziale di informazioni da parte degli utenti. Infatti, mentre durante i primi anni di vita del Web la quantità di informazioni caricate era generalmente bassa, con l'avvento del Web dinamico e con la possibilità data a chiunque di pubblicare contenuti si è assistito progressivamente a un aumento più che esponenziale dei dati immessi sulla rete.

Per comprendere come questa crescita sia stata – ed è tuttora – incredibilmente enorme e significativa, è sufficiente osservare i dati in *Figura 1*: l'immagine mostra uno “snapshot” della quantità di informazioni prodotte in media dai principali servizi della rete nell'arco di 60 secondi.

Il principale problema derivato dalla presenza di questa incredibile mole di dati è la maggiore difficoltà da parte degli utenti di reperire le informazioni e i contenuti

desiderati. Tale problema è molto sentito in particolare nell'utilizzo dei motori di ricerca.



Figura 1. Il Web in 60 secondi (anno 2014)

1.2.1. Motori di ricerca

Nell'ambito del World Wide Web, un motore di ricerca è un sistema automatico di *information retrieval* il cui scopo è quello di aiutare un utente a reperire le informazioni che desidera all'interno di un sistema informatico. Una ricerca è effettuata dall'utente attraverso un'apposita interfaccia, e i risultati sono solitamente restituiti nella forma di lista contenente collegamenti ipertestuali a documenti indicizzati all'interno del database del sistema.

I motori di ricerca rappresentano al momento l'unico strumento attraverso il quale un utente può cercare e reperire informazioni all'interno di più siti web.

1.2.1.1. *Information retrieval*

In generale, lo scopo dell'information retrieval è soddisfare il cosiddetto *bisogno informativo* di un utente, ovvero garantire che i risultati restituiti da una ricerca siano rilevanti per la richiesta effettuata dall'utente. Una richiesta è tradotta nella digitazione all'interno di un campo di input di un'espressione contenente una o più parole chiave, detta *query*.

L'aumento della quantità di informazioni presenti sul Web è un problema che ha condizionato fortemente i motori di ricerca. Gli algoritmi di information retrieval implementati dai principali servizi di ricerca sono stati potenziati e adeguati nel tempo per continuare a garantire una certa efficacia delle ricerche effettuate (in termini di soddisfacimento del bisogno informativo di un utente). Nonostante ciò, proprio perché il numero di documenti in cui cercare le informazioni è diventato sempre più grande, e i contenuti presenti sono contestualmente eterogenei fra loro, non sempre i risultati ottenuti sono pertinenti alle richieste dell'utente.

In generale, per quantificare la *bontà* di una ricerca, in termini di rilevanza dei risultati restituiti rispetto al bisogno informativo reale di un utente, si utilizzano due misure:

- *precision* (precisione), che indica la proporzione di documenti rilevanti fra tutti quelli recuperati da una ricerca;
- *recall* (recupero o richiamo), che indica la proporzione fra il numero di documenti rilevanti recuperati da una ricerca e il numero di tutti i documenti classificati come tali nel database del motore.

Precision e *recall* sono grandezze inversamente proporzionali fra loro: maggiore è la *precision* di una ricerca, minore sarà la *recall*, e viceversa. Ne consegue che realizzare motori di ricerca "perfetti" in grado di garantire *precision* e *recall* massimi è praticamente impossibile.

1.2.1.2. Limiti dei motori di ricerca

In generale, quindi, i problemi presenti intrinsecamente nei motori di ricerca impediscono spesso all'utente di ottenere le informazioni che desidera. Di seguito vengono brevemente discussi i principali limiti.

1. **Bontà di una ricerca.** Come si accennava poco sopra, realizzare motori di ricerca “perfetti” utilizzando le tecnologie attualmente esistenti è molto complesso. In relazione alla valutazione della bontà di una ricerca, possono capitare le seguenti situazioni “estreme”:
 - *recall* elevata e *precision* bassa, ovvero la ricerca estrae un gran numero di documenti, classificati rilevanti o non rilevanti nel database del sistema, ma solo una minima parte di questi risulta pertinente alla richiesta dell'utente;
 - *recall* bassa o nulla, ovvero una ricerca restituisce un numero molto ridotto di risultati o nessun risultato.
2. **Processo di ricerca.** I risultati restituiti dai motori di ricerca sono talvolta elenchi di migliaia di collegamenti a singoli documenti contenenti l'informazione cercata. L'utente non solo dovrà districarsi all'interno di questa lista per trovare i documenti più attinenti al suo bisogno informativo, ma dovrà anche ricercare “manualmente” l'informazione all'interno di ciascun documento selezionato. Da questo punto di vista, un processo di ricerca che coinvolge più richieste correlate fra loro può diventare molto lungo e complesso.
3. **Linguaggio e termini nella query di ricerca.** Un problema molto sentito nell'utilizzo dei motori di ricerca è dato dal linguaggio delle parole chiave contenute all'interno di una query. Se utilizziamo un'espressione con termini in lingua italiana per la nostra ricerca, il motore ci restituirà un insieme di risultati collegati a documenti contenenti quei termini. Eventuali documenti che riportano termini scritti in altre lingue riferiti allo stesso concetto non saranno estratti, in quanto un algoritmo di information retrieval non ha le

capacità di associare tra loro termini semanticamente correlati. Per lo stesso motivo, documenti contenenti termini sinonimi⁴ od omonimi⁵ delle parole chiave contenute nella query non saranno estratti.

- 4. Accesso alle informazioni.** Un motore di ricerca reperisce informazioni dai documenti in base ai contenuti dei siti indicizzati all'interno del proprio database. Per motivi di proprietà e di privacy, questi sistemi non hanno alcuna possibilità di accedere direttamente alle sorgenti dati da cui i server estraggono le informazioni generate dinamicamente nei documenti trasmessi all'utente. Analogamente, i risultati restituiti da una ricerca non sono facilmente accessibili da applicazioni esterne. Da questo punto di vista, quindi, i motori di ricerca possono essere considerati dei veri e propri sistemi isolati.

1.2.2. Web Service, *mash-up* e accesso alle informazioni

Un *Web Service* (servizio web) è un componente software progettato per supportare l'interoperabilità tra applicazioni web. Attraverso la definizione di un'interfaccia associata all'applicazione, un Web Service offre un insieme di operazioni che le applicazioni esterne possono attivare per compiere specifiche azioni, tra cui accedere alle informazioni del sistema.

I Web Service sono stati introdotti nel Web 2.0 allo scopo di permettere l'interazione tra applicazioni di sistemi diversi. Un caso particolare di utilizzo dei Web Service è dato dai cosiddetti *mash-up*. Un *mash-up* è un sito o un'applicazione che include dinamicamente informazioni provenienti da più sorgenti, allo scopo di creare un nuovo servizio.

I *mash-up* rappresentano una soluzione parziale ai problemi descritti al paragrafo

⁴ Due termini sono *sinonimi* fra loro se hanno lo stesso significato.

⁵ Due termini sono *omonimi* se sono scritti in maniera uguale ma presentano significati diversi.

precedente. A titolo d'esempio, si consideri il caso della prenotazione di un viaggio: il processo di prenotazione può coinvolgere diversi aspetti, tra cui la prenotazione del mezzo di trasporto per arrivare alla destinazione, la prenotazione dell'alloggio, la ricerca dei luoghi di interesse della meta visitata, e molto altro ancora. Per semplificare ricerche di questo tipo, negli ultimi anni sono nati diversi siti mash-up che, utilizzando le operazioni fornite da appositi Web Service, accedono alle informazioni di diverse sorgenti per combinarle fra loro e fornire un nuovo servizio che aiuti gli utenti a trovare più facilmente le informazioni che desidera relativamente a una sua necessità.

Nonostante i mash-up, e in generale i Web Service, rappresentino un passo avanti nella possibilità di accedere più facilmente a informazioni semanticamente correlate fra loro, il modo con cui tali informazioni sono accedute è ancora controllato e vincolato dalla specifica implementazione di un Web Service. In altre parole, non vi è alcuna possibilità di accedere direttamente ai dati originali per utilizzarli e manipolarli liberamente. Una possibile soluzione a questo problema potrebbe essere quella di impiegare tecnologie standard in modo tale che l'accesso alle informazioni avvenga in maniera libera e uniforme per tutte le applicazioni.

1.2.3. Contestualizzazione di un'informazione

Abbiamo visto che il Web nasce essenzialmente come un grande deposito di documenti consultabili dalle persone. Le informazioni contenute all'interno di tali documenti sono sempre state pensate per essere comprese esclusivamente dall'essere umano, quindi l'esperienza di navigazione di un utente all'interno del Web rappresenta il requisito fondamentale per permettere la comprensione dei contenuti e delle relazioni tra le informazioni. In tal senso, si dice che un essere umano è in grado di *contestualizzare*, ovvero di associare implicitamente un contesto a una determinata informazione.

A titolo d'esempio, si consideri il termine "pesca". Questo termine può essere associato sia al frutto del pesco, sia all'attività sportiva. Un utente umano che sta

navigando su un sito web di attrezzature sportive sa che il termine “pesca” si riferisce all’attività sportiva, e non al frutto. In altre parole, l’utente è in grado di associare implicitamente il significato corretto a un’informazione in base al contesto in cui essa si trova. Il termine “pesca” rappresenta in questo caso un tipico esempio di omonimia che l’utente umano è in grado di *disambiguare* facilmente.

La capacità di contestualizzare di una persona è utile anche per interpretare i risultati restituiti da un motore di ricerca: l’esperienza di navigazione, infatti, permette all’utente di contestualizzare correttamente le informazioni ottenute, accettando o scartando mentalmente i risultati in base al proprio bisogno informativo.

E per quanto riguarda le applicazioni? Ovviamente non è possibile fare lo stesso discorso: le applicazioni web, e in generale le macchine, non sono in grado di contestualizzare un’informazione. Questo problema rappresenta una delle cause principali per cui si è arrivati a definire un Web di tipo semantico, come vedremo a breve.

1.2.4. Collegamenti ipertestuali

I collegamenti ipertestuali hanno sempre rappresentato il vero punto di forza nella diffusione del Web su larga scala. La possibilità di “saltare” da un documento all’altro attraverso un singolo click rende più semplice all’utente la navigazione e la ricerca di informazioni all’interno dell’enorme quantità di documenti presenti sulla rete.

Nonostante ciò, anche i collegamenti ipertestuali presentano un limite significativo: mentre la loro capacità di collegare documenti è ormai affermata, non possiamo dire lo stesso per quanto riguarda la capacità di descrivere la *locazione* verso cui il collegamento porta. In altre parole, sarebbe utile se un collegamento ipertestuale non rimandasse solamente a una locazione, ma descrivesse il significato di quella locazione. Nel primo caso si parlerà di collegamento di tipo *sintattico*, nel secondo caso di collegamento con capacità *semantica*.

L'obiettivo di attribuire un significato alla destinazione di un link potrà essere raggiunto associando ad esempio un'etichetta a esso, in maniera tale che descriva esplicitamente la locazione da raggiungere.

1.2.5. Derivazione di conoscenza implicita

Finora si è parlato di informazioni esplicitamente contenute e “visibili” dagli utenti all'interno dei documenti web. Oltre a questo tipo di informazioni, molto spesso i documenti contengono informazioni implicite, non visibili, che l'utente può derivare mentalmente da quelle disponibili attraverso la sua capacità di contestualizzare. Si parla in questo caso di *derivazione di conoscenza implicita*.

Per quanto un utente umano possa essere abile a contestualizzare, egli non è in grado di ricavare tutta la conoscenza implicita nascosta all'interno dei documenti: le deduzioni logiche necessarie per derivare parte di queste informazioni possono essere effettuate esclusivamente dalle macchine attraverso l'utilizzo di tecnologie impiegate nel campo dell'intelligenza artificiale.

1.2.6. Problemi del Web attuale: un riepilogo

- 1. Motori di ricerca.** I motori di ricerca, nati per aiutare gli utenti a reperire le informazioni dai documenti web, presentano numerosi limiti e problemi che spesso impediscono all'utente di ottenere ciò che desidera.
- 2. Contestualizzazione delle informazioni.** Il Web è stato concepito come deposito di documenti consultabile esclusivamente da un utente umano, il quale è in grado di contestualizzare un'informazione grazie alla sua esperienza di navigazione, capacità che le macchine non possiedono.

3. **Accesso alle informazioni.** L'accesso diretto alle informazioni contenute all'interno delle sorgenti dati di un sito web è ancora controllato e vincolato.
4. **Semantica delle informazioni.** In generale, le applicazioni web non sono in grado né di attribuire un significato esplicito (semantica) alle informazioni, né di interpretare il contenuto dei documenti e i collegamenti tra i documenti.
5. **Derivazione di conoscenza implicita.** L'utente umano non sempre è in grado di derivare tutte le informazioni implicite contenute all'interno di un documento.

L'insieme di questi limiti mostra, nel complesso, quale elemento essenziale manca nel Web attuale: la **semantica** associata alle informazioni.

1.3. Stato dell'arte del Web semantico

1.3.1. Idee di base

In generale, il Web attuale può essere visto come un'enorme rete di documenti collegati fra loro da collegamenti ipertestuali sintattici. L'idea di base del Web semantico è che non solo i documenti dovrebbero essere collegati fra loro, ma anche le singole informazioni contenute all'interno di essi, attraverso la definizione di collegamenti ipertestuali di tipo semantico che attribuiscono un significato ai dati. Come detto in precedenza, la semantica delle informazioni e dei collegamenti tra le informazioni può essere fornita associando ad essi delle specifiche "etichette".

Secondo la definizione fornita dal suo ideatore, Tim Berners-Lee, il Web semantico rappresenta:

“un'estensione del Web corrente in cui le informazioni hanno un ben preciso significato e in cui computer e utenti lavorano in cooperazione”.

Da questa definizione possiamo evincere due caratteristiche fondamentali del Web semantico:

1. il Web semantico non si propone come un nuovo Web, ma come una evoluzione ed *estensione* di quello attuale in cui la componente semantica è aggiunta ai contenuti;
2. macchine e persone devono cooperare tra loro, nel senso che le applicazioni dovrebbero essere in grado di dare un supporto concreto agli esseri umani nell'esecuzione e nell'automazione di certe operazioni, a partire dalla comprensione del significato delle "etichette" attribuite alle informazioni contenute all'interno dei documenti web (in questo caso si parlerà di informazioni *machine-readable* o *machine-understandable*).

E' importante notare che quando si parla di "*informazioni comprensibili dalle applicazioni*" non si intende che le applicazioni sono in grado di comprendere il significato delle informazioni! Aggiungendo una componente semantica alle informazioni, le applicazioni diventano capaci di interpretarle a seconda di come è modellato il relativo dominio di riferimento.

Un obiettivo non contemplato nella definizione, ma che rappresenta uno dei punti centrali del Web semantico, è quello di incoraggiare le persone a pubblicare e rendere "aperte" le proprie informazioni sul Web, allo scopo di facilitarne la scoperta, lo scambio, il riutilizzo e la condivisione. Si parlerà in questo caso di *Linked Data*, argomento che verrà approfondito più avanti.

In generale, per garantire il raggiungimento di questi obiettivi, il Web semantico può essere realizzato introducendo sopra il Web attuale un *framework* di tecnologie web standard. Possiamo quindi definire formalmente il Web semantico come un *insieme di tecnologie standard che permette di realizzare il cosiddetto Web of Data*, ovvero una rete di informazioni collegate semanticamente fra loro e interpretabili da specifiche applicazioni.

1.3.2. La visione del Web semantico

Le opportunità attualmente offerte dal Web semantico sono numerose, tuttavia alcuni aspetti non sono stati ancora adeguatamente approfonditi. Il raggiungimento degli obiettivi individuati per realizzare il *Web of Data* potrebbe richiedere ancora molto tempo. Per questo motivo, più che di tecnologia, si parlerà di *visione* del Web semantico, che è possibile sintetizzare nei seguenti punti.

1. **Descrizione dei dati:** le informazioni del Web devono essere descritte formalmente attraverso specifiche tecnologie standard che permettano di attribuire ad esse un significato. La descrizione avviene attraverso l'utilizzo dei cosiddetti *metadati*, ovvero le "etichette" che descrivono i dati veri e propri.
2. **Interpretazione dei dati:** dati e metadati devono essere rappresentati in un formato che sia comprensibile e interpretabile da specifici programmi, detti *agenti semantici*.
3. **Inferenza delle informazioni:** gli agenti semantici devono essere in grado di eseguire alcune limitate *inferenze* sui dati per estrarre eventuale conoscenza implicita.

L'idea di base per realizzare un *Web of Data* è quella di definire innanzitutto un insieme di schemi strutturati che descrivano domini di informazioni, sia generici che specifici, attraverso l'utilizzo di tecnologie standard.

1.3.3. Architettura del Web semantico

L'architettura del Web semantico può essere rappresentata attraverso il cosiddetto *Semantic Web Stack* mostrato in [Figura 2](#). Questo modello illustra il framework di tecnologie standard necessarie a realizzare il *Web of Data*, e mostra come l'architettura sia costruita sopra il Web attuale (rappresentato dalla parte arancione dello stack).

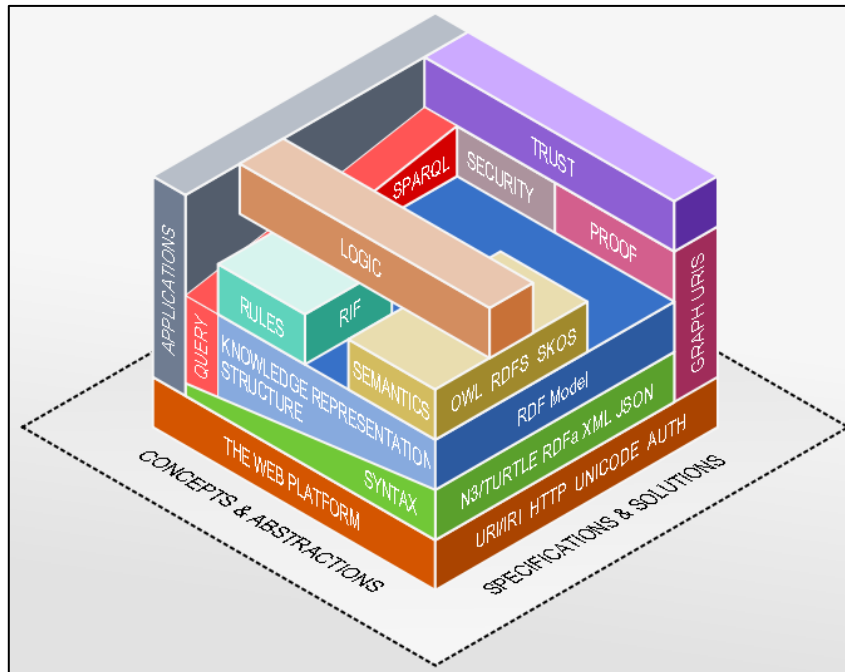


Figura 2. Il Semantic Web Stack

Il *Semantic Web Stack* è organizzato nei seguenti strati, a partire da quello più basso.

1. *Strato sintattico*: rappresenta le fondamenta del Web semantico.
2. *Strato semantico realizzato*: indica l'insieme delle tecnologie standard affermate necessarie a realizzare applicazioni semantiche.
3. *Strato semantico non realizzato*: individua l'insieme di tecnologie semantiche non ancora adeguatamente sviluppate.

Le tecnologie del primo e del secondo strato sono standardizzate e gestite dal *World Wide Web Consortium (W3C)*⁶. Le tecnologie del terzo livello, invece, rappresentano in generale idee che devono essere ancora implementate per realizzare concretamente la visione del Web semantico. Nei prossimi paragrafi si illustreranno solamente i primi due strati di tecnologie.

⁶ Il *World Wide Web Consortium* è l'organizzazione internazionale responsabile della standardizzazione dei linguaggi e dei protocolli del Web.

1.3.3.1. Lo strato sintattico

1.3.3.1.1. *Uniform Resource Identifier e Unicode*

Un *Uniform Resource Identifier* (**URI**) è una stringa che permette di identificare univocamente una risorsa generica. Gli URI rappresentano il mezzo fondamentale per individuare una risorsa nel Web semantico, così come gli URL rappresentano l'unico modo per identificare univocamente un documento all'interno del Web. In generale, una *risorsa* nel Web semantico può essere “qualsiasi cosa” che esprima un'informazione e sia referenziabile attraverso un URI.

Sullo stesso livello degli URI si ha *Unicode*, un sistema di codifica utilizzato per rappresentare e manipolare correttamente informazioni in lingue diverse. Questa componente dell'architettura è fondamentale per codificare traduzioni di un termine provenienti da idiomi diversi che ovviamente saranno collegate semanticamente fra loro.

1.3.3.1.2. *eXtensible Markup Language (XML)*

XML (*eXtensible Markup Language*) è il linguaggio di markup che permette di descrivere un insieme di dati all'interno di un documento attraverso l'uso di appositi tag di marcatura personalizzati. XML rappresenta la base sintattica attraverso la quale i domini di informazioni sono modellati al livello più alto dell'architettura del Web semantico.

Lo spazio dei nomi degli elementi contenuti in un documento XML costituisce un *vocabolario* per quel documento. Attraverso gli URI, un documento può referenziare vocabolari provenienti da diverse sorgenti XML. Se tali vocabolari contengono elementi con lo stesso nome, possono verificarsi problemi di ambiguità, risolvibili utilizzando i cosiddetti *namespace XML*. Un namespace fornisce un metodo per identificare univocamente gli elementi di un vocabolario, specificandone l'appartenenza attraverso un prefisso posto prima del nome dell'elemento stesso. Il

meccanismo dei namespace è fondamentale nel Web semantico per garantire il riutilizzo e l'integrazione di vocabolari provenienti da sorgenti diverse.

Il punto di forza di XML è dato dalla possibilità di produrre documenti che risultano comprensibili dall'uomo e interpretabili dalle applicazioni, tuttavia l'espressività del linguaggio è molto limitata e non adatta a definire in maniera corretta e completa la semantica di un dominio informativo. Serve quindi un framework specifico di tecnologie che permettano alle applicazioni di interpretare correttamente i dati.

1.3.3.2. Lo strato semantico

1.3.3.2.1. Resource Description Framework (RDF)

RDF (*Resource Description Framework*) è un framework di specifiche per la descrizione e la modellazione di informazioni relative a una risorsa. Una risorsa è descritta attraverso la definizione delle sue caratteristiche (o proprietà).

Il framework RDF si compone di due parti.

- 1. RDF model and syntax:** definisce le specifiche e la sintassi del modello per descrivere le risorse.
- 2. RDF Schema:** permette di definire vocabolari (o schemi) che specificano e vincolano la struttura delle risorse e le loro relazioni.

Il linguaggio utilizzato per modellare le informazioni associate alle risorse è basato sulla sintassi di XML.

RDF Data Model

La modellazione dei dati in RDF è effettuata attraverso la definizione di *statement*, ovvero asserzioni sulle risorse. Uno statement è formalmente rappresentato da una tripla nella forma

(soggetto, predicato, oggetto)

dove

- *soggetto* indica la risorsa,
- *predicato* indica una specifica proprietà o caratteristica della risorsa,
- *oggetto* indica il valore di tale proprietà.

Una proprietà permette quindi di collegare il soggetto della tripla con il relativo oggetto. Un oggetto può rappresentare a sua volta un soggetto, ovvero un'altra risorsa.

Le specifiche del modello RDF offrono un vocabolario di elementi predefiniti costituiti da classi e proprietà, utilizzabili per modellare le informazioni all'interno dei documenti RDF. Tale vocabolario può essere referenziato attraverso il meccanismo dei namespace XML discussi in precedenza. Tra gli elementi predefiniti, di fondamentale importanza è l'elemento *rdf:type*, che permette di definire il tipo di una risorsa, ovvero la sua appartenenza a una classe.

A titolo d'esempio, si consideri la seguente tripla RDF modellata secondo il formato RDF/XML:

```
<rdf:Description rdf:about="http://people.com/id/1375">
  <rdf:type rdf:resource="/myschema.rdf#Persona" />
  <s:Nome>Mario Rossi</s:Nome>
</rdf:Description>
```

La risorsa identificata dall'URI assoluto *http://people.com/id/1375* rappresenta una persona descritta da due proprietà:

1. *rdf:type*, a cui è associato come valore la risorsa identificata dall'URI relativo */myschema.rdf#Persona* per indicare che essa è istanza della classe *Persona* definita nello schema locale *myschema.rdf*;
2. *s:Nome*, a cui è associato come valore una stringa che indica il nome della persona.

Si noti l'utilizzo dei namespace per riferirsi agli elementi dei rispettivi vocabolari.

Un insieme di triple RDF può essere rappresentato da un grafo orientato etichettato, dove i nodi rappresentano soggetti e oggetti delle triple e gli archi individuano le rispettive proprietà. L'orientamento degli archi è sempre diretto verso l'oggetto. Poiché qualsiasi risorsa può essere descritta da una tripla, il Web semantico è spesso riferito con il nome di *Giant Global Graph* per indicare una rete universale di risorse collegate semanticamente fra loro.

Dal punto di vista pratico, le triple RDF possono essere serializzate in documenti utilizzando diversi formati. Oltre al più comune *RDF/XML* visto nell'esempio precedente, un formato ampiamente utilizzato è *Turtle*, grazie alla sua sintassi "leggera" e compatta.

RDF Schema

RDF Schema (abbreviato in **RDFS**) è un insieme di classi e proprietà che estende il vocabolario di elementi definiti nel modello di dati RDF al fine di renderlo più espressivo.

Mentre il modello RDF offre semplicemente un meccanismo per descrivere le informazioni, RDF Schema rappresenta la tecnologia che permette di definire la semantica vera e propria dei dati attraverso la specifica delle caratteristiche e delle relazioni di classi e proprietà, oltre alla modellazione delle gerarchie tra classi. In generale, quindi, RDFS permette di definire vocabolari RDF che specificano la struttura gerarchica e le proprietà delle risorse.

Fra i diversi vocabolari RDF rappresentati in RDFS, i più noti e utilizzati sono i seguenti:

- *Friend Of A Friend* (FOAF)
- *Dublin Core* (DC)
- *Schema.org*
- *Simple Knowledge Organization System* (SKOS)

1.3.3.2.2. *Web Ontology Language (OWL)*

OWL (*Web Ontology Language*) rappresenta una famiglia di linguaggi il cui scopo è quello di rappresentare la conoscenza di uno specifico dominio informativo attraverso la definizione e la realizzazione di **ontologie**. Formalmente, un'ontologia è una rappresentazione formale, condivisa ed esplicita della concettualizzazione di un certo dominio. In altre parole, realizzare un'ontologia significa creare un modello che descriva un dominio di informazioni e sia riusabile e condivisibile da altre persone e applicazioni.

Il più comune tipo di ontologia nel Web semantico presenta una tassonomia - definita attraverso schemi RDFS - e un insieme di regole di inferenza per estrarre conoscenza implicita dalle informazioni. OWL, quindi, estende ulteriormente l'espressività semantica di RDFS introducendo la modellazione di altri aspetti delle risorse, quali:

- le relazioni tra classi (es. disgiunzione, complemento ecc.);
- l'equivalenza tra classi;
- la cardinalità e i vincoli nei valori ammessi delle proprietà;
- i tipi di relazioni tra risorse (es. simmetrica, transitiva ecc.).

Per quanto riguarda la derivazione di informazioni implicite, il processo di inferenza è effettuato da appositi programmi chiamati *reasoner*, i quali derivano le informazioni basandosi su una rappresentazione delle logiche descrittive del modello.

1.3.3.2.3. *SPARQL Protocol and RDF Query Language*

SPARQL (acronimo ricorsivo di *SPARQL Protocol and RDF Query Language*) è un linguaggio di interrogazione utilizzato per estrarre informazioni modellate come triple RDF. Affinché sia possibile estrarre la conoscenza, le triple devono essere memorizzate in un database specifico, ad esempio un *triplestore*.

La sintassi di SPARQL è molto simile alla sintassi del linguaggio SQL utilizzato per le interrogazioni sui database relazionali.

1.3.4. Il progetto *Linked Open Data*

“*RAW, DATA, NOW!*”

Con questa celebre frase pronunciata nel 2009 durante una conferenza al TED⁷, Tim Berners-Lee esortava le persone, e in particolare gli enti pubblici, a pubblicare sul Web i propri dati per renderli disponibili a chiunque. Durante la stessa conferenza, Berners-Lee presentò *Linked Open Data (LOD)*, un progetto nato proprio allo scopo di incoraggiare le persone a rendere pubblici i dati e collegarli semanticamente tra loro (da cui la definizione di *linked data*).

In generale, affinché i dati siano *linked*, è necessario che vengano rispettate le seguenti regole:

1. utilizzare gli URI per identificare le risorse;
2. utilizzare URI basati sul protocollo HTTP in maniera tale che i dati siano referenziabili e reperibili sia dalle persone che dalle applicazioni semantiche;
3. pubblicare i dati in modelli RDF affinché siano interpretabili dalle macchine e interrogabili attraverso linguaggi come SPARQL;
4. includere riferimenti ad altre risorse esterne per migliorare la ricerca di informazioni collegate semanticamente fra loro.

Il grande obiettivo del progetto LOD è quello di estendere il Web attuale attraverso la pubblicazione di dataset aperti, riferiti a domini di informazioni sia di carattere generale che incentrati su un particolare tema. Questi dataset possono poi essere riutilizzati e referenziati in schemi e modelli RDF per evitare di dover ridefinire ogni volta concetti già espressi nelle ontologie e nei dataset pubblicati.

⁷ TED è un marchio di conferenze statunitensi, gestite dall'organizzazione privata non-profit *The Sapling Foundation*.

CAPITOLO 2

LINKED DATA RELATIVI AI BENI CULTURALI: PUBBLICAZIONE E UTILIZZO SUL WEB SEMANTICO

Con la diffusione delle idee alla base del progetto *Linked Open Data* si è assistito negli ultimi anni a una rapida ed estesa pubblicazione dei dati relativi ai domini informativi più disparati. Tra questi, il dominio di interesse a cui fa riferimento il progetto di tesi corrente è quello relativo ai beni del patrimonio culturale mondiale, indicato con il nome di *Cultural Heritage (Linked) Data*.

Attualmente, il lavoro di pubblicazione di dati culturali collegati sta coinvolgendo diverse istituzioni e diversi esperti del settore. Curatori di musei, bibliotecari, archivisti e ricercatori del mondo culturale, così come scienziati informatici e web designer, si sono mossi collettivamente verso questa direzione per portare a compimento un progetto lungo e di difficile realizzazione, ma potenzialmente significativo e utile. In generale, le informazioni relative ai beni del patrimonio culturale sono sintatticamente eterogenee, ricche di significato e collegabili fra loro. Tali informazioni sono prodotte in maniera distribuita dalle diverse istituzioni culturali, così come dagli enti pubblici e privati che operano nel settore.

In questo capitolo si fornirà una panoramica generale del mondo dei beni culturali che collettivamente costituiscono il patrimonio culturale umano. Successivamente vedremo come i dati culturali sono pubblicati e utilizzati nel Web 2.0, identificando i limiti delle soluzioni proposte e illustrando in contrapposizione un modello concettuale semantico, condiviso e generale che supporti la realizzazione di portali informativi semantici.

2.1. Il patrimonio culturale

Con il termine *patrimonio culturale* ci si riferisce a un insieme di cose di valore, dette *beni culturali*, che, per particolare rilievo storico, culturale o estetico, sono di interesse pubblico e costituiscono nello specifico la ricchezza di un luogo e della relativa popolazione. In altre parole, il patrimonio culturale rappresenta l'insieme di beni quali oggetti fisici, luoghi, tradizioni e conoscenze di una società che sono tramandati ed ereditati dal passato, mantenuti nel presente e preservati a beneficio delle generazioni future. Il termine *patrimonio* enfatizza il valore economico attribuito ai beni culturali che lo compongono in relazione alla loro artisticità e storicità.

I luoghi di cui i beni del patrimonio culturale costituiscono ricchezza possono essere paesi, città, nazioni o qualunque settore territoriale giuridicamente circoscritto. I beni culturali sono di interesse pubblico in quanto destinati alla fruizione collettiva, in altre parole tutti devono poter godere liberamente della visione del patrimonio e del sapere ad esso collegato (anche se, in realtà, non sempre questo è possibile).

I beni che costituiscono il patrimonio culturale possono essere classificati in tre aree distinte:

1. **beni culturali tangibili**, ovvero oggetti concreti come opere d'arte, libri, edifici e in generale oggetti fisici realizzati dall'uomo;
2. **beni culturali non tangibili**, ovvero fenomeni come tradizioni, conoscenze popolari, linguaggi, tecniche artigianali e in generale tutta la conoscenza tramandata oralmente;
3. **beni culturali naturali**, come ad esempio i paesaggi, le biodiversità e le geodiversità.

Il sapere collegato a un bene culturale può essere documentato in modi diversi a seconda del tipo:

1. i beni tangibili sono documentati principalmente attraverso l'utilizzo di metadati, ovvero informazioni che descrivono le proprietà e le caratteristiche del bene considerato;

2. i beni non tangibili e i beni naturali sono documentati attraverso l'utilizzo di descrizioni testuali, immagini, video e altri elementi multimediali simili.

I beni tangibili, così come la documentazione associata ad essi e alle altre tipologie di beni, sono raccolti in diversi spazi fisici e preservati da specifici enti pubblici o privati. I principali luoghi preposti alla loro conservazione sono le librerie, gli archivi e i musei, questi ultimi specializzati in specifiche aree come l'arte, l'archeologia, le storie naturali, la scienza e così via. Un ristretto insieme di beni culturali è mantenuto e conservato da associazioni culturali e da cittadini privati. L'insieme di beni culturali conservato da un'istituzione, da un ente o da un privato è detto *raccolta*.

L'attività di acquisizione e raccolta delle informazioni relative ai beni del patrimonio culturale è detta *catalogazione*. La catalogazione è fondamentale per la conoscenza, la diffusione e il mantenimento del sapere legato ai beni che compongono il patrimonio culturale. Questo processo è strutturato in un insieme di attività distinte che comprendono la schedatura, l'inventario, il censimento e la creazione di apposite basi di dati, accompagnate da un'analisi storica, artistica ed estetica del bene in esame. L'attività di catalogazione è strettamente correlata al rispetto dei principi di tutela del patrimonio culturale.

2.2. Pubblicazione e utilizzo sul Web 2.0

Fin dalla diffusione del Web 2.0 librerie, archivi e musei si sono adoperati per realizzare siti web adibiti a luoghi di esposizione dei beni appartenenti alle proprie raccolte. Tali esposizioni "virtuali" sono solitamente accompagnate da dettagliate descrizioni testuali, da immagini ed eventualmente da registrazioni audio e video relative al bene descritto.

Il problema principale di questi siti web deriva essenzialmente da uno dei limiti del Web 2.0 discussi nel capitolo precedente: un utente interessato a cercare informazioni relative a uno specifico bene culturale deve sapere dove è esposto e in quale raccolta

si trova. Inoltre, ciascun sito web utilizza interfacce proprie con differenti sistemi di ricerca, che l'utente deve imparare a utilizzare. Il processo di ricerca dell'informazione desiderata potrebbe quindi essere molto complesso e tedioso. Per superare questi problemi sono nati i cosiddetti *portali informativi culturali*.

2.2.1. Portali informativi culturali

Un portale informativo culturale è un sito *mash-up* che raccoglie informazioni sui beni culturali da diverse sorgenti dati, li combina e li presenta in maniera uniforme all'utente. Lo scopo di questi portali è tentare di offrire un servizio centralizzato globale che permetta agli utenti di reperire in maniera più rapida, semplice e organica le informazioni relative ai beni culturali conservati nelle diverse raccolte presenti sul territorio nazionale e internazionale. La ricerca è effettuata dal sistema sulle sorgenti dati delle istituzioni che hanno dato il permesso ad accedere alle proprie informazioni.

Si distinguono due tipologie di portali informativi:

- **portali tematici**, che esibiscono informazioni specifiche relative a un tema ben definito del patrimonio culturale.
- **portali generici**, che forniscono sistemi di ricerca, navigazione e visualizzazione di grandi moli di informazioni culturali generiche derivate da diverse sorgenti eterogenee.

La nostra attenzione sarà rivolta esclusivamente ai portali del secondo tipo. Tra questi, merita una doverosa citazione *Europeana*, una biblioteca digitale che riunisce milioni di contributi già digitalizzati in 30 lingue diverse provenienti da diverse istituzioni dei paesi membri dell'Unione Europea.

La ricerca delle informazioni all'interno di questi tipi di portali è favorita dalla presenza di interfacce *user-friendly* che offrono un limitato numero di filtri, detti *facet*, per restringere l'elenco dei risultati in base alle caratteristiche del bene culturale da cercare.

2.2.2. Accesso alle informazioni

I database a cui i portali informativi accedono per reperire le informazioni e combinarle fra loro sono strutture di proprietà delle relative istituzioni culturali. Ciascuna di esse modella le proprie basi di dati utilizzando specifici schemi e sistemi, e controlla le modalità di accesso ai contenuti attraverso procedure ad hoc. Per raggiungere lo scopo, i sistemi alla base dei portali informativi devono combinare e aggregare fra loro i dati presenti localmente su ciascuna sorgente per presentarli all'utente in maniera uniforme e unificata. Il processo di combinazione delle informazioni è detto *data integration*.

Esistono due strategie principali per realizzare la data integration. Le differenze tra i due tipi di approcci riguardano la fase del processo in cui l'integrazione avviene.

2.2.2.1. Federated search

Una prima strategia di integrazione dei dati è la cosiddetta *federated search*, conosciuta anche con il nome di *meta-ricerca*. Questa strategia consiste nell'effettuare l'integrazione dei dati durante la fase di elaborazione di una query. Nel dettaglio:

1. il sistema riceve una query e la invia a ogni servizio partecipante, trasformandola in base al sistema e allo schema utilizzati dal database di ciascuno di essi;
2. i risultati prodotti da ciascun servizio sono raccolti e aggregati;
3. l'insieme di risultati è presentato all'utente.

Il vantaggio principale nell'utilizzo di questa strategia è che i diversi sistemi di database coinvolti continuano ad operare in maniera indipendente. In altre parole, le istituzioni partecipanti devono soltanto mettersi d'accordo sul protocollo utilizzato per le interrogazioni, per il resto possono continuare a gestire senza alcun vincolo le proprie basi di dati. La logica di integrazione è quindi a carico del sistema di ricerca del portale.

Data la sua semplicità e flessibilità, la federated search è un approccio ampiamente sfruttato dai sistemi dei diversi portali informativi, non solo in ambito di beni culturali. La stessa *Europeana* utilizza questa strategia per combinare e presentare informazioni estratte dalle diverse raccolte di beni delle istituzioni partecipanti. In particolare, la federated search è utile quando si cercano tipi di oggetti simili descritti da metadati in comune, come ad esempio le raccolte di libri forniti dalle librerie.

Nonostante l'ampio utilizzo, la federated search presenta comunque un limite abbastanza significativo: poiché ciascun sistema di database elabora le query in arrivo in maniera totalmente indipendente dagli altri sistemi, le dipendenze globali, in termini di associazioni semantiche esistenti tra gli oggetti delle diverse raccolte, sono difficilmente identificabili. Questo limite viola in un certo senso l'obiettivo principale dei *Linked Data* secondo cui i dati devono essere aperti e collegati semanticamente fra loro. Il paradigma basato sulla federated search non può quindi essere utilizzato come sistema di ricerca nei portali informativi semantici.

2.2.2.2. *Data warehousing*

Un approccio alternativo alla federated search è la strategia basata sul *data warehousing*. Questa strategia consiste nell'effettuare l'integrazione dei dati in maniera statica, memorizzando le informazioni in un repository globale chiamato *data warehouse*. Nel dettaglio:

1. il sistema preleva e combina i dati dei diversi servizi locali partecipanti;
2. i dati combinati sono memorizzati nel *data warehouse*;
3. tutte le interrogazioni sono effettuate su di esso.

L'approccio basato sul data warehousing permette di superare il limite principale dei sistemi di ricerca basati su federated search attraverso l'utilizzo di ontologie e modelli concettuali semantici condivisi sui dati immagazzinati nel data warehouse. Questo approccio, tuttavia, richiede che il sistema esegua periodicamente un processo separato per la generazione dei dati nel repository, al fine di mantenerlo allineato con le informazioni presenti nei database locali. Ciò richiede quindi un maggior

coinvolgimento e coordinamento da parte delle istituzioni partecipanti rispetto ai sistemi basati su federated search.

2.3. Pubblicazione e utilizzo sul Web semantico: un modello concettuale condiviso

I principi e le tecnologie alla base del Web semantico illustrati nel capitolo precedente possono supportare la realizzazione dei cosiddetti *portali informativi semantici*, al fine di superare le limitazioni presenti nei portali del Web 2.0 e nei relativi sistemi di ricerca discussi finora. Per raggiungere questo scopo è necessario definire un modello concettuale ontologico, condiviso dalle diverse istituzioni culturali, in modo tale che le informazioni relative ai beni culturali possano essere pubblicate e utilizzate seguendo precisi standard.

2.3.1. Caratteristiche dei dati e interoperabilità

Per definire correttamente un modello concettuale è opportuno innanzitutto analizzare le caratteristiche peculiari dei contenuti⁸ associati ai beni culturali.

- **Multi-formato.** I contenuti sono presenti in varie forme: documenti testuali, immagini, registrazioni audio e video ecc.
- **Multi-argomento.** I contenuti possono riferirsi a diversi argomenti: arte, storia, tradizioni ecc.
- **Multi-lingua.** I contenuti sono disponibili in varie lingue.
- **Multi-cultura.** I contenuti sono collegati e interpretati in maniera diversa in base alla cultura e alle tradizioni di un paese, di una città o di una nazione.

⁸ Qui inteso come “insieme di dati”.

- **Multi-target.** I contenuti possono essere rivolti a diversi tipi di persone: esperti e non esperti del settore, giovani e anziani ecc.
- **Multi-istituzione.** I contenuti sono raccolti, mantenuti e pubblicati da diverse istituzioni e associazioni che operano nel settore culturale, così come da cittadini privati.

Un problema comune derivante da queste caratteristiche e dall'eterogeneità dei contenuti è la difficoltà di rendere le informazioni *mutuamente interoperabili*. In generale, un insieme di contenuti è **interoperabile** se i dati presenti possono essere ricercati, collegati semanticamente e presentati all'utente in maniera organica e standardizzata, superando i vincoli imposti dalle implementazioni ad hoc dei servizi delle varie istituzioni partecipanti. L'interoperabilità del modello concettuale dovrà essere garantita sia a livello sintattico, attraverso il collegamento di diversi formati di dati e di diverse raccolte, sia a livello semantico, tramite l'utilizzo di schemi di metadati comuni e di vocabolari condivisi per la descrizione delle informazioni.

La pubblicazione sul Web dei dati relativi ai beni culturali non rappresenta soltanto un problema tecnologico. Infatti:

- una buona parte dei contenuti potrebbe essere protetta da licenze e/o diritti d'autore;
- per questioni di privacy, non tutte le informazioni disponibili possono essere rese liberamente pubbliche;
- le informazioni pubblicate potrebbero essere non attendibili.

Poiché difficilmente è possibile ignorare o evitare questi problemi, l'obiettivo principale diventa quello di massimizzare la quantità dei contenuti culturali pubblicati e condivisi come Linked Data, al fine di diffondere il maggior volume di conoscenza possibile.

2.3.2. Modello concettuale per la descrizione dei dati

Il modello concettuale semantico definito per il dominio relativo ai beni del patrimonio culturale non deve soltanto supportare l'interoperabilità tra i dati, ma anche continuare a garantire la creazione indipendente e distribuita dei contenuti da parte degli autori delle istituzioni partecipanti al progetto.

Per raggiungere questo scopo, il modello proposto (vedi *Figura 3*) definisce una *infrastruttura ontologica* costituita da ontologie condivise e schemi di metadati comuni per la descrizione dei contenuti culturali. Il collegamento semantico tra le informazioni è così attuabile utilizzando i concetti e gli schemi definiti nelle ontologie condivise. L'aspetto fondamentale di questo approccio è che permette di garantire implicitamente una sorta di collaborazione e coordinamento tra gli autori che utilizzano tale modello per descrivere le informazioni, preservando comunque la possibilità di continuare a creare in maniera indipendente e distribuita i contenuti.

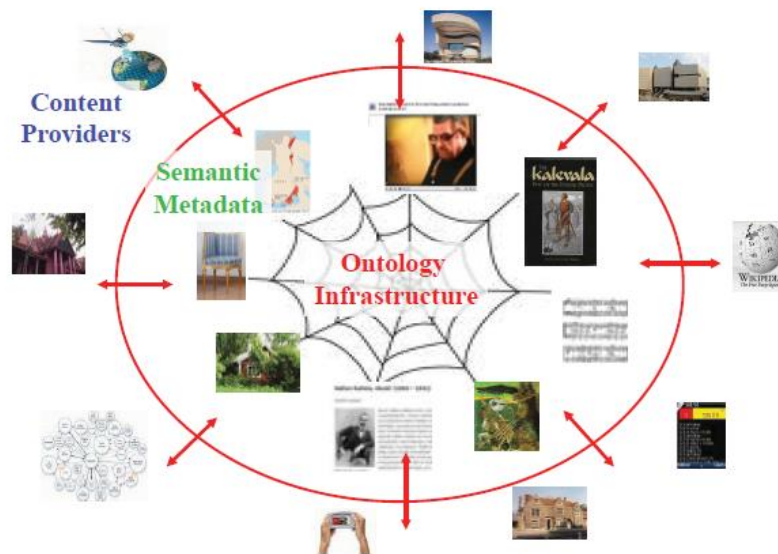


Figura 3. Rappresentazione del modello concettuale per la descrizione di Linked Data culturali

2.3.3. Portali informativi culturali semantici

Un *portale informativo semantico* è un sito mash-up all'apparenza simile a quelli realizzati per il Web 2.0, con la sostanziale differenza che la raccolta e la presentazione delle informazioni è effettuata utilizzando il framework di tecnologie standard fornite dal Web semantico.

2.3.3.1. Componenti di un portale semantico

In generale, un portale semantico realizzato sulla base del modello concettuale definito in precedenza deve essere costituito da tre componenti fondamentali.

- 1. Infrastruttura dei contenuti.** Comprende un insieme di ontologie *cross-domain*⁹ per la descrizione di concetti generici (persone, luoghi, eventi ecc.). Il collegamento tra i dati avviene attraverso la condivisione degli URI utilizzati per identificare i concetti.
- 2. Sistema di produzione dei contenuti.** Comprende un insieme di modelli e di linee guida per la creazione e la catalogazione uniforme dei contenuti. Offre inoltre modelli e protocolli per la raccolta delle informazioni provenienti dai sistemi di proprietà delle varie istituzioni e associazioni culturali.
- 3. Interfaccia semantica.** Rappresenta la parte del sistema visibile agli utenti finali. Comprende un sistema di ricerca semantica e di navigazione e visualizzazione dei contenuti. Sono presenti inoltre apposite API per abilitare l'accesso al sistema ad applicazioni web esterne, come ad esempio i portali web delle diverse istituzioni partecipanti.

⁹ Le ontologie coinvolte possono provenire da diversi domini amministrativi.

2.3.3.2. Benefici per gli utilizzatori

Un portale realizzato sulla base del modello concettuale semantico fornisce diversi benefici sia agli utenti finali che agli autori dei contenuti pubblicati.

Per quanto riguarda gli *utenti finali*, si individuano i seguenti benefici.

1. *Visione globale su contenuti eterogenei e distribuiti.* I contenuti prodotti dai diversi autori possono essere visualizzati da un singolo punto centralizzato, così come accade per i portali informativi del Web 2.0, con la sostanziale differenza che ora l'accesso alle informazioni avviene in maniera standard su un "contenitore" uniforme e omogeneo di informazioni.
2. *Aggregazione semantica delle informazioni.* In un processo di ricerca delle informazioni, il bisogno informativo di un utente viene soddisfatto associando semanticamente tra loro i dati estratti dalle diverse sorgenti.
3. *Ricerca semantica.* Le interfacce fornite dai portali possono essere arricchite di ulteriori filtri di ricerca basati su concetti e schemi definiti all'interno delle ontologie.
4. *Collegamenti semantici.* I contenuti possono essere navigati più intuitivamente attraverso la definizione esplicita della semantica attribuita ai collegamenti presenti tra i dati.

Per gli *autori* dei contenuti si individuano invece i seguenti benefici.

1. *Creazione di contenuti distribuiti.* Nei portali informativi tradizionali i contenuti sono tipicamente creati in maniera centralizzata attraverso l'utilizzo di appositi sistemi gestionali (i *Content Management System*). Questo approccio è poco flessibile, in quanto i contenuti, anche se distribuiti, sono creati senza alcun coordinamento tra gli autori. In contrapposizione, come abbiamo visto prima, l'utilizzo dell'infrastruttura ontologica proposta dal modello concettuale abilita implicitamente la collaborazione tra le diverse istituzioni partecipanti.

2. *Manutenzione automatica dei collegamenti.* Nei portali semantici, i collegamenti tra i dati sono creati e mantenuti automaticamente sulla base degli schemi di metadati e delle ontologie definiti nel modello.
3. *Canale di pubblicazione di contenuti condivisi.* L'obiettivo comune della pubblicazione dei dati culturali sul Web è diffondere la conoscenza tra il pubblico e tra chi opera nel settore della cultura. I portali semantici rappresentano in questo senso un "punto di ingresso" per la fruizione e la condivisione della suddetta conoscenza.
4. *Arricchimento semantico dei contenuti altrui.* Definire collegamenti semantici tra le informazioni abilita la possibilità di arricchire semanticamente i contenuti pubblicati da altri autori.
5. *Riuso dei contenuti.* I contenuti combinati e aggregati dai portali semantici possono essere riutilizzati da applicazioni e sistemi esterni.

CAPITOLO 3

UN'ONTOLOGIA PER RAPPRESENTARE I DATI DEI BENI CULTURALI NEL WEB SEMANTICO: INTRODUZIONE

Ora che abbiamo un quadro generale dello stato dell'arte del Web semantico e delle modalità di utilizzo delle relative tecnologie per la pubblicazione dei dati del patrimonio culturale, possiamo passare a introdurre il progetto su cui verte la trattazione corrente.

Prima di parlare del progetto vero e proprio e discutere le varie fasi che hanno portato alla sua realizzazione, si fornirà una breve panoramica del contesto in cui esso è nato e si è sviluppato, attraverso la descrizione del profilo dell'azienda committente, *3D Informatica*. Successivamente si passerà a discutere le motivazioni che hanno spinto alla necessità di tale progetto, e l'obiettivo che l'azienda si è proposta di raggiungere attraverso la sua realizzazione. Nei capitoli successivi si illustreranno nel dettaglio le varie fasi che hanno portato alla costruzione dell'ontologia finale.

3.1. Profilo dell'azienda committente

*3D Informatica*¹⁰ è una realtà aziendale attiva sul territorio italiano da oltre 30 anni, e ad oggi rappresenta una delle principali aziende informatiche del territorio specializzate nei settori dell'*Enterprise Document Management*, del *Record Management* e dell'*Information Retrieval*, discipline che hanno come obiettivo comune la gestione, il mantenimento e il reperimento di informazioni di diverso tipo e ambito.

La *mission* aziendale, così come espressa dall'organizzazione stessa, è la seguente:

“Il desiderio è sempre stato quello di dare risposte uniche nel loro genere e colmare con esse quello che trent'anni or sono era un considerevole vuoto: quello inerente la gestione delle informazioni e documenti ed ancor di più quello della gestione della conoscenza.”

E' proprio sulla base di questo desiderio di stare al passo con le innovazioni e i risultati delle ricerche nel panorama informatico – con riferimento in particolare alle modalità di gestione di informazioni strutturate, semi-strutturate e non strutturate – che ha preso piede l'idea del progetto in questione. Nello specifico, il contesto in cui esso va a collocarsi fa parte di una delle tipologie di soluzioni proposte e offerte dall'azienda a un preciso target di clienti: la gestione, il trattamento e la valorizzazione di informazioni relative all'ambito dei beni culturali.

Generalmente parlando, per la gestione e la conservazione nel tempo delle informazioni, l'azienda ha sviluppato internamente una piattaforma documentale, *Extraway® XML DB*, la quale comprende diverse soluzioni applicative. Tra queste vi è *Odysseus*, il software responsabile della catalogazione dei beni culturali secondo le normative vigenti stabilite dall'*Istituto Centrale per il Catalogo e la Documentazione (ICCD)*, di cui si parlerà nel prossimo capitolo.

¹⁰ <http://www.3di.it/>

3.2. Motivazioni

Attualmente l'applicazione *Odysseus* utilizza il formato XML per catalogare e rappresentare i dati dei beni culturali. Nello specifico, ciascun bene culturale è catalogato da *Odysseus* attraverso la compilazione di *schede di catalogo*, ovvero documenti il cui contenuto è strutturato secondo le regole sintattiche di XML.

Una delle caratteristiche chiave dell'utilizzo di questa tecnologia come strumento per la rappresentazione di informazioni è sicuramente l'*interoperabilità*, elemento fondamentale in un mondo che tenta di essere il più standard possibile come il Web. Tuttavia, come discusso in precedenza, la potenza espressiva di XML è molto limitata e poco adeguata a definire in maniera chiara e completa la semantica dei concetti di un dominio informativo. E' quindi abbastanza evidente quali siano le motivazioni che hanno spinto l'azienda committente a intraprendere il progetto in questione: le innovazioni tecnologiche emerse negli ultimi anni con il Web semantico hanno aperto una porta su un nuovo mondo, un mondo che, come si accennava, è in parte ancora inesplorato, ma le cui potenzialità, se sfruttate adeguatamente, possono essere davvero notevoli. L'utilizzo della sola tecnologia XML per la catalogazione e la gestione delle informazioni – di qualunque tipo esse siano – rappresenta quindi un vero e proprio ostacolo alla loro reale valorizzazione. Questo aspetto rappresenta il punto di partenza verso la realizzazione di un'ontologia formale che modelli semanticamente i dati relativi ai beni culturali.

3.3. Obiettivi e requisiti

Dalle motivazioni riportate sopra possiamo derivare facilmente l'obiettivo del progetto in questione: la possibilità di manipolare, utilizzare e collegare liberamente fra loro le informazioni sui beni culturali sulla base di un modello concettuale semantico di riferimento può aprire scenari davvero interessanti per l'intero mondo culturale e non

solo, favorendo la diffusione della conoscenza via Web sia a livello nazionale che internazionale. Proprio perché ora la prospettiva di utilizzo dei dati si sposta da uno scope locale e chiuso delle varie istituzioni culturali, a uno *globale e aperto* del Web, di fondamentale importanza sarà la possibilità di arricchire ulteriormente le informazioni relative a un bene culturale per rendere i contenuti più completi e garantire un maggiore grado di correttezza.

Un requisito fondamentale per la costruzione dell'ontologia in questione è quello di garantire un rigoroso allineamento con gli standard normativi stabiliti dall'ICCD per la rappresentazione delle informazioni. In altre parole, le modalità di definizione dei concetti all'interno dell'ontologia dovranno rispettare obbligatoriamente i vincoli definiti dalle normative ICCD.

Da questo requisito deriva un aspetto chiave del progetto: l'ontologia in questione non andrà a sostituire il sistema di catalogazione utilizzato finora per rappresentare le informazioni, ma ne costituirà semplicemente una sua estensione, ovvero una sorta di "infrastruttura" costruita sopra di esso. Questa caratteristica è perfettamente in linea con il principio generale del Web semantico secondo cui esso rappresenta essenzialmente un'estensione del Web tradizionale, che include in più la componente semantica.

3.4. Fasi di realizzazione del progetto

Il processo di costruzione dell'ontologia è stato suddiviso nelle seguenti fasi.

- 1. Analisi del dominio.** Comprende la descrizione del sistema di compilazione di una scheda di catalogo e l'analisi delle normative ICCD di interesse per le fasi successive.

2. **Mapping del livello inventariale.** Comprende la definizione di associazioni tra gli elementi descrittivi del livello inventariale¹¹ di una scheda di catalogo e i concetti ontologici definiti in vocabolari condivisi e modelli di riferimento preesistenti, tra cui principalmente *CIDOC-CRM*.
3. **Costruzione dell'ontologia.** Comprende la modellazione e la costruzione vera e propria dell'ontologia sulla base del mapping effettuato nella fase precedente.

Al termine della realizzazione dell'ontologia è stata sviluppata un'applicazione desktop minimale per la trasformazione delle schede di catalogo XML in documenti strutturati secondo le regole sintattiche di RDF, utilizzando l'ontologia costruita come modello di riferimento per la conversione dei documenti.

¹¹ Come si vedrà nel prossimo capitolo, il *livello di inventario* rappresenta il livello minimo di catalogazione che permette di rendere valida una scheda.

CAPITOLO 4

FASE 1: ANALISI DEL DOMINIO APPLICATIVO

Il passo preliminare propedeutico alla fase di costruzione della nostra ontologia consiste nell'analizzare e comprendere il sistema di compilazione di una scheda di catalogo così come definito dall'ICCD, l'*Istituto Centrale per il Catalogo e la Documentazione*. Verrà quindi descritto innanzitutto cos'è l'ICCD e di cosa si occupa nello specifico, dopodiché si passerà a riportare le indicazioni di carattere generale stabilite dai responsabili dell'ICCD per compilare una scheda.

Fatto questo, la trattazione entrerà nel vivo del progetto con la descrizione degli elementi di compilazione appartenenti al livello inventariale di catalogazione. Come vedremo, la catalogazione di un bene culturale può avvenire su tre livelli: livello di inventario, livello di precatalogo e livello di catalogo. Il livello di base è il livello di inventario, ovvero il livello minimo di compilazione delle voci di una scheda che permette di rendere quest'ultima valida.

La fase di analisi preliminare del dominio è di fondamentale importanza per l'intero progetto, in quanto rappresenta la base per la successiva fase di mapping degli elementi descrittivi delle normative ICCD in concetti ontologici, da cui poi sarà costruito il modello finale.

4.1. *L'Istituto Centrale per il Catalogo e la Documentazione (ICCD)*

L'ICCD¹² è un organo del *Ministero dei Beni e delle Attività Culturali e del Turismo* (MiBACT) che si occupa di definire gli standard e gli strumenti di catalogazione dei beni culturali. In particolare, l'ICCD:

1. gestisce il Catalogo generale del patrimonio archeologico, architettonico, storico-artistico e etnoantropologico nazionale;
2. elabora le metodologie di catalogazione e coordina le attività operative degli enti tecnici che la realizzano sul territorio;
3. tutela, conserva e valorizza le proprie collezioni di fotografia storica e di fotografia aerea, e realizza campagne di documentazione del patrimonio culturale;
4. assicura il coordinamento e promuove programmi di digitalizzazione del patrimonio culturale di competenza del MiBACT.

In sintesi, quindi, l'ICCD è l'organo responsabile della ricerca e dello sviluppo di strumenti e metodi per la diffusione della conoscenza, atti a tutelare e valorizzare il patrimonio artistico e culturale italiano.

Come accennato in precedenza, per la compilazione di schede di catalogo, l'ICCD ha rilasciato nel tempo diverse *normative* (dette anche *standard catalogafici*), ciascuna riferita a una precisa tipologia di bene culturale (opera d'arte, fotografia, reperto archeologico e così via). Tali normative definiscono modalità e vincoli per la compilazione delle voci di una scheda, e sono aggiornate e rilasciate periodicamente in nuove versioni. Tutte le normative ICCD presentano tra di loro elementi in comune per la descrizione di informazioni generiche di un bene culturale, in più definiscono elementi di descrizione esclusivi della particolare tipologia di bene da rappresentare.

¹² <http://www.iccd.beniculturali.it/>

4.2. Modalità di compilazione di una scheda¹³

Una normativa ICCD è un documento costituito da due parti principali:

1. *tracciato* (o struttura dei dati);
2. *norme di compilazione*.

Il tracciato di una scheda è organizzato in una serie di insiemi omogenei di informazioni chiamati **paragrafi**. Ciascun paragrafo contiene a sua volta altri elementi, detti **campi**. Un campo può essere:

- *strutturato*, se contiene a sua volta ulteriori campi, detti *sottocampi*;
- *semplice*, se rappresenta una singola voce da compilare (e quindi non contiene ulteriori sottocampi).

Paragrafi e campi strutturati sono detti *elementi contenitore*, in quanto il loro unico scopo è raggruppare campi e sottocampi. Le voci che vengono effettivamente compilate quando si redige una scheda di catalogo sono quelle relative ai campi semplici e ai sottocampi di un elemento contenitore.

Ciascun elemento descrittivo definito finora presenta specifiche **proprietà**.

1. *Lunghezza*. Indica il numero massimo di caratteri che è possibile utilizzare per compilare una voce.
2. *Ripetitività*. Indica che un elemento può essere ripetuto per registrare le diverse occorrenze di informazioni di uno stesso tipo.
3. *Obbligatorietà*. Indica che la compilazione di un certo elemento è obbligatoria.
4. *Vocabolario*. Indica che per la compilazione di un campo semplice o di un sottocampo è disponibile uno specifico strumento terminologico.
5. *Visibilità*. Indica il livello di visibilità in relazione alla possibilità o meno che un campo possa contenere dati riservati per motivi di privacy o di tutela.

¹³ Si tenga presente che le indicazioni normative riportate fanno ovviamente riferimento a una *gestione informatizzata* degli strumenti e delle procedure di catalogazione.

La proprietà di *obbligatorietà* si distingue in due tipi.

- *Obbligatorietà assoluta*. Segnalata da un asterisco (“*”), indica che la compilazione di un campo è obbligatoria affinché la scheda di catalogo possa essere ritenuta valida.
- *Obbligatorietà di contesto*. Segnalata da un asterisco tra parentesi tonde (“(*)”), indica che la compilazione di un campo è obbligatoria qualora il paragrafo o il campo strutturato in cui esso è contenuto venga compilato.

I due tipi di obbligatorietà possono assumere a loro volta carattere di *obbligatorietà alternativa*. Un’obbligatorietà alternativa indica che, in un gruppo di elementi obbligatori, il catalogatore può scegliere quale elemento compilare, in relazione a ciò che deve descrivere e/o alle informazioni a disposizione. In altre parole, in presenza di obbligatorietà alternativa, almeno uno dei campi appartenenti al gruppo di elementi obbligatori è compilato. Il gruppo di appartenenza è indicato attraverso l’utilizzo di numeri in progressione nel tracciato della scheda. Ad esempio, nel caso di obbligatorietà assoluta alternativa, il simbolo “*1” indica che l’elemento della scheda così contrassegnato appartiene al “gruppo 1” di elementi per i quali è prevista la compilazione alternativa.

Per quanto riguarda il *vocabolario*, si distinguono due tipologie.

- *Vocabolario chiuso*. Segnalato con il simbolo “C”, indica che il campo è compilabile utilizzando un elenco di termini predefinito che solamente l’ICCD, in caso, può estendere con ulteriori voci.
- *Vocabolario aperto*. Segnalato con il simbolo “A”, indica che il campo è compilabile utilizzando un elenco di termini che eventualmente il catalogatore può estendere con ulteriori voci.

Tutti i campi e i sottocampi per i quali non è definito alcun tipo di vocabolario sono considerati *a testo libero*, purché rispettino il vincolo dato dalla proprietà di lunghezza.

Infine, per quanto riguarda la proprietà di *visibilità*, si distinguono tre livelli.

1. *Livello basso di riservatezza* (1): indica che l'informazione contenuta nel campo a cui il valore è assegnato è liberamente fruibile da chiunque.
2. *Livello medio di riservatezza* (2): indica che il contenuto del campo a cui il valore è assegnato presenta dati personali riguardanti soggetti privati, e sono quindi riservati.
3. *Livello alto di riservatezza* (3): indica che l'informazione contenuta nel campo a cui il valore è assegnato presenta dati riservati che consentono la precisa localizzazione del bene.

A titolo di esempio, si riporta in *Tabella 1* parte del tracciato del paragrafo *SG* estratto dalla *normativa F (Fotografia) versione 3.00*. Questo paragrafo riporta tutti i dati utili all'individuazione e alla descrizione dei soggetti raffigurati in una fotografia. Si noti l'utilizzo delle proprietà per la specifica dei vincoli associati a un elemento.

acronimo ¹⁴			definizione	lun.	rip.	obb.	voc.	vis.
SG			SOGGETTO			*		
	SGT		SOGGETTO			*		
		SGTI	Identificazione	250	Sì	*		1
	SGL		TITOLO		Sì	*		
		SGLT	Titolo proprio	250		* ¹		1
		SGLA	Titolo attribuito	250		* ¹		1
		SGLS	Specifiche del titolo	250		*		1

Tabella 1. Parte del tracciato del paragrafo SG

¹⁴ L'acronimo rappresenta una sigla convenzionale che individua univocamente un elemento del tracciato. Viene utilizzato nelle procedure di trasferimento dei dati digitali fra sistemi eterogenei.

Dal punto di vista di una gestione informatizzata delle procedure di catalogazione, una scheda di catalogo può essere effettivamente redatta utilizzando diversi formati tecnologici. Nel caso dell'azienda committente, abbiamo visto che le schede gestite dal software *Odysseus* sono compilate utilizzando il formato XML.

4.3. La normativa F: analisi del livello di inventario

La **normativa F**¹⁵ relativa alle modalità di rappresentazione dei *beni di tipo fotografico* è la normativa assegnata al sottoscritto per la definizione del mapping dei campi inventariali e la successiva costruzione dell'ontologia finale. Per questioni di "allineamento" con le normative analizzate dai colleghi, la versione di riferimento è la *3.00*, sebbene l'ultima versione, rilasciata dall'ICCD a febbraio 2016, sia la *4.00*.

Le motivazioni che hanno spinto a scegliere il solo livello di inventario per la definizione del modello ontologico in questione sono presto dette: poiché il numero di elementi definiti dagli standard catalografici ICCD per descrivere un bene culturale è molto elevato, l'azienda committente ha deciso di considerare esclusivamente i paragrafi appartenenti a tale livello, in quanto sufficienti a rappresentare schede di catalogo valide. La fase di mapping successiva a quella corrente di analisi verrà quindi effettuata considerando solamente i campi del livello inventariale.

Come si riportava in precedenza, a ciascun collega tesista del progetto è stata assegnata una specifica normativa. In particolare, i colleghi hanno analizzato rispettivamente la *normativa OA* (riferita alle opere d'arte) e la *normativa RA* (riferita ai reperti archeologici). Per suddividere equamente il lavoro di analisi dei campi del livello inventariale, di comune accordo è stato selezionato e assegnato a ciascuno un

¹⁵ <http://www.iccd.beniculturali.it/index.php?it/473/standard-catalografici/Standard/10>

sottoinsieme di paragrafi con obbligatorietà assoluta da analizzare e da cui poi effettuare la successiva fase di mapping.

I **paragrafi in comune** analizzati dal sottoscritto sono i seguenti.

- **DT (cronologia):** informazioni specifiche sulla datazione del bene catalogato.
- **MT (dati tecnici):** informazioni relative alla tecnica di esecuzione, alla materia dei supporti primari¹⁶ e alle misure.
- **TU (condizione giuridica e vincoli):** indicazioni relative alla proprietà del bene.

Oltre ai paragrafi in comune, per la descrizione specifica di un bene fotografico, la normativa F presenta anche **paragrafi esclusivi** – o comunque in comune solo con un ristretto sottoinsieme di normative non analizzate per il progetto in questione. I paragrafi esclusivi sono riportati di seguito:

- **SG (soggetto):** riporta i dati utili alla individuazione e alla descrizione del/dei soggetto/i raffigurato/i nel bene catalogato.
- **UB (ubicazione e dati patrimoniali):** informazioni specifiche riguardanti l'ubicazione dell'opera.
- **AU (definizione culturale):** dati anagrafici relativi ai diversi autori responsabili del bene in esame.

Infine, in aggiunta ai paragrafi riportati sopra, sono stati analizzati anche sottocampi di paragrafi esaminati dai colleghi che riportavano obbligatorietà assoluta nella sola normativa F. I campi in questione appartengono ai seguenti paragrafi:

- **OG (oggetto):** informazioni per la definizione e l'identificazione dell'oggetto descritto nella scheda.
- **LC (localizzazione geografico-amministrativa):** dati relativi alla localizzazione geografico-amministrativa del bene nel momento in cui viene compilata la scheda.

¹⁶ In ambito fotografico, il *supporto primario* indica il materiale usato come supporto per lo strato immagine. Materiali comunemente utilizzati sono la carta, il metallo, il materiale plastico e il vetro.

Nei prossimi paragrafi si riporterà brevemente l'analisi effettuata su ciascuno dei paragrafi elencati sopra, a partire dalla descrizione dei paragrafi del livello inventariale in comune con le normative analizzate dai colleghi. Per ciascuno di essi si riporterà l'elenco di campi e sottocampi in esso contenuti che presentano obbligatorietà assoluta, una breve descrizione della semantica di ciascun campo, e alcuni esempi di compilazione.

4.3.1. Analisi dei paragrafi in comune

4.3.1.1. Il paragrafo *DT*

acronimo		definizione	lun.	rip.	obb.	voc.	vis.	
DT		CRONOLOGIA			*			
	DTZ	CRONOLOGIA GENERICA			*			
		DTZG	Secolo	50		*	C	1
	DTS	CRONOLOGIA SPECIFICA			*			
		DTSI	Da	15		*		1
		DTSF	A	15		*		1
	DTM	MOTIVAZIONE CRONOLOGIA			Sì	*		
		DTMM	Motivazione	50		*	C	1

Tabella 2. Tracciato del livello inventariale del paragrafo DT

Il paragrafo **DT** riporta informazioni specifiche sulla datazione del bene catalogato, con riferimento esclusivo alla data di esecuzione dell'oggetto in esame.

Relativamente a un bene di tipo fotografico, i dati sulla cronologia possono riguardare:

- il momento dello *scatto* (o dello sviluppo, se significativamente distante dal momento dello scatto), per negativi¹⁷ e unicum¹⁸;
- il momento dell'*acquisizione* dell'immagine, per fotografie virtuali;
- il momento della *realizzazione* della stampa, per la catalogazione di un positivo¹⁹.

Di seguito si riporta l'analisi dei sottocampi considerati.

- **DTZ (cronologia generica):** indicazione della fascia cronologica di riferimento.
 - *DTZG (secolo):* indicazione, in numeri romani, del/dei secolo/i in cui è stata eseguita l'opera in esame.
Es.: XIX
- **DTS (cronologia specifica):** datazione specifica in anni. Se l'anno preciso di esecuzione dell'opera è noto, i valori dei sottocampi *DTSI* e *DTSF* coincidono, in caso contrario indicheranno gli estremi cronologici dell'intervallo di tempo noto (o presunto) come periodo al quale si fa risalire l'esecuzione dell'opera.
 - *DTSI (da):* indicazione della data di esecuzione dell'opera, nel caso in cui si conosca precisamente, oppure dell'inizio dell'intervallo di tempo individuato come periodo di esecuzione dell'opera.
Es.: 1918/05
 - *DTSF (a):* indicazione della data di esecuzione dell'opera, corrispondente al valore del campo *DTSI* nel caso in cui si conosca precisamente, oppure della fine dell'intervallo di tempo individuato

¹⁷ Un *negativo* è un'immagine fotografica i cui valori tonali sono invertiti rispetto a quelli dell'immagine raffigurata.

¹⁸ Un *unicum* è un'immagine fotografica cosiddetta "unica", ottenuta cioè senza mediazione di negativi.

¹⁹ Un *positivo* è un'immagine fotografica ottenuta da un negativo, i cui valori tonali corrispondono a quelli del soggetto raffigurato.

come periodo di esecuzione.

Es.: 1918/05

- **DTM (motivazione cronologica):** indicazione sulle motivazioni e sulle fonti della datazione riportata nei campi precedenti.
 - *DTMM (motivazione):* indicazioni sul genere di fonte utilizzata o sul tipo di analisi adottata per stabilire la cronologia dell'opera.

Es.: analisi storica

4.3.1.2. Il paragrafo *MT*

acronimo		definizione	lun.	rip.	obb.	voc.	vis.	
MT		DATI TECNICI			*			
	MTX	Indicazione di colore	4		*	C	1	
	MTC	Materia e tecnica	70	Sì	*	A	1	
	MIS	MISURE		Sì	*			
		MISO	Tipo misure	50		*	A	1
		MISU	Unità di misura	5		*	C	1
		MISA	Altezza	6		* ³		1
		MISL	Larghezza	6		* ³		1
		MISS	Spessore	6		* ³		1
		MISD	Diametro	6		* ³		1

Tabella 3. Tracciato del livello inventariale del paragrafo *MT*

Il paragrafo **MT** contiene tutte le informazioni relative alla tecnica di esecuzione, alla materia dei supporti primari e alle misure del bene catalogato.

Di seguito si riporta l'analisi dei sottocampi considerati.

- **MTX (indicazione di colore):** indica se si tratta di fotografia (o in generale di un bene in cui tale proprietà è applicabile) in bianco/nero o a colori.

Es.: B/N [per fotografie in bianco/nero]

- **MTC (materia e tecnica):** indicazione del procedimento utilizzato per l'esecuzione dell'opera in esame e della materia del supporto primario.

Es.: stampa ai pigmenti/ carta

- **MIS (misure):** registra i dati relativi alle dimensioni dell'opera in esame, riferite principalmente al supporto primario in caso di catalogazione di un bene fotografico.

- *MISO (tipo misure):* indicazione dell'elemento di cui si specificano le misure nei campi obbligatori che seguono.

Es.: supporto primario

- *MISU (unità):* indicazione dell'unità di misura usata.

Es.: mm [millimetri]

- *MISA (altezza):* indica l'altezza massima del bene catalogato.

- *MISL (larghezza):* indica la larghezza massima del bene catalogato.

- *MISS (spessore):* indica lo spessore dell'elemento catalogato qualora questi rappresenti un oggetto contenitore come album, astucci, custodie ecc.

- *MISD (diametro):* indica il diametro massimo del bene catalogato.

Es.: 12

4.3.1.3. Il paragrafo TU

acronimo			definizione	lun.	rip.	obb.	voc.	vis.
TU			CONDIZIONE GIURIDICA E VINCOLI			*		
	CDG		CONDIZIONE GIURIDICA			*		
		CDGG	Indicazione generica	50		*	C	1
		CDGS	Indicazione specifica	250	Sì	*		2

Tabella 4. Tracciato del livello inventariale del paragrafo TU

Il paragrafo **TU** contiene le informazioni relative alla proprietà del bene.

Di seguito si riporta l'analisi dei sottocampi considerati.

- **CDG (condizione giuridica):** specifica le 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.
Es.: proprietà Ente pubblico territoriale
 - *CDGS (indicazione specifica):* indica l'esatta denominazione dell'Amministrazione, dell'Ente o del privato che hanno la proprietà del bene in esame. Qualora questi non siano noti, viene indicata la denominazione del detentore o del possessore.
Es.: Comune di Tivoli

4.3.2. Analisi dei paragrafi esclusivi

4.3.2.1. Il paragrafo SG

acronimo			definizione	lun.	rip.	obb.	voc.	vis.
SG			SOGGETTO			*		
	SGT		SOGGETTO			*		
		SGTI	Identificazione	250	Sì	*		1
	SGL		TITOLO		Sì	*		
		SGLT	Titolo proprio	250		* ¹		1
		SGLA	Titolo attribuito	250		* ¹		1
		SGLS	Specifiche del titolo	250		*		1

Tabella 5. Tracciato del livello inventariale del paragrafo SG

Il paragrafo **SG** contiene tutti i dati utili alla individuazione e alla descrizione del/dei soggetto/i raffigurato/i nel bene in esame.

Di seguito si riporta l'analisi dei sottocampi considerati.

- **SGT (soggetto):** informazioni relative al contesto figurativo, al contenuto generale dell'immagine.
 - *SGTI (identificazione):* individuazione del/dei soggetto/i raffigurato/i nell'opera in esame.
Es.: Lago di Garda – Vedute
- **SGL (titolo):** dati sul/sui titolo/i del bene catalogato.
 - *SGLT (titolo proprio):* indicazione del titolo assegnato dall'autore del bene catalogato o dalla tradizione.
 - *SGLA (titolo attribuito):* indicazione del titolo attribuito dal catalogatore, qualora il titolo proprio non sia presente.

- *SGLS (specifiche del titolo)*: indicazioni sulla fonte da cui è tratta l'informazione relativa al titolo riportato (proprio o attribuito).

Es.: dell'autore, stampato nell'immagine

4.3.2.2. Il paragrafo *UB*

acronimo		definizione	lun.	rip.	obb.	voc.	vis.
UB		UBICAZIONE E DATI PATRIMONIALI			*		
	UBF	UBICAZIONE FOTO			*		
		UBFP Fondo	150		*		2
		UBFC Collocazione	150		*		

Tabella 6. Tracciato del livello inventariale del paragrafo *UB*

Il paragrafo **UB** riporta informazioni specifiche riguardanti l'ubicazione del bene in esame, con i dati utili al suo reperimento e i riferimenti al fondo di appartenenza²⁰.

Di seguito si riporta l'analisi dei sottocampi considerati.

- **UBF (ubicazione foto)**: informazioni relative alla collocazione del bene catalogato nell'ambito della collezione (museo, biblioteca, archivio ecc.) a cui appartiene.
 - *UBFP (fondo)*: denominazione del fondo di appartenenza.
 - *UBFC (collocazione)*: sigla o segnatura indicante la collocazione fisica dell'opera all'interno della collezione o del fondo di appartenenza.

²⁰ Il fondo di appartenenza indica in questo caso una raccolta o collezione di fotografie, così chiamata perché creata sulla base di donazioni da parte di privati o per volontà specifica dell'ente che la detiene.

4.3.2.3. Il paragrafo AU

acronimo		definizione	lun.	rip.	obb.	voc.	vis.
AU		DEFINIZIONE CULTURALE			*		
	AUF	AUTORE DELLA FOTOGRAFIA		Sì	*		

Tabella 7. Tracciato del livello inventariale del paragrafo AU

Il paragrafo **AU** riporta i dati individuativo-anagrafici relativi ai diversi autori responsabili del bene in esame.

Di seguito si riporta l'analisi dell'unico sottocampo considerato.

- **AUF (autore della fotografia):** dati relativi ad ogni autore fotografo responsabile, a diverso titolo, del processo creativo dell'opera che si sta catalogando (ideatore, esecutore della ripresa, ecc.).

4.3.3. Analisi dei campi esclusivi di paragrafi in comune

4.3.3.1. Il paragrafo OG

acronimo		definizione	lun.	rip.	obb.	voc.	vis.
OG		OGGETTO			*		
	OGT	OGGETTO			*		
		OGTB Natura biblioteconomica dell'oggetto	25		*	C	1
	QNT	QUANTITA'			*		
		QNTN Numero oggetti/elementi	10		*		1

Tabella 8. Tracciato dei campi esclusivi del paragrafo OG nella normativa F

Il paragrafo **OG** contiene le informazioni che consentono la definizione e l'identificazione del bene descritto nella scheda, in relazione alla sua tipologia e al contesto in cui viene considerato dal punto di vista catalografico.

Di seguito si riporta l'analisi dei sottocampi considerati.

- **OGT (oggetto):** informazioni che consentono la precisa e corretta individuazione, sia tipologica che morfologica, del bene.
 - *OGTB (natura biblioteconomica dell'oggetto):* individuazione del contesto biblioteconomico cui va riferito il bene catalogato.
Es.: s [serie editoriale]
- **QNT (quantità):** informazioni specifiche sulla consistenza numerica del bene catalogato e, nel caso di un elemento di un oggetto complesso o di un'opera composita (es. album di fotografie), la sua posizione in relazione all'insieme.
 - *QNTN (numero oggetti/elementi):* quantità, in cifre, delle opere uguali schedate unitariamente con l'indicazione, in caso di oggetto complesso o composito, del numero degli elementi componenti. Qualora il bene

catalogato sia un oggetto complesso o un'opera composita, al numero si farà seguire la quantità in cifre degli elementi componenti.

Es.: 1 (5) [per un'opera composita, es. 'panorama', in 5 pezzi]

4.3.3.2. Il paragrafo LC

acronimo			definizione	lun.	rip.	obb.	voc.	vis.
LC			LOCALIZZAZIONE GEOGRAFICO- AMMINISTRATIVA			*		
	LDC		COLLOCAZIONE SPECIFICA			*		
		LDCN	Denominazione	80		*		3
		LDCM	Denominazione raccolta	70		*		2

Tabella 9. Tracciato dei campi esclusivi del paragrafo LC nella normativa F

Il paragrafo **LC** registra i dati relativi alla localizzazione geografico-amministrativa del bene nel momento in cui viene compilata la scheda.

Di seguito si riporta l'analisi dei sottocampi considerati.

- **LDC (collocazione specifica):** informazioni specifiche sulla collocazione del bene catalogato.

- *LDCN (denominazione):* nome proprio completo, o denominazione corrente, o titolo del contenitore.

Es.: Chiesa della Ss. Annunziata

- *LDCM (denominazione raccolta):* nome ufficiale per esteso della raccolta o tipologia e denominazione della collezione cui appartiene l'opera. Per le opere di proprietà privata viene indicato il cognome e il nome del proprietario preceduto dalla dizione "Collezione privata".

Es.: via Vetulonia, 324

CAPITOLO 5

FASE 2: MAPPING DEL LIVELLO INVENTARIALE

La fase che rappresenta il fulcro dell'intero progetto e il punto di partenza per la costruzione effettiva del modello ontologico è la fase di definizione di un *mapping* tra gli elementi descrittivi del livello inventariale analizzato nella fase precedente e le entità di tipo ontologico derivate da modelli e schemi di metadati (quindi vocabolari e ontologie) già esistenti e pubblicati nel Web semantico.

Il modello concettuale di riferimento proposto dall'azienda committente e utilizzato per effettuare gran parte del mapping è il modello *CIDOC-CRM*, che sarà analizzato nei prossimi paragrafi. Poiché durante la selezione delle entità ontologiche è emersa talvolta la necessità di utilizzare concetti più specifici di quelli proposti da tale modello, nei prossimi paragrafi, oltre all'analisi di *CIDOC-CRM*, sarà riportata anche una panoramica veloce degli schemi pubblici considerati per tale scopo.

Dal punto di vista operativo, la fase di mapping è stata svolta secondo le attività elencate di seguito:

1. analisi e comprensione delle caratteristiche del modello *CIDOC-CRM* e dei concetti espressi;
2. mapping iniziale delle informazioni contenute all'interno dei paragrafi in comune del livello inventariale delle normative ICCD con i concetti definiti in *CIDOC-CRM*;
3. confronto con i colleghi tesisti e con il responsabile dell'azienda committente sul mapping effettuato;

4. correzioni emerse durante il confronto e completamento del mapping dei paragrafi in comune;
5. mapping delle informazioni contenute all'interno dei paragrafi esclusivi della normativa assegnata a ciascun tesista.

Il mapping di tutto il livello inventariale è stato inizialmente riportato su un foglio elettronico Excel locale e successivamente trasferito su un foglio online condiviso tra i tesisti, al fine di facilitare le comunicazioni e avere un documento unico e organico.

Vediamo quindi innanzitutto quali sono gli schemi e i modelli di riferimento utilizzati per modellare le informazioni delle normative ICCD e costruire il mapping.

5.1. Schemi e modelli di riferimento

5.1.1. *CIDOC Conceptual Reference Model (CRM)*

*CIDOC*²¹ *Conceptual Reference Model (CIDOC-CRM)* è un'ontologia formale realizzata allo scopo di abilitare e facilitare l'integrazione, la mediazione e lo scambio di informazioni eterogenee relative ai beni culturali. Più precisamente, l'obiettivo principale di questo modello ontologico è fornire un insieme di definizioni semantiche utili a trasformare dati eterogenei derivati da sorgenti locali in un insieme coerente di informazioni globali, fruibili quindi da tutti.

E' importante tenere conto che *CIDOC-CRM* non rappresenta un'ontologia specifica relativa al dominio dei beni culturali, ma un modello concettuale che descrive formalmente la semantica generale sottostante ai dati culturali memorizzati nelle varie sorgenti locali mantenute dalle istituzioni. In altre parole, *CIDOC-CRM* non definisce alcuna terminologia presente nel dominio dei beni culturali, ma ne descrive le

²¹ International Committee for Documentation.

caratteristiche e le loro relazioni ad un alto livello di astrazione, classificandone i concetti e le relative proprietà. Per questo motivo, *CIDOC-CRM* è un modello che può favorire altamente l'interoperabilità tra le informazioni modellate in termini di standardizzazione nello scambio di dati tra sistemi informativi eterogenei.

Per realizzare il mapping sulla base dei concetti espressi da *CIDOC-CRM* è stata utilizzata la *versione 5.1 di maggio 2013* fornita dall'azienda committente, sebbene l'ultima versione rilasciata sia la 6.2 di maggio 2015. Gli elementi offerti dalla versione 5.1 sono suddivisi in *90 classi di entità*, per esprimere concetti, e *152 proprietà* (e relative inverse), per esprimere relazioni tra i concetti. Una classe del modello *CIDOC-CRM* è identificata dalla lettera "E" seguita da un numero progressivo (es. *E22*). Analogamente, una proprietà è identificata dalla lettera "P" seguita da un numero progressivo (es. *P4*). Entità e proprietà sono organizzate rispettivamente in gerarchie di classi e gerarchie di proprietà, similmente a come avviene in RDF Schema.

A titolo di esempio, in *Figura 4* viene mostrata una gerarchia delle classi principali offerte da *CIDOC-CRM*.

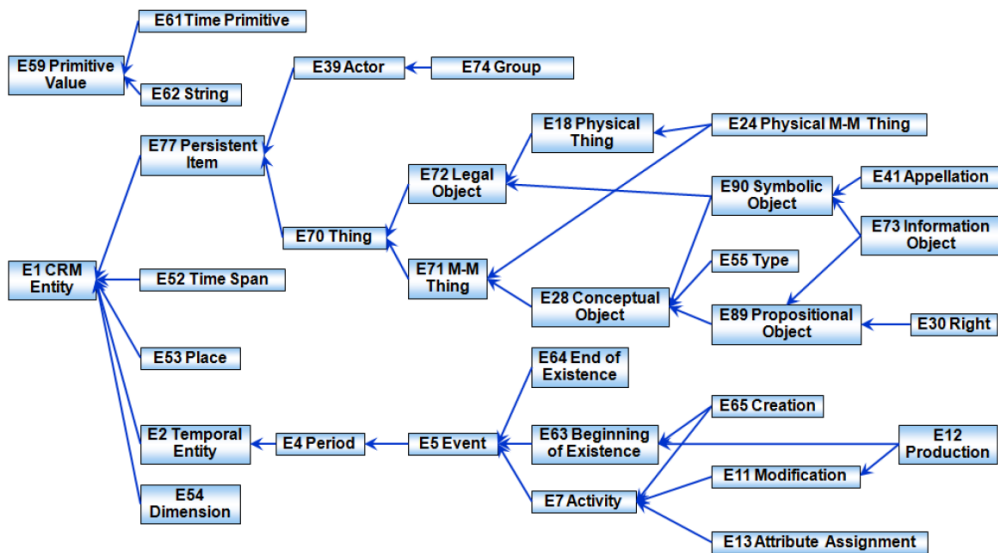


Figura 4. Gerarchia delle principali classi offerte dal modello *CIDOC-CRM*

Similmente a come avviene nei modelli di programmazione orientati agli oggetti, l'utilizzo delle classi in *CIDOC-CRM* è concretamente effettuato attraverso la definizione di *istanze* (o individui) per la classe considerata. Nel caso della nostra ontologia, ad esempio, vedremo che un bene culturale generico descritto in una scheda di catalogo può essere rappresentato da un'istanza della classe *E22 Man-Made Object*, che identifica in generale tutti gli artefatti creati dall'uomo.

CIDOC-CRM è al giorno d'oggi utilizzata in diverse organizzazioni e istituzioni. L'organismo responsabile dell'aggiornamento e del mantenimento del modello è l'ICOM, l'*International Committee for Documentation*²².

5.1.2. *Dublin Core (DC)*

Il *Dublin*²³ *Core (DC)* è un sistema di metadati e un framework costituito da un insieme di elementi essenziali ai fini della descrizione di qualsiasi materiale digitale accessibile dal Web. Le motivazioni di questo progetto sono derivate dalla necessità, avanzata da bibliotecari, archivisti, editori e sviluppatori software, di creare un insieme di strumenti condivisi per l'accesso alle risorse digitali.

Gli elementi di descrizione offerti dal Dublin Core costituiscono nel complesso il *DC Metadata Element Set*²⁴ (**DCES**). Tale insieme, formato da 15 elementi (o proprietà), è riportato in *Tabella 10*.

Tutti gli elementi all'interno del DCES hanno cardinalità *0..n*, ovvero rappresentano proprietà opzionali che possono modellare uno o più valori. Per arricchire ulteriormente la semantica delle informazioni descritte, gli elementi in DCES possono

²² <http://icom.museum/the-committees/international-committees/international-committee/international-committee-for-documentation>

²³ Dublin non si riferisce alla capitale irlandese, ma alla città di Dublin in Ohio, dove ha sede il quartier generale dell'organizzazione OCLC (*Online Computer Library Center*)

²⁴ <http://dublincore.org/documents/dces>

title	creator	subject	description	publisher
contributor	date	type	format	identifier
source	language	relation	coverage	rights

Tabella 10. Dublin Core Metadata Element Set (DCES)

essere estesi in maniera completamente interoperabile specializzando quelli esistenti o introducendone di nuovi. Tra le estensioni raccomandate dallo stesso Dublin Core vi è *DCMI Metadata Terms*²⁵, che introduce 40 nuovi elementi al DCES di base. DCMI Metadata Terms rappresenta al giorno d'oggi uno dei vocabolari RDF più popolari, e le sue specifiche sono totalmente compatibili con le idee alla base del Web semantico e del progetto LOD.

Nella costruzione della nostra ontologia, il modello Dublin Core verrà utilizzato talvolta per rappresentare i contenuti effettivi dei campi delle schede di catalogo.

5.1.3. *FRBR-aligned Bibliographic Ontology (FaBiO)*

Come abbiamo visto, il modello Dublin Core nasce principalmente per descrivere informazioni relative alle risorse digitali, tra cui anche libri e in generale documenti del patrimonio bibliotecario. Sebbene il Dublin Core sia particolarmente flessibile per descrivere questi tipi di oggetti, il suo limite principale è dato dall'impossibilità di poter distinguere tra il concetto di *expression* (espressione) di un oggetto e il concetto di *manifestation* (manifestazione) di quest'ultimo.

Per comprendere la differenza tra i due concetti, si consideri a titolo d'esempio l'Odissea di Omero. Questa opera esiste in diverse espressioni, ovvero la versione originale in greco più le varie traduzioni in altre lingue. Ciascuna di queste espressioni

²⁵ <http://dublincore.org/documents/dcmi-terms>

può essere “manifestata” in diverse forme: libro cartaceo, documento digitale, e così via.

Per quanto comunque il Dublin Core continui a rimanere un modello altamente valido, esso non è in grado di descrivere tali caratteristiche. Per questo motivo è stato sviluppato *Functional Requirements for Bibliographic Records (FRBR)*, un framework e una famiglia di standard per la rappresentazione di metadati concettuali nel dominio degli elementi bibliografici.

Tra gli standard offerti da FRBR, di particolare interesse per la costruzione della nostra ontologia è il modello omonimo **FRBR**. Questo modello distingue in maniera più netta i concetti di espressione e manifestazione discussi in precedenza, introducendo anche i concetti di *work* (opera) e *item* (oggetto).

Si riassume di seguito la semantica di ciascun concetto.

- Un *work* rappresenta l’idea artistica o intellettuale di un’entità astratta che può diventare potenzialmente reale e tangibile. L’idea della storia raccontata nell’Odissea da Omero costituisce in questo senso un *work*.
- Una *expression* rappresenta la realizzazione concreta ed effettiva di un *work*. La prima copia dell’Odissea costituisce in tal senso una particolare *expression* dell’idea iniziale. Da un *work* possono derivare più *expression*.
- Una *manifestation* rappresenta la forma fisica o digitale di una *expression*, Una specifica pubblicazione dell’Odissea in una data lingua costituisce, ad esempio, una *manifestation*. Ciascuna *expression* può essere incorporata in diverse *manifestation*.
- Un *item* rappresenta un esemplare di *manifestation*, ad esempio la copia in una specifica lingua dell’Odissea distribuita in una determinata libreria. Un *item* può rappresentare l’esemplare di una e una sola *manifestation*.

In Figura 5 vengono mostrate le relazioni tra i concetti appena descritti. Le frecce a singola punta indicano relazioni uno-a-uno, mentre le frecce a doppia punta indicano relazioni uno-a-molti.

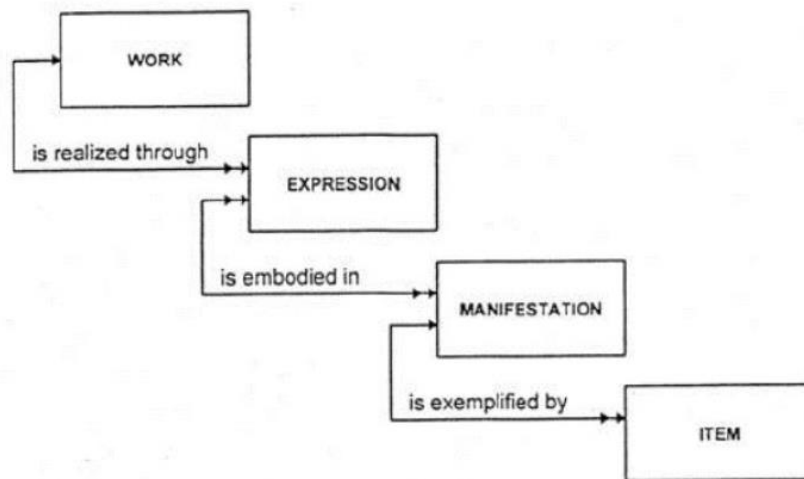


Figura 5. Relazioni tra i concetti espressi nel modello FRBR

In riferimento all'ontologia di progetto, lo schema effettivo utilizzato per il mapping di alcuni concetti, ispirato al modello FRBR appena descritto, è *FRBR-aligned Bibliography Ontology (FaBiO)*. In *Figura 6* viene riportata una rappresentazione delle principali entità modellate da *FaBio* relative ai quattro concetti discussi sopra. Si noti come le entità sono collegate tra loro attraverso apposite proprietà.

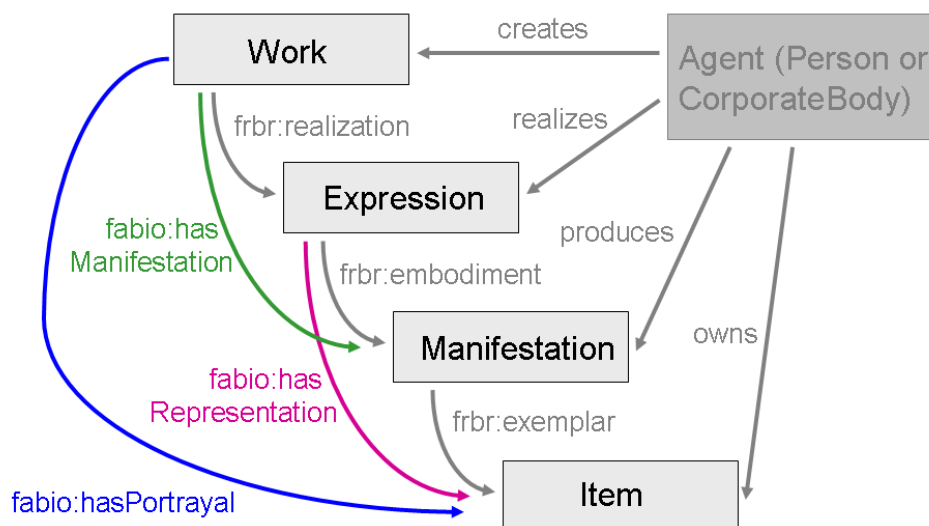


Figura 6. Rappresentazione dei concetti FRBR e delle loro relazioni nell'ontologia FaBiO

5.1.4. Altri schemi e modelli secondari

Si riportano di seguito ulteriori schemi e modelli di riferimento secondari utilizzati durante la fase di mapping delle informazioni.

1. *Time Interval (TI)*²⁶. Semplice ontologia in formato OWL utilizzata per modellare in maniera più espressiva concetti relativi a intervalli temporali.
2. *Historical Context Ontology (HiCo)*²⁷. Ontologia il cui obiettivo è descrivere il contesto storico di un bene culturale. Più precisamente, *HiCo* è utilizzata allo scopo di rappresentare in maniera significativa le asserzioni e le interpretazioni relative al contesto storico-culturale di un bene. In generale, un contesto di questo tipo può riguardare sia aspetti descritti esplicitamente nel bene in esame, sia aspetti impliciti derivati da esso attraverso apposite interpretazioni, come ad esempio la derivazione della corrente artistica dall'analisi di un dipinto. Per la costruzione della nostra ontologia, la scelta di utilizzare *HiCo* deriva dalla necessità di rappresentare informazioni relative alle interpretazioni attribuite a una particolare caratteristica di un bene culturale, ad esempio le motivazioni che hanno portato ad attribuire una specifica data di realizzazione al bene in questione. In generale, lo scopo di queste interpretazioni è comprendere in maniera chiara e completa il contenuto del bene in esame.
3. *PROV Ontology (PROV-O)*²⁸. PROV è un framework che consente di descrivere in maniera strutturata l'origine e la provenienza di artefatti e dei relativi dati generati in sistemi e contesti eterogenei. L'ontologia derivata da PROV, *PROV-O*, fornisce un insieme di classi, proprietà e vincoli che possono essere utilizzati per la rappresentazione e lo scambio di questi tipi di informazioni nel Web semantico. In riferimento all'ontologia di progetto,

²⁶ <http://www.ontologydesignpatterns.org/cp/owl/timeinterval.owl>

²⁷ <http://hico.sourceforge.net/index.html>

²⁸ <https://www.w3.org/TR/prov-o/>

PROV-O è stata scelta per dare un significato più preciso ad alcune informazioni.

5.2. Mapping del livello inventariale

Nelle prossime pagine si riporterà il mapping generale definito per il livello inventariale dei paragrafi e dei campi analizzati nella fase precedente, suddiviso in tre categorie:

1. mapping dei paragrafi in comune con le normative dei colleghi tesisti;
2. mapping dei campi esclusivi della normativa F assegnata al sottoscritto;
3. mapping dei campi esclusivi della normativa F nei paragrafi in comune.

Prima di cominciare con la definizione vera e propria del mapping è stato necessario stabilire, di comune accordo con i colleghi tesisti, l'entità ontologica che potesse riferirsi all'elemento generico "bene culturale" descritto nelle schede di catalogo, a cui poi collegare le diverse informazioni contenute nei campi. La scelta è ricaduta sull'entità *E22 Man-Made Object* del modello CIDOC-CRM. Tale classe comprende tutti gli oggetti fisici creati dall'uomo, ed è quindi adatta a rappresentare un bene culturale tangibile, che sia esso una fotografia, un reperto archeologico o un'opera d'arte. In realtà, come si riportava in precedenza, il concetto di "bene culturale" include anche entità astratte o naturali, come tradizioni e paesaggi. In un'eventuale estensione futura dell'ontologia che comprenda anche la rappresentazione di informazioni relative a beni culturali di questo tipo, si dovrà selezionare un'entità più adeguata. In ogni caso, per il progetto corrente, la maggior parte delle informazioni sarà descritta da proprietà collegate a istanze della classe *E22*.

Oltre a *E22*, un'entità di CIDOC-CRM spesso utilizzata per modellare certe informazioni è la *E55 Type*, la quale comprende istanze di concetti derivati da termini di tesauri e vocabolari controllati. Questa entità è quindi utile per descrivere particolari proprietà o caratteristiche di un'istanza riferite a gerarchie di termini, come potrebbe

essere, ad esempio, il tipo di misura adottato per stabilire le dimensioni di un bene. Con riferimento alle informazioni contenute nei campi delle schede di catalogo, l'entità *E55* è stata scelta ogniqualvolta la normativa prevedesse per quel campo l'utilizzo di strumenti terminologici specifici. Ciascuna istanza di *E55* è collegata dalla proprietà *P2 has type (is type of)*.

Il mapping del livello inventariale sarà descritto riportando le entità ontologiche scelte per rappresentare una specifica informazione di un paragrafo o di un campo, e le motivazioni che hanno condotto a tale scelta. Al termine della descrizione di ciascuna sotto-fase di mapping verrà riportata una tabella contenente entità e proprietà selezionate per mappare paragrafi e campi di quella categoria, e alcuni esempi di triple RDF in formato Turtle come riferimento per la rappresentazione delle informazioni nel Web semantico.

Nota. Il modello di riferimento principale utilizzato per definire il mapping è CIDOC-CRM. I nomi di entità e proprietà selezionate da altri modelli od ontologie saranno riportati introducendo un prefisso prima del nome, per distinguere tali concetti da quelli espressi in CIDOC-CRM.

5.2.1. Mapping dei paragrafi in comune

5.2.1.1. Il paragrafo *DT*

Remark. Il paragrafo *DT* riporta informazioni specifiche sulla datazione del bene catalogato.

Poiché per i beni culturali di tipo fotografico le informazioni riportate dipendono dalla tipologia di immagine fotografica, l'utilizzo della sola entità *E22* non è sufficiente a modellare i dati contenuti all'interno di questo paragrafo. E' necessario quindi selezionare entità "intermedie" più adeguate a rappresentare questa tipologia di bene.

- Per i *positivi*, la datazione riportata è relativa al momento della stampa. Si utilizzerà allora l'entità *E12 Production* di CIDOC-CRM, che comprende tutte

quelle attività che producono con successo uno o più oggetti con carattere di unicità, nel senso che non dovrebbero esistere altri oggetti con caratteristiche simili o identiche ad esso.

- Per i *negativi*, gli *unicum* e le *fotografie virtuali*, la datazione riportata è relativa al momento dello scatto o dell'acquisizione. Si utilizzerà allora l'entità **E65 Creation** di CIDOC-CRM, che definisce gli eventi in cui un oggetto è creato, come nel caso ad esempio delle fotografie virtuali che inizialmente non rappresentano altro che un insieme di bit contenuti in una memoria di massa.

Dato che le fotografie di tipo virtuale non rappresentano beni culturali tangibili, le informazioni relative ad esse dovranno essere modellate utilizzando un'entità più specifica. La scelta è ricaduta sull'entità **E28 Conceptual Object** di CIDOC-CRM.

In generale, un'istanza di bene culturale rappresentata dall'entità **E22** è collegata a un'istanza di **E12** attraverso la proprietà **P108 has produced (was produced by)**, che identifica gli oggetti creati dall'uomo derivati da un'attività di produzione, come nel caso della stampa. Un'istanza di bene culturale rappresentato dall'entità **E28**, invece, è collegata a un'istanza di **E65** attraverso la proprietà **P94 has created (was created by)**.

Per collegare un'istanza di **E65** o di **E12** rispettivamente alla tipologia di attività o di evento che rappresentano, si può utilizzare l'entità **E55** per indicare rispettivamente l'attività di stampa e l'attività di scatto o acquisizione.

5.2.1.1.1. Il campo DTZ

Remark. Il campo **DTZ** riporta la fascia cronologica di riferimento del bene catalogato.

L'unico sottocampo di **DTZ** da considerare è **DTZG**, che identifica il secolo in cui l'opera in questione è stata eseguita o, più specificamente, in cui la fotografia è stata stampata, scattata o acquisita. L'informazione di questo campo può essere modellata associando un'istanza di **E12** o di **E65** a un'istanza dell'entità **E52 Time-Span** di

CIDOC-CRM, attraverso la proprietà *P4 has time-span (is time-span of)*. L'entità *E52*, infatti, comprende concetti di estensioni temporali astratte, ovvero intervalli di tempo che hanno un inizio, una fine e una durata precisi (o perlomeno approssimati).

5.2.1.1.2. Il campo *DTS*

Remark. Il campo *DTS* riporta la datazione specifica, in anni, del bene catalogato.

I sottocampi considerati per *DTS* sono *DTSI* e *DTSF*, che rappresentano rispettivamente l'anno di inizio e l'anno di fine, in un range di anni, in cui l'opera è stata eseguita o, più specificatamente, in cui l'immagine fotografica è stata stampata, scattata o acquisita. Entrambe le informazioni contenute in questi campi possono essere modellate associando un'istanza di *E52* al datatype XML *xsd:dateTime*, utilizzando le proprietà definite all'interno dell'ontologia TI. In particolare:

- per il campo *DTSI* si utilizzerà la proprietà *ti:hasIntervalStartDate*;
- per il campo *DTSF* si utilizzerà la proprietà *ti:hasIntervalEndDate*.

Per arricchire ulteriormente la semantica di questo campo, nel caso in cui i valori di *DTSI* e *DTSF* siano diversi si potrebbe pensare di concatenarli al fine di generare un unico valore che esprima il range di anni in cui l'opera è stata eseguita o, più specificatamente, in cui la fotografia è stata stampata, scattata o acquisita. A tale scopo, si può fare riferimento a tale valore utilizzando nuovamente l'entità *E52*.

5.2.1.1.3. Il campo *DTM*

Remark. Il campo *DTM* riporta indicazioni sulle motivazioni e sulle fonti della datazione espressa nei campi precedenti.

L'unico sottocampo di *DTM* da considerare è *DTMM*, che descrive nello specifico il criterio interpretativo utilizzato per stabilire la cronologia del bene in esame.

Al fine di standardizzare i tipi di criteri riportati nelle schede e facilitare l'integrazione e il riuso dei dati, si è scelto di modellare le informazioni di questo campo selezionando

alcune entità e proprietà dalle ontologie HiCo e PROV-O. La classe principale di riferimento è *hico:InterpretationAct*, le cui istanze individuano formalmente una situazione in cui un agente²⁹ generico estrae informazioni utili dal “contenuto” di un oggetto. Ciascuna istanza di *hico:InterpretationAct* può essere associata a un’istanza di *E12* o *E65* attraverso la proprietà *prov:wasGeneratedBy*, per specificare come l’interpretazione sia stata derivata dal bene in esame.

Ciascuna istanza di *hico:InterpretationAct* può essere arricchita di significato specificando ulteriori proprietà collegate ad esso, tra cui:

- la proprietà *hico:hasInterpretationType* per indicare il tipo di interpretazione utilizzata (ad esempio “attribuzione di una datazione”), modellato attraverso un’istanza della classe *hico:InterpretationType*;
- la proprietà *hico:hasInterpretationCriterion* per indicare il tipo di criterio utilizzato (ad esempio “analisi formale”, istanza della classe *hico:InterpretationCriterion*) in riferimento alla motivazione riportata nel campo, che dovrebbe ricondursi a un vocabolario ben preciso.

5.2.1.2. Il paragrafo *MT*

Remark. Il paragrafo *MT* riporta tutte le informazioni relative alla tecnica di esecuzione, alla materia dei supporti primari e alle misure del bene catalogato.

Le informazioni contenute all’interno del paragrafo *MT* fanno riferimento a una particolare manifestazione del bene in esame. Nel caso di una fotografia, ad esempio, una sua manifestazione potrebbe essere il supporto primario (ovvero il positivo).

Per questo motivo, oltre all’utilizzo dell’entità di base *E22*, si può arricchire ulteriormente la semantica delle informazioni contenute nei sottocampi di *MT*

²⁹ Inteso come ente o individuo che si determini attraverso una data azione od operazione.

introducendo la classe *fabio:AnalogManifestation* dell'ontologia FaBiO, la quale modella in generale manifestazioni analoghe di espressioni disponibili in forma fisica o digitale.

5.2.1.2.1. Il campo *MTX*

Remark. Il campo *MTX* indica se il bene catalogato per cui tale caratteristica è applicabile è in bianco/nero o a colori.

L'informazione contenuta in *MTX* può essere modellata associando un'istanza di *E22* o di *fabio:AnalogManifestation* a un'istanza della classe ***E26 Physical Feature*** di CIDOC-CRM, la quale comprende concetti di caratteristiche riferite a uno specifico oggetto fisico, ad esempio, come nel caso corrente, la tipologia di colore del supporto primario di una fotografia. L'istanza dell'entità di base scelta può essere collegata a un'istanza di *E26* attraverso la proprietà ***P56 bears feature (is found on)***.

5.2.1.2.2. Il campo *MTC*

Remark. Il campo *MTC* riporta l'indicazione del procedimento utilizzato per l'esecuzione dell'opera in esame e della materia del supporto primario.

L'informazione contenuta in *MTC* può essere modellata associando un'istanza di *E22* o di *fabio:AnalogManifestation* a un'istanza della classe ***E57 Material*** di CIDOC-CRM, la quale comprende i concetti di materiale. Per riferirsi a un'istanza di *E57* si può utilizzare la proprietà ***P45 consists of (is incorporated in)***.

Poiché *E57* è sottoclasse di *E55*, le sue istanze dovrebbero fare riferimento a un vocabolario controllato ben definito.

5.2.1.2.3. Il campo *MIS*

Remark. Il campo *MIS* registra i dati relativi alle dimensioni dell'opera in esame, con

referimento principale al supporto primario in caso di catalogazione di un bene fotografico.

Per specificare in maniera adeguata le diverse tipologie di misure descritte dalle varie occorrenze³⁰ del campo *MIS*, all'entità di base scelta viene associata un'istanza della classe ***E54 Dimension*** di CIDOC-CRM, la quale comprende in generale proprietà quantificabili e misurabili attraverso specifici strumenti. Per riferirsi a un'istanza di *E54* si può utilizzare la proprietà ***P43 has dimension (is dimension of)***.

I sottocampi considerati per il campo *MIS* sono i seguenti.

- *MISO*. L'informazione contenuta in questo campo può essere modellata associando a un'istanza dell'entità *E28* un'istanza della medesima classe, in quanto tale campo indica il tipo di parte del bene a cui si riferiscono le misurazioni. Per collegare le due istanze si può utilizzare la proprietà ***P106 is composed of (forms part of)***.
- *MISU*. Per specificare l'unità di misura adottata per stabilire le dimensioni della parte in esame si può utilizzare la proprietà ***P91 has unit (is unit of)***. Tale proprietà permette di collegare l'istanza di *E54* definita in precedenza a un'istanza di ***E58 Measurement Unit*** di CIDOC-CRM, la quale rappresenta una specializzazione di *E55* introdotta per modellare i tipi di unità di misura. Ad esempio, la dimensione "altezza" può essere misurata utilizzando come unità di misura il millimetro.
- *MISA*, *MISL*, *MISS* e *MISD*. Le informazioni contenute in ciascuno di questi campi possono essere modellate associando all'istanza di *E54* definita in precedenza le seguenti proprietà:
 - *P2*, riferita a un'istanza di *E55*, per definire il tipo di misura adottata (es. millimetro);
 - ***P90 has value***, riferita a un'istanza di ***E60 Number***, per specificare il valore della misura effettuata sulla dimensione in esame.

³⁰ Si ricorda che il campo *MIS* è ripetitivo.

5.2.1.3. Il paragrafo *TU*

Remark. Il paragrafo *TU* contiene informazioni relative alla proprietà del bene.

L'unico campo di *TU* da considerare è *CDG*, i cui sottocampi specificano a loro volta tutte le informazioni relative all'attuale proprietà o detenzione del bene catalogato.

Per modellare correttamente tali informazioni, è necessario innanzitutto associare all'entità di base *E22* un'istanza della classe *E8 Acquisition* di CIDOC-CRM, la quale comprende i concetti di trasferimento di proprietà legali tra persone, queste ultime nella forma di istanze delle classe *E39 Actor*. L'entità di base e l'istanza di *E8* considerata possono essere collegate tra loro attraverso la proprietà *P24 transferred title of (changed ownership through)*, la quale permette di identificare oggetti coinvolti in un'acquisizione, come nel nostro caso un bene culturale.

I sottocampi da considerare per il campo *CDG* sono i seguenti.

- *CDGG*. L'informazione contenuta in questo campo specifica il tipo di proprietà del bene in esame, e può quindi essere modellata associando all'istanza di *E8* considerata un'istanza della classe *E55* che faccia riferimento a un vocabolario controllato.
- *CDGS*. L'informazione contenuta in questo campo specifica il nome o il titolo dell'attuale proprietario del bene, e può quindi essere modellata associando all'istanza di *E8* considerata un'istanza della classe *E39* attraverso la proprietà *P52 has current owner (is current owner of)*.

5.2.1.4. Sintesi del mapping

	SOGGETTO	PROPRIETA'	OGGETTO
DT			
	<i>E12 Production</i>	<i>P108 has produced (P108i was produced by)</i>	<i>E22 Man-Made Object</i>
		<i>P2 has type (P2i is type of)</i>	<i>E55 Type</i>
	<i>E65 Creation</i>	<i>P94 has created (P94i was created by)</i>	<i>E28 Conceptual Object</i>
		<i>P2 has type (P2i is type of)</i>	<i>E55 Type</i>
DTZ			
<i>DTZG</i>	<i>E12 Production</i>	<i>P4 has time-span (P2i is time-span)</i>	<i>E52 Time-Span</i>
	<i>E65 Creation</i>		
DTS			
<i>DTSI</i>	<i>E52 Time-Span</i>	<i>ti:hasIntervalStartDate</i>	<i>xsd:dateTime</i>
<i>DTSF</i>	<i>E52 Time-Span</i>	<i>ti:hasIntervalEndDate</i>	<i>xsd:dateTime</i>
DTM			
<i>DTMM</i>	<i>E12 Production</i>	<i>prov:wasGeneratedBy</i>	<i>hico:InterpretationAct</i>
	<i>E65 Creation</i>		
	<i>hico:InterpretationAct</i>	<i>hico:hasInterpretationType</i>	<i>hico:InterpretationType</i>
		<i>hico:hasInterpretationCriterion</i>	<i>hico:InterpretationCriterion</i>

MT			
<i>MTX</i>	<i>E22 Man-Made Object</i>	<i>P56 bears feature (P56i is found on)</i>	<i>E26 Physical Feature</i>
	<i>fabio: AnalogManifestation</i>		
<i>MTC</i>	<i>E22 Man-Made Object</i>	<i>P45 consists of (P45i is incorporated in)</i>	<i>E57 Material</i>
	<i>fabio: AnalogManifestation</i>		
MIS			
	<i>E22 Man-Made Object</i>	<i>P43 has dimension (P43i is dimension of)</i>	<i>E54 Dimension</i>
	<i>fabio: AnalogManifestation</i>		
<i>MISO</i>	<i>E22 Man-Made Object</i>	<i>P106 is composed of (forms part of).</i>	<i>E22 Man-Made Object</i>
	<i>fabio: AnalogManifestation</i>		<i>fabio: AnalogManifestation</i>
<i>MISU</i>	<i>E54 Dimension</i>	<i>P91 has unit (P91i is unit of)</i>	<i>E58 Measurement Unit</i>
<i>MISA</i> <i>MISL</i> <i>MISS</i> <i>MISD</i>	<i>E54 Dimension</i>	<i>P2 has type (P2i is type of)</i>	<i>E55 Type</i>
		<i>P90 has value</i>	<i>E60 Number</i>
TU			
	<i>E8 Acquisition</i>	<i>P24 transferred title of (P24i changed ownership through)</i>	<i>E22 Man-Made Object</i>

CDG			
<i>CDGG</i>	<i>E8 Acquisition</i>	<i>P2 has type (P2i is type of)</i>	<i>E55 Type</i>
<i>CDGS</i>	<i>E8 Acquisition</i>	<i>P52 has current owner (P52i is current owner of)</i>	<i>E39 Actor</i>

Tabella 11. Mapping dei paragrafi in comune del livello inventariale

5.2.1.5. Esempi di utilizzo

DT	
DTZ	
<i>DTZG</i>	<pre> :photo-72486-positive-print a crm:E12_Production ; crm:P108_has_produced :photo-72486-positive-item ; crm:P2_has_type :photograph-print ; crm:P4_has_time_span :xx-century . :xx-century a crm:E52_Time-Span . </pre>
DTS	
<i>DTSI</i>	<pre> :1926post-1940ca a crm:E52_Time-Span ; ti:hasIntervalStartDate "1926-01-01T00:00:00"^^xsd:dateTime . </pre>
<i>DTSF</i>	<pre> :1926post-1940ca a crm:E52_Time-Span ; ti:hasIntervalEndDate "1940-01-01T00:00:00"^^xsd:dateTime . </pre>
DTM	
<i>DTMM</i>	<pre> :photo-72486-positive-print prov:wasGeneratedBy :photo-72486-positive-date-attribution . :photo-72486-positive-date-attribution a hico:InterpretationAct, prov:Activity ; hico:hasInterpretationType :zeri-preferred-dating ; hico:hasInterpretationCriterion :formal-analysis . :zeri-preferred-dating a hico:InterpretationType . :formal-analysis a hico:InterpretationCriterion . </pre>

MT	
<i>MTX</i>	:photo-72486- positive a crm:E22_Man-Made_Object , fabio:AnalogManifestation ; crm:P56_bears_feature :black-and-white . :black-and-white a crm:E26_Physical_Feature .
<i>MTC</i>	:photo-72486-positive a crm:E22_Man-Made_Object , fabio:AnalogManifestation ; crm:P45_consists_of :gelatin-silver . :gelatin-silver a crm:E57_Material ;
MIS	
<i>MISO</i>	:photo-72486-positive a crm:E22_Man-Made_Object , fabio:AnalogManifestation ; crm:P106_is_composed_of :photo-72486-primary-support . :photo-72486-primary-support a crm:E22_Man-Made_Object , fabio:AnalogManifestation .
<i>MISU</i>	:photo-72486-positive a crm:E22_Man-Made_Object , fabio:AnalogManifestation ; crm:P43_has_dimension :height-194mm . :height-194mm a crm:E54_Dimension ; crm:P91_has_unit :mm . :mm a crm:E58_Measurement_Unit .
<i>MISA</i>	:photo-72486-positive a crm:E22_Man-Made_Object , fabio:AnalogManifestation ; crm:P43_has_dimension :height-194mm . :height-194mm a crm:E54_Dimension ; crm:P2_has_type :height ; crm:P90_has_value "194" ; crm:P91_has_unit :mm .
<i>MISL</i>	:photo-72486-positive a crm:E22_Man-Made_Object , fabio:AnalogManifestation ; crm:P43_has_dimension :width-250mm . :width-250mm a crm:E54_Dimension ; crm:P2_has_type :width ; crm:P90_has_value "250" ; crm:P91_has_unit :mm .

<i>MISS</i>	<pre> :photo-72486-positive a crm:E22_Man-Made_Object , fabio:AnalogManifestation ; crm:P43_has_dimension :thickness-0-05mm . :thickness-0-05mm a crm:E54_Dimension ; crm:P2_has_type :thickness ; crm:P90_has_value "0.05" ; crm:P91_has_unit :mm . </pre>
<i>MISD</i>	<pre> :photo-72486-positive a crm:E22_Man-Made_Object , fabio:AnalogManifestation ; crm:P43_has_dimension :diameter-300mm . :diameter-300mm a crm:E54_Dimension ; crm:P2_has_type :diameter ; crm:P90_has_value "300" ; crm:P91_has_unit :mm . </pre>
TU	
CDG	
<i>CDGG</i>	<pre> :photo-72486-acquisition a crm:E8_Acquisition ; crm:P24_transferred_title_of :photo-72486-positive-item ; crm:P2_has_type :public-ownership . </pre>
<i>CDGS</i>	<pre> :photo-72486-positive-item crm:P52_has_current_owner :university-of-bologna . :university-of-bologna a crm:E40_Legal_Body³¹ . </pre>

Tabella 12. Esempi di utilizzo degli elementi mappati

³¹ L'entità *E40 Legal Body* di CIDOC-CRM è una sottoclasse di *E39 Actor*.

5.2.2. Mapping dei paragrafi esclusivi

5.2.2.1. Il paragrafo SG

Remark. Il paragrafo SG contiene tutti i dati utili alla individuazione e alla descrizione del/dei soggetto/i raffigurato/i nel bene in esame.

Per modellare le informazioni contenute all'interno del paragrafo SG si utilizzerà come entità di base la classe *E28 Conceptual Object*, in quanto tali informazioni fanno riferimento a beni culturali che possono includere anche oggetti non tangibili, come ad esempio le immagini fotografiche virtuali.

5.2.2.1.1. Il campo SGT

Remark. Il campo SGT riporta informazioni relative al contesto figurativo, al contenuto generale dell'immagine.

L'unico sottocampo di SGT da considerare è SGTI, il cui contenuto riporta il nome del soggetto o dei soggetti raffigurati nel bene in esame. L'informazione di questo campo può essere modellata associando un'istanza dell'entità di base *E28* a un'istanza dell'entità *fabio:SubjectTerm* dell'ontologia FaBiO, attraverso la proprietà *fabio:hasSubjectTerm*. La classe *fabio:SubjectTerm* comprende concetti riferiti a vocabolari controllati, utilizzati in generale per descrivere il soggetto, il significato o il contenuto di una particolare entità.

5.2.2.1.2. Il campo SGL

Remark. Il campo SGL riporta dati relativi al/ai titolo/i del bene catalogato.

I sottocampi da considerare per il campo SGL sono i seguenti.

- *SGLT* e *SGLA*. Entrambi i campi specificano un tipo di titolo (proprio o attribuito) assegnato all'opera in esame. I relativi contenuti possono quindi

essere modellati utilizzando le medesime proprietà ed entità. Per ciascuno di essi si assocerà l'istanza di *E28* considerata a un'istanza della classe *E35 Title* di CIDOC-CRM, la quale comprende in generale concetti di nomi assegnati ad artefatti. Le due istanze possono essere collegate tra loro attraverso la proprietà *P102 has title (is title of)*. Per definire adeguatamente il tipo di titolo descritto, si può associare l'istanza di *E35* considerata a un'istanza della classe *E55 Type*.

- *SGLS*. L'informazione contenuta in questo campo specifica la fonte o l'entità che ha fornito il/i titolo/i dell'opera in esame. Analogamente al campo *DTMM* del paragrafo *DT*, il contenuto di *SGLS* può essere modellato facendo riferimento ai concetti definiti all'interno delle ontologie HiCo e PROV-O. L'istanza di *E35* considerata in uno dei campi precedenti³² viene collegata a un'istanza di *hico:InterpretationAct* tramite la proprietà *prov:wasGeneratedBy* per specificare come l'interpretazione sia stata derivata dal bene in esame. Ciascuna istanza di *hico:InterpretationAct* può essere arricchita di significato specificando ulteriori proprietà collegate ad esso:
 - la proprietà *hico:hasInterpretationType*, per indicare il tipo di interpretazione utilizzata (modellato attraverso un'istanza della classe *hico:InterpretationType*);
 - la proprietà *hico:hasInterpretationCriterion*, per indicare il tipo di criterio utilizzato (istanza della classe *hico:InterpretationCriterion*) in riferimento alla motivazione riportata nel campo, che dovrebbe ricondursi a un vocabolario ben preciso.

5.2.2.2. Il paragrafo *UB*

Remark. Il paragrafo *UB* riporta informazioni specifiche riguardanti l'ubicazione del bene in esame.

³² I campi *SGLT* e *SGLA* presentano un'obbligatorietà assoluta alternativa.

L'unico campo di *UB* da considerare è *UBF*, i cui sottocampi specificano a loro volta tutte le informazioni relative alla collocazione del bene catalogato nell'ambito della collezione a cui appartiene.

Per modellare correttamente tali informazioni, è necessario innanzitutto fare riferimento alla classe *fabio:WorkCollection* dell'ontologia FaBiO. Tale classe rappresenta in generale insiemi di *work* FRBR, ed è quindi adatta a rappresentare collezioni di beni culturali generici. Un'istanza della classe *fabio:WorkCollection* può essere collegata a una o più istanze della classe *fabio:Work* – o della classe analoga *E90 Symbolic Object* – attraverso la proprietà *P106 is composed of (forms part of)* di CIDOC-CRM. In questo modo è possibile modellare archivi di beni culturali come gerarchie, dove l'istanza di *fabio:WorkCollection* considerata rappresenta l'archivio (o fondo di appartenenza) e le singole istanze di *fabio:Work* rappresentano ciascuna un bene culturale.

I sottocampi da considerare per il campo *UBF* sono i seguenti.

- *UBFP*. L'informazione contenuta in questo campo specifica il nome dell'archivio che contiene il bene in esame, e può quindi essere modellata associando all'istanza di *fabio:WorkCollection* considerata un'istanza della classe *E35* attraverso la proprietà *P102 has title (is title of)*.
- *UBFC*. L'informazione contenuta in questo campo riporta una sigla relativa alla collocazione fisica del bene all'interno della collezione, e può quindi essere modellata associando innanzitutto un'istanza di *fabio:AnalogItem* a un'istanza della classe *E53 Place* di CIDOC-CRM, la quale comprende concetti di luoghi fisici. Le due istanze possono essere collegate tra loro attraverso la proprietà *P55 has current location (currently holds)*. In questo modo viene modellata la collocazione fisica del bene all'interno della collezione. Per riferirsi all'identificatore, l'istanza di *E53* in questione può essere collegata a un'istanza della classe *E42 Identifier* di CIDOC-CRM attraverso la proprietà *P48 has preferred identifier (is preferred identifier of)*. La classe *E42*, infatti, comprende concetti di stringhe e codici assegnati a istanze generiche di entità appartenenti al modello CIDOC-CRM.

5.2.2.3. Il paragrafo AU

Remark. Il paragrafo AU riporta i dati anagrafici relativi ai diversi autori responsabili del bene in esame.

Non sono presenti campi semplici con obbligatorietà assoluta all'interno del paragrafo AU. L'unico campo da considerare è il campo strutturato ripetitivo AUF, il quale riporta i dati relativi ad ogni autore fotografo responsabile, a diverso titolo, del processo creativo dell'opera in esame. Per modellare i diversi sottocampi di AUF si può quindi pensare di collegare un'istanza della classe E65 Creation (riferita all'evento che ha portato alla creazione del bene catalogato) a un'istanza della classe E21 Person di CIDOC-CRM, utilizzando la proprietà P14 carried out by (performed).

5.2.2.4. Sintesi del mapping

	SOGGETTO	PROPRIETA'	OGGETTO
SG			
SGT			
<i>SGTI</i>	<i>E28 Conceptual Object</i>	<i>fabio:hasSubjectTerm</i>	<i>fabio:SubjectTerm</i>
SGL			
<i>SGLT</i>	<i>E28 Conceptual Object</i>	<i>P102 has title (P102i is title of)</i>	<i>E35 Title</i>
<i>SGLA</i>	<i>E35 Title</i>	<i>P2 has type (P2i is type of)</i>	<i>E55 Type</i>
<i>SGLS</i>	<i>E35 Title</i>	<i>prov:wasGeneratedBy</i>	<i>hico:InterpretationAct</i>
	<i>hico:InterpretationAct</i>	<i>hico:hasInterpretationType</i>	<i>hico:InterpretationType</i>

		<i>hico:hasInterpretationCriterion</i>	<i>hico:InterpretationCriterion</i>
UB			
UBF			
	<i>fabio:WorkCollection</i>	<i>P106 is composed of (P106i forms part of)</i>	<i>E90 Symbolic Object</i>
			<i>fabio:Work</i>
<i>UBFP</i>	<i>fabio:WorkCollection</i>	<i>P102 has title (P102i is title of)</i>	<i>E35 Title</i>
<i>UBFC</i>	<i>fabio:AnalogItem</i>	<i>P55 has current location (P55i currently holds)</i>	<i>E53 Place</i>
	<i>E53 Place</i>	<i>P48 has preferred identifier (P48i is preferred identifier of)</i>	<i>E42 Identifier</i>
AU			
AUF			
	<i>E65 Creation</i>	<i>P14 carried out by (P14i performed)</i>	<i>E21 Person</i>

Tabella 13. Mapping dei paragrafi esclusivi del livello inventariale della normativa F

5.2.2.5. Esempi di utilizzo

SG	
SGT	
<i>SGTI</i>	<code>:photo-72486 a crm:E28_Conceptual_Object ; fabio:hasSubjectTerm :jesus-baptism . :jesus-baptism a fabio:SubjectTerm .</code>
SGL	

<i>SGLT</i>	<pre> :photo-72486 a crm:E28_Conceptual_Object ; crm:P102_has_title :jesus-baptism-verrocchio . :jesus-baptism-verrocchio a crm:E35_Title ; crm:P2_has_type :proper-title . </pre>
<i>SGLA</i>	<pre> :photo-72486 a crm:E28_Conceptual_Object ; crm:P102_has_title :florence-uffizi-jesus-baptism-verrocchio . :florence-uffizi-jesus-baptism-verrocchio a crm:E35_Title ; crm:P2_has_type :attributed-title . </pre>
<i>SGLS</i>	<pre> :jesus-baptism-verrocchio a crm:E35_Title ; prov:wasGeneratedBy :photo-72486-title-attribution . :photo-72486-title-attribution a hico:InterpretationAct ; hico:hasInterpretationType :zeri-preferred-attribution ; hico:hasInterpretationCriterion :cataloguer-choice . :zeri-preferred-attribution a hico:InterpretationType . :cataloguer-choice a hico:InterpretationCriterion . </pre>
UB	
UBF	
<i>UBFP</i>	<pre> :folder-leonardo a fabio:Work , crm:E90_Symbolic_Object ; crm:P106i_forms_part_of :zeri-photo-archive . :zeri-photo-archive a fabio:WorkCollection ; crm:P102_has_title :zeri-photo-archive-1 . :zeri-photo-archive-1 a crm:E35_Title . </pre>
<i>UBFC</i>	<pre> :folder-leonardo-item a fabio:AnalogItem ; crm:P55_has_current_location :folder-leonardo-item-collocation . :folder-leonardo-item-collocation a crm:E53_Place ; crm:P48_has_preferred_identifier :PI_0360-1-11 . :PI_0360-1-11 a crm:E42_Identifier . </pre>
AU	
AUF	
	<pre> :photo-72486-creation a crm:E65_Creation ; crm:P14_carried_out_by :brogi . :brogi a crm:E21_Person ; </pre>

Tabella 14. Esempi di utilizzo degli elementi mappati

5.2.3. Mapping dei campi esclusivi di paragrafi in comune

5.2.3.1. Il paragrafo *OG*

Remark. Il paragrafo *OG* contiene tutte le informazioni che consentono la definizione e l'identificazione del bene descritto nella scheda, in relazione alla sua tipologia e al contesto in cui viene considerato dal punto di vista catalografico.

5.2.3.1.1. Il campo *OGT*

Remark. Il campo *OGT* contiene tutte le informazioni che consentono la precisa e corretta individuazione, sia tipologica che morfologica, del bene.

L'unico sottocampo di *OGT* da considerare è *OGTB*, il cui contenuto riporta il codice identificativo del contesto biblioteconomico cui è riferito il bene catalogato. L'informazione di questo campo può essere modellata selezionando innanzitutto un'istanza a scelta dall'entità *E28 Conceptual Object* di CIDOC-CRM o dalle entità *fabio:Work* o *fabio:WorkCollection* dell'ontologia FaBiO, a seconda del codice riportato nel campo:

- un codice pari a “*m*” identifica un lavoro di tipo monografico³³, e in questo caso è possibile utilizzare una qualsiasi istanza di una delle tre entità sopra elencate;
- un codice pari a “*s*” o “*c*” identifica rispettivamente una serie o un lavoro complesso, e in questo caso si potrà utilizzare esclusivamente un'istanza della classe *fabio:WorkCollection*.

Per modellare l'informazione riportata nel campo, si collegherà l'istanza dell'entità scelta a un'istanza della classe *E55 Type*.

³³ Per lavoro monografico si intende un'opera la cui trattazione è relativa a uno specifico argomento.

5.2.3.1.2. Il campo *QNT*

Remark. Il campo *QNT* contiene informazioni specifiche sulla consistenza numerica del bene catalogato e, nel caso di un elemento di un oggetto complesso o di un'opera composita (es. album di fotografie), la sua posizione in relazione all'insieme.

L'unico sottocampo di *QNT* da considerare è *QNTN*, il cui contenuto riporta il numero di parti che compongono il bene descritto o l'opera composita. Poiché il concetto di "numero di parti" di un oggetto rappresenta una caratteristica fisica di quest'ultimo, l'informazione presente in *QNTN* può essere modellata selezionando innanzitutto un'istanza a scelta tra l'entità *E22 Man-Made Object* di CIDOC-CRM e le entità *fabio:AnalogItem* o *fabio:ItemCollection*, a seconda che l'informazione si riferisca rispettivamente a un bene singolo o composito. Per modellare il dato riportato, si collegherà l'istanza dell'entità scelta a un'istanza della classe *E60 Number*, utilizzando la proprietà *P57 has number of parts* di CIDOC-CRM.

5.2.3.2. Il paragrafo *LC*

Remark. Il paragrafo *LC* registra tutti i dati relativi alla localizzazione geografico-amministrativa del bene nel momento in cui viene compilata la scheda.

L'unico campo di *LC* da considerare è *LDC*, i cui sottocampi specificano a loro volta informazioni relative alla collocazione fisica del bene catalogato.

I sottocampi in questione sono i seguenti.

- *LDCN*. L'informazione contenuta in questo campo indica sia l'attuale custode del bene in esame, sia il contenitore fisico nel quale il bene risiede. Tale informazione può quindi essere modellata associando innanzitutto un'istanza dell'entità *E22* o della classe *fabio:AnalogItem* a un'istanza della classe *E40 Legal Body* attraverso la proprietà *P50 has current keeper (is current keeper of)* di CIDOC-CRM. La classe *E40*, infatti, comprende istituzioni o gruppi di persone che hanno ottenuto un riconoscimento legale come gruppo, e che, in generale, rappresentano i proprietari o custodi legali di un certo oggetto.

L'istanza di *E40* considerata può essere collegata a sua volta a un'istanza di *E53 Place* attraverso la proprietà *P74 has current or former residence (is current or former residence of)* di CIDOC-CRM, per specificare il luogo in cui l'istituzione o il gruppo legale risiede. Nel caso in cui il sottocampo *LDCS* del paragrafo corrente non sia compilato, l'istanza di *E53* considerata indicherà il luogo in cui il bene culturale è attualmente conservato. Tale informazione può essere modellata collegando bene culturale e luogo attraverso la proprietà *P55 has current location (currently holds)*.

- *LDCM*. L'informazione contenuta in questo campo specifica il nome della collezione in cui risiede il bene in esame. Tale informazione può quindi essere modellata associando innanzitutto un'istanza di *fabio:WorkCollection* a un'istanza di *E28 Conceptual Object* o di *fabio:Work* attraverso la proprietà *P106 is composed of (forms part of)*, al fine di collegare la collezione al bene in esame. Il nome della collezione può essere modellato associando l'istanza di *fabio:WorkCollection* selezionata a un'istanza di *E35* attraverso la proprietà *P102 has title (is title of)*.

5.2.3.3. Sintesi del mapping

	SOGGETTO	PROPRIETA'	OGGETTO
OG			
OGT			
<i>OGTB</i>	<i>E28 Conceptual Object</i>	<i>P2 has type (P2i is type of)</i>	<i>E55 Type</i>
	<i>fabio:Work</i>		
	<i>fabio:WorkCollection</i>		
QNT			
<i>QNTN</i>	<i>E22 Man-Made Object</i>	<i>P57 has number of parts</i>	<i>E60 Number</i>
	<i>fabio:AnalogItem</i>		
	<i>fabio:ItemCollection</i>		
LC			
LDC			
<i>LDCN</i>	<i>E22 Man-Made Object</i>	<i>P50 has current keeper (P50i is current keeper of)</i>	<i>E40 Legal Body</i>
	<i>fabio:AnalogItem</i>		
	<i>E40 Legal Body</i>	<i>P74 has current or former residence (P74i is current or former residence of)</i>	<i>E53 Place</i>
	<i>E22 Man-Made Object</i>	<i>P55 has current location (P55i currently holds)</i>	<i>E53 Place</i>
<i>fabio:AnalogItem</i>			
<i>LDCM</i>	<i>fabio:WorkCollection</i>		<i>E28 Conceptual Object</i>

		<i>P106 is composed of (P106i forms part of)</i>	<i>fabio:Work</i>
		<i>P102 has title (P102i is title of)</i>	<i>E35 Title</i>

Tabella 15. Mapping dei campi esclusivi di paragrafi in comune

5.2.3.4. Esempi di utilizzo

OG	
OGT	
<i>OGTB</i>	<code>:album-12015 a fabio:WorkCollection ; crm:P2_has_type :complex .</code>
QNT	
<i>QNTN</i>	<code>:album-12015-item a fabio:ItemCollection , crm:E22_Man-Made_Object ; crm:P57_has_number_of_parts "3" .</code>
LC	
LDC	
<i>LDCN</i>	<code>:ex-convent-santa-cristina a crm:E53_Place . :zeri-foundation a crm:E40_Legal_Body ; crm:P74_has_current_or_former_residence :ex-convent-santa-cristina ; crm:P50i_is_current_keeper_of :photo-72486-positive-item . :photo-72486-positive-item a crm:E22_Man-Made_Object, fabio:AnalogItem .</code>
<i>LDCM</i>	<code>:photo-72486 a E28_Conceptual_Object, fabio:Work ; crm:P106i_forms_part_of :zeri-photo-archive . :zeri-photo-archive a fabio:WorkCollection ; crm:P102_has_title :zeri-photo-archive-1 .</code>

Tabella 16. Esempi di utilizzo degli elementi mappati

CAPITOLO 6

FASE 3: REALIZZAZIONE DELLE ONTOLOGIE *CICDO E CICDO-F*

La terza e ultima fase di progetto riguarda la costruzione vera e propria dell'ontologia finale, derivata dagli elementi modellati durante la fase di mapping precedente.

Di comune accordo con i colleghi tesisti, l'ontologia di base realizzata è stata chiamata **CICDO**, acronimo che sta per *Central Institute for Cataloguing and Documentation Ontology*.

Dal punto di vista operativo, l'ontologia in questione è stata interamente costruita utilizzando il software open-source *Protégé*. In particolare, sulla base del mapping definito per i paragrafi in comune, ciascun tesista ha realizzato per conto proprio un'ontologia parziale relativa alle informazioni dei paragrafi ad egli assegnati. Una volta concluso questo lavoro, le tre ontologie prodotte sono state “integrate” fra di loro al fine di avere un'unica ontologia che potesse rappresentare il livello inventariale dei paragrafi in comune delle normative considerate. Successivamente, ciascun tesista ha realizzato separatamente un'ulteriore ontologia che modellasse le informazioni relative ai paragrafi e ai campi esclusivi della propria normativa. Questa scelta è stata dettata dal fatto che l'ontologia di base deve rappresentare esclusivamente il “core” del modello concettuale che si vuole costruire per modellare le informazioni relative ai beni culturali. Tutti i concetti esclusivi di una o più tipologie di beni devono essere rappresentati in un'ontologia a parte, al fine di estendere in maniera logica, coerente e naturale l'ontologia di base. Nel caso delle informazioni relative ai beni fotografici assegnati al sottoscritto, l'ontologia prodotta è stata chiamata **CICDO-F**.

Nel capitolo corrente verrà fornita innanzitutto una breve descrizione del software di supporto *Protégé* utilizzato per costruire le diverse ontologie, dopodiché si andranno ad illustrare, in apposite tabelle, le classi e le proprietà create per modellare i concetti definiti durante la fase di mapping. La descrizione degli elementi sarà effettuata considerando prima l'ontologia “core” *CICDO* e successivamente *CICDO-F*.

6.1. Il software *Protégé*

*Protégé*³⁴ è un editor open-source gratuito che permette di semplificare la costruzione e la realizzazione di ontologie del Web semantico attraverso l'utilizzo di apposite interfacce e finestre grafiche per la definizione degli elementi ontologici. La versione utilizzata per costruire le ontologie di progetto è la *5.1.0*, rilasciata a *ottobre 2016*.

Protégé integra al suo interno un *reasoner semantico*³⁵ che è possibile lanciare al fine di verificare la consistenza dell'ontologia costruita e derivare eventuali informazioni implicite sulla base dei concetti espressi nel modello. Il software offre diversi tipi di *reasoner*, tra questi si è scelto di utilizzare l'*HermiT 1.3.8.413* per validare i modelli ontologici costruiti.

In generale, *Protégé* permette di definire concetti e relazioni tra concetti rispettivamente nella forma di classi e proprietà, così come definite dai principi alla base delle tecnologie del Web semantico.

La costruzione di un'ontologia da parte di *Protégé* avviene attraverso la produzione di file in formato *OWL* contenenti i concetti modellati in forma testuale. Nel caso delle ontologie di progetto *CICDO* e *CICDO-F*, le informazioni all'interno del documento sono state serializzate utilizzando il formato *RDF Turtle*.

³⁴ <https://protege.stanford.edu/>

³⁵ Un *reasoner semantico* è una componente software in grado di inferire conseguenze logiche da un insieme di fatti asseriti, detti *assiomi*, modellati tramite appositi formalismi all'interno di un'ontologia.

6.2. Costruzione delle ontologie

Il primo passo per costruire le ontologie di progetto è importare tutte le ontologie, gli schemi e i modelli necessari a rappresentare i diversi concetti modellati durante la fase di mapping. *Protégé* dà la possibilità di specificare quali schemi pubblicati sul Web devono essere utilizzati richiedendo semplicemente all'utente di fornire il corrispondente URL. E' possibile eventualmente fare riferimento anche a schemi presenti in file locali, ma poiché l'obiettivo del progetto in questione (e in generale del Web semantico) è produrre un modello ontologico condiviso e riusabile a livello globale, tutti gli schemi sono stati importati facendo riferimento esclusivamente ai rispettivi URL.

La strategia principale utilizzata per modellare le classi e le proprietà tra classi definite nel mapping è stata quella di evitare il riutilizzo di concetti già presenti, se non strettamente necessario. Questa scelta ha permesso di dare maggiore espressività alla semantica dei concetti modellati. Dal punto di vista pratico, tale strategia è stata applicata definendo i concetti modellati come *sottoclassi* o *sottoproprietà* rispettivamente delle classi e delle proprietà riportate nel mapping.

Prima di passare a illustrare i concetti definiti nelle ontologie prodotte, è necessario fare una distinzione tra i due tipi di proprietà che è possibile modellare in *Protégé*.

- **Object Property (OP).** Una *object property* permette di mettere in relazione due individui, siano essi di classi distinte o della stessa classe. Ad esempio, la proprietà *P52 has current owner (is current owner of)* di CIDOC-CRM è una *object property*, in quanto mette in relazione istanze della classe *E18 Physical Thing* con istanze della classe *E39 Actor*.
- **Data Property (DP).** Una *data property* permettere di mettere in relazione un individuo con un valore di tipo primitivo. Ad esempio, la proprietà *P90 has value* di CIDOC-CRM è una *data property*, in quanto mette in relazione un'istanza della classe *E54 Dimension* con un valore numerico.

In generale, sappiamo che una proprietà descrive una caratteristica specifica di un elemento generico. L'elemento a cui viene riferita la proprietà è detto *domain*, mentre l'elemento che rappresenta il valore di tale proprietà è indicato con la parola *range*. Per alcune proprietà, il dominio o il range potrebbero non essere specificati perché riferiti a più di una classe.

6.2.1. Ontologia *CICDO*

Nota. Nelle tabelle di questo paragrafo saranno riportate soltanto le classi e le proprietà modellate relativamente ai paragrafi in comune assegnati al sottoscritto.

6.2.1.1. Definizione delle classi

Classe <i>CICDO</i>	Superclasse	Descrizione
<i>Cultural Heritage (Conceptual Object)</i>	<i>E28 Conceptual Object fabio:Work</i>	Classe dei beni culturali intesi come oggetti concettuali (i.e. non tangibili o riferiti a paesaggi o tradizioni).
<i>Cultural Heritage (Item Collection)</i>	<i>E22 Man-Made Object fabio:ItemCollection</i>	Classe di insiemi di beni culturali intesi come oggetti tangibili.
<i>Cultural Heritage (Physical Object)</i>	<i>E22 Man-Made Object fabio:AnalogItem</i>	Classe dei beni culturali intesi come oggetti fisici e tangibili.
<i>Cultural Heritage (Work Collection)</i>	<i>E28 Conceptual Object fabio:WorkCollection</i>	Classe di insiemi di beni culturali intesi come oggetti concettuali (i.e. non tangibili o riferiti a paesaggi o tradizioni).

<i>Cultural Heritage Physical Part</i>	<i>E22 Man-Made Object</i>	Classe delle parti da cui può essere composto un bene culturale.
<i>Creation Type</i>	<i>E55 Type</i>	Classe dei tipi di attività di creazione.
<i>Production Type</i>	<i>E55 Type</i>	Classe dei tipi di attività di produzione.
<i>Scheda</i>	<i>E31 Document</i>	Classe delle schede di catalogo.
<i>(CDGG) Acquisition Type</i>	<i>E55 Type</i>	Classe dei tipi di acquisizione di un bene culturale.
<i>(CDGS) Current Owner</i>	<i>E39 Actor</i>	Classe dei proprietari di un bene culturale.
<i>(DTMM) Date Attribution</i>	<i>prov:Activity hico:InterpretationAct</i>	Classe dei criteri di interpretazione per l'attribuzione di date.
<i>(DTZG) Century Span</i>	<i>E52 Time-Span</i>	Classe dei secoli.
<i>(MISA) Height Dimension</i>	<i>E54 Dimension</i>	Classe delle dimensioni di tipo altezza.
<i>(MISD) Diameter Dimension</i>	<i>E54 Dimension</i>	Classe delle dimensioni di tipo diametro.
<i>(MISL) Width Dimension</i>	<i>E54 Dimension</i>	Classe delle dimensioni di tipo larghezza.
<i>(MISS) Thickness Dimension</i>	<i>E54 Dimension</i>	Classe delle dimensioni di tipo spessore.
<i>(MISU) Measurement Type</i>	<i>E55 Type</i>	Classe dei tipi di misure.

(MTC) Material	<i>E57 Material</i>	Classe dei materiali.
(MTX) Colour	<i>E26 Physical Feature</i>	Classe dei colori.

Tabella 17. Classi definite per l'ontologia CICDO

Oltre all'introduzione delle nuove classi descritte in tabella, sono state modellate le seguenti **equivalenze**:

- *Cultural Heritage (Conceptual Object)* ↔ *fabio:Work*
- *Cultural Heritage (Physical Object)* ↔ *fabio:analogManifestation*

6.2.1.2. Definizione delle proprietà

6.2.1.2.1. Object property

<i>changed ownership through</i>	<i>Superproprietà: P24 changed ownership through</i> <i>Domain: Cultural Heritage (Physical Object)</i> <i>Range: E8 Acquisition</i> <i>Proprietà inversa: transferred title of</i>
<i>has created</i>	<i>Superproprietà: P94 has created</i> <i>Domain: E65 Creation</i> <i>Range: Cultural Heritage (Conceptual Object)</i> <i>Proprietà inversa: was created by</i>
<i>has creation type</i>	<i>Superproprietà: P2 has type</i> <i>Domain: E65 Creation</i> <i>Range: Creation Type</i> <i>Proprietà inversa: is creation type of</i>

<i>has dimension</i>	<p><i>Superproprietà:</i> P43 has dimension</p> <p><i>Domain:</i> Cultural Heritage (Physical Object)</p> <p><i>Range:</i> E54 Dimension</p> <p><i>Proprietà inversa:</i> is dimension of</p>
<i>has dimension type</i>	<p><i>Superproprietà:</i> P2 has type</p> <p><i>Domain:</i> E54 Dimension</p> <p><i>Range:</i></p> <p><i>Proprietà inversa:</i> is dimension type of</p>
<i>has identifier</i>	<p><i>Superproprietà:</i> P48 has preferred identifier</p> <p><i>Domain:</i> Cultural Heritage (Physical Object)</p> <p><i>Range:</i> (CD) Codes³⁶</p> <p><i>Proprietà inversa:</i> identifies</p>
<i>has produced</i>	<p><i>Superproprietà:</i> P108 has produced</p> <p><i>Domain:</i> E12 Production</p> <p><i>Range:</i> Cultural Heritage (Physical Object)</p> <p><i>Proprietà inversa:</i> was produced by</p>
<i>has production type</i>	<p><i>Superproprietà:</i> P2 is type of</p> <p><i>Domain:</i> E12 Production</p> <p><i>Range:</i> Production Type</p> <p><i>Proprietà inversa:</i> is production type of</p>
<i>is creation type of</i>	<p><i>Superproprietà:</i> P2 is type of</p> <p><i>Domain:</i> Creation Type</p> <p><i>Range:</i> E65 Creation</p> <p><i>Proprietà inversa:</i> has creation type</p>

³⁶ Modellata dal collega responsabile del mapping della normativa RA.

<i>is dimension of</i>	<p>Superproprietà: P43 is dimension of</p> <p>Domain: E54 Dimension</p> <p>Range: Cultural Heritage (Physical Object)</p> <p>Proprietà inversa: has dimension</p>
<i>is dimension type of</i>	<p>Superproprietà: P2 is type of</p> <p>Domain:</p> <p>Range: E54 Dimension</p> <p>Proprietà inversa: has dimension type</p>
<i>is production type of</i>	<p>Superproprietà: P2 is type of</p> <p>Domain: Production Type</p> <p>Range: E12 Production</p> <p>Proprietà inversa: has production type</p>
<i>transferred title of</i>	<p>Superproprietà: P24 transferred title of</p> <p>Domain: E8 Acquisition</p> <p>Range: Cultural Heritage (Physical Object)</p> <p>Proprietà inversa: changed ownership through</p>
<i>was created by</i>	<p>Superproprietà: P94 was created by</p> <p>Domain: Cultural Heritage (Conceptual Object)</p> <p>Range: E65 Creation</p> <p>Proprietà inversa: has created</p>
<i>was produced by</i>	<p>Superproprietà: P108 was produced by</p> <p>Domain: Cultural Heritage (Physical Object)</p> <p>Range: E12 Production</p> <p>Proprietà inversa: has produced</p>
<i>(CDGG) has acquisition type</i>	<p>Superproprietà: P2 has type</p> <p>Domain: Cultural Heritage (Physical Object)</p> <p>Range: (CDGG) Acquisition Type</p> <p>Proprietà inversa: (CDGG) is acquisition type of</p>

<p>(CDGG) <i>is acquisition type</i></p>	<p><i>Superproprietà: P2 is type of</i> <i>Domain: (CDGG) Acquisition Type</i> <i>Range: Cultural Heritage (Physical Object)</i> <i>Proprietà inversa: (CDGG) has acquisition type</i></p>
<p>(CDGS) <i>has current owner</i></p>	<p><i>Superproprietà: P52 has current owner</i> <i>Domain: E8 Acquisition</i> <i>Range: (CDGS) Current Owner</i> <i>Proprietà inversa: (CDGS) is current owner of</i></p>
<p>(CDGS) <i>is current owner of</i></p>	<p><i>Superproprietà: P52 is current owner of</i> <i>Domain: (CDGS) Current Owner</i> <i>Range: E8 Acquisition</i> <i>Proprietà inversa: (CDGS) has current owner</i></p>
<p>(DTMM) <i>attributed</i></p>	<p><i>Superproprietà: prov:influenced</i> <i>Domain: (DTMM) Date Attribution</i> <i>Range: -</i> <i>Proprietà inversa: (DTMM) was attributed by</i></p>
<p>(DTMM) <i>was attributed by</i></p>	<p><i>Superproprietà: prov:wasGeneratedBy</i> <i>Domain: -</i> <i>Range: (DTMM) Date Attribution</i> <i>Proprietà inversa: (DTMM) attributed</i></p>
<p>(DTZG) <i>has century span</i></p>	<p><i>Superproprietà: P4 has time-span</i> <i>Domain: -</i> <i>Range: (DTZG) Century Span</i> <i>Proprietà inversa: (DTZG) is century span of</i></p>
<p>(DTZG) <i>is century span of</i></p>	<p><i>Superproprietà: P4 is time-span of</i> <i>Domain: (DTZG) Century Span</i> <i>Range: -</i> <i>Proprietà inversa: (DTZG) has century span</i></p>

<p>(MISO) <i>forms part of</i></p>	<p><i>Superproprietà: P106 forms part of</i> <i>Domain: Cultural Heritage Physical Part</i> <i>Range: Cultural Heritage (Physical Object)</i> <i>Proprietà inversa: (MISO) is composed of</i></p>
<p>(MISO) <i>is composed of</i></p>	<p><i>Superproprietà: P106 is composed of</i> <i>Domain: Cultural Heritage (Physical Object)</i> <i>Range: Cultural Heritage Physical Part</i> <i>Proprietà inversa: (MISO) forms part of</i></p>
<p>(MISU) <i>has measurement type</i></p>	<p><i>Superproprietà: P2 has type</i> <i>Domain: E16 Measurement</i> <i>Range: (MISU) Measurement Type</i> <i>Proprietà inversa: (MISU) is measurement type of</i></p>
<p>(MISU) <i>is measurement type of</i></p>	<p><i>Superproprietà: P2 is type of</i> <i>Domain: (MISU) Measurement Type</i> <i>Range: E16 Measurement</i> <i>Proprietà inversa: (MISU) has measurement type</i></p>
<p>(MTC) <i>has material</i></p>	<p><i>Superproprietà: P2 has type</i> <i>Domain: Cultural Heritage (Physical Object)</i> <i>Range: (MTC) Material</i> <i>Proprietà inversa: (MTC) is material of</i></p>
<p>(MTC) <i>is material of</i></p>	<p><i>Superproprietà: P2 is type of</i> <i>Domain: (MTC) Material</i> <i>Range: Cultural Heritage (Physical Object)</i> <i>Proprietà inversa: (MTC) has material</i></p>
<p>(MTX) <i>has colour</i></p>	<p><i>Superproprietà: P56 bears feature</i> <i>Domain: Cultural Heritage (Physical Object)</i> <i>Range: (MTX) Colour</i> <i>Proprietà inversa: (MTX) is colour of</i></p>

(MTX) <i>is colour of</i>	<i>Superproprietà: P56 is found on</i> <i>Domain: (MTX) Colour</i> <i>Range: Cultural Heritage (Physical Object)</i> <i>Proprietà inversa: (MTX) has colour</i>
-------------------------------------	--

Tabella 18. Object property definite per l'ontologia CICDO

6.2.1.2.2. Data property

(DTSF) <i>has end date</i>	<i>Superproprietà: ti:hasIntervalEndDate</i> <i>Domain: (DTZG) Century Span</i> <i>Range: xsd:dateTime</i>
(DTSI) <i>has start date</i>	<i>Superproprietà: ti:hasIntervalStartDate</i> <i>Domain: (DTZG) Century Span</i> <i>Range: xsd:dateTime</i>

Tabella 19. Data property definite per l'ontologia CICDO

6.2.2. Ontologia CICDO-F

6.2.2.1. Definizione delle classi

Classe CICDO	Superclasse	Descrizione
<i>Photograph</i> <i>(Conceptual Object)</i>	<i>Cultural Heritage</i> <i>(Conceptual Object)</i>	Classe delle fotografie intese come oggetti concettuali (i.e. non tangibili o riferiti a paesaggi o tradizioni).

<i>Photograph (Physical Object)</i>	<i>Cultural Heritage (Physical Object)</i>	Classe delle fotografie intese come oggetti fisici e tangibili.
<i>Photograph Collection (Conceptual Object)</i>	<i>Cultural Heritage (Work Collection)</i>	Classe di insiemi di fotografie intese come oggetti concettuali (i.e. non tangibili o riferiti a paesaggi o tradizioni).
<i>(AUF) Photograph Author</i>	<i>E21 Person</i>	Classe degli autori responsabili della creazione di una fotografia.
<i>(LDCM) Collection Name</i>	<i>E35 Title</i>	Classe dei nomi delle raccolte a cui una fotografia appartiene.
<i>(LDCN) Current Keeper</i>	<i>E40 Legal Body</i>	Classe dei proprietari o custodi legali di una fotografia.
<i>(LDCN) Current Keeper Residence</i>	<i>E53 Place</i>	Classe dei luoghi in cui i custodi legali di una fotografia risiedono.
<i>(LDCN) Current Location</i>	<i>E53 Place</i>	Classe dei luoghi in cui una fotografia è conservata.
<i>(OGTB) Librarianship Context</i>	<i>E55 Type</i>	Classe dei tipi di contesti biblioteconomici di una fotografia.
<i>(SGL) Photo Title</i>	<i>E35 Title</i>	Classe dei titoli attribuiti a una fotografia.
<i>(SGL) Title Type</i>	<i>E55 Type</i>	Classe dei tipi di titoli assegnati a una fotografia.
<i>(SGLS) Title Attribution</i>	<i>prov:Activity hico:InterpretationAct</i>	Classe dei criteri di interpretazione per l'attribuzione di titoli a una fotografia.

(SGTI) <i>Photo Subject</i>	<i>fabio:SubjectTerm</i>	Classe dei soggetti raffigurati in una fotografia.
(UBFC) <i>Physical Collection Placing</i>	<i>E53 Place</i>	Classe delle collocazioni fisiche di una fotografia all'interno di una collezione.
(UBFC) <i>Physical Collection Identifier</i>	<i>E42 Identifier</i>	Classe delle sigle attribuite alle collocazioni fisiche.
(UBFP) <i>Archive Name</i>	<i>E35 Title</i>	Classe dei nomi degli archivi a cui una fotografia appartiene.

Tabella 20. Classi definite per l'ontologia CICDO-F

6.2.2.2. Definizione delle proprietà

6.2.2.2.1. Object property

(author) has created	<p><i>Superproprietà: P14 performed</i></p> <p><i>Domain: (AUF) Photograph Author</i></p> <p><i>Range: E65 Creation</i></p> <p><i>Proprietà inversa: created by (author)</i></p>
created by (author)	<p><i>Superproprietà: P14 carried out by</i></p> <p><i>Domain: E65 Creation</i></p> <p><i>Range: (AUF) Photograph Author</i></p> <p><i>Proprietà inversa: (author) has created</i></p>

<p><i>is composed of photo</i></p>	<p><i>Superproprietà: P106 is composed of</i> <i>Domain: Photograph Collection (Conceptual Object)</i> <i>Range: Photograph (Conceptual Object)</i> <i>Proprietà inversa: is photo of collection</i></p>
<p><i>is photo of collection</i></p>	<p><i>Superproprietà: P106 forms part of</i> <i>Domain: Photograph (Conceptual Object)</i> <i>Range: Photograph Collection (Conceptual Object)</i> <i>Proprietà inversa: is composed of photo</i></p>
<p><i>(LDCM)</i> <i>has collection name</i></p>	<p><i>Superproprietà: P102 has title</i> <i>Domain: Photograph Collection (Conceptual Object)</i> <i>Range: (LDCM) Collection Name</i> <i>Proprietà inversa: is collection name of</i></p>
<p><i>(LDCN)</i> <i>has current keeper</i></p>	<p><i>Superproprietà: P50 has current keeper</i> <i>Domain: Photograph (Physical Object)</i> <i>Range: (LDCN) Current Keeper</i> <i>Proprietà inversa: (LDCN) is current keeper of</i></p>
<p><i>(LDCN)</i> <i>has current location</i></p>	<p><i>Superproprietà: P55 has current location</i> <i>Domain: Photograph (Physical Object)</i> <i>Range: (LDCN) Current Location</i> <i>Proprietà inversa: (LDCN) is current location of</i></p>
<p><i>(LDCN)</i> <i>has current residence</i></p>	<p><i>Superproprietà: P74 has current or former residence</i> <i>Domain: (LDCN) Current Keeper</i> <i>Range: (LDCN) Current Keeper Residence</i> <i>Proprietà inversa: (LDCN) is current residence of</i></p>
<p><i>(LDCM)</i> <i>is collection name of</i></p>	<p><i>Superproprietà: P102 is title of</i> <i>Domain: (LDCM) Collection Name</i> <i>Range: Photograph Collection (Conceptual Object)</i> <i>Proprietà inversa: has collection name</i></p>

<p>(LDCN) <i>is current keeper of</i></p>	<p><i>Superproprietà: P50 is current keeper of</i> Domain: (LDCN) Current Keeper Range: Photograph (Physical Object) Proprietà inversa: (LDCN) has current keeper</p>
<p>(LDCN) <i>is current location of</i></p>	<p><i>Superproprietà: P55 currently holds</i> Domain: (LDCN) Current Location Range: Photograph (Physical Object) Proprietà inversa: (LDCN) has current location</p>
<p>(LDCN) <i>is current residence of</i></p>	<p><i>Superproprietà: P74 is current or former residence of</i> Domain: (LDCN) Current Keeper Residence Range: (LDCN) Current Keeper Proprietà inversa: (LDCN) has current residence</p>
<p>(OGTB) <i>has librarianship context</i></p>	<p><i>Superproprietà: P2 has type</i> Domain: - Range: (OGTB) Librarianship Context Proprietà inversa: (OGTB) is librarianship context of</p>
<p>(OGTB) <i>is librarianship context of</i></p>	<p><i>Superproprietà: P2 is type of</i> Domain: (OGTB) Librarianship Context Range: - Proprietà inversa: (OGTB) has librarianship context</p>
<p>(SGL) <i>has photo title</i></p>	<p><i>Superproprietà: P102 has title</i> Domain: Photograph (Conceptual Object) Range: (SGL) Photo Title Proprietà inversa: (SGL) is photo title</p>
<p>(SGL) <i>is photo title of</i></p>	<p><i>Superproprietà: P102 is title of</i> Domain: (SGL) Photo Title Range: Photograph (Conceptual Object) Proprietà inversa: (SGL) has photo title</p>

<p>(SGL) <i>has title type</i></p>	<p><i>Superproprietà: P2 has type</i> <i>Domain: (SGL) Photo Title</i> <i>Range: (SGL) Title Type</i> <i>Proprietà inversa: (SGL) is title type of</i></p>
<p>(SGL) <i>is title type of</i></p>	<p><i>Superproprietà: P2 is type of</i> <i>Domain: (SGL) Title Type</i> <i>Range: (SGL) Photo Title</i> <i>Proprietà inversa: (SGL) has title type</i></p>
<p>(SGLS) <i>attributed</i></p>	<p><i>Superproprietà: prov:influenced</i> <i>Domain: (SGLS) Title Attribution</i> <i>Range: (SGL) Photo Title</i> <i>Proprietà inversa: (SGLS) was attributed by</i></p>
<p>(SGLS) <i>was attributed by</i></p>	<p><i>Superproprietà: prov:wasGeneratedBy</i> <i>Domain: (SGL) Photo Title</i> <i>Range: (SGLS) Title Attribution</i> <i>Proprietà inversa: (SGLS) attributed</i></p>
<p>(SGTI) <i>has photo subject</i></p>	<p><i>Superproprietà: fabio:hasSubjectTerm</i> <i>Domain: Photograph (Conceptual Object)</i> <i>Range: (SGTI) Photo Subject</i> <i>Proprietà inversa: -</i></p>
<p>(UBFC) <i>has physical collection identifier</i></p>	<p><i>Superproprietà: P48 has preferred identifier</i> <i>Domain: (UBFC) Physical Collection Placing</i> <i>Range: (UBFC) Physical Collection Identifier</i> <i>Proprietà inversa: (UBFC) is physical collection identifier of</i></p>
<p>(UBFC) <i>has physical collection placing</i></p>	<p><i>Superproprietà: P55 has current location</i> <i>Domain: Photograph (Physical Object)</i> <i>Range: (UBFC) Physical Collection Placing</i> <i>Proprietà inversa: (UBFC) is physical collection placing of</i></p>

(UBFC) <i>is physical collection identifier of</i>	<i>Superproprietà: P48 is preferred identifier of</i> Domain: (UBFC) Physical Collection Identifier Range: (UBFC) Physical Collection Placing Proprietà inversa: (UBFC) has physical collection identifier
(UBFC) <i>is physical collection placing of</i>	<i>Superproprietà: P55 currently holds</i> Domain: (UBFC) Physical Collection Placing Range: Photograph (Physical Object) Proprietà inversa: (UBFC) has physical collection placing
(UBFP) <i>has archive name</i>	<i>Superproprietà: P102 has title</i> Domain: Photograph Collection (Conceptual Object) Range: (UBFP) Archive Name Proprietà inversa: (UBFP) is archive name of
(UBFP) <i>is archive name of</i>	<i>Superproprietà: P102 is title of</i> Domain: (UBFP) Archive Name Range: Photograph Collection (Conceptual Object) Proprietà inversa: (UBFP) has archive name

Tabella 21. Object property definite per l'ontologia CICDO-F

6.2.2.2.2. Data property

(QNTN) <i>has number of parts</i>	Superproprietà: P57 has number of parts Domain: - Range: xsd:integer
---	---

Tabella 22. Data property definite per l'ontologia CICDO-F

CAPITOLO 7

UNA PROPOSTA DI APPLICAZIONE PER LA CONVERSIONE DI SCHEDE ICCD IN DOCUMENTI RDF

Su proposta avanzata dall'azienda committente, al termine della costruzione dell'ontologia è stata progettata e sviluppata un'applicazione minimale per la conversione delle schede di catalogo ICCD in formato XML, gestite dal software *Odysseus* dell'azienda, in documenti strutturati secondo le regole sintattiche di RDF, sulla base dei concetti modellati all'interno dell'ontologia *CICDO*.

L'applicazione in questione è stata sviluppata utilizzando *Java* come linguaggio per l'implementazione ed *Eclipse Neon.3* come ambiente di sviluppo. Le modalità di conversione dal formato XML al formato RDF sono state definite formalmente attraverso l'utilizzo della tecnologia *XSLT*, di cui se ne discuteranno gli aspetti di base nel prossimo paragrafo.

Grazie allo sviluppo di questa applicazione, non solo si è potuto disporre di uno strumento utile per la conversione delle schede di catalogo in un formato accettato e interpretabile dalle tecnologie del Web semantico, ma si è anche potuto verificare la coerenza e la consistenza dell'ontologia realizzata rispetto ai concetti definiti per essa, attraverso l'esecuzione di diversi test di conversione su schede di catalogo reali fornite dall'azienda committente.

Oltre a fare una breve panoramica sugli aspetti di base di *XSLT*, in questo capitolo verrà riportata una descrizione a livello funzionale dell'applicazione sviluppata, senza scendere eccessivamente in dettagli di tipo tecnologico. Al termine della discussione

verrà mostrato, a titolo di esempio, un caso di conversione di un'informazione contenuta all'interno di una scheda di catalogo in una tripla RDF.

7.1. XSLT come tecnologia di conversione

XSLT (*eXtensible Stylesheet Language Transformations*) è un linguaggio nato allo scopo di permettere la trasformazione di documenti XML in altri documenti XML o in documenti con formati diversi. E' una delle componenti che costituiscono *XSL*, il linguaggio di descrizione dei fogli di stile per i documenti XML. Per questo motivo, un file in formato XSLT è spesso riferito anche con il nome di *foglio di stile XSLT*.

XSLT è diventato uno standard web del *W3C* nel novembre 1999, e la sua potenza espressiva permette di produrre documenti strutturati in formati testuali diversi in maniera semplice e controllata.

7.1.1. Modello di trasformazione

La componente software di XSLT responsabile della trasformazione dei documenti XML è il *processore XSLT*. Come riportato in precedenza, si distinguono due casi specifici di conversione:

1. da un documento XML a un altro documento XML;
2. da un documento XML a un documento strutturato secondo altri formati.

Nel secondo caso, il processore XSLT è in grado di generare esclusivamente file di tipo testuale, quindi anche documenti RDF. Questa tecnologia, dunque, ben si adatta allo scopo della nostra applicazione.

Dal punto di vista funzionale, un processore XSLT opera nel modo seguente.

- 1. Input.** Il processore riceve in input due file: il documento XML da convertire e un file XSLT, il quale definisce la semantica per la trasformazione dei contenuti del documento XML.
- 2. Elaborazione.** Seguendo la struttura gerarchica del documento di partenza, il processore elabora ciascun nodo secondo le modalità di trasformazione definite all'interno del file XSLT. In particolare, quando il processore trova una corrispondenza (*matching*), applica le relative regole di trasformazione al contenuto del nodo individuato.
- 3. Output.** Al termine delle conversioni, il processore produce un nuovo documento nel formato specificato, lasciando comunque intatto il documento XML di partenza.

Per effettuare la trasformazione, il processore XSLT valuta il documento XML in input come un insieme di nodi strutturati ad albero. Il matching di ciascun elemento viene definito all'interno del file XSLT attraverso la specifica del percorso del corrispondente nodo da trasformare, strutturato secondo il formato definito da *XPath*³⁷.

In generale, le regole di conversione riportate all'interno di un file XSLT sono formalizzate attraverso la definizione di un insieme di modelli, detti *template*. Le regole contenute all'interno di un template sono descritte dal punto di vista dichiarativo; in altre parole, i template nei file XSLT non specificano sequenze di comandi da eseguire per effettuare la conversione dei documenti XML, ma definiscono semplicemente la struttura che il nuovo documento dovrà avere.

In [Figura 7](#) viene illustrato il modello di trasformazione discusso finora.

³⁷ *XPath* è un linguaggio della famiglia XML che permette di individuare e localizzare con precisione i nodi all'interno di un documento XML.

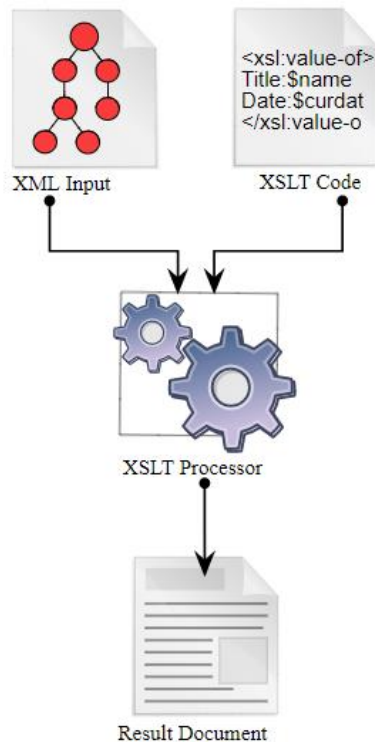


Figura 7. Modello di trasformazione di XSLT

7.1.2. Elementi sintattici di interesse

In questo paragrafo verranno elencati e discussi brevemente gli elementi sintattici principali utilizzati per definire il contenuto del file XSLT dell'applicazione in questione.

- **Template.** L'elemento `<xsl:template>` è utilizzato per costruire template. Un template può essere associato esplicitamente a un nodo del documento XML di input attraverso la definizione di un attributo `match`. Il valore di tale attributo è riportato secondo il formato definito da *XPath*, e indica il percorso del nodo a cui fare riferimento per la trasformazione nel nuovo formato. Il contenuto dell'elemento `<xsl:template>` riporta le regole di trasformazione effettive del nodo individuato dal percorso definito all'interno dell'attributo

match. Un valore pari a “/” indica che le regole di conversione faranno riferimento all’intero documento.

- **Valore di un nodo.** Il valore di un nodo può essere estratto dal documento di input e visualizzato nel nuovo documento attraverso la specifica dell’elemento `<xsl:value-of>`. Utilizzando l’attributo `select`, si può specificare il percorso del nodo da cui estrarre il valore.
- **Selezione multipla.** L’elemento `<xsl:for-each>` permette di selezionare uno o più elementi da un insieme di nodi, secondo il percorso specificato all’interno dell’attributo `select`. Questo elemento risulta particolarmente utile nel caso dell’applicazione in questione per trasformare correttamente i campi delle schede ICCD che presentano proprietà di ripetitività.
- **Variabili.** L’elemento `<xsl:variable>` permette di definire variabili, qui intese in realtà come vere e proprie costanti il cui valore può essere riutilizzato all’interno di un template. Il nome di una variabile è specificato attraverso l’attributo `name`, mentre il valore è riportato internamente ai tag.

7.2. Architettura e strategie di sviluppo

7.2.1. Struttura del processo di conversione

L’applicazione in questione è stata progettata con l’idea di realizzare una sorta di *Web Service*, fornendo quindi un’interfaccia che espone pubblicamente un insieme di operazioni offerte dall’applicazione.

Il processo di conversione di una scheda di catalogo in un documento RDF è strutturato in tre parti, ciascuna corrispondente a una delle operazioni esposte dall’interfaccia.

1. **Caricamento file XML.** Il file da convertire, il cui nome è specificato in input, viene prelevato da una cartella predefinita.

2. **Parsing del file caricato.** Il file caricato viene validato al fine di verificare che sia sintatticamente corretto e che rispetti gli standard definiti nella normativa relativa al tipo di bene catalogato nella scheda.
3. **Trasformazione del file.** Se la validazione è andata a buon fine, il file XML in input viene convertito in un documento RDF secondo le regole di trasformazione definite all'interno del template contenuto in un apposito file XSLT.

7.2.2. Architettura generale

Dal punto di vista architetturale, l'applicazione è stata organizzata logicamente in tre package³⁸ Java distinti.

1. *Package core.* Contiene le componenti core dell'applicazione, ovvero l'interfaccia delle operazioni e la logica di implementazione del programma.
2. *Package resources.* Contiene le configurazioni del programma.
3. *Package util.* Contiene componenti di supporto al core del programma, tra cui un sistema di log.

Le operazioni di caricamento e parsing del documento XML sono state implementate utilizzando apposite librerie fornite dal linguaggio Java. Il sistema di log integrato nell'applicazione permette di memorizzare in un apposito file tutti gli errori e i messaggi generati dal programma durante la sua esecuzione. Tale sistema è utile in particolare per verificare se la conversione delle schede avviene in maniera corretta.

Per gestire i file di input e output sono state predisposte apposite cartelle:

- una cartella contenente i *documenti XML* da convertire;
- una cartella contenente i *documenti RDF* generati dall'applicazione;

³⁸ Un *package* è un meccanismo utilizzato in Java per organizzare le componenti di un'applicazione in gruppi di classi logicamente correlate.

- una cartella contenente un insieme di *file XSD*³⁹ corrispondenti alle normative prodotte dall'ICCD, utilizzati al momento della validazione del documento XML per verificare che esso rispettivi gli standard normativi.

Come si accennava, le regole di trasformazione dei documenti XML sono specificate all'interno di un apposito template XSLT, di cui si discuterà il contenuto nel prossimo paragrafo.

7.2.3. Il template XSLT

Il file XSLT predisposto per l'applicazione è stato costruito coerentemente all'insieme di concetti modellati nell'ontologia *CICDO*.

Per descrivere la struttura che il documento RDF da produrre dovrà avere è stato definito un unico template. Nel complesso, possiamo suddividere logicamente la struttura del file XSLT in tre sezioni.

1. **Importazione degli schemi.** Attraverso la definizione dei rispettivi URL, tutti gli schemi, i modelli e le ontologie utilizzate all'interno dell'ontologia *CICDO* vengono riportati all'inizio del file.
2. **Inizializzazione.** Per minimizzare la ripetitività del codice all'interno del template, sono state definite due variabili: l'URL dell'ontologia *CICDO*, utilizzato per riferirsi ai relativi termini, e l'identificatore univoco del bene catalogato nella scheda in input, derivato dalla concatenazione dei valori dei campi contenuti all'interno del campo *NCT* del paragrafo *CD*.
3. **Template e regole di trasformazione.** All'interno del template sono state modellate le varie triple RDF corrispondenti ai concetti rappresentati nell'ontologia *CICDO*. La tripla principale rappresenta il bene culturale

³⁹ *XML Schema Definition* è il linguaggio di schema standard corrente per XML, che permette di definirne la struttura e i tipi di dati consentiti.

catalogato nella scheda, cui segue la definizione di tutte le triple derivate dalle entità collegate ad essa.

7.2.4. Un esempio di conversione

A titolo d'esempio, si riporta di seguito un caso di conversione di un campo di una scheda relativa a un bene fotografico nella corrispondente tripla RDF, attraverso le modalità di trasformazione definite all'interno del file XSLT. Il campo in questione è il campo semplice *MTX* del paragrafo *MT*.

```
...
<MT>
  <MTX>B/N</MTX>
</MT>
...
```

Scheda XML: campo MTX da convertire

```
...
<xsl:variable name="ontologyURI">
  http://www.semanticweb.org/ontologies/cultural_heritage/CICDO
</xsl:variable>
<xsl:variable name="uniqueIdentifier"
  select="concat (/* /CD/NCT/NCTR/text(), /* /CD/NCT/NCTN/text())"
/>
...
<rdf:Description rdf:about="{ $ontologyURI }#{ $uniqueIdentifier }">
  <rdf:type rdf:resource=
    "{ $ontologyURI }#Cultural_Heritage_Physical_Object" />
  ...
  <cicdo:has_colour
    rdf:resource="{ $ontologyURI }#{ $uniqueIdentifier }-MTX" />
  ...
</rdf:Description>
```

```

...
<xsl:variable name="MTX" select="*/MT/MTX/text()" />

<rdf:Description
  rdf:about="{ $ontologyURI }#{ $uniqueIdentifier }-MTX">
  <rdf:type rdf:resource="{ $ontologyURI }#Colour" />
  <cicdo:has_colour_value>{ $MTX }</cicdo:has_colour_value>
</rdf:Description>

```

Template XSLT: definizione delle regole di trasformazione

```

...
<rdf:Description rdf:about="http://www.semanticweb.org/
  cultural_heritage/CICDO/CICDO#03">
  <rdf:type rdf:resource="http://www.semanticweb.org/
  cultural_heritage/CICDO/
  CICDO#Cultural_Heritage_Physical_Object" />
  ...
  <cicdo:has_colour
    rdf:resource="http://www.semanticweb.org/
    cultural_heritage/CICDO/CICDO#03-MTX" />
  ...
</rdf:Description>
...
<rdf:Description rdf:about="http://www.semanticweb.org/
  cultural_heritage/CICDO/CICDO#03-MTX">
  <rdf:type rdf:resource="http://www.semanticweb.org/
  cultural_heritage/CICDO/CICDO#Colour" />
  <cicdo:has_colour_value>B/N</cicdo:has_colour_value>
</rdf:Description>

```

Documento RDF: triple relative al campo MTX

7.2.5. Osservazioni

L'applicazione realizzata rappresenta soltanto una proposta e un punto di partenza per realizzare applicazioni di conversione più avanzate. Sebbene tale applicazione raggiunga perfettamente il suo scopo (trasformare schede XML in documenti RDF), al momento non è fornita di alcuna interfaccia grafica. Inoltre, essa non è ancora utilizzabile come un vero e proprio Web Service, in quanto le operazioni implementate vengono semplicemente eseguite in sequenza all'avvio dell'applicazione.

In secondo luogo, su richiesta della stessa azienda committente, l'applicazione è stata parametrizzata per permettere anche la conversione di schede di catalogo che non aderiscono rigorosamente alle specifiche degli standard catalografici ICCD. Questa scelta è stata dettata dal fatto che l'obiettivo principale è avere al momento uno strumento semplice e minimale che converta le schede di catalogo, a prescindere che queste rispettino o meno i vincoli normativi dell'ICCD.

I parametri impostati per “bypassare” la validazione sono i seguenti:

- 1. Controllo delle obbligtorietà.** Qualora alcuni campi con obbligtorietà assoluta non siano presenti nella scheda XML passata in input al programma, l'applicazione non si deve bloccare, ma deve memorizzare messaggi di avviso nel file di log che riportino i campi obbligatori assenti.
- 2. Standardizzazione del nome del nodo radice.** Qualora il nome del nodo radice, che rappresenta l'intera scheda, sia diverso dal valore “*scheda*”, l'applicazione non si deve bloccare, ma deve trasformare il nome del nodo nel nome accettato dalla normativa.
- 3. Attributi.** Tutti gli attributi presenti negli elementi delle schede vengono rimossi attraverso una procedura ad hoc, in quanto le normative non modellano attributi all'interno dei nodi.

CONCLUSIONI E SVILUPPI FUTURI

Le ontologie descritte nel corso della trattazione corrente, come ribadito in diversi punti, rappresentano semplici proposte di modelli, da cui però le istituzioni del mondo culturale possono trarre notevole potenziale. Tuttavia, il lavoro da fare è ancora parecchio.

Innanzitutto, l'ontologia di base *CICDO* è in grado attualmente di rappresentare solo le informazioni relative al livello inventariale delle normative ICCD. Ciò permette agli utilizzatori del modello di descrivere qualsiasi scheda di catalogo ritenuta valida dagli standard catalografici, tuttavia non dà la possibilità di rappresentare le informazioni dei restanti livelli di catalogazione. Da questo punto di vista, l'ontologia proposta è ancora incompleta.

Oltre al livello di completezza, è necessario considerare anche il discorso della manutenzione dell'ontologia. Le normative ICCD, infatti, vengono aggiornate periodicamente dagli organi responsabili attraverso il rilascio di nuove versioni, di conseguenza anche per le stesse schede di catalogo è necessario verificare ad ogni aggiornamento che siano conformi ad esse. La stessa normativa F è attualmente alla versione 4.00, mentre altre normative, tra cui quelle analizzate dai colleghi tesisti, sono ancora alla versione 3.00. L'allineamento delle schede di catalogo ai rilasci delle nuove versioni richiede dunque anche un allineamento degli elementi del modello ontologico ad esse. D'altro canto, anche lo stesso modello concettuale di riferimento *CIDOC-CRM* viene aggiornato periodicamente con l'introduzione di nuove entità e nuove proprietà o la modifica di quelle già presenti. L'allineamento dell'ontologia dovrà quindi essere garantito in entrambe le direzioni: sia nei confronti dei nuovi rilasci delle normative ICCD, sia rispetto agli aggiornamenti che coinvolgono *CIDOC-CRM*.

In generale, comunque, l'ontologia *CICDO* qui proposta, e le sue estensioni relative ai beni fotografici, alle opere d'arte e ai reperti archeologici, rappresentano nel complesso un'ottima base e uno spunto per la rappresentazione di informazioni culturali nel formato accettato dal Web semantico.

Oltre alla necessità di estendere il modello ontologico proposto per accogliere anche i concetti relativi alle informazioni dei restanti livelli di catalogazione, una sua estensione futura potrebbe considerare l'introduzione di entità e proprietà di altri schemi e modelli concettuali, siano essi general-purpose o domain-specific. Questo permetterà non solo di arricchire maggiormente la semantica delle informazioni rappresentate, ma potrà favorire ulteriormente la possibilità di derivare conoscenza implicita dalle informazioni già presenti.

In generale, l'enorme potenza espressiva del modello ontologico proposto risiede nella possibilità di rappresentare le informazioni contenute all'interno delle schede di catalogo indifferentemente dal formato con cui esse sono rappresentate. Sebbene il formato di riferimento considerato in diversi punti della trattazione è XML, *CICDO* e le sue estensioni permettono di descrivere tali informazioni a prescindere dalle modalità di rappresentazione utilizzate. Inoltre, come riportato all'inizio della trattazione corrente, nonostante le ontologie in questione facciano riferimento a standard catalografici riconosciuti solo a livello nazionale, esse rappresentano comunque un modello ampiamente valido anche per un suo utilizzo a livello internazionale.

Infine, anche per quanto riguarda l'applicazione di conversione proposta c'è molto lavoro da fare. Un primo passo può essere sicuramente quello di rendere l'applicazione un Web Service, in modo tale che le modalità di utilizzo siano rese il più possibile standard. In secondo luogo, l'applicazione potrà essere stesa in maniera tale da accogliere anche regole di trasformazione relative alle normative non considerate in questo lavoro e alle informazioni degli altri livelli di catalogazione riconosciuti dall'ICCD.

BIBLIOGRAFIA E SITOGRAFIA

1. “World Wide Web”, *Wikipedia, l’enciclopedia libera*,
https://en.wikipedia.org/wiki/World_Wide_Web
2. Vizzari Giuseppe (2015), *Evoluzione del Web: dal Web 1.0 al Web 2.0*,
<https://www.slideshare.net/giuseppegvizzari/11-levoluzione-del-web>
3. Fusco Guido (2015), *Evoluzione del Web*,
<http://www.corso-informatica-web.com/wp-content/uploads/2015/03/Evoluzione-del-Web.pdf>
4. “Web statico”, *Wikipedia, l’enciclopedia libera*,
https://it.wikipedia.org/wiki/Web_statico
5. “Web dinamico”, *Wikipedia, l’enciclopedia libera*,
https://it.wikipedia.org/wiki/Web_dinamico
6. “Web 2.0”, *Wikipedia, l’enciclopedia libera*,
https://it.wikipedia.org/wiki/Web_2.0
7. Piacenza Maurizio, *Dal WEB al WEB 3.0*,
<https://reteparcosud.wordpress.com/le-imprese/comunicazione/piacenza-maurizio/dal-web-1-0-al-web-3-0/>
8. Qmee (2014), “*Online in 60 seconds [Infographic] – A Year Later*”
<http://blog.qmee.com/online-in-60-seconds-infographic-a-year-later/>
9. “Search engine (computing)”, *Wikipedia, l’enciclopedia libera*,
[https://en.wikipedia.org/wiki/Search_engine_\(computing\)](https://en.wikipedia.org/wiki/Search_engine_(computing))
10. “Indicizzazione (motori di ricerca)”, *Wikipedia, l’enciclopedia libera*,
[https://it.wikipedia.org/wiki/Indicizzazione_\(motori_di_ricerca\)](https://it.wikipedia.org/wiki/Indicizzazione_(motori_di_ricerca))
11. “Precisione e recupero”, *Wikipedia, l’enciclopedia libera*,
https://it.wikipedia.org/wiki/Precisione_e_recupero

12. Turetta Giulio (2006), “*Cosa sono i Web Service*”,
<http://www.html.it/pag/16448/cosa-sono-i-web-service/>
13. “Web Service”, *Wikipedia, l'enciclopedia libera*,
https://it.wikipedia.org/wiki/Web_service
14. “Mash-up (informatica)”, *Wikipedia, l'enciclopedia libera*,
[https://it.wikipedia.org/wiki/Mash-up_\(informatica\)](https://it.wikipedia.org/wiki/Mash-up_(informatica))
15. Buongiorno Davide (2009), *Web Semantico: lo stato dell'arte*,
http://tesi.cab.unipd.it/23559/1/TESI_Buongiorno_Davide_322428IF.pdf
16. “Semantic Web”, *Wikipedia, l'enciclopedia libera*,
https://en.wikipedia.org/wiki/Semantic_Web
17. Grigoris Antoniou, Frank van Harmelen, *A Semantic Web Primer*
MIT Press, USA, 2004
18. “Il Web semantico, il futuro del Web”, *Websemantico.org*
<http://www.websemantico.org>
19. Ceravolo Paolo (2006), “*Cos'è e a cosa serve il Web Semantico*”,
<http://www.html.it/articoli/cos-e-a-cosa-serve-il-web-semantico-1/>
20. Stancari Stefano (2015), “*Il Web Semantico, il futuro del Web*”,
<http://www.thismarketerslife.it/digital/il-web-semantico-il-futuro-del-web/>
21. “Semantic Web Stack”, *Wikipedia, l'enciclopedia libera*,
https://en.wikipedia.org/wiki/Semantic_Web_Stack
22. “Resource Description Framework”, *Wikipedia, l'enciclopedia libera*,
https://en.wikipedia.org/wiki/Resource_Description_Framework
23. “RDF Schema”, *Wikipedia, l'enciclopedia libera*,
https://it.wikipedia.org/wiki/RDF_Schema
24. “Web Ontology Language”, *Wikipedia, l'enciclopedia libera*,
https://en.wikipedia.org/wiki/Web_Ontology_Language

25. “Linked data”, *Wikipedia, l’enciclopedia libera*,
https://en.wikipedia.org/wiki/Linked_data
26. Eero Hyvönen (2012), *Publishing and Using Cultural Heritage Linked Data on the Semantic Web*, Morgan & Claypool Publishers
27. “Patrimonio culturale”, *Wikipedia, l’enciclopedia libera*,
https://it.wikipedia.org/wiki/Patrimonio_culturale
28. “Europeana”, *Wikipedia, l’enciclopedia libera*,
<https://it.wikipedia.org/wiki/Europeana>
29. “Catalogazione dei beni culturali”, *Wikipedia, l’enciclopedia libera*,
https://it.wikipedia.org/wiki/Catalogazione_dei_beni_culturali
30. Istituto Centrale per il Catalogo e la Documentazione (2014),
Scheda F - Fotografia versione 3.00
31. Daquino Marilena (2016), *F Entry to RDF: Mapping the F Entry Italian content standard for cataloguing photographs to RDF*
DOI 10.6084/m9.fgshare.3175273
32. Patrick Le Boeuf, Martin Doerr, Christian Emil Ore, Stephen Stead (2013),
Definition of the CIDOC Conceptual Reference Model
33. “Semantic reasoner”, *Wikipedia, l’enciclopedia libera*,
https://en.wikipedia.org/wiki/Semantic_reasoner
34. “XSLT”, *Wikipedia, l’enciclopedia libera*,
<https://en.wikipedia.org/wiki/XSLT>
35. “XSLT introduction”, *W3Schools.com*
https://www.w3schools.com/xml/xsl_intro.asp