

ECODIGIT

Ecosistema Digitale per la Fruizione e la Valorizzazione
dei Beni e delle Attività Culturali della Regione Lazio

D3.2v2 Modello di ingresso

Acronimo Progetto:

EcoDigit

Titolo Progetto:

**Ecosistema digitale per la fruizione
e la valorizzazione dei beni e delle
attività culturali della regione Lazio**

D3.2

Work Package:	WP3 Task 3.2	
Deliverable Dovuto il:	2 Aprile 2019	
Inizio Progetto:	2 Ottobre 2018	
Durata Progetto:	15 Mesi	
Reponsabile Deliverable:	Valentina Presutti valentina.presutti@cnr.it	
Versione:	2.0	
Stato:	Versione Finale	
Autore:	Luigi Asprino	ISTC-CNR
	Ludovica Marinucci	ISTC-CNR
	Andrea Giovanni Nuzzolese	ISTC-CNR
	Valentina Presutti	ISTC-CNR
Altri contribuenti al lavoro riportato nel deliverable:	Marialuisa Mongelli	ENEA
	Antonio Budano	INFN
	Massimo Mecella	RM1
	Maria Prezioso	RM2
	Marco Canciani	RM3
	Giovanni Fiorentino	UNITUS
Reviewer:	Miguel Ceriani	RM1
	Massimo Mecella	RM1

Per citare questo documento si prega di utilizzare il seguente record bibliografico

Luigi Asprino, Ludovica Marinucci, Andrea Giovanni Nuzzolese, and Valentina Presutti. *D3.2 Modello di ingresso*. Deliverable Progetto EcoDigit. 2019

Revisioni

Versione	Data	Modificata da	Commento
v 0.1	18/3/2019	Ludovica Marinucci	Creazione documento
v 0.2	20/3/2019	Ludovica Marinucci	Prima Versione Documento
v 0.3	26/3/2019	Massimo Mecella	Revisione
v 1.0	2/4/2019	Ludovica Marinucci	Versione Finale
v 1.1	24/09/2019	Ludovica Marinucci	Creazione documento
v 1.2	26/09/2019	Ludovica Marinucci	Prima Versione Documento
v 1.3	28/09/2019	Valentina Presutti e Andrea Nuzzolese	Prima revisione
v 1.4	29/09/2019	Ludovica Marinucci e Luigi Asprino	Seconda versione
v 1.5	30/09/2019	Massimo Mecella	Revisione finale
v 2.0	1/10/2019	Ludovica Marinucci e Luigi Asprino	Versione Finale

Executive Summary

Le cinque università statali del Lazio in rete con CNR, ENEA e INFN si candidano a costituire il Centro di Eccellenza del Distretto Tecnologico per i beni e le attività Culturali (DTC) del Lazio. La mission del Centro di Eccellenza è costituire un centro di aggregazione ed integrazione di competenze nel settore delle tecnologie per i beni e le attività culturali. In questo contesto, il progetto **EcoDigit-Ecosistema digitale per la fruizione e la valorizzazione dei beni e delle attività culturali del Lazio** ha l'obiettivo di arricchire il sistema Anagrafe delle Competenze del DTC con una piattaforma middleware che faciliti l'integrazione di nuove sorgenti di dati e consenta la pubblicazione e il riuso di servizi per la valorizzazione e la fruizione del patrimonio culturale del Lazio.

Il presente documento, dal titolo **D3.2v2 Modello di ingresso**, è il risultato delle attività previste dal **Task 3.2**, che ha l'obiettivo di definire il modello di integrazione, ovvero l'insieme di pratiche e standard tecnologici che una sorgente dovrà rispettare per entrare in EcoDigit. Il documento è una estensione della precedente versione **D3.2 Modello di ingresso** e riporta una revisione del modello fatta sulla base della esperienza della Proof-of-Concept [1].

Il deliverable è composto da una sezione introduttiva che descrive gli obiettivi principali delle attività del Task 3.2 e la sua connessione con le attività degli altri task del progetto. Segue una seconda sezione relativa alla presentazione dello stato dell'arte sulle ontologie e i Linked Open Data relativi in particolare alla valorizzazione del patrimonio culturale, seguita da un ricognizione generale dei possibili riusi di ontologie già esistenti e utilizzate per integrare i dataset censiti nel Lazio nel corso delle attività del Task 3.1. In seguito, si propone un'analisi tecnica che arriva a un maggior dettaglio degli allineamenti semantici possibili tra gli schemi di dati della *Science & Technology Digital Library* (S&TDL) del CNR come esempio di modello di integrazione nel sistema EcoDigit. Infine, al fine di soddisfare requisiti emersi dallo studio delle sorgenti e non soddisfatti dalle ontologie disponibili allo stato dell'arte presenteremo le nuove ontologie che sono state sviluppate all'interno del contesto del progetto.

Indice

1	Introduzione	6
1.1	Obiettivi del Work Package	6
1.2	Obiettivo del deliverable	6
1.3	Relazione con le altre attività del progetto	6
2	Definizione del modello	7
2.1	Metodologia	7
2.2	Analisi dei dataset censiti nel Lazio	7
2.3	Stato dell'arte	10
2.4	Integrazione della S&TDL in EcoDigit	11
2.5	Ontologie Sviluppate	15
2.5.1	Dettagli implementativi	15
2.5.2	Ontologia delle Organizzazione DTC	17
2.5.3	Ontologia delle Esperienze e delle Competenze e Ontologia delle Valutazioni	17
2.5.4	Ontologia dei Progetti	21
2.5.5	Categorie di Contenuti del Prototipo	22
2.5.6	Modello per Oggetti	22
2.5.7	Localizzazione	24
2.5.8	Modello per la Provenance dei Dati	26
3	Conclusioni	27

Lista delle Tabelle

1	Esempio di allineamenti per le Persone della S&TDL.	13
2	Esempio di allineamenti per per le Organizzazioni della S&TDL.	13
3	Allineamenti scheda Prodotti	15
4	Competency question dell'ontologia delle Esperienze e delle Competenze. . .	19
5	Competency question dell'ontologia delle Valutazioni.	20
6	Competency Question soddisfatte dall'Ontologia Project	21

1 Introduzione

1.1 Obiettivi del Work Package

Il **WP3 “Modelli, metodi e strumenti per l’aggregazione di sorgenti”** si occupa dei contenuti di EcoDigit. In esso sono curati: le sorgenti dei dati, i modelli e le tecniche per la loro integrazione e standardizzazione. Queste tecniche sono basate su formati standard permettono di definire formalmente la semantica dei dati.

Obiettivo di questo WP è analizzare, progettare e sviluppare metodologie e strumenti per aggregare risorse e poli distinti sul territorio (indicate genericamente come sorgenti), che includano archivi, dati strutturati e database, patrimoni fotografici e multimediali in generale, biblioteche digitali, ecc.

1.2 Obiettivo del deliverable

Questo deliverable descrive le attività del **Task 3.2 “Definizione del modello di integrazione di una sorgente”**. Il modello di integrazione è l’insieme di pratiche e standard tecnologici che una sorgente dovrà rispettare per entrare in EcoDigit. Tale modello sui principi e tecniche di integrazione proprie del Semantic Web, considerando anche tecnologie proprie delle openAPI e tutto ciò che viene comunemente classificato come Open Data.

Obiettivo del presente deliverable è la ricognizione e l’analisi delle tecniche e degli strumenti con cui una sorgente non compatibile possa diventarlo, al fine di offrire un sistema concettuale utile successivamente a tutte le sorgenti. Tale attività si pone come preliminare alla definizione di una prima versione del modello di ingresso, ovvero degli standard che una sorgente deve rispettare per essere aggregata nel Middleware DTC.

1.3 Relazione con le altre attività del progetto

Il T3.2 si pone in relazione di dipendenza con le attività previste dagli altri task del WP3. Preliminare è il D3.1 relativo al censimento dei sorgenti potenziali, grazie al quale è stato possibile produrre una prima versione del modello di integrazione che successivamente sarà revisionato sulla base sia dell’identificazione degli strumenti di supporto di una sorgente (D3.3) per aderire al modello di ingresso sia della *Proof-of-Concept* (D3.4).

Inoltre, secondo gli obiettivi previsti dalla Ms3, le attività di questo task sono in stretta relazione con le attività del WP2, in particolare relative alla validazione delle interfacce per l’aggregazione delle sorgenti, definite nel WP2, rispetto ai modelli di aggregazione, definiti nel corso delle attività del WP3.

2 Definizione del modello

2.1 Metodologia

Il modello di dati e metadati, che le sorgenti devono rispettare per poter essere acquisite dal middleware EcoDigit, si basa su metodi e tecniche di “metadattazione” e di Semantic Web, comprendendo anche tecnologie proprie delle openAPI e tutto ciò che viene comunemente classificato come Open Data.

Per la sua elaborazione, si è proceduto a un’analisi delle sorgenti censite nel Lazio al fine di evidenziare tanto i domini di conoscenza coperti dalle sorgenti quanto il dettaglio dei vari campi che il modello deve rappresentare. Parallelamente, per i domini emersi da questa attività è stato effettuato uno studio dello stato dell’arte al fine di identificare modelli standard per la modellazione di dati e metadati. La descrizione del processo di formulazione di un primo censimento dei dataset presenti nel Lazio è contenuta nel deliverable D3.1 [5]; mentre una prima descrizione dei contenuti e dei formati dei dataset censiti, associando i possibili riusi di ontologie scelte, è stata pubblicata in una prima versione di questo deliverable e verrà rivisitata in questo documento sulla base dell’esperienza della Proof-of-Concept del Task T3.3. Per ognuno di essi sono stati ricercati gli schemi concettuali considerati standard di riferimento per la modellazione dei dati inerenti a un certo dominio.

Successivamente, quando i modelli esistenti allo stato dell’arte sono stati ritenuti non in grado di rappresentare semanticamente campi peculiari presenti nei dataset di input, è stata effettuata una modellazione ex novo, utilizzando una metodologia di ingegneria ontologica [2], basata su un’estensione di *eXtreme Design* (XD) [3]. XD è un metodo di progettazione agile di ontologie che si basa sul riuso di *Ontology Design Patterns* (ODP) [9] al fine di risolvere problemi di modellazione ontologica noti e ricorrenti. Questo è ad esempio il caso di alcune ontologie che sono presentate nella sezione 2.5.

2.2 Analisi dei dataset censiti nel Lazio

La descrizione del processo di formulazione di un primo censimento dei dataset presenti nel Lazio è contenuta nel deliverable **D3.1** (cfr. [5]). In questa sezione, si delinea una breve descrizione dei contenuti e dei formati dei dataset censiti, associando i possibili riusi di ontologie scelte per ciascuno dei gruppi di dataset di seguito elencati.

Personale e Prodotti della ricerca di Università e Enti di ricerca. Questi dataset contengono le informazioni relative al personale e ai prodotti della ricerca, ovvero pubblicazioni scientifiche, e sono quindi compatibili con gli allineamenti semantici in particolare delle ontologie FOAF, CERIF, BIBO e SPAR.

- *ART - Anagrafe della Ricerca dell’Università di Roma Tor Vergata (RM2).* Il contenuto

dei dati è relativo a: (i) pubblicazioni in vari ambiti disciplinari, ovvero articoli, monografie, capitoli di libro, interventi a convegno, brevetti, tesi di dottorato, rapporti tecnici, ecc.; (ii) informazioni anagrafiche e curriculum di ricercatori e professori dell'ateneo; (iii) attività e progetti di ricerca finanziata. I dati sono indicizzati per tipologia e lo standard di metadatozione è il Dublin Core qualified. Sono presenti vari formati di export dei dati (ad es. MODS, BibTex, ISI, CSV, etc, bibtex, ris, endnote, refworks, excel) ed è possibile l'integrazione con diversi database internazionali relativi all'indicizzazione dei prodotti della ricerca (quali, ad esempio, Scopus, WOS, ORCID, Scopus, Web of Science, ORCID, ecc.).

- *UGOV Risorse Umane e IRIS della Sapienza Università di Roma (RM1)*. Si tratta di informazioni relative al personale della Sapienza (ruolo, dipartimento, area scientifica), che sono collegate a informazioni anagrafiche e fiscali e all'archivio delle pubblicazioni. IRIS presenta lo standard MODS per i metadati e il sistema è integrato con U-GOV Cineca, Scopus e Web of Science. I dati sono esposti via API o in formato OAI-PMH ed utilizzati dai servizi interni Sapienza. I formati di export dei dati sono: bibtex, MODS, RIS, EndNOTE, RefWorks, csv; mentre per UGOV sono: SQL, Excel e csv.
- *La Science & Technology Digital Library del CNR*. Si tratta di dati relativi a: (i) persone, alle loro attività e alle loro competenze; (ii) prodotti della ricerca (relativi alle persone inserite del database), ovvero pubblicazioni, prodotti di letteratura grigia, oggetti multi-mediali, ecc.); (iii) organizzazioni (a cui le persone afferiscono), ovvero i vari Istituti del CNR a livello nazionale; (iv) progetti di ricerca, il cui responsabile è indicizzato nelle persone) finanziati in vari ambiti; (v) dati della ricerca, ovvero riferimenti ai dataset ottenuti tramite progetti di ricerca. L'export dei dati è in formato XML.

Prodotti nell'ambito dei beni culturali. Questo gruppo di dataset contiene informazioni su prodotti che afferiscono alla tipologia di bene culturale con la sua relativa documentazione e analisi storica. L'ontologia di riferimento da utilizzare è ArCo con le relative ontologie relative al patrimonio culturale ad essa collegate, ad es. CIDOC-CRM, EDR, ecc (cfr. Sezione **2.3**).

- *Repository Archeologia e Calcolatori*. Si tratta degli articoli pubblicati nella rivista Archeologia e Calcolatori e suoi supplementi. La repository è accessibile tramite protocollo OAI-PMH, dai portali CulturalItalia e DARIAH-IT. Lo standard di metadatozione utilizzato è il Dublin Core, mentre il formato di export è XML.
- *Greek Literary Hands of the Roman Period*. Questa repository è interfacciata con i database delle principali collezioni di materiali papiracei e collegata con i portali trismegistos.org e papyri.info, che consistono in due raccolte di papiri letterari greci di età imperiale databili su basi non esclusivamente paleografiche, con immagini digitali e descrizioni paleografiche.
- *Iscrizioni Greche di Antinoupolis*. Si tratta di materiali archeologici con descrizione analitica ed edizione di tutte le iscrizioni greche e copte rinvenute nel corso degli scavi

della Missione Italiana dell'Istituto Papirologico "G. Vitelli" ad Antinoupolis. Le schede sono corredate da riproduzioni digitali ad alta risoluzione e talvolta multispettrali.

- *Omeliari in scrittura beneventana*. La repository consiste di una catalogazione dettagliata dei manoscritti liturgici, delle omelie e dei testi in essi contenuti. La banca dati si interfaccia con la Banca-dati BMB. Bibliografia dei manoscritti in scrittura beneventana¹.
- *MaGI. Manoscritti greci d'Italia*. Censimento uniforme dei dati identificativi essenziali di tutti i manoscritti greci conservati sul territorio italiano, accompagnati da almeno una riproduzione fotografica, adeguatamente indicizzata, per ciascuna unità codicologica e da una bibliografia aggiornata. Le descrizioni prevedono, ove possibile, il recupero delle informazioni fornite dai cataloghi e dalle pubblicazioni specialistiche esistenti, sistematicamente integrato dalla visione autoptica dei manoscritti. L'export dei dati è possibile in formato TXT, PDF e JPG.
- *Repository Museo Facile dell'Università di Cassino*. Si tratta di prodotti della ricerca consistenti nella descrizione teorica e storica di opere e monumenti artistici tramite documentazione fotografica e video.
- *UCI machine learning repository*. Il dataset è stato estratto da 800 immagini della Bibbia di Avila, cosiddetta "Atlantica", prodotta durante il XII secolo tra Italia e Spagna. Ogni record contiene 10 features corrispondenti a ritagli di immagini di gruppi di 4 righe consecutive.

Prodotti di ricerca con dati 3D e/o georiferiti nell'ambito dei beni culturali. Questi dataset hanno una particolare tipologia di dati sia per contenuto, ovvero relativo al campo dei beni culturali, ma anche per il tipo di formato 3d e georiferito che supportano. Oltre alle ontologie sovraespresse sul patrimonio culturale, è possibile riutilizzare per l'Arpenteur Ontology per i beni architettonici, l'ontologia CLV sugli indirizzi e i luoghi, allineata al vocabolario controllato per i dati georeferenziati WGS84 Geo Positioning (geo) vocabulary.

- *Repository E3S - ENEA Staging, Storage, Sharing System*). Si tratta di una piattaforma volta alla protezione, conservazione e valorizzazione del patrimonio culturale, che è trasparente al formato dei dati, ovvero in grado di gestire i file di diversi formati, quali: immagini, filmati, documenti excel, modelli numerici dalla fotogrammetria (fem), file c3D di un sistema optoelettronico di tipo motion capture 3D. I formati di export possibile sono: csv, c3d e jpeg.
- *STeMA-VAS dell'Università Tor Vergata*. Si tratta di prodotti della ricerca consistenti in dati georiferiti con contenuto statistico, puntuale, lineare e areale, ecc. volto alla pianificazione territoriale sostenibile con procedura VAS (Valutazione Ambientale Strategica). I formati di export sono: shape, xls, DBF, jpg.

¹<http://edu.let.unicas.it/bmb/>

- *Datasets del DIPSA dell'Università RomaTre.* Contengono prodotti della ricerca rappresentati da: modelli, rilievi 3D, ricostruzioni virtuali, immagini ed elaborazioni grafiche, analisi storica, documentazione d'archivio ed elaborazioni digitali, testi e pubblicazioni scientifiche realizzate in seno al Dipartimento di Architettura dell'Università Roma Tre. I formati di export sono: 3D dwg, obj, pts, 3dmax.

2.3 Stato dell'arte

Le tecnologie semantiche, e in particolare i Linked Open Data (LOD), sono state ampiamente sfruttate con successo nel campo del patrimonio culturale al fine di migliorare l'accesso e l'esperienza di fruizione ed esplorazione dei beni culturali da parte dei cittadini, così come di facilitare la reperibilità, l'integrazione e l'arricchimento dei dati sul patrimonio culturale [6]. Infatti, il paradigma dei LOD è utilizzato per collegare dati provenienti da diverse istituzioni culturali, aumentando così la possibilità di raggiungere i dati culturali disponibili nel Web of Data. L'interconnessione dei contenuti delle organizzazioni collaboratrici ha anche contribuito ad arricchire le informazioni in modo efficace e finalizzato alla valorizzazione del patrimonio culturale [10]. La collaborazione tra organizzazioni culturali ha portato anche allo sviluppo collaborativo di ontologie che descrivono il patrimonio culturale a livello internazionale, ad esempio CIDOC-Conceptual Reference Model (CRM) [7], in modo tale che i requisiti di interoperabilità semantica potessero essere soddisfatti all'interno dei loro sistemi. Inoltre, l'uso di ontologie comuni ha facilitato lo scambio di dati e la creazione di enormi librerie digitali, ad esempio l'Europeana Data Model (EDM) [11].

Relativamente al patrimonio culturale italiano, dalla collaborazione tra il Semantic Technology Laboratory (STLab) dell'Istituto di Scienze e Tecnologie della Cognizione (ISTC) del CNR e l'Istituto Centrale per il Catalogo e la Documentazione (ICCD) del MiBAC è stata creata l'ontologia ArCo (Architettura della Conoscenza)² al fine di pubblicare come Linked Open Data i dati dell'ICCD sui beni culturali italiani, circa 800.000 schede di catalogo, estratte dal sistema informativo SIGECweb³ e riversate nel DB Catalogo Generale dei Beni Culturali⁴, ognuna contenente la descrizione di uno specifico bene culturale. Partendo da questo vasto repertorio catalografico, ArCo si propone come uno standard *de facto*, compatibile e allineato con ontologie già esistenti e adottate, per la rappresentazione dei dati sul patrimonio culturale italiano. Infatti, la complessità dei dati modellati da ArCo ha portato a creare nuovi moduli ontologici e, al contempo, a fare un riuso: (i) *indiretto*, per mezzo di allineamenti, di risorse già esistenti e adottate, come CIDOC-CRM⁵, EDM⁶, FEntry⁷ e OAEntry⁸, relative alle fotografie e alle opere d'arte, e i vocabolari controllati del Getty Research Institute⁹; (ii) *diretto*

²<http://dati.beniculturali.it/progetto-arco-architettura-della-conoscenza/>

³<http://www.iccd.beniculturali.it/it/sigec-web>

⁴<http://catalogo.beniculturali.it>

⁵<http://cidoc-crm.org/>

⁶<https://pro.europeana.eu/page/edm-documentation>

⁷<http://www.essepuntato.it/2014/03/fentry>

⁸<http://purl.org/emmedi/oaentry>

⁹<http://www.getty.edu/research/tools/vocabularies/index.html>

di classi e proprietà dai moduli *core* di OntoPiA¹⁰, una rete di ontologie e vocabolari controllati per la Pubblica Amministrazione, in particolare l'ontologia Cultural-ON¹¹ che modella gli eventi culturali e gli istituti e luoghi della cultura [12].

Ulteriori ontologie di riferimento per la definizione del modello di integrazione del sistema EcoDigit sono: la *FOAF*¹² che rappresenta le informazioni anagrafiche sulle persone in generale; la *CERIF*¹³ che specifica le informazioni relative ai ruoli, alle competenze, ai progetti e ai prodotti nell'ambito specifico della ricerca scientifica; la *DOAP* (Description Of A Project)¹⁴ che descrive i progetti in generale; la *Org Ontology*¹⁵ che descrive le organizzazioni in generale ed è integrata ed estesa dall'ontologia *COV*¹⁶ di OntoPiA; la *BIBO*¹⁷ e le *SPAR Ontologies*¹⁸, relative al dominio delle pubblicazioni e dei riferimenti bibliografici; l'*Arpenteur Ontology*¹⁹, dedicata ai beni archeologici e architettonici; il *WGS84 Geo Positioning (geo) vocabulary*²⁰ per allineare i dati georeferenziati e il *Dublin Core Metadata*²¹ espresso in RDF Schema, che consiste di concetti generali e di alto livello.

2.4 Integrazione della S&TDL in EcoDigit

Al fine di valutare la capacità dei modelli disponibili allo stato dell'arte di rappresentare informazioni contenute in una delle sorgenti del D3.1 [5] si è proceduto alla mappatura degli schemi della *Science & Technology Digital Library* sulle ontologie elencate nelle Sezioni 2.3 e 2.2. In particolare, si intende proporre un'analisi tecnica che arriva a un maggior dettaglio degli allineamenti semantici possibili tra le ontologie selezionate e gli schemi di dati della *Science & Technology Digital Library* (S&TDL)²² del CNR, che consiste in un sistema informativo integrato in grado di gestire grandi volumi di documenti, dati e informazioni, volto allo sviluppo di iniziative di digital library, repository, sistemi di digital preservation, ecc. Attualmente la S&TDL gestisce centinaia di migliaia di risorse informative, corrispondenti a milioni di oggetti digitali, suddivisibili in: (i) *prodotti della ricerca*, ovvero datasets e contenuti digitalizzati di interesse storico e culturale; (ii) *attività di ricerca*, ovvero progetti, istituzioni e attori operanti nei settori della ricerca e sviluppo (R&S) e dei beni archivistici e culturali.

L'analisi di questo dataset, scelto per la sua varietà di contenuti, è stata possibile gra-

¹⁰<https://github.com/italia/daf-ontologie-vocabolari-controllati>

¹¹<http://dati.beniculturali.it/cis>

¹²<http://xmlns.com/foaf/spec/>

¹³<https://www.eurocris.org/ontologies/cerif/>

¹⁴<http://usefulinc.com/ns/doap#>

¹⁵<https://www.w3.org/TR/vocab-org/>

¹⁶<https://w3id.org/italia/onto/COV>

¹⁷<http://bibliographic-ontology.org/>

¹⁸<http://www.sparontologies.net/ontologies>

¹⁹<http://www.arpenteur.org/ontology/Arpenteur.owl>

²⁰http://www.w3.org/2003/01/geo/wgs84_pos

²¹<http://www.dublincore.org/specifications/dublin-core/dcmi-terms/>

²²<https://dl.cnr.it/>

zie alla collaborazione con il team del progetto *Anagrafe delle Competenze*²³ che ha reso disponibile il dettaglio degli schemi XML dei dati nella Repository della S&TDL. Le schede XML sono relative in particolare alle sezioni della S&TDL: Persone, Prodotti, Progetti, Organizzazioni e Dati della ricerca.

Persone. Per rappresentare i dati relativi alle persone della S&TDL sono state riusate le seguenti ontologie:

- *FOAF* per rappresentare le seguenti informazioni: nome, cognome, immagine profilo, indirizzo email, indirizzo della homepage, eventuali link ad account detenuti dalla persona (come account ORCID, profilo Google Scholar, profilo Scopus etc.). Inoltre, tramite la proprietà `foaf:made` è possibile associare alla persona le proprie pubblicazioni.
- *OntoPiA L0*²⁴ per fornire una breve descrizione della persona in linguaggio naturale.
- *W3C's Org Ontology*²⁵ per fornire le informazioni riguardo all'affiliazione della persona ad una certa organizzazione.
- *OntoPiA COV*²⁶ per rappresentare l'appartenenza (come membro) di una persona ad una certa organizzazione specificando anche il ruolo tramite *OntoPiA RO*²⁷.
- *OntoPiA COV*²⁸ per descrivere il livello di istruzione di una certa persona (e.g. laurea magistrale, dottorato etc.).

Tuttavia non è stato possibile trovare allo stato dell'arte ontologie capaci di descrivere le seguenti informazioni: (i) le esperienze formative di una persona (e.g. la frequentazione di una certa scuola con il conseguente ottenimento di una certificazione). (ii) le lingue conosciute da una persona che permetta di indicare ad esempio: il livello di una persona in una certa lingua (rispetto ad una capacità del linguaggio, come lettura, scrittura etc.), l'eventuale conseguimento da parte di una persona di un attestato che certifichi la capacità della persona. (iii) i progetti a cui una persona ha partecipato.

Accedendo al seguente indirizzo²⁹ è possibile ottenere un esempio di risorsa della S&TDL modellata seguendo i modelli sopraelencati.

A titolo di esemplificativo, alcuni allineamenti dei campi principali dello schema XML delle Persone della S&TDL con le relative classi delle ontologie scelte sono riportati in **Tabella 1**.

²³<https://www.dtclazio.it/anagrafe-delle-competenze>

²⁴<https://w3id.org/italia/onto/10>

²⁵<http://www.w3.org/ns/org#>

²⁶<https://w3id.org/italia/onto/COV>

²⁷<https://w3id.org/italia/onto/RO>

²⁸<https://w3id.org/italia/onto/CPV>

²⁹<https://raw.githubusercontent.com/ecodigit/workspace/master/examples/Persona.rdf.xml>

Campo file XML della S&TDL	Classe con prefisso Ontologia
firstName	foaf:firstName
lastName	foaf:lastName
email	foaf:mbox
contacts	Ontologia SM ³⁰ di OntoPiA
profile	RO:Role

Table 1: Esempio di allineamenti per le Persone della S&TDL.

Campo file XML della S&TDL	Classe con prefisso Ontologia
denominazione	foaf:name
sigla	foaf:acronym
indirizzo	CLV:Address
telefono	SM:TelephoneType(voce)
responsabile	OntoPiA RO
websiteurl	foaf:homepage

Table 2: Esempio di allineamenti per per le Organizzazioni della S&TDL.

Organizzazioni. Per rappresentare relativi alle organizzazioni dalla S&TDL sono state riusate le seguenti ontologie:

- *W3C ORG Organization* e *FOAF* per rappresentare le seguenti informazioni relative ad una organizzazione: nome, acronimo, email, indirizzo del sito web, sede, classificazione della organizzazione, membri dell'organizzazione, ed eventuali sotto-organizzazioni.
- *OntoPiA COV* per associare una organizzazione con una URL che punta ad una immagine raffigurante il logo dell'organizzazione.
- *OntoPiA IO* per fornire una descrizione in linguaggio naturale della organizzazione.
- *OntoPiA CLV* per descrivere l'indirizzo della sede di una organizzazione.
- *OntoPiA RO* per descrivere il ruolo che una certa persona ha all'interno di una organizzazione.

Tuttavia, non è stato possibile trovare allo stato dell'arte una ontologia per: (i) classificare le organizzazioni secondo le categorie del Centro di Eccellenza (e.g. Rete di Laboratori, Laboratorio etc.). (ii) specificare la "mission" di una organizzazione.

La **Tabella 2** riporta un estratto degli allineamenti della scheda XML delle Organizzazioni e le relative classi delle ontologie.

Progetti. Le uniche ontologie disponibili allo stato dell'arte per la modellazione dei Progetti sono DOAP e CERIF. Tuttavia entrambe sono risultate non mantenute. Per questo motivo si

è proceduto alla realizzazione di una nuova ontologia per i progetti che è diventata parte del network di Ontologie OntoPiA. Questa ontologia verrà presentata nella sezione 2.5.

Prodotti. Per rappresentare relativi ai prodotti dalla S&TDL sono state riusate le seguenti ontologie:

1. *FABIO*³¹ e *BIBO*³² per classificare il prodotto. Qualora il prodotto ricada in una classe non coperta da queste due ontologie sono state definite delle linee guida per definire dei nuovi tipi derivandoli da quelli usati in S&TDL. Queste indicazioni sono contenute in una pagina Wiki del workspace del progetto³³.
2. *BIBO* è anche usata per: associare ad una pubblicazione una lista ordinata di autori; per associare alla pubblicazione il proprio abstract; per indicare la pagina iniziale e la pagina finale dell'articolo; fornire DOI e ISBN di una pubblicazione; per associare ad una pubblicazione i suoi editori.
3. *Datacite*³⁴ delle rete di ontologie SPAR per specificare gli identificativi di una pubblicazione. Datacite permette di specificare gli identificati secondo gli schemi più comuni: arxivni, bibcodeni, doini, ean13ni, eissnni, handleni, infourini, isbnni, issnni, istcni, lissnni, local resource identifier schemeni, lsidni, nihmsidni, ocini, piini, pmcidni, pmidni, purlni, sicini, upcni, urini, urlni, urnniark, arxiv, bibcode, doi, ean13, eissn, handle, infouri, isbn, issn, istc, lissn, local resource identifier scheme, lsid, nihmsid, oci, pii, pmcid, pmid, purl, sici, upc, uri, url, urn.
4. *Prism*³⁵ per indicare le parole chiave associate ad una certa pubblicazione.
5. *DC/DCTerms*³⁶ per esprimere l'estensione di una pubblicazione, per associare ad una pubblicazione il relativo *publisher*; per specificare il fatto che una pubblicazione è parte di un'altra pubblicazione; per specificare la data della pubblicazione; titolo e abstract di una pubblicazione; la lingua della pubblicazione.

Alcuni allineamenti dello schema XML dei Prodotti con le relative classi e proprietà delle ontologie sopradescritte sono riportati nella **Tabella 3**. Un esempio completo di una pubblicazione modellata secondo queste indicazioni è disponibile al seguente link³⁷.

³¹<http://purl.org/spar/fabio/>

³²<http://purl.org/ontology/bibo/>

³³<https://github.com/ecodigit/workspace/wiki/Tipi-CNR-DL>

³⁴<http://purl.org/spar/datacite>

³⁵<http://prismstandard.org/namespaces/basic/2.0/>

³⁶<https://www.dublincore.org/specifications/dublin-core/dcmi-terms/>

³⁷<https://raw.githubusercontent.com/ecodigit/workspace/master/examples/Article.rdf.xml>

Campo file XML della S&TDL	Classe con prefisso Ontologia
titolo	dc:title
anno	dc:date
lingue	dc:language
abstract	bibo:abstract
autori	bibo:authorList
editori	bibo:editorList
isbn	bibo:isbn

Table 3: **Allineamenti scheda Prodotti**

2.5 Ontologie Sviluppate

Durante la realizzazione della Proof-of-Concept (cf. Deliverable D3.4 [1]) è emersa l'insufficienza dei modelli disponibili allo stato dell'arte nel modellare alcune caratteristiche delle sorgenti selezionate. Si è quindi proceduto nell'estendere i modelli esistenti catturando i requisiti derivanti dal lavoro di Proof-of-Concept. In questa sezione descriveremo il processo realizzativo di questo ontologie ed effettueremo una panoramica dei nuovi modelli sviluppati.

2.5.1 Dettagli implementativi

URI permanente EcoDigit. Uno dei principi degli standard del Semantic Web e Linked Data è identificare entità (e.g. individui, proprietà, concetti ecc.) attraverso delle URI. Le applicazioni che maneggiano Linked Data hanno bisogno di specificare i dati usando delle URI (Uniform Resource Identifier) stabili, quindi che non cambino nel tempo (cioè non deve accadere che si usi un'altra URI per identificare lo stesso concetto) e che vengano mantenute (cioè, accedendo a quella URI deve esserci una "risposta" da parte di un server). Una delle soluzioni allo stato dell'arte per rendere le URI persistenti è usare un servizio come w3id.org. Questo servizio permette di registrare nuove URI che nel tempo vengono garantite come permanenti. Le nuove URI possono essere quindi fatte puntare verso una qualsiasi risorsa di cui si vuole creare l'URI permanente. Quindi, per far sì che le nuove ontologie fossero definite usando degli URI permanenti abbiamo registrato su w3id.org la URI:

`https://w3id.org/ecodigit`

Questa URI viene usata per come prefisso per i concetti definiti nelle nuove ontologie (e.g. `eas:Certification`³⁸) o per identificare individui della base di conoscenza creata nella Proof-of-Concept (e.g. `person:luigi.asprino`³⁹).

³⁸<https://w3id.org/ecodigit/ontology/eas/Certification>

³⁹<https://w3id.org/ecodigit/person/luigi.asprino>

Workspace del progetto EcoDigit <http://ecodigit.dtlazio.it>

76 commits 1 branch 0 releases 2 contributors

Branch: master New pull request Find File Clone or download

File/Folder	Description	Last Commit
bin	zip deleted	4 months ago
examples	Ruolo	2 months ago
ontology	aggiornamento modello persona e eas	2 months ago
src	Object Transformer	14 days ago
templates	Object Transformer	14 days ago
README.md	readme	2 months ago
pom.xml	Object Transformer	14 days ago

Workspace del progetto Ecodigit

Questo repository contiene il workspace del progetto EcoDigit.

Ontologie

I dati acquisiti e pubblicati da EcoDigit saranno strutturati secondo ontologie raccomandate da enti nazionali (e.g. Agld) e internazionali (e.g. W3C) o riconosciute come standard de-facto (e.g. [OntoPiA](#), [FOAF](#), [Dublin Core](#), [W3C's Organization Ontology](#), [Bibliographic Ontology](#) ecc.). Dove la specificità dei dati da rappresentare non consentirà un riuso complete dei modelli standard, verranno definite delle nuove ontologie che estenderanno le ontologie esistenti dove necessario.

Figure 1: Homepage del Workspace di EcoDigit.

Workspace. Al fine di supportare il processo di sviluppo delle ontologie e del software della Proof-of-Concept è stato realizzato un Workspace per il progetto EcoDigit⁴⁰. In particolare Questo workspace contiene un repository usato per lo sviluppo e la pubblicazione delle ontologie (cfr. Figura 1) è disponibile su GitHub⁴¹.

Documentazione delle Ontologie. Accedendo alle URI delle singole ontologie è possibile generare dinamicamente la documentazione delle ontologie stesse. Per fare questo è stato usato LODE [13].

⁴⁰<https://github.com/ecodigit>

⁴¹<https://github.com/ecodigit/workspace>

2.5.2 Ontologia delle Organizzazioni DTC

Come già anticipato nella Sezione 2.4, le ontologie disponibili allo stato dell'arte non forniscono costrutti per (i) classificare le organizzazioni secondo le categorie definite dal Centro di Eccellenza DTC Lazio (e.g. Rete di Laboratori, Laboratorio etc.); (ii) specificare la "mission" di una organizzazione. Per soddisfare questi requisiti abbiamo definito una nuova ontologia, chiamata *Ontologia delle Organizzazioni EcoDigit*, disponibile alla URI <https://w3id.org/ecodigit/ontology/organization> (il prefisso definito per l'ontologia `ecodigit-org` ed è associato al valore <https://w3id.org/ecodigit/ontology/organization/>). L'ontologia estende le ontologie FOAF, OntoPiA COV, W3C Org definendo un vocabolario condiviso di termini per la descrizione delle organizzazioni che partecipano al Centro di Eccellenza - DTC Lazio. In particolare, l'ontologia definisce la classe `ecodigit-org:EnteDiRicerca`, sottoclasse di `org:Organization`, descritta come "organizzazione con il compito di svolgere attività di ricerca scientifica nei principali settori di sviluppo delle conoscenze e delle loro applicazioni in ambito tecnico-scientifico". Questa classe viene specializzata da:

- `ecodigit-org:Dipartimento`, una divisione di un ente di ricerca;
- `ecodigit-org:UnitaDiRicerca`, un ente di ricerca che hanno la funzione di coordinamento tecnico scientifico negli ambiti di propria competenza e di integrazione delle tecnologie e conoscenze multidisciplinari che caratterizzano l'unicità e la ricchezza del Centro;
- `ecodigit-org:LaboratorioDiRicerca`, un istituto di ricerca creato per operare e promuovere la ricerca in uno o più ambiti della scienza.
- `ecodigit-org:ReteDiLaboratori` raggruppamento per dei laboratori di ricerca per tematica di competenza o tipologia di tecnologia sviluppata. Per questa classe è stata definita una restrizione che impone che una rete di laboratori abbiamo almeno un laboratorio come membro.

Inoltre l'ontologia definisce una datatype property, cioè `ecodigit-org:mission` che permette di specificare in linguaggio naturale la "mission" di una certa organizzazione.

2.5.3 Ontologia delle Esperienze e delle Competenze e Ontologia delle Valutazioni

Come già anticipato nella Sezione 2.4, le ontologie disponibili allo stato dell'arte non sono capaci di descrivere le seguenti informazioni: (i) le esperienze formative di una persona (e.g. la frequentazione di una certa scuola con il conseguente ottenimento di una certificazione). (ii) le lingue conosciute da una persona che permetta di indicare ad esempio: il livello di una persona in una certa lingua (rispetto ad una capacità del linguaggio, come lettura, scrittura etc.), l'eventuale conseguimento da parte di una persona di un attestato che certifichi la capacità della persona. Per soddisfare questi requisiti abbiamo definito due nuove ontologie:

1. *Ontologia delle esperienze e delle competenze* che ha lo scopo di definire un vocabolario condiviso di termini per la descrizione delle esperienze e competenze di una persona. L'ontologia estende OntoPiA-CPV, OntoPiA-COV, FOAF e BIBO. L'ontologia è disponibile alla URI <https://w3id.org/ecodigit/ontology/eas>, e ha namespace <https://w3id.org/ecodigit/ontology/eas/>, associato al prefisso `ecodigit-eas`. L'ontologia importa l'ontologia delle valutazioni descritta di seguito.
2. *Ontologia delle valutazioni* che ha lo scopo di definire un vocabolario di termini per la descrizione di qualsiasi cosa abbiamo una valutazione associata espressa rispetto una certa scala. L'ontologia è disponibile alla URI <https://w3id.org/ecodigit/ontology/grade>, e ha namespace <https://w3id.org/ecodigit/ontology/grade/>, associato al prefisso `grade`.

Ontologia delle Esperienze e delle Competenze. La Figura 2 mostra il diagramma dell'ontologia delle Esperienze e delle Competenze. L'ontologia soddisfa i requisiti espressi dalle competency question elencate in Tabella 4. L'ontologia definisce il concetto di Training (sottoclasse di `TI:TimeIndexedEvent` di OntoPiA TI) che è fornito da (i.e. `providedBy`) da una certa Organizzazione (i.e. `foaf:Organization`) ed è frequentata da (i.e. `isAttendedBy`) una persona (i.e. `foaf:Person`). Un Training può includere opzionalmente un esame, rappresentato dalla classe `Examination`, e può essere associato ad una `Certification` a fronte della frequentazione del Training da parte della persona. L'esame (i.e. `Examination`), è anch'esso un `TI:TimeIndexedEvent` che associato ad una valutazione tramite la object property `grade:hasGrade` e ad una certificazione (i.e. `Certification`) che può essere rilasciata a fronte del superamento dell'esame da parte della persona coinvolta nel training. Un esame può avere come outcome una tesi (`bibo:Thesis`). L'ontologia definisce due tipi di certificazioni: certificazioni istituzionali (i.e. `InstitutionalCertification`) rilasciate da (i.e. `issuedBy`) una organizzazione (i.e. `foaf:Organization`); e auto certificazioni (rappresentate dalla classe `SelfCertification`). In entrambi i casi la certificazione è associata ad una data di emissione (`issueDate`) e attesta un certo livello di abilità espresso tramite le classi `Level` o `LevelInActivity` (ad esempio questo permettere di esprimere il livello di una certa persona in una certa attività linguistica rispetto alla scala CEFR⁴²). Infine, l'ontologia permette di descrivere le abilità di una persona (`ExpertiseAndSkill`), come ad esempio le abilità linguistiche (`LanguageSkill`). Sia `ExpertiseAndSkill` che `Level` e `LevelInActivity` sono definite come sottoclassi di `l0:Characteristic` dell'ontologia I0 di OntoPiA.

Ontologia delle Valutazioni. La Figura 3 mostra il diagramma dell'ontologia delle Valutazioni. L'ontologia soddisfa i requisiti espressi dalle competency question elencate in Tabella 5. L'ontologia definisce il concetto di voto (i.e. `Grade`), specificato come sottoclasse di `LiteralReification`. Questa classe rappresenta le reificazioni di valori letterali ed è una

⁴²https://it.wikipedia.org/wiki/Quadro_comune_europeo_di_riferimento_per_la_conoscenza_delle_lingue

ID	Competency Question
CQ1	Quali sono le esperienze formative di una persona?
CQ2	Quali certificazioni sono rilasciate ad una persona a fronte di una sua partecipazione ad una esperienza formatica?
CQ3	Quale organizzazione ha erogato una certa esperienza formativa?
CQ4	Quali sono le expertise di una persona?
CQ5	Quali sono le lingue conosciute da una persona?
CQ6	Quali esami una persona ha dovuto superare, e con quale voto, per concludere una esperienza formativa?
CQ7	Quali lingue conosce una persona?
CQ8	Quale è il livello di conoscenza di una lingua di una persona rispetto al framework CEFR?
CQ9	Quali sono le expertise certificate di una certa persona? E quali invece sono auto-dichiarate?

Table 4: Competency question dell'ontologia delle Esperienze e delle Competenze.

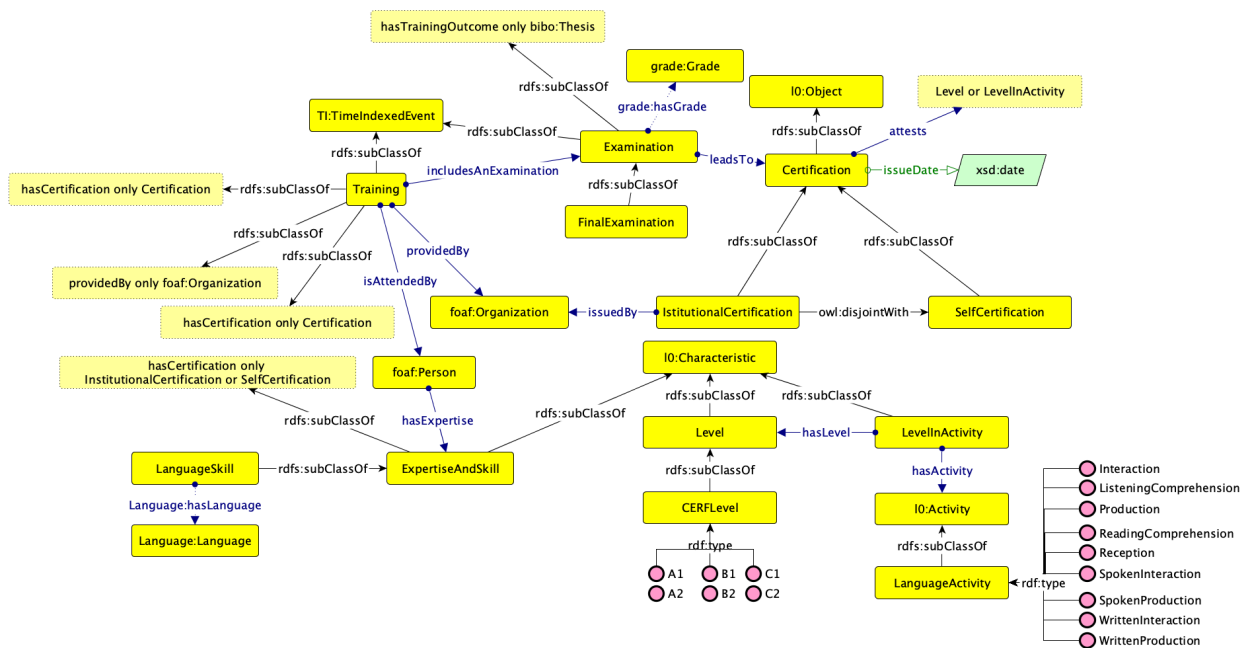


Figure 2: Diagramma dell'Ontologia delle Esperienze e delle Competenze.

ID	Competency Question
CQ1	Quale è il voto di una certa cosa?
CQ2	Rispetto a quale scala è espresso un certo voto?
CQ3	Quali sono il valore massimo e il valore minimo della scala di un certo voto?

Table 5: Competency question dell'ontologia delle Valutazioni.

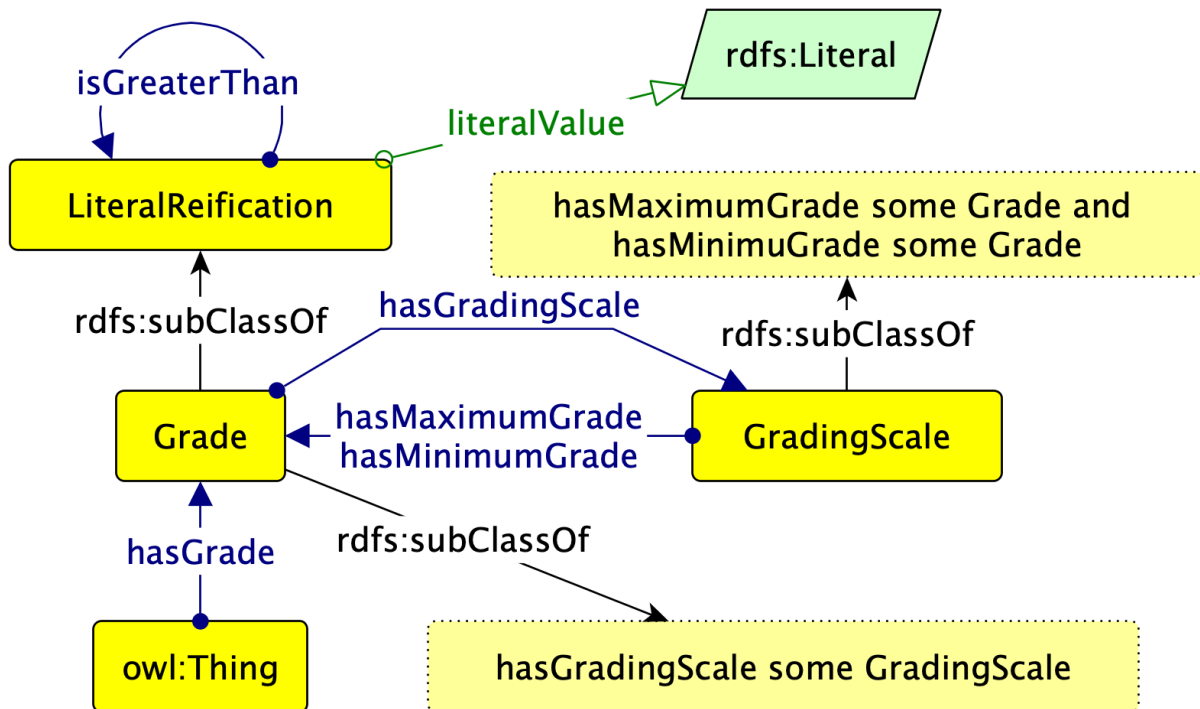


Figure 3: Diagramma dell'Ontologia delle Valutazioni.

implementazione dell'Ontology Design Pattern "Literal Reification"⁴³ [8]. Questo pattern permette di istanzianziare degli individui che rappresentano un valore letterale (individuo e valore letterale da esso rappresentato sono associati dalla proprietà `literalValue`). Un `Grade` è associato obbligatoriamente ad una scala, cioè un individuo della classe `GradingScale`. Una scala ha almeno un voto minimo (`hasMinimumGrade`) e almeno un voto massimo (`hasMaximumGrade`). Infine, l'ontologia permette di definire associare un voto ad una qualsiasi cosa (i.e. `owl:Thing`) tramite la proprietà `hasGrade`.

⁴³<http://purl.org/spar/literal>

ID	Competency Question
R1	Qual è il titolo di un certo progetto?
R2	Qual è l'acronimo di un certo progetto?
R3	Qual è l'URL del sito web di un certo progetto?
R4	Qual è il CUP del progetto (Codice Unico Identificativo)?
R5	Quali sono le organizzazioni coinvolte in un certo progetto?
R6	Quale organizzazione finanzia un certo progetto?
R7	Qual è la descrizione di un certo progetto?
R8	A quali progetti una persona ha partecipato?
R9	Qual è la data di inizio/fine di un certo progetto?
R10	Con quale ruolo (e.g. Coordinatore) una organizzazione partecipa in un certo progetto?
R11	Qual è l'area geografica (o le aree) in cui è localizzato il progetto?
R12	Quali sono i progetti che hanno un CUP di tipo "master"? E quelli di tipo normale e collegato? (questi dati li ho visti dal progetto OpenCUP - http://opencup.gov.it/opendata dove ci sono anche i metadati)
R13	Qual è il costo totale del progetto?

Table 6: **Competency Question soddisfatte dall'Ontologia Project**

2.5.4 Ontologia dei Progetti

Come già discusso nella Sezione 2.4, le ontologie, come CERIF o DOAP, disponibili allo stato dell'arte per il dominio dei progetti di ricerca sono risultate insufficienti per rappresentare la complessità dei requisiti del dominio e poco mantenute. Al fine di rappresentare i dati relativi ai progetti di ricerca abbiamo proceduto, in collaborazione con AGID (Agenzia per l'Italia Digitale), allo sviluppo di nuova ontologia. Il diagramma dell'ontologia dei progetti è mostrata in Figura 4. L'ontologia è entrata a far parte del network di ontologie OntoPiA. L'URI dell'ontologia è <https://w3id.org/italia/onto/Project> e il namespace <https://w3id.org/italia/onto/Project/> è stato assegnato al prefisso `Project`. L'ontologia soddisfa i requisiti delineati dalle competency question elencate in Tabella 6.

L'ontologia definisce la classe `PublicInvestmentProject` che racchiude tutti i progetti ad investimento pubblico. La classe è definita come sottoclasse di `10:Object` e ha a sua volta due sottoclassi disgiunte tra loro `ProcurementProject` e `PublicResearchProject` che rappresentano rispettivamente i progetti di approvvigionamento e i progetti di ricerca. Ad i progetti pubblici è associato, tramite la object property `hasUniqueProjectCode`, uno (e uno solo) CUP (cioè, l'identificativo unico del progetto). Un progetto può essere associato ad una o più entità coinvolte a diverso titolo tramite la proprietà `involvesEntity`. Il progetto può avere uno o più agenti partecipanti tramite la proprietà `hasParticipantAgent` e un ente finanziatore, cioè una organizzazione (i.e. `COV:Organization`) associata tramite la object property `projectFunder`. Inoltre un progetto ha: una durata, specificata tramite l'ontologia TI di OntoPiA; un'estensione spaziale, specificata tramite OntoPiA CLV; un contact point, specificato riusando l'ontologia OntoPiA SM; una descrizione (`10:description`); un titolo,

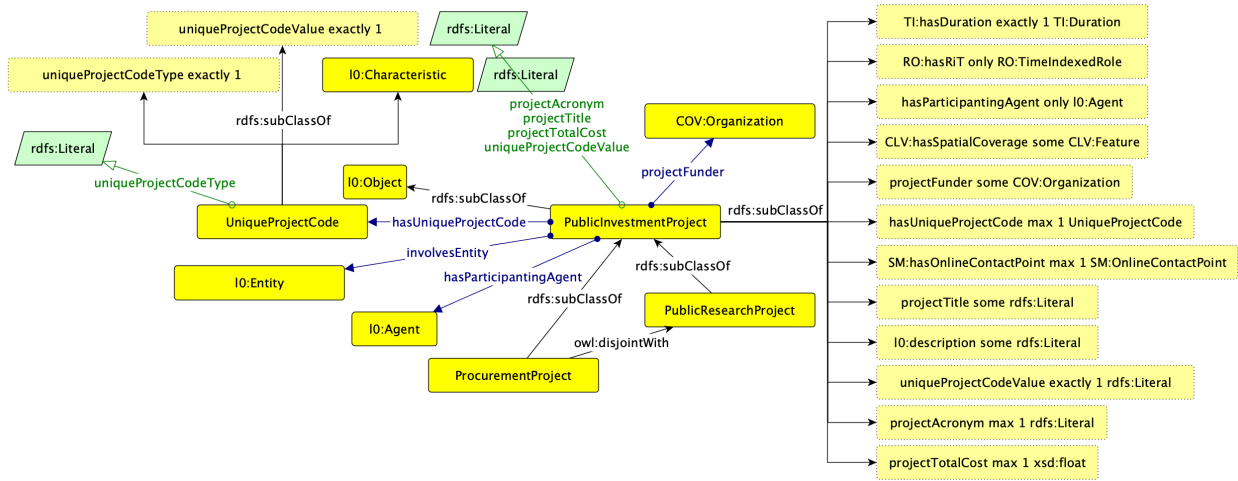


Figure 4: Diagramma dell'Ontologia dei Progetti.

i.e. `projectTitle`; un acronimo, `projectAcronym` e un costo totale (i.e. `projectTotalCost`).

2.5.5 Categorie di Contenuti del Prototipo

Al fine di garantire che la fruizione dei contenuti del prototipo possa essere organizzata secondo delle categorie relative al dominio della formazione (cf. Task 4.4) sono state sviluppate, usando il vocabolario SKOS⁴⁴, cinque tassonomie di concetti. Quattro di queste tassonomie intendono modellare la classificazione SSD (Settore-Scientifico Disciplinare) usata nelle università italiane. Le tassonomie sono disponibili online ai seguenti indirizzi: Area Disciplinare⁴⁵, Disciplina⁴⁶, Settore Affine⁴⁷, Tematica⁴⁸. La quinta tassonomia invece intende classificare i contenuti a seconda della tipologia di oggetto⁴⁹.

2.5.6 Modello per Oggetti

Il concetto di Oggetto nel dominio del progetto racchiude un'ampia varietà di tipologie di oggetti (cf. Sezione 2.5.5). Per molte di queste tipologie di oggetti sono disponibili modelli allo stato dell'arte per cui si ritiene che il riuso di tali modelli possa migliorare la fruizione dei contenuti. Tuttavia, si ritiene che una vista comune a tutti gli oggetti contenuti nella base di conoscenza possa migliorare la ricerca e la fruizione dei contenuti integrati nel sistema. Ovviamente, affinché questa vista sia comune a tutti gli oggetti della base di conoscenza, questa

⁴⁴<https://www.w3.org/2004/02/skos/>

⁴⁵https://w3id.org/ecodigit/resource/area_disciplinare

⁴⁶<https://w3id.org/ecodigit/resource/disciplina>

⁴⁷https://w3id.org/ecodigit/resource/settore_affine

⁴⁸<https://w3id.org/ecodigit/resource/tematica>

⁴⁹<https://w3id.org/ecodigit/resource/tipologia>

vista sarà di alto livello, cioè tutte le tipologie di oggetti saranno trattate come “Oggetto” generico, e catturerà solo un sotto-insieme di attributi comuni a tutte le tipologie di oggetto. Quello che verrà realizzato sarà quindi un modello multi-vista, cioè per ogni oggetto saranno disponibili più viste: una *generale* in cui verranno specificati un sotto-insieme di attributi comuni a tutti gli oggetti (e.g. nome, autore, data di creazione ecc.), e una *specifica* in cui verrà usato il modello specifico per quella tipologia di oggetto per descrivere le peculiarità di quella tipologia (e.g. DOI se si tratta di una pubblicazione o orientamento lungo asse Z se si tratta di un modello 3D).

Come detto, la metodologia seguita per la definizione del modello tende, per quanto possibile, di riusare ontologie esistenti allo stato dell’arte. In questa sezione forniremo delle indicazioni di come usare le ontologie disponibili allo stato dell’arte per implementare il modello multi-vista. Per la descrizione del modello useremo come esempio la modellazione dei dati riguardanti una pubblicazione, cioè [12], risorsa consultabile online al link⁵⁰. La descrizione del modello accompagnata da snippet di codice RDF è disponibile al link⁵¹.

Vista Generale. In questo paragrafo forniremo i dettagli sulla vista generale per gli oggetti.

- *Tipo.* Tutti gli oggetti devono essere specificati come risorse appartenenti alla classe `DUL:Object`.
- *Attributi Core.* Per specificare gli più comuni attributi degli oggetti si è scelto di usare il vocabolario Dublin Core. Tutti gli attributi, a meno del titolo e della descrizione, sono opzionali e potrebbero non applicarsi per delle particolari entità (e.g. entità senza identificativo univoco). Gli attributi base per un oggetto sono il titolo, l’identificativo, la descrizione dell’oggetto, i creatori dell’oggetto (elencati con una stringa), la lingua dell’oggetto, la data di creazione.
- *Altri attributi opzionali.* Altri attributi possono essere specificati come segue.
 1. *Coverage.* La proprietà coverage associata ad un oggetto viene usata per associare all’oggetto un entità spaziale a cui l’oggetto si riferisce. Ad esempio, l’articolo [12] si riferisce al territorio italiano. Un oggetto può essere associato anche a più entità spaziali.
 2. *Subject.* La proprietà subject permette di specificare l’argomento dell’oggetto. Per rappresentare il subject è preferibile usare risorse (invece che stringhe) provenienti da una fonte nota (e.g. vocabolario controllato o dataset ben conosciuto come DBpedia). Nell’esempio disponibile online, la pubblicazione [12] viene associata alle entità di DBpedia che definiscono i concetti di “Patrimonio Culturale” e “Linked Data”. La object property subject permette anche di classificare l’oggetto secondo gli schemi di classificazione del prototipo EcoDigit (cf. Sezione 2.5.5).

⁵⁰<https://w3id.org/ecodigit/work/Lodi2017Semantic>

⁵¹<https://github.com/ecodigit/workspace/wiki/Oggetto>

3. *Rights holder*. La proprietà rights holder permette di specificare la persona o l'organizzazione (in generale un agente) che detiene i diritti sull'oggetto. Nell'esempio disponibile online la pubblicazione Lodi2017Semantic viene associata all'organizzazione Springer che ne detiene i diritti.
4. *Relation*. La proprietà relation permette di specificare altri oggetti con cui l'oggetto descritto ha un qualche tipo di relazione. Nell'esempio disponibile online la pubblicazione Lodi2017Semantic viene associata alla pubblicazione Carriero2019 [4].
5. *URL*. La proprietà URL permette di specificare l'URL dell'oggetto descritto nel caso in cui l'oggetto sia un oggetto digitale (e.g. nel caso di modello 3D la URL specifica l'indirizzo da cui è scaricabile l'oggetto stesso), o una risorsa scaricabile associata all'oggetto descritto (e.g. nel caso in cui l'oggetto descriva un'area archeologica la URL potrebbe puntare ad una foto dell'area) o una realizzazione di esso (e.g. nel caso di un articolo potrebbe puntare al PDF che è la realizzazione dell'articolo). Nell'esempio online la pubblicazione Lodi2017Semantic viene associata alla URL che punta al PDF che ne è una sua realizzazione.
6. *Localizzazione*. Le indicazioni riguardo alla localizzazione di un oggetto rispetto ad un sistema di riferimento sono riportate in una sezione dedicata 2.5.7.
7. *Creator*. Nel caso in cui l'opera descritta nell'oggetto sia frutto di un lavoro creativo di un certo creator è possibile utilizzare la proprietà `frbr:creator` per associare l'oggetto con alla risorsa che identifica un suo creator. Nell'esempio la pubblicazione Lodi2017Semantic viene associata con le risorse che identificano gli autori dell'articolo. Un creator è sempre un agente (e.g. Persona o Organizzazione).
8. *Contributor*. Un contributor è un agente (e.g. una persona o una organizzazione) che ha contribuito alla creazione dell'oggetto descritto.
9. *Immagine*. E' possibile associare ad un oggetto una URL ad una immagine digitale che ne fornisce una anteprima o una descrizione.

Vista Speciale. Per la vista speciale si può usare una qualsiasi ontologia di dominio, ad esempio nel caso di Lodi2017Semantic si possono usare BIBO e le ontologie SPAR. Nella definizione della vista speciale alcuni attributi già specificati nella vista generale potranno essere eventualmente specializzati (ad esempio una pubblicazione avrà sia un `dc:identifier` che un `bibo:doi` per specificare il DOI). Questo creerà ridondanza dei dati ma permetterà ai dati di essere interrogati secondo le due viste.

2.5.7 Localizzazione

In questa sezione verranno fornite una serie di linee guida che definiscono come associare alle entità (e.g. Organizzazione o Oggetto) una localizzazione geografica espressa o con un indirizzo o attraverso un particolare sistema geodetico di riferimento.

Le indicazioni fornite in questa pagina sono conformi: alle direttive OGC per la codifica dei dati geografici in RDF e ai vocabolari definiti dall'AgID per le Pubbliche Amministrazioni⁵².

Le indicazioni, accompagnate da snippet di codice di esempio, sono disponibili online alla pagina⁵³. In questa sezione mostreremo quali sono i diversi modi per fornire una geolocalizzazione di una risorsa di esempio, cioè <https://example.org/resource>.

Localizzazione tramite indirizzo. Per localizzare un oggetto usando un indirizzo usare il modulo ontologico CLV (Core Location Vocabulary) di OntoPiA. Ad esempio:

```
<rdf:Description rdf:about="https://example.org/resource">
  <CLV:hasAddress>
    <CLV:Address>
      <CLV:fullAddress xml:lang="it">Piazzale Aldo Moro, 5,
        00185 Roma, Italia</CLV:fullAddress>
    </CLV:Address>
  </CLV:hasAddress>
</rdf:Description>
```

Localizzazione con Coordinate Puntuali. Per localizzare un oggetto con coordinate puntuali usare le ontologie Location Core Vocabulary, OGC's Simple Features e Geo Sparql.

```
<rdf:Description rdf:about="https://example.org/resource">
  <locn:geometry>
    <sf:Point>
      <geo:asWKT
        rdf:datatype= "http://www.opengis.net/ont/geosparql#wktLiteral">
        <![CDATA[<http://www.opengis.net/def/crs/OGC/1.3/CRS84>
          POINT(12.514480809352563 41.90375430114942)]]>
      </geo:asWKT>
    </sf:Point>
  </locn:geometry>
</rdf:Description>
```

Localizzazione con Coordinate Lineari . Per localizzare un oggetto con coordinate lineari usare le ontologie Location Core Vocabulary, OGC's Simple Features e Geo SPARQL.

```
<rdf:Description rdf:about="https://example.org/resource">
  <locn:geometry>
```

⁵²<http://www.opengis.net/doc/IS/geosparql/1.0>

⁵³<https://github.com/ecodigit/workspace/wiki/Localizzazione>

```

<sf:LineString>
  <geo:asWKT
    rdf:datatype= "http://www.opengis.net/ont/geosparql#wktLiteral">
    <![CDATA[<http://www.opengis.net/def/crs/OGC/1.3/CRS84>
    LINESTRING(12.510866848581259 41.9041536910628,
    12.512240139596884 41.90118317328348)]]>
  </geo:asWKT>
</sf:LineString>
</locn:geometry>
</rdf:Description>

```

Localizzazione con Coordinate Areali. Per localizzare un oggetto con coordinate lineari usare le ontologie Location Core Vocabulary, OGC's Simple Features e Geo Sparql.

```

<rdf:Description rdf:about="https://example.org/resource">
  <locn:geometry>
    <sf:Polygon>
      <geo:asWKT
        rdf:datatype= "http://www.opengis.net/ont/geosparql#wktLiteral">
        <![CDATA[<http://www.opengis.net/def/crs/OGC/1.3/CRS84>
        POLYGON((12.510990042879826 41.904068340163064,
        12.512320418551212 41.901193643852224,
        12.513908286288029 41.90042703631114,
        12.51837148208881 41.90202412496926,
        12.519058127596622 41.902790713336486,
        12.51562490005756 41.90556951901655,
        12.510990042879826 41.904068340163064)) ]]>
      </geo:asWKT>
    </sf:Polygon>
  </locn:geometry>
</rdf:Description>

```

2.5.8 Modello per la Provenance dei Dati

Con il termine *Provenance* si intendono tutte quelle informazioni associate ad una certa risorsa che descrivono come quella risorsa è stata generata. Queste informazioni sono associate a qualsiasi tipo di risorsa (sia essa di tipo Persona, Organizzazione, Progetto ecc.). Per descrivere le informazioni di provenance associate ad una certa entità è stata adottata l'ontologia PROV-O⁵⁴.

Alla pagina⁵⁵ sono disponibili delle indicazioni dettagliate sul modello per la provenance

⁵⁴<http://www.w3.org/ns/prov#>

⁵⁵<https://github.com/ecodigit/workspace/wiki/Provenance>

dei dati. Come esempio verrà descritto come associare le informazioni di provenance alla risorsa `example:resource`.

Innanzitutto bisogna dichiarare `example:resource` come appartenente alla classe `prov:Entity`. In seguito, usando la proprietà `prov:wasGeneratedBy` bisognerà associare all'entità una risorsa che descrive l'attività (i.e. un individuo della classe `prov:Activity`) che ha portato alla definizione della risorsa `example:resource`. All'attività è associato a:

1. Una descrizione tramite `rdfs:comment` in cui viene descritta in maniera generica l'attività che ha portato alla creazione del dato (e.g. "Attività di inserimento manuale compiuta da Mario Rossi").
2. L'agente che ha compiuto l'attività tramite la object property `prov:wasAssociatedWith`. L'agente può essere descritto come `foaf:Agent`.
3. Eventualmente, una o più fonti, specificate con la proprietà `prov:used` (indicate con delle URL) usate per produrre i dati.
4. Eventualmente, l'inizio e la fine dell'attività di produzione dei dati, specificate rispettivamente con le proprietà `prov:startedatTime` e `prov:endedAtTime`.

3 Conclusioni

Questo è il risultato di parte delle attività previste dal **Task 3.2** che ha l'obiettivo di definire il modello di integrazione, ovvero l'insieme di pratiche e standard tecnologici che una sorgente dovrà rispettare per entrare in EcoDigit. Il documento è una estensione della precedente versione **D3.2 Modello di ingresso** e riporta una revisione del modello fatta sulla base della esperienza della Proof-of-Concept [1].

References

- [1] Luigi Asprino, Ludovica Marinucci, Andrea Giovanni Nuzzolese, Valentina Presutti, Massimo Mecella, and Miguel Ceriani. *D3.4 Proof-of-Concept*. Deliverable Progetto EcoDigit. 2019.
- [2] Eva Blomqvist, Karl Hammar, and Valentina Presutti. “Engineering Ontologies with Patterns - The eXtreme Design Methodology”. In: *Ontology Engineering with Ontology Design Patterns*. Volume 25. Studies on the Semantic Web. IOS Press, 2016.
- [3] Eva Blomqvist, Valentina Presutti, Enrico Daga, and Aldo Gangemi. “Experimenting with eXtreme Design”. In: *Proc. of EKAW 2010* (Lisbon, Portugal). Volume 6317. Springer, 2010, pages 120–134.
- [4] Valentina Anita Carriero, Aldo Gangemi, Maria Letizia Mancinelli, Ludovica Marinucci, Andrea Giovanni Nuzzolese, Valentina Presutti, and Chiara Veninata. “ArCo: The Italian Cultural Heritage Knowledge Graph”. In: *The Semantic Web – ISWC 2019*. Edited by Chiara Ghidini, Olaf Hartig, Maria Maleshkova, Vojtěch Svátek, Isabel Cruz, Aidan Hogan, Jie Song, Maxime Lefrançois, and Fabien Gandon. Cham: Springer International Publishing, 2019, pages 36–52. ISBN: 978-3-030-30796-7.
- [5] Miguel Ceriani and Massimo Mecella. *D3.1 Report sul Censimento*. Deliverable Progetto EcoDigit. 2019.
- [6] Chris Dijkshoorn, Lora Aroyo, Jacco van Ossenbruggen, and Guus Schreiber. “Modeling cultural heritage data for online publication.” In: *Applied Ontology* 13.4 (2018), pages 255–271.
- [7] Martin Doerr. “The CIDOC Conceptual Reference Module: An Ontological Approach to Semantic Interoperability of Metadata”. In: *AI Magazine* 24.3 (2003), pages 75–92.
- [8] Aldo Gangemi, Silvio Peroni, and Fabio Vitali. “Literal reification”. In: *Proceedings of WOP 2010* (2010), pages 65–66.
- [9] Aldo Gangemi and Valentina Presutti. “Ontology Design Patterns”. In: *Handbook on Ontologies*. Edited by Steffen Staab and Rudi Studer. International Handbooks on Information Systems. Springer, 2009, pages 221–243.
- [10] Eero Hyvönen. “Semantic Portals for Cultural Heritage”. In: *Handbook on Ontologies*. Edited by Steffen Staab and Rudi Studer. International Handbooks on Information Systems. Springer, 2009, pages 757–778.
- [11] Antoine Isaac and Bernhard Haslhofer. “Europeana Linked Open Data - data.europeana.eu”. In: *Semantic Web* 4.3 (2013), pages 291–297.
- [12] Giorgia Lodi, Luigi Asprino, Andrea Giovanni Nuzzolese, Valentina Presutti, Aldo Gangemi, Diego Reforgiato Recupero, Chiara Veninata, and Annarita Orsini. “Semantic Web for cultural heritage Valorisation”. In: *Data Analytics in Digital Humanities*. Springer, 2017, pages 3–37.

- [13] Silvio Peroni, David Shotton, and Fabio Vitali. “Tools for the automatic generation of ontology documentation: A task-based evaluation”. In: *International Journal on Semantic Web and Information Systems (IJSWIS)* 9.1 (2013), pages 21–44.