

# ECODIGIT

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

## **D3.2 Modello di ingresso**

Acronimo Progetto:

Titolo Progetto:

**EcoDigit**

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

---

## D3.2

<b>Work Package:</b>	WP3 Task 3.2	
<b>Deliverable Dovuto il:</b>	2 Aprile 2019	
<b>Inizio Progetto:</b>	2 Ottobre 2018	
<b>Durata Progetto:</b>	15 Mesi	
<b>Reponsabile Deliverable:</b>	Valentina Presutti valentina.presutti@cnr.it	
<b>Versione:</b>	1.0	
<b>Stato:</b>	Versione Finale	
<b>Autore:</b>	Luigi Asprino	ISTC-CNR
	Ludovica Marinucci	ISTC-CNR
	Andrea Giovanni Nuzzolese	ISTC-CNR
	Valentina Presutti	ISTC-CNR
<b>Altri contribuenti al lavoro riportato nel deliverable:</b>	Marialuisa Mongelli	ENEA
	Antonio Budano	INFN
	Massimo Mecella	RM1
	Maria Prezioso	RM2
	Marco Canciani	RM3
	Giovanni Fiorentino	UNITUS
<b>Reviewer:</b>	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

## 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 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 **D3.2 Modello di ingresso**, è un primo 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 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. Infine, 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.

## Indice

<b>1</b>	<b>Introduzione</b>	<b>6</b>
1.1	Obiettivi del Work Package . . . . .	6
1.2	Obiettivo del deliverable . . . . .	6
1.3	Relazione con le altre attività del progetto . . . . .	6
<b>2</b>	<b>Definizione del modello</b>	<b>7</b>
2.1	Stato dell'arte . . . . .	7
2.2	Analisi dei dataset censiti nel Lazio . . . . .	8
2.3	Integrazione della S&TDL in EcoDigit . . . . .	10

## Lista delle Tabelle

1	<b>Allineamenti scheda Persone</b> . . . . .	12
2	<b>Allineamenti scheda Organizzazioni</b> . . . . .	13
3	<b>Allineamenti scheda Progetti</b> . . . . .	14
4	<b>Allineamenti scheda Prodotti</b> . . . . .	15
5	<b>Allineamenti scheda Dati della ricerca</b> . . . . .	16

## 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, basata su formati aperti e semantici.

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 si basa su metodi e tecniche di "metadattazione" e di 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 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 [2]. 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 [4]. 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) [3], 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) [5].

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)<sup>1</sup> 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 SIGECweb<sup>2</sup> e riversate nel DB Catalogo Generale dei Beni Culturali<sup>3</sup>, 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-CRM<sup>4</sup>, EDM<sup>5</sup>, FEntry<sup>6</sup> e OAEntry<sup>7</sup>, relative alle fotografie e alle opere d'arte, e i vocabolari controllati del Getty Research Institute<sup>8</sup>; (ii) *diretto* di classi e proprietà dai moduli *core* di OntoPIA<sup>9</sup>, una rete di ontologie e vocabolari controllati per la Pubblica Amministrazione, in particolare l'ontologia Cultural-ON<sup>10</sup> che modella gli eventi culturali e gli istituti e luoghi della cultura [6].

<sup>1</sup><http://dati.beniculturali.it/progetto-arco-architettura-della-conoscenza/>

<sup>2</sup><http://www.iccd.beniculturali.it/it/sigec-web>

<sup>3</sup><http://catalogo.beniculturali.it>

<sup>4</sup><http://cidoc-crm.org/>

<sup>5</sup><https://pro.europeana.eu/page/edm-documentation>

<sup>6</sup><http://www.essepuntato.it/2014/03/fentry>

<sup>7</sup><http://purl.org/emmedi/oaentry>

<sup>8</sup><http://www.getty.edu/research/tools/vocabularies/index.html>

<sup>9</sup><https://github.com/italia/daf-ontologie-vocabolari-controllati/tree/master/Ontologie>

<sup>10</sup><http://dati.beniculturali.it/lodview/cis/.html>



Ulteriori ontologie di riferimento per la definizione del modello di integrazione del sistema EcoDigit sono: la *FOAF*<sup>11</sup> che rappresenta le informazioni anagrafiche sulle persone in generale; la *CERIF*<sup>12</sup> 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)<sup>13</sup> che descrive i progetti in generale; la *Org Ontology*<sup>14</sup> che descrive le organizzazioni in generale ed è integrata ed estesa dall'ontologia *COV*<sup>15</sup> di *OntoPiA*; la *BIBO*<sup>16</sup> e le *SPAR Ontologies*<sup>17</sup>, relative al dominio delle pubblicazioni e dei riferimenti bibliografici; l'*Arpenteur Ontology*<sup>18</sup>, dedicata ai beni archeologici e architettonici; il *WGS84 Geo Positioning (geo) vocabulary*<sup>19</sup> per allineare i dati georeferenziati e il *Dublin Core Metadata*<sup>20</sup> espresso in *RDF Schema*, che consiste di concetti generali e di alto livello.

## 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. [1]). 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 metadattazione 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.).

<sup>11</sup><http://xmlns.com/foaf/spec/>

<sup>12</sup><https://www.eurocris.org/ontologies/cerif/>

<sup>13</sup><http://usefulinc.com/ns/doap#>

<sup>14</sup><https://www.w3.org/TR/vocab-org/>

<sup>15</sup><https://github.com/italia/daf-ontologie-vocabolari-controllati/tree/master/Ontologie/COV>

<sup>16</sup><http://bibliographic-ontology.org/>

<sup>17</sup><http://www.sparontologies.net/ontologies>

<sup>18</sup><http://www.arpenteur.org/ontology/Arpenteur.owl>

<sup>19</sup>[http://www.w3.org/2003/01/geo/wgs84\\_pos](http://www.w3.org/2003/01/geo/wgs84_pos)

<sup>20</sup><http://www.dublincore.org/specifications/dublin-core/dcmi-terms/>

- *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 2.1).

- *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 metadattazione 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 beneventana<sup>21</sup>.
- *MaGI. Manoscritti greci d'Italia*. Censimento uniforme dei dati identificativi essenziali di tutti i manoscritti greci conservati sul territorio italiano, accompagnati da almeno una

<sup>21</sup><http://edu.let.unicas.it/bmb/>

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 sovraesposte 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.
- *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 Integrazione della S&TDL in EcoDigit

In questa sezione, si intende proporre un'analisi tecnica che arriva a un maggior dettaglio degli allineamenti semantici possibili tra gli schemi di dati della *Science & Technology Dig-*

*ital Library* (S&TDL)<sup>22</sup> 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 grazie alla collaborazione con il team del progetto *Anagrafe delle Competenze*<sup>23</sup> 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<sup>24</sup>, Prodotti<sup>25</sup>, Progetti<sup>26</sup>, Organizzazioni<sup>27</sup> e Dati della ricerca<sup>28</sup>.

Considerata la natura di questi dati, che hanno come fornitori di contenuti vari enti culturali e di ricerca (quali, ad esempio, il *Consiglio Nazionale delle Ricerche*, l'*Archivio Centrale dello Stato*, l'*Istituto Luigi Sturzo*, ecc.) sono state scelte due ontologie principali che descrivono concetti di alto livello:

- (i) *FOAF*<sup>29</sup> rappresenta le informazioni anagrafiche sulle persone in generale (ad esempio, foaf:Person, foaf:name, foaf:familyName, foaf:age, ecc.) messe in relazione con attività ed enti (ad esempio, foaf:Project, foaf:Organization, foaf:Group, ecc.);
- (ii) *CERIF*<sup>30</sup> specifica le informazioni relative ai ruoli, alle competenze, ai progetti e ai prodotti nell'ambito della ricerca scientifica (ad esempio, cerif:Person, cerif:Project, cerif:Organization, cerif:Product, ecc.).

Questi due modelli semantici, coprendo aspetti generali del dominio di riferimento, necessitano di ulteriori specificazioni presenti in altre ontologie che descrivono nel dettaglio domini specifici (ad esempio, quello dei social media, dei riferimenti bibliografici, delle organizzazioni pubbliche e private, ecc.). Di seguito vengono riportate le altre ontologie che rappresentano nella maniera più completa possibile ciascuno schema XML della S&TDL.

**Persone.** La mappatura di questa tipologia di dati è quasi del tutto esaurita dall'allineamento con classi e proprietà appartenenti alle ontologie sopramenzionate, in particolare in merito

<sup>22</sup><https://dl.cnr.it/>

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

<sup>24</sup><https://bit.ly/2TZGd7V>

<sup>25</sup><https://bit.ly/2UF68yR>

<sup>26</sup><https://bit.ly/2JxrxBY>

<sup>27</sup><https://bit.ly/2W9J0ZE>

<sup>28</sup><https://bit.ly/2TVAvDZ>

<sup>29</sup><http://xmlns.com/foaf/spec/>

<sup>30</sup><https://www.eurocris.org/ontologies/cerif/>

Campo file XML della S&TDL	Classe con prefisso Ontologia
<DOCTYPE/>	cerif:Person (subClassOf foaf:Person)
<firstName/>	foaf:firstName
<lastName></lastName>	foaf:lastName
<email/>	cerif:ElectronicAddress
<contacts/>	SM:OnlineContactPoint
<profile/>	RO:Role
<education/>	cerif:Qualification
<work/>	foaf:workInfoHomepage
<knownLanguages/>	curriculum:Language
<skills/>	curriculum:FieldOfStudy

Table 1: **Allineamenti scheda Persone**

al concetto di *Persona*, modellato dalla classe `cerif:Person` (subClassOf `foaf:Person`). A questa classe sono collegate informazioni inerenti alla persona in generale (come, ad esempio, nome e cognome, rappresentati rispettivamente dalle proprietà `foaf:firstName` e `foaf:lastName`) e alla persona che ricopre un ruolo in un ente culturale e/o di ricerca (ad esempio, l'informazione relativa all'email istituzionale, contenuta nel campo XML `<email/>`, è allineata alla classe `cerif:ElectronicAddress` oppure le informazioni relative ai titoli di studio, previste dal campo XML `<education/>`, sono allineate con la classe `cerif:Qualification`). Per allineare ulteriori informazioni specifiche si possono utilizzare classi e proprietà appartenenti a ontologie tra loro collegate in una vera e propria network chiamata *OntoPia*<sup>31</sup>, che è definita come modello standard per la Pubblica Amministrazione italiana, quali: l'ontologia RO<sup>32</sup> per la specifica dei ruoli delle persone (ad esempio, il campo XML `<profile/>`, l'ontologia SM<sup>33</sup> per la modellazione dei social media e delle informazioni di contatto digitali (ad esempio, il campo XML `<contacts/>` è allineato con la classe `SM:OnlineContactPoint`), e l'ontologia top-level IO<sup>34</sup> che consente di collegare tutte le ontologie sopra elencate abilitando così la rete di ontologie. Inoltre, la *Curriculum Ontology*<sup>35</sup> della BBC rappresenta un valido strumento per mappare la descrizione, ad esempio, delle competenze linguistiche (`<knownLanguages/>`) e delle competenze relative a un certo ambito di studio (`<skills/>`), rispettivamente, con le classi `curriculum:Language` e `curriculum:FieldOfStudy`.

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**.

**Organizzazioni.** Anche questi dati sono rappresentati dalle classi e proprietà delle due ontologie sopramenzionate, FOAF e CERIF, in particolare relativamente al concetto di *Organiz-*

<sup>31</sup><https://github.com/italia/daf-ontologie-vocabolari-controllati>

<sup>32</sup><https://github.com/italia/daf-ontologie-vocabolari-controllati/tree/master/Ontologie/RO>

<sup>33</sup><https://github.com/italia/daf-ontologie-vocabolari-controllati/tree/master/Ontologie/SM>

<sup>34</sup><https://github.com/italia/daf-ontologie-vocabolari-controllati/tree/master/Ontologie/IO>

<sup>35</sup><https://www.bbc.co.uk/ontologies/curriculum>

Campo file XML della S&TDL	Classe con prefisso Ontologia
<DOCTYPE/>	cerif:Organization (subClassOf foaf:Organization)
<denominazione/>	COV:legalName
<sigla/>	cerif:acronym
<indirizzo/>	CLV:Address
<telefono/>	SM:TelephoneType(voce)
<missione/>	COV:mainFunction
<responsabile/>	RO:Role
<attivita_ricerca/>	COV:ActivityType
<websiteurl/>	SM:WebSite

Table 2: Allineamenti scheda Organizzazioni

zazione, modellato dalla classe `cerif:Organization` (subClassOf `foaf:Organization`). A questa classe l'ontologia CERIF collega la proprietà `cerif:acronym` che descrive le informazioni relative alla sigla del nome dell'ente preso in considerazione (ad esempio, CNR), come richiesto dal campo XML `<sigla/>`. Per l'allineamento semantico di ulteriori informazioni è particolarmente adatta l'ontologia COV<sup>36</sup>, anch'essa appartenente alla network Ontopia e relativa alle organizzazioni pubbliche e private. I principali campi XML che sono stati allineati sono: `<denominazione/>`, relativo alle informazioni sul nome dell'ente, con la proprietà `COV:legalName`; `<missione/>`, relativo alle informazioni sulla funzione ufficiale dell'ente, con la proprietà `COV:mainFunction`; `<attivita_ricerca/>`, relativo alle informazioni che descrivono il tipo di attività dell'ente, con la classe `<COV:ActivityType/>`. Inoltre, altre ontologie di Ontopia utili per gli allineamenti sono: l'ontologia CLV<sup>37</sup> sugli indirizzi e i luoghi, per le informazioni, ad esempio, sulla sede fisica dell'ente (`<indirizzo/>`), che sono associate alla classe `CVL:Address`; l'ontologia sopramenzionata SM sui social media per le informazioni relative, ad esempio, al recapito telefonico dell'ente (`SM:TelephoneType(voce)`) o al sitoweb ufficiale (`SM:WebSite`); la sopramenzionata ontologia RO sui ruoli di cui si utilizza la classe `RO:Role` per il campo XML `<responsabile/>`.

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

**Progetti.** Questo tipo di dati sono in gran parte catturati dalla modellazione del concetto di *Progetto* nella classe `cerif:Project` (subClassOf `foaf:Project`), che è collegata, ad esempio, alle proprietà del nome (`cerif:name`), delle date di inizio (`cerif:startDate` e di fine (`cerif:startEnd`), e del riassunto sintetico (`cerif:abstract`) del progetto. Inoltre, ulteriori informazioni sono allineate con due delle sopramenzionate ontologie di Ontopia: RO, per quanto riguarda il ruolo di responsabile del progetto (`RO:Role`) e IO, per quanto riguarda l'argomento generale di cui tratta il progetto (`IO:Topic`). A questa mappatura si aggiunge

<sup>36</sup><https://github.com/italia/daf-ontologie-vocabolari-controllati/tree/master/Ontologie/COV>

<sup>37</sup><https://github.com/italia/daf-ontologie-vocabolari-controllati/tree/master/Ontologie/CLV>

Campo file XML della S&TDL	Classe con prefisso Ontologia
<DOCTYPE/>	cerif:Project (subClassOf foaf:Project)
<titolo/>	cerif:name
<anno_inizio/>	cerif:startDate
<anno_fine/>	cerif:endDate
<abstract/>	cerif:abstract
<responsabile/>	RO:Role
<tematiche/>	l0:Topic
<obiettivi/>	deo:Motivation
<stato_dell_arte/>	deo:RelatedWork

Table 3: Allineamenti scheda Progetti

una delle SPAR Ontologies<sup>38</sup>, relative al dominio delle pubblicazioni e dei riferimenti bibliografici, ovvero l'ontologia DEO (Discourse Element Ontology)<sup>39</sup>, che fornisce il vocabolario degli elementi retorici di cui è costituito un documento. Infatti, considerando il progetto come un documento scritto articolato di varie parti, il campo XML <obiettivi/> è allineato alla classe deo:Motivation, mentre il campo <stato\_dell\_arte/> è allineato alla classe deo:RelatedWork.

Nella **Tabella 3** si trovano esempi di allineamenti tra campi XML dei Progetti e relative ontologie.

**Prodotti.** Le informazioni contenute nella scheda dei *Prodotti* consistono in pubblicazioni scientifiche, rappresentabili dalla classe cerif:Publication, che è sottoclasse di bibo:Document, appartenente all'ontologia BIBO<sup>40</sup> relativa alla descrizione dei riferimenti bibliografici. Per questo tipo di dati, infatti, è necessario aggiungere ontologie che descrivono i diversi aspetti e componenti della documentazione scritta. Particolarmente rilevanti sono due delle SPAR ontologies, ovvero: l'ontologia DoCO<sup>41</sup>, che importa l'ontologia DEO sopra-mentzionata e fornisce il vocabolario per la descrizione delle parti, sia strutturali che retoriche, costituenti un documento scritto; l'ontologia FaBiO<sup>42</sup>, che importa l'ontologia di meta-livello FRBR<sup>43</sup> e descrive le entità pubblicate o potenzialmente pubblicabili in edizioni scientifiche. Quest'ultima è la più utilizzata: ad esempio, il campo XML <anno/>, relativo all'anno di pubblicazione, è allineata alla proprietà fabio:hasPublicationDate; <lingua/>, che rappresenta la lingua della pubblicazione, è allineato a fabio:hasLanguage; il campo relativo alle informazioni sull'editore è associato a fabio:hasPublisher; il campo relativo all'ISBN (International Standard Book Number) è allineato a fabio:hasISBN. Inoltre, dell'ontologia DoCO

<sup>38</sup><http://www.sparontologies.net/ontologies>

<sup>39</sup><https://sparontologies.github.io/deo/current/deo.html>

<sup>40</sup><http://bibliographic-ontology.org/>

<sup>41</sup><https://sparontologies.github.io/doco/current/doco.html>

<sup>42</sup><https://sparontologies.github.io/fabio/current/fabio.html>

<sup>43</sup><https://sparontologies.github.io/frbr/current/frbr.html>

Campo file XML della S&TDL	Classe con prefisso Ontologia
<DOCTYPE/>	cerif:Publication (subClassOf bibo:Document)
<titolo/>	cerif:title
<anno/>	fabio:hasPublicationDate
<lingue/>	fabio:hasLanguage
<abstract/>	cerif:abstract
<autori/>	doco:ListOfAuthors
<affiliazioni/>	doco:ListOfOrganizations
<editori/>	fabio:hasPublisher
<isbn/>	fabio:hasISBN

Table 4: **Allineamenti scheda Prodotti**

sono utilizzate, ad esempio, le classi: `doco:ListOfAuthors` per il campo XML `<autori/>`, ovvero gli autori della pubblicazione che sono schedati come Persone nella Repository della S&TDL, e `doco:ListOfOrganizations` per il campo `<affiliazioni/>`, relativo all'ente di appartenenza dell'autore.

Alcuni allineamenti dello schema XML dei Prodotti con le relative classi e proprietà delle ontologie sopradescritte sono riportati nella **Tabella 4**.

**Dati della ricerca.** Le informazioni contenute in queste schede della S&TDL sono relative a progetti di ricerca che hanno avuto come risultato la creazione di dataset georiferiti. Quindi, la classe di riferimento è `fabio:Dataset` che ha relazioni con: le proprietà dell'ontologia CERIF sopramenzionata, per esempio per quanto riguarda il titolo del dataset (`cerif:title`) oppure la sua data di inizio (`cerif:startDate`) ecc.; le proprietà dell'ontologia top level I0 di Ontopia sopradescritta per quanto riguarda i campi relativi alla descrizione generica (`10:description`) e dell'argomento (`10:Topic`) del dataset. Particolarmente rilevanti in questo contesto sono i campi relativi alla geolocalizzazione dei dati che sono rappresentabili tramite l'ontologia CLV sopramenzionata: ad esempio, il campo `<localizzazione/>` è allineato alla classe `CLV:Geometry`, che rappresenta la geometria (ovvero la rappresentazione spaziale in termini di coordinate geometriche) di un'entità spaziale; ad essa è collegata la proprietà `CLV:coordinateSystem`, che indica le coordinate del sistema di riferimento utilizzato ed è quindi capace di rappresentare le informazioni contenute nel campo XML `<tipo_rappresentazione_spaziale/>`.

Nella **Tabella 5** si trovano gli allineamenti della scheda XML dei Dati della ricerca con le relative ontologie.



Campo file XML della S&TDL	Classe con prefisso Ontologia
<DOCTYPE/>	fabio:Dataset
<titolo/>	cerif:title
<localizzazione/>	CLV:Geometry
<topic/>	l0:Topic
<descrizione/>	l0:Description
<websiteurls/>	SM:WebSite
<periodo_inizio/>	cerif:startDate
<periodo_fine/>	cerif:endDate
<diritti/>	fabio:hasRights
<tipo_rapresentazione_spaziale/>	CLV:coordinateSystem

Table 5: Allineamenti scheda Dati della ricerca

## References

- [1] Miguel Ceriani and Massimo Mecella. *D3.1 Report sul Censimento*. Deliverable Progetto EcoDigit. 2019.
- [2] 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.
- [3] Martin Doerr. “The CIDOC Conceptual Reference Module: An Ontological Approach to Semantic Interoperability of Metadata”. In: *AI Magazine* 24.3 (2003), pages 75–92.
- [4] Eero Hyvönen. “Semantic Portals for Cultural Heritage”. In: *Handbook on Ontologies*. 2009, pages 757–778.
- [5] Antoine Isaac and Bernhard Haslhofer. “Europeana Linked Open Data - data.europeana.eu”. In: *Semantic Web* 4.3 (2013), pages 291–297.
- [6] 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.