
Catalogo Nazionale Dati per l'Interoperabilità Semantica

Release versione-corrente

Team per la Trasformazione Digitale

27 mag 2026

1	Premesse	3
1.1	Struttura della presente guida	3
1.2	Sviluppo collaborativo della Guida al Catalogo	3
1.3	La semantica dei dati della PA	4
1.4	Cos'è il Catalogo	4
1.5	Scopo del Catalogo	5
1.6	Riferimenti normativi	5
1.7	Contenuti del Catalogo	6
1.7.1	Creazione e gestione dei repository per le risorse semantiche	7
1.7.2	Quando chiedere semantic stewardship ad Istat	7
1.8	Scelta degli identificativi univoci nel Web	7
2	Funzionamento generale	9
2.1	Attori	9
2.2	Avvio della contribuzione	10
2.3	Raccolta e memorizzazione dei dati semantici	10
2.4	Esposizione dei contenuti semantici	10
3	Come contribuire	11
3.1	Attività propedeutiche alla contribuzione al Catalogo	11
3.1.1	Aggiunta di nuove risorse	11
3.1.2	Richiesta di modifica di risorse già in Catalogo	12
3.2	Richiesta formale di contribuzione	13
3.3	Contribuzione continua al Catalogo	13
4	Come utilizzare le risorse	15
4.1	Ricerca di risorse semantiche	15
4.2	Scheda della risorsa semantica	18
4.2.1	Ontologia	18
4.2.2	Vocabolario controllato	18
4.2.3	Schema	21
4.2.4	Richiesta di aggiornamento di asset semantici esistenti	21
5	Manuale operativo	23
5.1	Identificativi univoci delle risorse	23
5.1.1	Introduzione al redirect con w3id.org	24
5.1.2	Processo di pubblicazione degli htaccess	24

5.1.3	Contenuto dei file htaccess e readme	25
5.1.4	Esempi di strumenti a supporto dei test	29
5.2	Indicazioni su modellazione e metadatazione degli asset semantici	30
5.2.1	Nuovo vocabolario controllato	30
5.2.2	Nuova ontologia	31
5.2.3	Nuovi schemi dati	33
5.3	Indicazioni su aggiornamento di asset semantici esistenti	34
5.3.1	Modifiche a vocabolari controllati esistenti	35
5.3.2	Modifiche a ontologie esistenti	35
5.3.3	Modifiche a schemi dati	35
5.4	Istruzioni su come predisporre il repository in cui pubblicare le risorse semantiche	35
5.4.1	Struttura di base del repository	36
5.4.2	Contenuto del repository	36
5.4.3	Nomi delle cartelle e dei file	36
5.4.4	Proiezioni in CSV dei vocabolari controllati	36
5.4.5	Versionamento	37
5.4.6	Approfondimenti sugli schemi dati	38
5.4.7	Esempi	38
6	Crediti	47

La presente guida è uno strumento dedicato alle Pubbliche Amministrazioni che vogliono fruire e/o contribuire all'alimentazione dei contenuti semantici del Catalogo, e più in generale ai potenziali utenti del National Data Catalog raggiungibile al sito schema.gov.it¹.

Per le indicazioni su come esplorare la guida, fai riferimento alla sezione [struttura della guida](#).

Se sei nuovo nel mondo della semantica dei dati, e hai bisogno di un'introduzione generale sul tema, fai riferimento al documento [Introduzione alla semantica dei dati e del Web Semantico](#)².

La guida è aggiornata a Giugno 2024

¹ <https://schema.gov.it>

² <https://teamdigitale.github.io/dati-semantic-doc-introduzione-semantic/>

La sezione contiene informazioni sulla struttura e sullo sviluppo collaborativo della presente guida, e sul contesto, lo scopo e il contenuto del Catalogo.

1.1 Struttura della presente guida

La presente guida, dopo una breve introduzione sul ruolo della semantica dei dati nell'ambito dell'interscambio informativo tra le pubbliche amministrazioni, fornisce indicazioni ai Contributori su come contribuire all'alimentazione del Catalogo, descrive le modalità di fruizione delle risorse semantiche tramite il portale [schema.gov.it](https://www.schema.gov.it)³ e dettaglia in un manuale operativo gli aspetti tecnici a supporto della contribuzione e comprensione del contenuto del Catalogo.

Se sei un Contributore ti consigliamo di leggere approfonditamente [Come contribuire](#) per comprendere il processo di iscrizione al Catalogo, e il [Manuale operativo](#) per reperire le indicazioni tecniche a supporto dello stesso.

Se sei uno sviluppatore di API/e-services, un ricercatore o un qualsiasi utente interessato a fruire dei contenuti semantici ti consigliamo di consultare [Come utilizzare le risorse](#).

Se vuoi contribuire al miglioramento e integrazione della presente guida, ti consigliamo di leggere [Sviluppo collaborativo della guida al Catalogo](#)

1.2 Sviluppo collaborativo della Guida al Catalogo

La presente guida è un documento pubblico, e chiunque può partecipare al processo di revisione e aggiornamento aprendo issue (per le discussioni) e pull request (per le proposte) sul [repository sorgente](#)⁴ che ne ospita i contenuti in file `.rst`.

Le modifiche al contenuto possono essere applicate direttamente tramite gli strumenti editor e preview integrati in Github; una volta completate, occorrerà fare commit delle modifiche, creare una pull request e attendere la revisione della stessa da parte degli amministratori del repository.

³ <https://www.schema.gov.it/>

⁴ <https://github.com/teamdigitale/dati-semantic-guida-ndc-docs>

Qui⁵ è disponibile una guida alla sintassi *reStructuredText*.

Altre risorse per l'editing in formato *reStructuredText*:

- Editor per il testo⁶
- Editor per le tabelle⁷
- Altro⁸

1.3 La semantica dei dati della PA

L'obiettivo di migliorare i servizi resi al cittadino mediante il pieno sfruttamento del patrimonio informativo pubblico rende sempre più chiara l'esigenza di “*abilitare lo sviluppo di una concreta interoperabilità semantica tra Pubbliche Amministrazioni a livello nazionale e transfrontaliero*” (cfr. [Linee Guida per l'Interoperabilità Semantica attraverso i Linked Open Data, Commissione di coordinamento SPC - 2012](#))⁹. L'interoperabilità semantica è un collegamento nativo tra i dati attraverso la definizione di un loro significato esplicito, condiviso ed espresso in un linguaggio formale. Solo assicurando l'interoperabilità semantica si rende possibile, nell'ambito di un interscambio di dati, una condivisione immediata delle informazioni e della conoscenza; viceversa, senza interoperabilità semantica, non vi è certezza che tutti i soggetti partecipanti allo scambio informativo concordino sull'attribuire a un determinato dato un unico significato.

Con specifico riferimento ai dati della PA, la semantica assume caratteri peculiari dettati dal fatto che le relative entità concettuali cui i dati si riferiscono (ad es. persona anagraficamente residente, persona fiscalmente residente, reddito, contribuente, imprenditore, società, azienda ecc.) in molti casi devono essere definite con il filtro dell'ordinamento giuridico che le ha create: ad esempio, imprenditore, società e azienda non sono entità del mondo naturale, ma tre distinte entità giuridiche la cui semantica è contenuta nelle relative fonti normative. Del resto, i dati pubblici vengono prodotti dai relativi procedimenti amministrativi regolati dal principio di legalità.

Nell'ambito di una strategia ispirata ai linked open data (LOD), inoltre, la condivisione delle informazioni non solo è immediata, ma anche automatica nella misura in cui il significato non è solo consultabile da persone ma anche elaborabile da software per realizzare applicazioni innovative, anche capaci di scoprire nuova conoscenza grazie alla navigazione dei collegamenti e al ragionamento sul significato attribuito ai dati (*automatic reasoning*).

In un contesto in cui la Piattaforma Digitale Nazionale dati (PDND) abilita l'interscambio automatico di dati tra le pubbliche amministrazioni, il Catalogo persegue l'interoperabilità semantica dei dati della PA al fine di rendere preventivamente consapevoli i fruitori della PDND circa il significato dei dati erogati con gli e-service.

Nel processo di perseguimento dell'interoperabilità semantica tramite linked open data, si distinguono, dunque, due componenti: la preliminare analisi concettuale dei dati condotta alla luce dell'impianto normativo di riferimento e finalizzata a rendere disponibile la conoscenza e, dall'altra, l'organizzazione (vocabolari controllati) e la rappresentazione (ontologie) della conoscenza così acquisita al fine di renderla *machine readable*.

Per approfondire la semantica dei dati fare riferimento [qui](#)¹⁰.

1.4 Cos'è il Catalogo

La trasformazione digitale della Pubblica Amministrazione è tra le priorità dell'UE e, per la sua realizzazione, è necessario, dunque, perseguire l'interoperabilità nell'ambito del settore pubblico. L'interoperabilità consente la con-

⁵ <http://docutils.sourceforge.net/docs/user/rst/quickref.html>

⁶ <http://rst.ninjs.org>

⁷ <http://truben.no/table/>

⁸ <http://docutils.sourceforge.net/docs/user/links.html#editors>

⁹ https://www.agid.gov.it/sites/default/files/repository_files/documentazione_trasparenza/cdc-spc-gdl6-interoperabilitasemopendata_v2.0_0.pdf

pdf

¹⁰ <https://teamdigitale.github.io/dati-semantic-doc-introduzione-semantica/>

divisione delle informazioni e delle conoscenze garantendo che i dati possano essere scambiati senza criticità. Tra i livelli di interoperabilità, l'*European Interoperability Framework* (EIF 2017) individua anche la dimensione semantica ossia il requisito che garantisce che il formato e il significato dei dati e delle informazioni scambiate siano preservati e compresi in qualsiasi scambio tra le parti. L'obiettivo è il riutilizzo del patrimonio informativo pubblico al fine di garantire l'attuazione del principio *once only* nei rapporti tra pubbliche amministrazioni e destinatari dei relativi servizi. Una simile prospettiva si concretizza con la predisposizione di *e-service* finalizzati, appunto, ad allungare il ciclo di vita dei dati amministrativi fino al loro riutilizzo tramite interscambio sulla PDND. Nel quadro nazionale, le Linee Guida per l'interoperabilità tecnica delle pubbliche amministrazioni (determinazione AGID n. 547/2021) individuano le tecnologie e gli standard che le pubbliche amministrazioni devono tenere in considerazione durante la realizzazione dei propri sistemi informatici, al fine di permettere il coordinamento informativo e informatico dei dati del settore pubblico.

Nell'ambito degli investimenti del Piano Nazionale di Ripresa e Resilienza, il Dipartimento per la Trasformazione Digitale, come **soggetto titolare** dell'intervento "Catalogo nazionale dati" M1C1 sub investimento 1.3.1 PNRR, e l'Istat, come **soggetto attuatore**, hanno stipulato un accordo per la realizzazione del Catalogo Nazionale per l'interoperabilità semantica. L'Istat garantisce il coordinamento e la realizzazione operativa del Catalogo; le pubbliche amministrazioni sono responsabili delle risorse semantiche (ontologie, vocabolari controllati e schemi dati) relative ai propri dati e pubblicate nel Catalogo.

Il Catalogo, accessibile tramite schema.gov.it, è frutto della collaborazione tra la Presidenza del Consiglio dei ministri, l'Istat, AGID, ISTC-CNR e PagoPA.

1.5 Scopo del Catalogo

Il Catalogo intende contribuire e agevolare l'interoperabilità tra le basi di dati di enti diversi con la progettazione e implementazione di una semantica condivisa che supporta la definizione di servizi digitali. Uno dei suoi possibili utilizzi è quello di abilitare la ricerca e il riuso di risorse semantiche (i.e., ontologie, schemi dati e vocabolari controllati) per lo sviluppo di API (*Application Programming Interface*) sulla PDND, che siano semanticamente e sintatticamente interoperabili grazie a standard specifici come RDF e i servizi REST, rispettivamente, contenuti nel Catalogo.

Un prerequisito per la costruzione del catalogo è la creazione di rappresentazioni formali del significato dei dati della PA fondata su un'analisi dei concetti utilizzati e che consideri attentamente la particolare natura degli stessi. Una parte preponderante del patrimonio informativo pubblico è costituita da dati generati nello svolgimento di procedimenti amministrativi che corrispondono, dunque, ad entità anche dell'ordinamento giuridico e relative relazioni tra esse. In tal caso, la chiave di lettura necessaria per le analisi concettuali dei dati amministrativi sono le classificazioni giuridiche.

Tale metodo consente di perseguire un'armonizzazione semantica anche in un contesto "federato" come quello del Catalogo, in cui vengono raccolte risorse semantiche gestite dai singoli enti, che designeremo come Contributori al Catalogo. Questo approccio garantisce la condivisione del significato dei dati e, quindi, di un appropriato livello semantico di interoperabilità.

1.6 Riferimenti normativi

L'interoperabilità semantica nella pubblica amministrazione è un obiettivo dell'UE e nazionale. A livello europeo, la Commissione ha raccolto alcune raccomandazioni nello *European Interoperability Framework*¹¹, punto di riferimento per fonti normative UE come la Direttiva (UE) 2019/1024 sugli open data¹², il Regolamento (UE) 2023/138 sui dati di elevato valore¹³, il Regolamento (UE) 2018/1724 sullo sportello digitale unico¹⁴ e il Regolamento (UE) 2022/1463

¹¹ https://ec.europa.eu/isa2/sites/default/files/eif_brochure_final.pdf

¹² <https://eur-lex.europa.eu/eli/dir/2019/1024/oj/ita>

¹³ https://eur-lex.europa.eu/legal-content/IT/TXT/?uri=uriserv%3AOJ.L_.2023.019.01.0043.01.ITA&toc=OJ%3AL%3A2023%3A019%3ATOC

¹⁴ <https://eur-lex.europa.eu/legal-content/IT/TXT/?uri=CELEX%3A32018R1724>

sul sistema tecnico per l'applicazione del principio "una tantum"¹⁵; nel quadro nazionale, è il CAD¹⁶ a tracciare le direttrici normative sulle cui basi vengono adottati atti programmatici come il Piano triennale per l'informatica¹⁷, nel quale si indica che, per favorire lo scambio di informazioni tra pubbliche amministrazioni, è necessario definire una data governance per armonizzare e standardizzare codici e nomenclature ricorrenti, identificare e definire modelli di dati (ontologie e vocabolari controllati) condivisi. In questo modo, il Catalogo favorisce l'uniformità dei dati trasversali, come persone, organizzazioni, servizi e luoghi, modellando dati e metadati di domini applicativi specifici.

Ulteriori riferimenti per il Catalogo sono:

- Linee Guida sull'interoperabilità tecnica delle Pubbliche Amministrazioni¹⁸
- Linee Guida per l'Interoperabilità Semantica attraverso i Linked Open Data¹⁹
- Linee guida di design per i siti internet e i servizi digitali della PA²⁰

1.7 Contenuti del Catalogo

Il Catalogo raccoglie in schema.gov.it²¹ le risorse semantiche già pubblicate dalle singole pubbliche amministrazioni, di cui gli stessi enti sono gli unici titolari, mantenendo così la responsabilità in merito al contenuto. Le tipologie di risorse semantiche supportate dal Catalogo sono:

- **ontologie**, ovvero la rappresentazione formale, condivisa ed esplicita, di un dominio della conoscenza;
- **vocabolari controllati**, liste, tassonomie, glossari e tesauri di termini e codici utilizzati per valorizzare concetti, indicizzare e recuperare informazioni;
- **schemi di API** creati dagli enti per descrivere, riferendosi alle ontologie e ai vocabolari controllati, le interfacce di programmazione delle applicazioni, il cui elenco è fornito dalla PDND.

Per ulteriori dettagli sulle classificazioni delle risorse semantiche, fare riferimento [qui](#)²².

Il Catalogo può essere esteso con nuove risorse, così da aumentare le possibilità di riuso dei concetti semantici, ad esempio, nell'ambito di sviluppo degli e-service erogati dalle pubbliche amministrazioni, e conseguentemente favorire la crescita dell'interoperabilità semantica nel settore pubblico.

È buona pratica, prima di proporre nuove risorse semantiche, esplorare le risorse nel Catalogo tramite portale schema.gov.it²³. Nel caso in cui il dominio di appartenenza delle risorse che il Contributore vuole modellare sia già presente a catalogo, ma vi sia l'esigenza di rappresentare una nuova entità e/o proprietà di un'entità presente in esso, il Contributore potrà valutare di estendere quanto presente per rispondere alle proprie esigenze di modellazione oppure «l'opportunità di definire/aggiornare delle entità e/o proprietà a livello nazionale» (cfr. [Linee Guida sull'interoperabilità tecnica delle Pubbliche Amministrazioni](#)²⁴). Per ulteriori dettagli in merito, fare riferimento alla sezione dedicata.

¹⁵ <https://eur-lex.europa.eu/legal-content/IT/TXT/?uri=CELEX:32022R1463>

¹⁶ <https://www.normattiva.it/atto/caricaDettaglioAtto?atto.dataPubblicazioneGazzetta=2005-05-16&atto.codiceRedazionale=005G0104&atto.articolo.numero=0&atto.articolo.sottoArticolo=1&atto.articolo.sottoArticolo=0&qId=&tabID=0.5596495678426472&title=lbl.dettaglioAtto>

¹⁷ <https://www.agid.gov.it/it/agenzia/piano-triennale>

¹⁸ https://www.agid.gov.it/sites/default/files/repository_files/linee_guida_interoperabilit_tecnica_pa.pdf

¹⁹ https://www.agid.gov.it/sites/default/files/repository_files/documentazione_trasparenza/cdc-spc-gdl6-interoperabilitasemopendata_v2.0_0.pdf

²⁰ <https://docs.italia.it/italia/design/lg-design-servizi-web/it/versione-corrente/index.html>

²¹ <https://www.schema.gov.it/>

²² <https://teamdigitale.github.io/dati-semantic-doc-introduzione-semantic/docs/introduzione-alla-semantic-dei-dati.html#dare-un-significato-alle-cose-del-mondo>

²³ <http://schema.gov.it/>

²⁴ https://www.agid.gov.it/sites/default/files/repository_files/linee_guida_interoperabilit_tecnica_pa.pdf

1.7.1 Creazione e gestione dei repository per le risorse semantiche

Considerato il funzionamento di harvesting di Schema, ovvero il processo di acquisizione dei contenuti semantici e conseguente caricamento sul Catalogo (rif. [Raccolta e memorizzazione dei dati semantici](#)) è necessario che ogni Contributore crei o abbia già a disposizione un repository da cui esporre le proprie risorse semantiche che saranno, quindi, raccolte in [schema.gov.it](#) come descritto nel [manuale operativo](#).

1.7.2 Quando chiedere semantic stewardship ad Istat

In qualità di Soggetto Attuatore del Progetto, l'Istat offre alle pubbliche amministrazioni interessate a pubblicare risorse semantiche nel Catalogo, una semantic stewardship finalizzata alla creazione di risorse in linea con i propri standard. In questo caso, il Contributore sceglie di affidare all'Istat l'analisi concettuale del proprio dominio di interesse e la conseguente rappresentazione in modelli di dati oppure definire concettualmente, secondo quanto indicato in [La semantica dei dati della PA e Scopo del Catalogo](#), la semantica del proprio dominio e richiedere la modellazione ontologica e/o la definizione di vocabolari controllati e/o la definizione di schemi dati.

Tale modalità è da scegliere nel caso in cui il Contributore abbia necessità di supporto per la creazione delle risorse semantiche relative ai propri dati. In questo caso, il Contributore si interfacerà direttamente con l'Istat al fine di individuare i contenuti da pubblicare sul Catalogo; le attività di strutturazione sintattica e semantica dei contenuti saranno curate da l'Istat, col supporto del Contributore. Il Contributore, in quanto responsabile dei contenuti semantici relativi ai propri dati, dovrà approvare, nel caso di ontologie, le definizioni dei concetti e delle loro relazioni, mentre, nel caso di vocabolari controllati, dovrà fornire la classificazione ed eventualmente la sua gerarchia nel caso di vocabolari più articolati in un formato strutturato (ad esempio, in file `.csv`) che riporti: il livello di gerarchia, la relazione gerarchica padre-figlio, la codifica, il lemma e la descrizione delle singole voci di classificazione in italiano, in inglese e in altre lingue se necessario per i fini del dominio in oggetto. Invece, nel caso di un e-service, dovrà fornire documentazione descrittiva del servizio e nello specifico dell'input e dell'output dello stesso.

Riguardo all'individuazione del dominio degli URI e del repository, il Contributore sarà supportato dall'Istat nella scelta tra le alternative descritte in [Scelta degli identificativi univoci nel web](#).

1.8 Scelta degli identificativi univoci nel Web

Per individuare univocamente tutti gli elementi (entità, attributi di entità, relazioni tra entità) che compongono le risorse semantiche che il Contributore intende creare e registrare nel catalogo, occorre definire degli URI/IRI persistenti nel tempo (rif. [manuale operativo](#)).

A tal fine, il Contributore può scegliere una tra le seguenti opzioni:

- **Utilizzare il servizio w3id.org del W3C Permanent Identifier Community Group²⁵**, per la registrazione di identificativi permanenti a partire dai quali reindirizzare verso URL specifici come spiegato nel [manuale operativo](#). In questo caso le opzioni possono essere:
 - `w3id/<dominio_specifico>`: il Contributore avrà piena autonomia nella gestione degli URI. Il Consorzio di società che amministra w3id.org ha definito la seguente pratica: «la pratica attuale [di w3id.org] è quella di rivendicare un nome di directory di primo livello e aggiungere identificatori di secondo livello specifici del progetto. Non esiste un elenco o una politica ufficiale per gli identificatori riservati. Tuttavia, gli amministratori possono rifiutare richieste di identificatori troppo generici, che potrebbero causare confusione, inappropriati o offensivi o che potrebbero comunque essere necessari per future espansioni del servizio”. Questa soluzione è stata adottata dal [progetto sui dati aperti della zootecnia italiana LEO²⁶](#).

²⁵ <https://w3id.org>

²⁶ <https://w3id.org/leo/>

- w3id.org/italia: il Contributore dovrà ricevere l'approvazione a contribuire al dominio "Italia" da parte dei relativi amministratori (DTD-Dipartimento per la Trasformazione Digitale) non dovendo, al tempo stesso, provvedere all'organizzazione dei reindirizzamenti; in questo caso, il Contributore beneficerà delle soluzioni di content negotiation e URI dereferentiation già incluse. La scelta del dominio w3id.org/italia comporta, inoltre, necessariamente l'archiviazione delle proprie risorse semantiche nell'"[apposito repository](#)"²⁷ gestito dal DTD; dunque, ogni operazione (inserimento, aggiornamento, modifica) da compiersi in tale repository github è subordinata all'approvazione dei suoi amministratori. Questa soluzione è stata adottata per alcune ontologie di respiro nazionale come l'ontologia Learning del Ministero dell'Università e della Ricerca o l'ontologia RPO sulla popolazione residente del Ministero dell'Interno.
- w3id.org/italia/<dominio_specifico>: il Contributore dovrà ricevere l'approvazione degli amministratori di w3id.org/italia (DTD – Dipartimento per la Trasformazione Digitale) per la denominazione degli identificatori di secondo livello, ossia la denominazione nelle URI del dominio specifico. La denominazione del dominio specifico dovrà rispettare quanto indicato nelle [Linee guida open data](#)²⁸ (sezione 7.1.3 pagina 114, 115) adottata da Agid con Determinazione n. 183/2023. In questo caso, il contributore dovrà configurare e gestire in autonomia il repository che conterrà le risorse semantiche del proprio dominio. Ad esempio, questa soluzione è stata adottata da [INPS](#)²⁹, [INAIL](#)³⁰, [Regione Lombardia](#)³¹. Inoltre, il Contributore potrà scegliere se utilizzare la soluzione di *URI dereferentiation* gestita dal Catalogo, o se usare una soluzione gestita in autonomia, così come nel caso di [ISPRA](#)³².
- **Utilizzare URI registrate in un proprio dominio** per gestire autonomamente gli identificatori delle risorse semantiche da pubblicare su schema.gov.it: in questo caso, il Contributore garantisce la persistenza degli identificatori nel tempo essendo completamente autonomo nella gestione degli URI. Questa opzione è stata adottata dal [Ministero della Cultura \(MiC\)](#)³³ nella definizione delle risorse semantiche dell'Ontologia Cultural-ON dei Luoghi della Cultura e degli Eventi Culturali.

²⁷ <https://github.com/italia/daf-ontologie-vocabolari-controllati>

²⁸ https://www.agid.gov.it/sites/default/files/repository_files/lg-open-data_v.1.0_1.pdf

²⁹ <https://w3id.org/italia/social-security/>

³⁰ <https://w3id.org/italia/work-accident/>

³¹ <https://w3id.org/italia/lombardia/>

³² <https://w3id.org/italia/env/>

³³ <http://dati.beniculturali.it/cis/>

Funzionamento generale

La presente sezione si focalizza sulla descrizione e comprensione del funzionamento generale del Catalogo.

2.1 Attori

Al fine di offrire una panoramica più chiara, la sequenza di utilizzo del Catalogo è sintetizzata nella figura seguente:



Fig. 2.1: Schema di interazione da parte dei diversi attori

Gli attori coinvolti sono i seguenti:

- **Contributori:** soggetti di cui all'art. 2, comma 2, CAD che alimentano il Catalogo con i propri asset semantici, oppure loro rappresentanti, anche privati;

- **Amministratori del Catalogo:** svolgono attività di gestione e amministrazione del Catalogo sia in termini di piattaforma che di alimentazione;
- **Utenti finali:** fruiscono dei contenuti semantici esposti dal Catalogo per lo sviluppo delle proprie API.

2.2 Avvio della contribuzione

L'alimentazione del Catalogo si basa su un processo, l'harvester, che ha come sorgenti dati i repository di risorse semantiche opportunamente configurati. Per contribuire all'alimentazione del Catalogo, i Contributori creano un proprio repository dove pubblicare le risorse semantiche destinate ad essere raccolte dall'harvester del Catalogo e inviano successivamente a info@schema.gov.it una richiesta di attivazione del processo di harvesting da parte del Catalogo. Completate le fasi indicate in [Come contribuire](#), le nuove risorse verranno raccolte nel Catalogo.

Anche nel caso di richiesta di semantic stewardship ad Istat, la richiesta deve essere inoltrata a info@schema.gov.it.

2.3 Raccolta e memorizzazione dei dati semantici

La fase di raccolta e memorizzazione dei dati semantici è affidata a un processo periodico di harvesting (tramite la componente **harvester del back-end** del Catalogo). Il processo estrae i dati contenuti nei repository configurati come sorgenti, e li memorizza su appositi server a supporto delle funzionalità di ricerca, esplorazione e fruizione del Catalogo.

2.4 Esposizione dei contenuti semantici

Gli asset semantici raccolti dall'harvester sono resi disponibili sul portale schema.gov.it, in modo da abilitarne la fruizione. Il portale integra funzionalità di ricerca testuale, filtri preimpostati, visualizzatori integrati nelle schede degli asset semantici, così come descritto in [Ricerca di risorse semantiche](#).

Come contribuire

La sezione descrive il processo di contribuzione a cui i soggetti individuati in *Attori* possono prendere parte.

3.1 Attività propedeutiche alla contribuzione al Catalogo

Le attività propedeutiche alla contribuzione al Catalogo da parte di un Contributore sono:

- Effettuare un'analisi delle risorse da creare, oppure delle possibili proposte di modifica da effettuare per risorse già pubblicate sul catalogo;
- Individuare la modalità di contribuzione più consona al caso specifico sulla base delle informazioni contenute nelle sotto-sezioni seguenti eventualmente richiedendo supporto agli Amministratori del Catalogo;
- Effettuare le ulteriori azioni di competenza descritte nei seguenti flowchart.

Diagramma 1 delle attività propedeutiche alla contribuzione

Il seguente diagramma mostra le ulteriori attività necessarie per poter aggiornare/creare asset semantici:

Diagramma 2 delle attività propedeutiche alla contribuzione

3.1.1 Aggiunta di nuove risorse

Di seguito il dettaglio delle attività propedeutiche del Contributore per l'aggiunta di nuove risorse sul Catalogo:

- Se vuoi **registrare URI sotto il** `w3id.org/italia`:
 - Nel caso in cui si abbiano a disposizione risorse da pubblicare caratterizzate da un elevato livello di generalità e riutilizzabilità su scala nazionale, allora si potranno registrare URI sotto `w3id.org/italia`, beneficiando dei meccanismi di *URI dereferentiation* e *Content Negotiation* già implementate. A tale scopo, fai riferimento alle istruzioni contenute nel *repository Italia*³⁴ e alle indicazioni rese disponibili nel *manuale operativo* su come modellare e metadattare le risorse.

³⁴ <https://github.com/italia/daf-ontologie-vocabolari-controllati>

- Nel caso in cui le risorse da pubblicare non siano caratterizzate da un elevato livello di generalità e riutilizzabilità su scala nazionale, allora si potranno registrare URI sotto `w3id.org/italia/<dominio_specifico>`. In tal caso, le istruzioni sono le seguenti:
 - * Per la **modellazione e metadatazione**, fai riferimento alle indicazioni differenziate per tipologia di risorsa (vocabolario controllato, ontologia o schema dati) contenute nel [manuale operativo](#);
 - * Per la **predisposizione del repository**, fai riferimento alle indicazioni riguardo la struttura del repository da creare e che verrà registrato tra le sorgenti del Catalogo. I file richiesti, il versionamento e ulteriori dettagli sulle risorse semantiche sono contenute nel [manuale operativo](#);
 - * Per l'**implementazione del redirect degli URI stabili**, fai riferimento alla soluzione descritta nel [manuale operativo](#);
 - * Per il **test dei requisiti tecnici per l'harvesting delle risorse semantiche**, fermo restando che il Contributore è responsabile dei contenuti pubblicati nel proprio repository, è necessario verificare che le risorse semantiche soddisfino i requisiti tecnici richiesti per l'avvio della fase di harvesting da parte del Catalogo. Per supportare al meglio i Contributori in tale processo, gli Amministratori del Catalogo sono al lavoro su una pagina web "**Strumenti di validazione**", che suggerirà per ciascun use-case il validatore più consono da poter utilizzare. In particolare, (i) per i file `index.ttl` degli schemi dati e per le ontologie verrà indicato il validatore in fase di sviluppo da parte degli Amministratori del Catalogo; (ii) per i vocabolari controllati si potrà utilizzare il [validatore DCAT-AP_IT sviluppato da AGID](#)³⁵ ignorando eventuali warning ed errori sulla presenza della classe `dcatapit:Catalog` e sull'uso della proprietà `owl:versionInfo` quando più lingue vengono specificate per la proprietà. In aggiunta, verrà indicato il validatore delle OpenAPI ([Italian API Guidelines Checker](#)³⁶), ovvero per i file `.yaml`. I controlli implementati dal validatore, attualmente in sviluppo da parte degli Amministratori, saranno un sottoinsieme di quelli eseguiti in fase di harvesting dalla piattaforma Catalogo; in particolare, i controlli verificheranno la presenza dei metadati mandatori nel file `turtle` e la validità dei prefissi rispetto alle relative ontologie. Infine, per un test di visualizzazione e di correttezza delle risorse semantiche rispetto ai requisiti tecnici per l'harvesting espressi nel [manuale operativo](#), è possibile richiedere, utilizzando la mail info@schema.gov.it, un primo aggiornamento nell'ambiente di test del Catalogo e, al termine della fase di test, richiedere l'harvesting in produzione.
- Se vuoi **registrare URI in domini come** `w3id.org/<dominio_specifico>`:
 - Segui le istruzioni contenute nell'elenco al punto precedente. Per l'attività di implementazione del redirect su URI stabili, puoi far riferimento alla [guida ufficiale pubblicata dal w3id](#)³⁷. Inoltre, per la creazione dei file di configurazione del redirect, puoi considerare a **titolo esemplificativo** le istruzioni contenute nel [manuale operativo](#).
- Se vuoi **registrare URI in domini proprietari**:
 - Segui autonomamente tutte le specifiche richieste.
- Se vuoi **chiedere semantic stewardship a Istat**:
 - Invia una richiesta di contribuzione al Catalogo utilizzando la mail info@schema.gov.it.

3.1.2 Richiesta di modifica di risorse già in Catalogo

Se vuoi **suggerire una modifica a un contenuto semantico già esistente nel Catalogo**, fai riferimento al [manuale operativo](#).

³⁵ https://portaledati3-130.dati.gov.it:3030/dcat-ap_validator.html

³⁶ <https://italia.github.io/api-oas-checker/>

³⁷ <https://w3id.org/>

3.2 Richiesta formale di contribuzione

Una volta individuata la modalità di contribuzione più opportuna, e portate a termine le relative attività descritte nelle precedenti sezioni, contatta gli Amministratori del Catalogo per richiedere formalmente di contribuire utilizzando la mail info@schema.gov.it. Una volta avvenuto il contatto, verrà organizzato un kick-off tra Amministratori e Contributore. È in corso di implementazione, da parte degli Amministratori del Catalogo, una Pagina Contatti che integrerà appositi wizard per fornire una procedura guidata ed automatizzata di iscrizione ai Contributori.

3.3 Contribuzione continua al Catalogo

Una volta che i Contributori hanno aderito al Catalogo, questi hanno la possibilità di espandere continuamente il proprio insieme di risorse semantiche all'interno del Catalogo. Questo significa che possono contribuire con delle nuove in diversi momenti successivi all'adesione iniziale.

Tuttavia, è importante considerare che tutte le modifiche apportate al proprio repository principale dopo l'adesione iniziale sono integrate automaticamente dall'harvester ogni volta che viene attivato. Di conseguenza, ogni modifica effettuata nel repository principale verrà incorporata in produzione all'attivazione successiva dell'harvester.

Per mitigare questo aspetto, i Contributori possono scegliere di caricare le risorse semantiche che non considerano stabili o complete in un branch separato, e non nel branch master. Solo quando le ritengono stabili e complete, possono eseguire il merge o il push nel branch master per far sì che vengano integrate in produzione e, di conseguenza, raccolte nel Catalogo.

Per le modifiche alle proprie risorse, al fine di agevolare il processo di caricamento tramite l'harvester su Schema.gov.it, il Titolare provvede a darne preventiva comunicazione a info@schema.gov.it 10 giorni prima della data prevista per il caricamento nel proprio repository al fine di consentire l'effettuazione di test di verifica da parte degli Amministratori.

Come utilizzare le risorse

Schema.gov.it è il portale del Catalogo che abilita i Contributori, oltreché qualsiasi cittadino, a ricercare, consultare e riusare gli asset semantici (ontologie, schemi dati e vocabolari controllati).

In particolare, il portale del Catalogo offre funzionalità di ricerca e strumenti integrati di visualizzazione e fruizione delle risorse allo scopo di semplificare le attività di riuso delle stesse, ad es. l'integrazione dei contenuti semantici all'interno degli e-service.

Il portale del Catalogo è pubblico, ovvero consultabile da chiunque, con scopi anche diversi tra i vari fruitori. Di seguito alcuni esempi di utente tipo del Catalogo:

- **sviluppatore di e-service:** può beneficiare del Catalogo riutilizzando i concetti semantici al fine di sviluppare e-service che siano interoperabili non solo sintatticamente, ma anche semanticamente;
- **ricercatore:** può beneficiare del Catalogo riutilizzando ontologie, vocabolari e schemi per i propri scopi di ricerca, riducendo i tempi di realizzazione di eventuali sperimentazioni;
- **contributore del Catalogo:** il beneficio è quello di poter individuare eventuali ontologie, vocabolari o schemi strettamente correlati, se non sovrapposti, ai contenuti semantici da pubblicare. In questo modo, piuttosto che creare ex-novo risorse semantiche ridondanti tra loro, è possibile estendere quelle esistenti, riducendo anche il carico di lavoro annesso.

Le successive sotto-sezioni hanno lo scopo di descrivere le funzionalità di ricerca e fruizione dei contenuti esposte dal portale del Catalogo.

4.1 Ricerca di risorse semantiche

La ricerca degli asset semantici può essere effettuata direttamente sull'homepage del Catalogo utilizzando l'apposito *textbox*, oppure selezionando sulle diverse categorie o tipologie degli asset semantici, come illustrato nella figura sottostante.

La ricerca tramite *textbox* permette agli utenti del Catalogo una rapida individuazione delle risorse desiderate, fornendo in input le **parole chiave** di interesse; in particolare, la *textbox* è supportata da appositi indici che abilitano la ricerca testuale per titolo, *keyword*, termini contenuti nella descrizione, titolare e concetti principali della risorsa.



Fig. 4.1: Homepage di schema.gov.it – Ricerca asset semantici



Fig. 4.2: Homepage di schema.gov.it – Ricerca asset semantici (**Esplora il catalogo**)

In aggiunta alla precedente modalità di ricerca, è possibile cliccare su **Esplora il catalogo** e di conseguenza accedere ad una pagina di ricerca avanzata che ne estende le funzionalità di filtraggio delle risorse.



Fig. 4.3: Catalogo – Ricerca asset semantici a partire da **Esplora il catalogo**

In particolare, è possibile filtrare le risorse per:

- **parole chiave:** la funzionalità di ricerca è analoga a quella descritta per la homepage, ma in questo caso può essere coadiuvata dalle successive.
- **tipologia:** tutte (opzione di default), ontologia, vocabolario controllato, schema.
- **categoria:** ad esempio Energia, Ambiente, Trasporti.
- **titolare:** ad esempio INPS, ISPRA, AIA.

L'utente può configurare uno, più o tutti i filtri per poi cliccare il tasto **OK** e ottenere i risultati della propria ricerca.

Inoltre, l'utente può rimuovere filtri applicati nella ricerca precedente servendosi delle filter chip mostrate appena sotto gli elementi di configurazione della ricerca, e/o aggiungere nuovi valori ai filtri, per poi cliccare **OK** e ottenere i nuovi risultati di ricerca.

Per ciascun risultato di ricerca, l'utente visualizza le anteprime delle risorse individuate (nessuna se il filtraggio è troppo restrittivo); per ciascuna di esse, è possibile cliccare il titolo e accedere alla relativa scheda di dettaglio. Ulteriori dettagli sulle schede dettaglio degli asset semantici sono forniti nelle sotto-sezioni seguenti.

4.2 Scheda della risorsa semantica

4.2.1 Ontologia

La scheda di dettaglio di un'ontologia fornisce informazioni quali il titolo, la descrizione, e mostra due tasti che permettono di accedere alle due funzionalità **SPARQL** e **Vai al sorgente**.

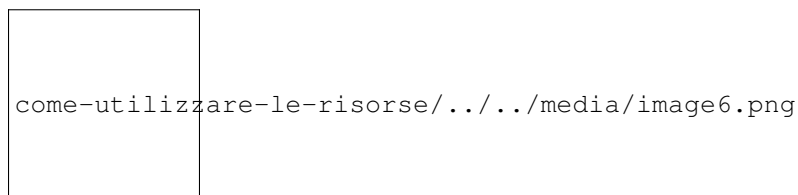


Fig. 4.4: Esempio schermata della scheda di un'ontologia

In particolare:

- Il tasto **SPARQL** permette di accedere all'*endpoint SPARQL* ed effettuare interrogazioni sull'ontologia, interagendo col *Graph Store* del Catalogo.

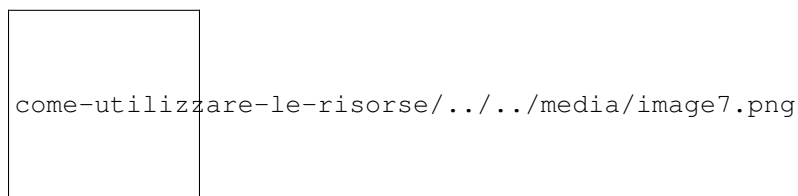


Fig. 4.5: Interfaccia per l'esecuzione di interrogazioni mediante *SPARQL Query Editor*

- Il tasto **Vai al sorgente** permette di accedere al repository sorgente della risorsa semantica, dove eventualmente l'utente può aprire nuove issue.

Nella scheda della risorsa è presente anche una tabella che contiene i dettagli dell'ontologia. Tra i vari elementi cliccabili in tabella vi è il campo `URI`, che permette di accedere al visualizzatore *lode* - un visualizzatore *HTML* di ontologie espresse in *RDFS* e *OWL* - dove l'utente ha la possibilità di esaminare varie caratteristiche dell'ontologia, tra cui le classi, le proprietà, nonché informazioni di carattere generale quali la versione corrente, gli autori e altre modalità di visualizzazione.

4.2.2 Vocabolario controllato

La scheda di dettaglio dei vocabolari controllati contiene pressoché gli stessi elementi descritti per le ontologie.

Per i vocabolari controllati è presente un tasto aggiuntivo, **API**, che permette di accedere alla *Swagger UI* e quindi fruire del vocabolario controllato tramite un'apposita *API*.

Anche in questo caso, è presente una tabella che contiene i dettagli della risorsa semantica, con un'importante differenza: il campo `URI` rimanda non più al visualizzatore *lode*, bensì *lodview* - un software aperto di dereferenziazione di `URI` che consente di navigare i dati via Web - dove l'utente ha la possibilità di esaminare varie caratteristiche del vocabolario controllato correlate al linguaggio *RDF*.



Fig. 4.6: Esempio visualizzatore *lode* di un'ontologia



Fig. 4.7: Esempio schermata della scheda di un vocabolario controllato



Fig. 4.8: Interfaccia *Swagger UI* per vocabolario controllato



Fig. 4.9: Esempio visualizzatore *lodview* di un vocabolario controllato

4.2.3 Schema

La scheda di dettaglio per gli schemi dati mostra il titolo, la descrizione e il tasto **Vai al sorgente** con funzionamento analogo a quello descritto per ontologie e vocabolari controllati.

In aggiunta, è presente una tabella che contiene, oltre alle informazioni di base della risorsa, una visualizzazione *Swagger UI* completamente integrata, e che abilita la fruizione dello schema dati mediante la visualizzazione estesa dei vari campi della sezione `schemas`.



Fig. 4.10: Esempio di tabella contenuta in una scheda per gli schemi dati

4.2.4 Richiesta di aggiornamento di asset semantici esistenti

A partire dalla scheda di dettaglio di qualsiasi asset semantico nel Catalogo, è possibile cliccare su un tasto **Vai al sorgente** per essere indirizzati sul repository git che contiene i dati su cui è stato effettuato l'harvesting. In tal modo, è possibile non solo consultare i codici sorgenti delle risorse semantiche, ma anche aprire eventuali issue nel caso in cui siano stati rilevati errori sui relativi contenuti semantici.

Nel caso di richiesta di integrazioni, il Contributore o l'utente che apre la issue dovrà fare riferimento alle indicazioni tecniche fornite nel [manuale operativo](#).

Il manuale operativo dettaglia le attività tecniche correlate al processo di contribuzione al catalogo.

5.1 Identificativi univoci delle risorse

Nel contesto del Web semantico, c'è necessità di definire i cosiddetti «URI stabili/permanenti», ovvero:

- che rimangano stabili per un tempo indeterminato, in quanto fanno riferimento a entità che devono essere permanenti e univoche;
- che permettano di essere risolti su una pagina web o simili per avere informazioni;
- nella forma `https://{{dominio}}/{{tipo_risorsa}}/{{concetto}}/{{codice_riferimento}}`, dove:
 - il **dominio**, il cui valore dipende dalla scelta effettuata dai Contributori (rif. [Scelta degli identificativi univoci nel web](#));
 - il **tipo di risorsa** può essere, ad esempio, uno dei seguenti valori:
 - * `onto`: per rappresentare tutto il mondo delle ontologie
 - * `controlled-vocabulary`: per rappresentare tutto il mondo dei vocabolari controllati
 - * `data`: per specificare gli URI dei dati collegati (linked data) che sono creati come istanze dei modelli ontologici
 - il **concetto** è lo specifico concetto che si istanzia nei dati o nome dell'ontologia;
 - il **codice di riferimento** è il codice univoco per identificare univocamente la «cosa» descritta.

Per ulteriori indicazioni in merito, fare riferimento alle [Linee Guida Open Data di AgID](#)³⁸.

Ciò detto, i Contributori possono seguire integralmente le indicazioni nella presente sezione, sia in termini di configurazione dei file di redirect che di definizione di URI, nel caso in cui vogliano pubblicare risorse semantiche in un proprio repository e **non abbiano già implementato una propria soluzione per le URI stabili**. Al contrario, se i

³⁸ https://www.agid.gov.it/sites/default/files/repository_files/lg-open-data_v.1.0_1.pdf

Contributori hanno già implementato una propria soluzione per le URI stabili, potranno continuare ad utilizzarla anche nel contesto del Catalogo, senza dover seguire le indicazioni nella presente sezione.

5.1.1 Introduzione al redirect con w3id.org

w3id.org³⁹ è una soluzione del W3C Permanent Identifier Community Group che permette l'aggiunta o modifica di identificativi permanenti a partire dai quali reindirizzare verso URL specifici; il processo si basa sull'aggiunta di una o più cartelle nel [repository git del w3id](https://github.com/w3id/w3id)⁴⁰ le quali contengono i file `.htaccess` e file `README.md`, eventualmente organizzati in sottocartelle.

Nel caso del Catalogo si può richiedere la registrazione di un proprio dominio su w3id creando una cartella dedicata oppure facendo riferimento alla [cartella Italia del GIT del w3id](#)⁴¹, nella quale possono essere aggiunte le sottocartelle, ciascuna per una particolare area tematica, contenenti i file di reindirizzamento, eventualmente organizzati in ulteriori sottocartelle, ed il file `README.md`; successivamente, le stesse saranno gestite autonomamente e con interfacciamento diretto verso w3id.org dai referenti indicati nei file `README.md`. La cartella e le relative sottocartelle, i file `.htaccess` e i file `README.md` sono creati sul [repository git del w3id](https://github.com/w3id/w3id)⁴² dal Contributore.

Una nota importante è che l'efficacia delle regole di redirect descritte nei seguenti paragrafi è subordinata all'applicazione delle indicazioni offerte dalla presente guida per la creazione ed organizzazione del repository contenente le risorse semantiche, con particolare riferimento all'uguaglianza che deve persistere tra il nome di ciascuna cartella e il nome dei relativi file, a meno dell'estensione, e il nome con il quale viene referenziata la risorsa semantica (ontologia) all'interno del proprio URI.

5.1.2 Processo di pubblicazione degli htaccess

Fork del repository git del w3id

Il primo passo per la registrazione dei vari redirect è il *fork* in locale del [repository del w3id](https://github.com/w3id/w3id)⁴³.

Gli identificativi permanenti (URI) verranno definiti sulla base del percorso nel quale saranno inseriti i vari file `htaccess`. Nel caso si vogliano registrare URI sotto la cartella Italia, allora il percorso della cartella `root` per `<nome-cartella>` dovrà essere `italia/<nome-cartella>/` dunque il namespace degli URI definiti nella stessa sarà `w3id.org/italia/<nome-cartella>/`.

Aggiunta della cartella

Nel repository in locale creato a partire dalla fork occorrerà creare la cartella sotto `italia` oppure direttamente sotto la cartella `root`, e che conterrà l'alberatura di sottocartelle e dei relativi file `.htaccess`, oltre al file `README.md`, la cui creazione e configurazione è a carico del Contributore. A tal proposito, in coda alla sezione sono disponibili una guida alla creazione dei file e alcuni esempi di possibili strumenti da utilizzare per i test del *redirect*.

Di seguito è fornito un esempio di alberatura della cartella sotto `/italia`:

```
italia
|--nome-cartella
|  |--controlled-vocabulary
|  |  |--.htaccess
|  |--data
```

(continues on next page)

³⁹ <https://w3id.org/>

⁴⁰ <https://github.com/perma-id/w3id.org/blob/master/italia/readme.md>

⁴¹ <https://github.com/perma-id/w3id.org/tree/master/italia>

⁴² <https://github.com/perma-id/w3id.org/tree/master/italia>

⁴³ <https://github.com/perma-id/w3id.org>

(continua dalla pagina precedente)

```
| | |-.htaccess
| |--onto
| | |-.htaccess
| | README.md
```

In tal modo, verranno definiti i seguenti URI stabili:

- `w3id.org/italia/<nome-cartella>/controlled-vocabulary`
- `w3id.org/italia/<nome-cartella>/data`
- `w3id.org/italia/<nome-cartella>/onto`

Il `<nome-cartella>` è molto importante dato che dovrà essere inserito in specifici parametri descritti di seguito, e sarà sempre utilizzato per riferirsi all'insieme delle risorse semantiche del Contributore nell'ambito della configurazione dei file di redirect.

Le regole di redirect associate a ciascun URI sono definite nei relativi file `.htaccess`. Il file `README.md` contiene i nominativi dei referenti unitamente ai loro riferimenti e-mail e di github. Questi riferimenti curano la gestione della cartella e dei relativi file a seguito dell'approvazione di **Italia**. È opportuno prendere come esempio il file `README.md` sotto la cartella `italia`.

Creazione della pull request

Una volta modificato il repository GIT in locale, si crea una pull request; nel caso in cui si stiano registrando URI sotto Italia, occorre indicare come reviewer della pull request i contatti presenti nel file `readme` sotto Italia⁴⁴, che procederanno all'analisi e conseguente validazione della stessa.

Il merge sul branch master verrà effettuato in ogni caso direttamente dal w3id e determinerà la pubblicazione definitiva dei nuovi URI e relative regole di redirect.

5.1.3 Contenuto dei file htaccess e readme

Di seguito è data una descrizione del file `.htaccess` per ciascuna tipologia di risorsa, da cui il Contributore può prendere spunto per creare i propri file di redirect. Gli esempi sono calati nella casistica in cui il Contributore voglia iscrivere le proprie URI sotto `w3id.org/italia/<dominio_specifico>`, voglia fruire delle soluzioni di *URI dereferentiation* implementate in Schema, e abbia rispettato le indicazioni sulla creazione del repository sorgente per le proprie risorse semantiche descritte in [Istruzioni su come predisporre il repository](#). In casi diversi rispetto al precedente, il Contributore dovrà adeguare opportunamente le regole di redirect descritte di seguito.

controlled-vocabulary

È buona norma creare il file `.htaccess` da inserire nella sottocartella `.../<nome-cartella>/controlled-vocabulary` prendendo come esempio quello contenuto nella `cartella italia/controlled-vocabulary`⁴⁵.

Esso contiene codice scritto sulla base delle Direttive Apache, e permette di gestire le richieste HTTP in base al valore dell'header `Accept` e di `SYNTAX`. A seconda del valore, gli URL vengono riscritti in modo diverso o reindirizzati a URL esterni. La specifica azione di riscrittura o reindirizzamento dipende dalla combinazione di `Accept` e `SYNTAX`.

Di seguito viene data una descrizione delle direttive di esempio, alle quali sono modificati i riferimenti degli URL di atterraggio, oltre all'eventuale modifica/integrazione delle regole al fine di meglio adattarsi al git del Contributore:

⁴⁴ <https://github.com/perma-id/w3id.org/blob/master/italia/readme.md>

⁴⁵ <https://github.com/perma-id/w3id.org/blob/master/italia/controlled-vocabulary/.htaccess>

```
Header set Access-Control-Allow-Origin *
```

Questa riga imposta l'header `Access-Control-Allow-Origin` su `*`, consentendo a qualsiasi dominio di accedere alle risorse sul server tramite richieste *Ajax* o da altri domini diversi.

```
Options +FollowSymLinks
```

Questa riga abilita l'opzione `FollowSymLinks`, che permette al server di seguire i collegamenti simbolici (*symlink*) all'interno del *file system*.

```
RewriteEngine on
```

Questa riga attiva il motore di riscrittura degli URL di Apache (*mod_rewrite*), che permette di manipolare gli URL delle richieste HTTP.

```
SetEnvIf Accept ^.*text/turtle.* SYNTAX=ttl
SetEnvIf Accept ^.*application/json.* SYNTAX=json
SetEnvIf Accept ^.*application/csv.* SYNTAX=csv
SetEnvIf Accept ^.*text/csv.* SYNTAX=csv
SetEnvIf Accept ^.*text/html.* SYNTAX=html
```

Queste righe impostano una variabile di ambiente chiamata `SYNTAX` in base all'header `Accept` della richiesta HTTP. Questo viene utilizzato per determinare il tipo di sintassi richiesto nella risposta. Queste righe sono modificate a seconda dei formati dei file presenti nelle proprie cartelle nel repository git.

```
SetEnvIf Request_URI ^.*$ ROOT_URL=<url-git>
```

Imposta la variabile di ambiente `ROOT_URL` con un URL fisso. L'URL inserito è quello del proprio repository git che punta alla cartella dei vocabolari controllati (in formato `https://raw.githubusercontent.com/<...>`).

```
RewriteCond %{ENV:SYNTAX} ^(ttl|json|csv)$
RewriteRule ^([a-zA-Z-0-9]+)(/?)$ %{ENV:ROOT_URL}$1/latest/$1.%{ENV:SYNTAX} [R=303,L]
```

Definisce la regola di riscrittura dell'URL nel caso in cui il tipo file richiesto sia `ttl`, `json` o `csv` (questi ultimi sono configurati sulla base dei tipi file presenti nel repository sorgente).

```
RewriteCond %{ENV:SYNTAX} ^html$
RewriteRule ^(.+)$ https://schema.gov.it/lodview/<nome-cartella>/controlled-
↳vocabulary/$1 [R=303,L]
RewriteRule ^(.+)/(.+)/(.+)$ https://schema.gov.it/lodview/<nome-cartella>/controlled-
↳vocabulary/$1/$2/$3 [R=303,L]
```

Le precedenti condizioni si applicano solo quando `SYNTAX` è `html`, oppure in tutti gli altri casi non gestiti dalle precedenti condizioni. Riscrivono gli URL in modo diverso, reindirizzando a URL esterni basati su modelli specifici. Al posto di `<nome-cartella>` occorre inserire il nome della cartella tematica aggiunta sotto `/italia` nel git del `w3id` con la quale ci si riferisce al particolare insieme di risorse semantiche.

onto

È buona norma creare il file `.htaccess` da inserire nella sottocartella `<nome-cartella>/onto` a partire da quello contenuto nella [cartella italia/onto](#)⁴⁶.

Esso contiene codice scritto sulla base delle Direttive Apache, e permette di gestire le richieste HTTP in base al valore dell'header `Accept` e di `SYNTAX`. A seconda del valore, le URL vengono riscritti in modo diverso o reindirizzati a URL esterni. La specifica azione di riscrittura o reindirizzamento dipende dalla combinazione di `Accept` e `SYNTAX`.

Di seguito viene data una descrizione delle direttive di esempio, alle quali sono modificati i riferimenti degli URL di atterraggio, oltre all'eventuale modifica/integrazione delle regole al fine di meglio adattarsi al git del Contributore:

```
Header set Access-Control-Allow-Origin *
```

Questa riga imposta l'header `Access-Control-Allow-Origin` su `*`, consentendo a qualsiasi dominio di accedere alle risorse sul server tramite richieste *Ajax* o da altri domini diversi.

```
Options +FollowSymLinks
```

Questa riga abilita l'opzione `FollowSymLinks`, che permette al server di seguire i collegamenti simbolici (*symlink*) all'interno del *file system*.

```
RewriteEngine on
```

Questa riga attiva il motore di riscrittura degli URL di Apache (*mod_rewrite*), che permette di manipolare gli URL delle richieste HTTP.

```
SetEnvIf Accept ^.*application/rdf\+xml.* SYNTAX=rdf
SetEnvIf Accept ^.*application/rdf\+xml.* SYNTAX=owl
SetEnvIf Accept ^.*application/n-triples.* SYNTAX=n3
SetEnvIf Accept ^.*text/turtle.* SYNTAX=ttl
SetEnvIf Accept ^.*text/html.* SYNTAX=html
```

Queste righe impostano una variabile di ambiente chiamata `SYNTAX` in base all'header `Accept` della richiesta HTTP. Questo viene utilizzato per determinare il tipo di sintassi richiesto nella risposta. Queste righe sono modificate a seconda dei formati dei file presenti nelle proprie cartelle nel repository delle risorse semantiche.

```
SetEnvIf Request_URI ^.*$ ROOT_URL=<url-git>
```

Imposta la variabile di ambiente `ROOT_URL` con un URL fisso. L'URL inserito è quello del proprio repository che punta alla cartella delle ontologie (in formato `https://raw.githubusercontent.com/<...>`).

```
RewriteCond %{ENV:SYNTAX} ^(rdf|ttl|owl|n3)$
RewriteRule ^([a-zA-Z-_0-9]+)(/)?$ %{ENV:ROOT_URL}$1/latest/$1.%{ENV:SYNTAX} [R=303,L]
```

Definisce la regola di riscrittura dell'URL nel caso in cui il tipo file richiesto sia `rdf`, `ttl`, `owl` o `n3` (questi ultimi sono configurati sulla base dei tipi file presenti nel repository sorgente).

```
RewriteCond %{ENV:SYNTAX} ^html$
```

(continues on next page)

⁴⁶ <https://github.com/perma-id/w3id.org/blob/master/italia/onto/.htaccess>

(continua dalla pagina precedente)

```
RewriteRule ^(.+) (/.+)$ https://schema.gov.it/lodview/<nome-cartella>/onto/$1$2_
↳ [R=303,L]

RewriteCond %{ENV:SYNTAX} ^html$

RewriteRule ^(.+)/$ https://schema.gov.it/lode/extract?url=https://w3id.org/italia/
↳<nome-cartella>/onto/$1 [R=303,L]

RewriteCond %{ENV:SYNTAX} ^html$

RewriteRule ^(.+)$ https://schema.gov.it/lode/extract?url=https://w3id.org/italia/
↳<nome-cartella>/onto/$1 [R=303,L]
```

Le precedenti condizioni si applicano solo quando SYNTAX è html. Riscrivono gli URL in modo diverso, reindirizzando a URL esterni basati su modelli specifici. Al posto di <nome-cartella> occorre inserire il nome della cartella tematica aggiunta sotto /italia nel git del w3id con la quale ci si riferisce al particolare insieme di risorse semantiche.

data

È bene che il file .htaccess da inserire nella sottocartella ../<nome-cartella>/data sia essere creato a partire da quello contenuto nella [cartella italia/data](#)⁴⁷.

Esso contiene codice scritto sulla base delle Direttive Apache, e permette di configurare il server Apache per consentire l'accesso da qualsiasi dominio alle risorse del server, impostare una variabile di ambiente ROOT_URL con un valore fisso, e quindi riscrivere tutte le richieste in modo che includano ROOT_URL prima dell'URI richiesto.

Di seguito viene data una descrizione delle direttive di esempio, alle quali sono modificati i riferimenti degli URL di atterraggio, oltre all'eventuale modifica/integrazione delle regole al fine di meglio adattarsi al git del Contributore:

```
Header set Access-Control-Allow-Origin *
```

Questa riga imposta l'header Access-Control-Allow-Origin su *, consentendo a qualsiasi dominio di accedere alle risorse sul server tramite richieste *Ajax* o da altri domini diversi.

```
Options +FollowSymLinks
```

Questa riga abilita l'opzione FollowSymLinks, che permette al server di seguire i collegamenti simbolici (*symlink*) all'interno del *file system*.

```
RewriteEngine on
```

Questa riga attiva il motore di riscrittura degli URL di Apache (*mod_rewrite*), che permette di manipolare gli URL delle richieste HTTP.

```
SetEnvIf Request_URI ^.*$ ROOT_URL=https://schema.gov.it/lodview/<nome-cartella>/data/
```

Questa riga imposta una variabile di ambiente chiamata ROOT_URL, dove al posto di <nome-cartella> occorre inserire il nome della cartella tematica aggiunta sotto /italia nel git del w3id con la quale ci si riferisce al particolare insieme di risorse semantiche.

```
RewriteRule ^(.*)$ %{ENV:ROOT_URL}$1 [R=303,L]
```

⁴⁷ <https://github.com/perma-id/w3id.org/blob/master/italia/data/.htaccess>

Questa riga è una regola di riscrittura degli URL. Ogni richiesta che arriva al server verrà riscritta in modo da includere il valore di `ROOT_URL` prima dell'URI richiesto. Il flag `[R=303,L]` indica che la risposta HTTP sarà un redirect temporaneo (codice di stato 303) e che questa è l'ultima regola da applicare.

```
RewriteRule ^(.*)/$ %{ENV:ROOT_URL}$1 [R=303,L]
```

Questa è una regola di riscrittura simile alla precedente, ma si applica solo alle richieste che terminano con una barra. Anche in questo caso, la risposta sarà un redirect temporaneo con codice di stato 303.

README.md

Per la creazione del file `README.md` è possibile far riferimento

- all'esempio fornito da w3id.org stessa⁴⁸;
- al file creato sotto la cartella `italia`⁴⁹.

In ogni caso, nella sezione **contatti** del file `readme` occorre descrivere le finalità di utilizzo degli URI e indicare i nominativi dei referenti specificando il loro contatto github e possibilmente un'e-mail.

5.1.4 Esempi di strumenti a supporto dei test

Il Contributore è responsabile della scrittura ed eventuale correzione dei file `htaccess` che vengono pubblicati sul `w3id`; pertanto, è tenuto a verificarne la correttezza.

A titolo di esempio, un possibile approccio per testare gli `htaccess` prima della pubblicazione sul `w3id` potrebbe basarsi sull'installazione di un *server Apache*, in un *container* o macchina virtuale, e la configurazione dei file `htaccess` all'interno di esso. Questo consentirebbe di eseguire test approfonditi sui redirect a partire dal server di test.

Per i test successivi alla pubblicazione su `w3id` dei file `htaccess` e dell'harvesting sul Catalogo, un'opzione può essere l'utilizzo di `cURL` per verificare la correttezza delle regole di redirect. In particolare, in ciascuno dei file `htaccess` sono definite una o più regole di redirect basate sul contenuto (es. `html`, `rdf`, `turtle`, ecc.). Per testare tutte le regole, occorre innanzitutto individuare URI utili a stressare le regole di redirect contenute in tutti i file `htaccess` (`onto`, `controlled-vocabulary` e `data`). Successivamente, per ciascuna tipologia di risorsa occorre testare, con l'URI individuato, tutti i possibili contenuti gestiti dal relativo file `htaccess`, e verificare che il redirect sia quello atteso. In caso contrario, il Contributore dovrà aprire una pull request sul git del `w3id` al fine di correggere il file `htaccess`.

La generica riga di comando in input da inserire nel prompt dei comandi è la seguente:

```
curl [URI] --header "Accept: [Content type]"
```

Di seguito viene fornito un esempio di utilizzo di `cURL` per verificare la correttezza dei redirect nel caso di una ontologia.

Nel caso d'esempio, l'URI fornito è il seguente: <https://w3id.org/italia/work-accident/onto/core/>, mentre il contenuto richiesto è `text/html`, ovvero uno di quelli gestiti dal relativo file `htaccess`⁵⁰ per le ontologie. Il risultato della `cURL` mostra come stato `http` il valore `303_See_Other`, che indica che l'indirizzamento avviene con successo, e come indirizzo di atterraggio quello costruito dall'apposita regola di redirect nel file `htaccess`, come atteso.

⁴⁸ <https://github.com/perma-id/w3id.org/blob/master/examples/README.md>

⁴⁹ <https://github.com/perma-id/w3id.org/blob/master/italia/readme.md>

⁵⁰ <https://github.com/perma-id/w3id.org/blob/master/italia/work-accident/onto/htaccess>

```
manuale-operativo/../../media/image13.png
```

Fig. 5.1: Prompt dei comandi- cURL per verifica redirect

5.2 Indicazioni su modellazione e metadatazione degli asset semantici

Di seguito si riportano le linee guida di base per l'aggiunta di un nuovo vocabolario controllato, una nuova ontologia e un nuovo schema dati.

5.2.1 Nuovo vocabolario controllato

Il vocabolario controllato è un dataset aperto e, pertanto, è metadato seguendo le linee guida nazionali sui dati aperti, adottando cioè il profilo nazionale di metadatazione [DCAT-AP_IT](#)⁵¹ e una licenza aperta.

Nel contesto delle iniziative per promuovere le politiche di valorizzazione del patrimonio informativo pubblico, AGID (Agenzia per l'Italia Digitale) ha collaborato con un Gruppo di Lavoro composto da amministrazioni centrali e locali per definire il profilo nazionale dei metadati ([DCAT-AP_IT](#)).

Questo profilo permette la documentazione dei dati di tipo aperto nel Catalogo, in linea con la specifica [DCAT-AP \(1.1\)](#)⁵², definita nell'ambito del [programma ISA](#)⁵³ della Commissione Europea.

È reso, dunque, disponibile da parte di AGID un [validatore di file aderente al profilo DCAT-AP_IT](#)⁵⁴: può essere utilizzato dal Contributore al fine di verificare il rispetto delle specifiche in oggetto, ricordando che eventuali errori segnalati su `dcatapit:Catalog` e su `owl:versionInfo` possono essere ignorati.

Inoltre, è necessario aggiungere al dataset il metadato `ndc:keyConcept` definito nell'[ontologia NDC](#)⁵⁵. Il *key concept* è il concetto principale o chiave a cui il vocabolario controllato si riferisce. Ad esempio, per il dataset <https://w3id.org/italia/controlled-vocabulary/classifications-for-public-services/authentication-type> il concetto chiave, quindi il valore della proprietà `ndc:keyConcept`, è `authentication-type` (il tipo di autenticazione). Questo metadato è necessario per abilitare i meccanismi di harvesting del Catalogo.

Essendo un vocabolario controllato un dataset per organizzare la conoscenza, esso è rappresentato utilizzando lo standard web di riferimento per i sistemi di organizzazione della conoscenza, ossia l'ontologia [SKOS](#)⁵⁶.

In particolare, il Contributore:

- definisce il `ConceptScheme` di [SKOS](#) che rappresenta tutto il vocabolario controllato collegandolo ai singoli concetti di primo livello mediante la proprietà `skos:hasTopConcept`. Un vocabolario controllato può essere infatti una lista di codici (*code list*), una tassonomia (*taxonomy*), oppure qualcosa di più complesso come un tesaurus. La proprietà `skos:hasTopConcept` collega tutti i singoli concetti in una lista di codici, i concetti di primo livello negli altri casi più articolati;

⁵¹ <https://docs.italia.it/italia/daf/linee-guida-cataloghi-dati-dcat-ap-it/stabile/index.html>

⁵² <https://joinup.ec.europa.eu/collection/semic-support-centre/solution/dcat-application-profile-data-portals-europe/release/11>

⁵³ https://ec.europa.eu/isa2/home_en/

⁵⁴ https://portaledati3-130.dati.gov.it:3030/dcat-ap_validator.html

⁵⁵ <https://w3id.org/italia/onto/NDC>

⁵⁶ <https://www.w3.org/TR/skos-primer/>

- per ciascun concetto, inoltre, specifica le seguenti proprietà:
 - `skos:notation` con l'identificativo univoco associato al concetto;
 - `skos:prefLabel` con l'etichetta principale di riferimento per il concetto definita con almeno il tag in italiano. È opportuno aggiungere anche la lingua inglese, soprattutto se il vocabolario è utilizzato in un contesto di interoperabilità anche transfrontaliera. Si ricorda che l'etichetta che rappresenta il valore di questa proprietà deve essere esplicita e non contenere abbreviazioni di testo che la renderebbe comprensibile solo da parte dell'ente che propone il vocabolario. È opportuno che l'etichetta principale sia nella forma singolare, lasciando eventualmente le forme plurali in un'etichetta alternativa rappresentata con la proprietà `skos:altLabel`;
 - `skos:inScheme` che assume come valore l'URI del vocabolario controllato che si sta definendo.
- in presenza di tassonomie o tesauri indica anche i concetti di livello inferiore e superiore attraverso le proprietà `skos:narrower` e `skos:broader` nonché di quanti livelli si compone il vocabolario mediante la proprietà `skos:numberOfLevels`, il cui valore indica il numero di livelli gerarchici previsti dal vocabolario controllato.

È opportuno che altre proprietà *SKOS* siano aggiunte ai vari concetti, come per esempio `skos:definition` quando una definizione formale del concetto è nota, `skos:note` per aggiungere qualsiasi altra nota, qualora esistente.

Qualora un concetto sia allineato semanticamente in maniera più o meno forte ad analoghi concetti definiti in altri vocabolari controllati esistenti nel web (es. `schema.org`, `EU authority list`) è buona norma utilizzare le proprietà `skos:exactMatch` (i due concetti sono di fatto la stessa cosa), `skos:relatedMatch` (i due concetti sono tra loro relazionati, nel senso più generale), `skos:closeMatch` (i due concetti sono molto simili ma non esattamente la stessa cosa) al fine di creare collegamenti tra diversi dataset dei vocabolari controllati (linked open data).

5.2.2 Nuova ontologia

Nel caso di proposta di una nuova ontologia, il Contributore rispetta le seguenti indicazioni:

- consegnare un'ontologia che non presenta errori sintattici. Non saranno valutate ontologie che non riescono a essere aperte con successo da software di editing di ontologie come per esempio *Protégé*;
- l'ontologia è almeno serializzata in RDF/Turtle (serializzazione utilizzata nella procedura di harvesting di `schema.gov.it`);
- l'ontologia è metadadata attraverso l'ontologia `ADMS-AP_IT`⁵⁷. Questo profilo di metadatazione è specifico per ontologie e si basa su `DCAT-AP_IT`. È questa metadatazione che consente di abilitare l'harvesting dell'ontologia in `schema.gov.it`. Per verificare la correttezza della metadatazione, i Contributori avranno a disposizione un modulo di pre-harvesting raggiungibile da un'apposita pagina web, così come espresso nel workflow in [Attività propedeutiche alla contribuzione al Catalogo](#);
- si verifica con dei ragionatori automatici, presenti come plug-in nei principali editor di sviluppo di ontologie come ad esempio *Protégé* o *Eddy*, che l'ontologia prodotta non abbia inconsistenze semantiche e che quello che si vuole modellare sia correttamente inferito dal ragionatore;
- le classi e proprietà di ogni tipo delle ontologie sono sempre annotate con almeno le proprietà `rdfs:label` e `rdfs:comment` in italiano. Qualora l'ontologia abbia una potenzialità in un contesto transfrontaliero, si fornisce anche la versione in lingua inglese delle stesse proprietà. È opportuno inoltre specificare la proprietà di annotazione `rdfs:isDefinedBy` con l'URI dell'ontologia;
- riutilizzare direttamente tutte le ontologie già esistenti in `schema.gov.it` qualora questo si applichi alla nuova modellazione. Questo consente anche di riutilizzare pattern di modellazione già implementati che, da letteratura scientifica, hanno dimostrato di fornire maggiori garanzie di interoperabilità semantica tra modellazioni diverse. A titolo d'esempio: se si deve definire un concetto di persona fisica oppure di organizzazione (pubblica o privata

⁵⁷ <https://www.schema.gov.it/semantic-assets/details?uri=https%3A%2F%2Fw3id.org%2Fitalia%2Fonto%2FADMS>

che sia) ma anche concetti legati al tempo o a ruoli di agenti che agiscono su oggetti del dominio è necessario riutilizzare rispettivamente le ontologie *CPV* (Persone), *COV* (Organizzazioni), *TI* (tempo), *RO* (ruoli);

- si ricorda che il riutilizzo diretto di concetti e/o proprietà già esistenti comporta l'ereditarietà di tutti i vincoli di cardinalità (restrizioni *OWL*) già definiti nelle ontologie di schema.gov.it;
- qualora ci sia necessità di aggiungere altre proprietà o vincoli di cardinalità a concetti già esistenti in ontologie già disponibili, o si richiede tale aggiunta agli Amministratori del Catalogo per agire direttamente nelle ontologie di [schema](http://schema.gov.it), oppure si definisce un proprio concetto che potrà diventare sottoclasse del concetto già esistente in ontologie esistenti di schema.gov.it;
- è raccomandato allineare la nuova ontologia all'ontologia di **livello top nazionale IO**. Questa raccomandazione ha alcuni vantaggi: 1) aiuta a comprendere meglio cosa si sta modellando. Se si modella un'entità che è **concetto** questa non può essere anche un **evento**; 2) l'ontologia *IO* aiuta a collegare tutte le ontologie tra loro creando quindi una grossa rete di ontologie nazionali. Il collegamento tra modelli concettuali abilita il collegamento tra i dati rappresentati con quei modelli; 3) *IO* è molto utile per verificare, attraverso l'inferenza semantica, eventuali inconsistenze semantiche in un'ottica più generale di rete più che di singola ontologia;
- è possibile aggiungere vincoli di cardinalità mediante restrizioni *OWL*, ma è bene non vincolare troppo le varie definizioni per dare più possibilità di riutilizzo alle ontologie anche in altri potenziali contesti, visto la valenza nazionale che nuove ontologie possono assumere con la pubblicazione in schema.gov.it. A tal proposito si suggerisce di definire restrizioni *OWL* come sottoclassi della classe a cui si riferiscono così da specificare una condizione necessaria ma non sufficiente e valutare attentamente se il vincolo di cardinalità è di fatto sempre stringente (costrutti come *some*, *exactly 1*) o meno (*max 1*, ecc.). Qualora ci siano vincoli di cardinalità più stringenti dal punto di vista applicativo, è bene che il Contributore consideri la possibilità di rilassare alcune restrizioni *OWL* definite nell'ontologia e creare a parte un vero e proprio profilo applicativo mediante regole *SHACL*, standard web pubblicato dal W3C. Questa pratica, tra l'altro, è quella adottata da alcuni paesi europei (es. Belgio) e dalla Commissione Europea stessa nel contesto di iniziative di interoperabilità semantica quali i *core vocabulary*, l'ontologia *ePO* sul'e-procurement, l'ontologia *ELM* – *European Learning Model*;
- è opportuno modularizzare il più possibile le ontologie, più che creare ontologie che contengono la rappresentazione di svariati domini/tipologie di dati. L'evidente vantaggio della modularizzazione è quello di riuscire a gestire in modo più agevole eventuali evoluzioni future che potrebbero anche seguire diverse frequenze di aggiornamento delle diverse tipologie di dato;
- è opportuno utilizzare *ontology design pattern* e un approccio *agile* alla modellazione con rilasci più frequenti e definizioni di versioni anche instabili dell'ontologia pian piano raffinate con requisiti nuovi fino alla versione finale. Gli *ontology design pattern* sono soluzioni di modellazione già disponibili, riusabili ed efficaci che possono essere specializzati o direttamente applicati nel dominio da modellare e che risolvono problemi di modellazione ricorrenti (es. un qualcosa che cambia nel tempo). Essi aiutano a ridurre l'arbitrarietà nel design dell'ontologia e consentono di ridurre gli errori di modellazione e quindi migliorare la qualità delle ontologie. Si ricorda, come prima menzionato, che già le ontologie esistenti implementano *ontology design pattern* che si devono riutilizzare (es. ruoli nel tempo, oggetti che variano nel tempo, valori, ecc.);
- è buona norma allineare l'ontologia anche ad altre ontologie esistenti nel Web dei Dati, qualora questo sia applicabile;
- è possibile corredare l'ontologia di una rappresentazione grafica dei concetti e delle loro relazioni. A tale scopo i Contributori sono liberi di utilizzare strumenti di loro preferenza. Solo a titolo d'esempio si possono citare diagrammi di rappresentazione grafica che utilizzano la notazione [UML](http://www.uml.org/)⁵⁸, che possono essere prodotti con strumenti quali *diagrams.net*, *Visual Paradigm*, *StarUML* oppure che usano notazioni tecniche specifiche di disegno ontologico come per esempio [Graffoo](https://essepuntato.it/graffoo/)⁵⁹ (che è possibile abilitare con strumenti come [yEd](https://www.yworks.com/products/yed)⁶⁰ oppure

⁵⁸ <http://www.uml.org/>

⁵⁹ <https://essepuntato.it/graffoo/>

⁶⁰ <https://www.yworks.com/products/yed>

[draw.io](#)⁶¹) o [Graphol](#)⁶² (che è possibile utilizzare attraverso strumenti come [Eddy](#)⁶³).

5.2.3 Nuovi schemi dati

I nuovi schemi dati devono utilizzare le risorse semantiche, come ontologie e vocabolari controllati, qualora siano catalogati in [schema.gov.it](#) e siano pertinenti rispetto ai dati descritti dagli schemi.

Da un lato le risorse semantiche sono un valido aiuto per la modellazione degli schemi: è infatti possibile creare una descrizione dei dati secondo uno dei formati del paradigma *linked-data* (ad esempio [JSON-LD](#)⁶⁴), quindi ricavare lo schema sintetizzando e semplificando. Si andranno quindi a includere soltanto le componenti variabili e i loro identificativi abbreviati, eliminando il più possibile le parti ridondanti e gli annidamenti. In particolare, per i vocabolari controllati saranno oggetto di trasferimento soltanto le componenti variabili delle URI.

Se invece si parte da uno schema dati già definito, è importante indicare come poter convertire in modo non ambiguo i dati trasferiti dal loro formato al formato del paradigma *linked-data*.

Siccome le specifiche OpenAPI prevedono che il formato di scambio dati sia JSON, il formato del paradigma *linked-data* più compatibile è JSON-LD, che aggiunge la possibilità di inserire alcune specifiche per la descrizione semantica, in particolare:

- il `@context` che permette di definire sia le abbreviazioni comuni, che i concetti semantici a cui associare ciascun attributo
- il `@type` che permette di associare a ciascun oggetto il proprio oggetto semantico corrispondente

Inoltre, è possibile definire gli attributi di tipo `@id` che fanno riferimento concetti esterni, che possono essere altri oggetti o elementi di un vocabolario controllato.

Gli schemi dati per essere sottoposti al processo di harvesting debbono contenere due file: il file di metadati in formato RDF/Turtle, e il modello dello schema dati in formato *OpenAPI*.

In particolare, il file di metadati deve avere l'estensione `.ttl` e un nome specifico, ossia `index.ttl`.

Il file `index.ttl`, come per le ontologie, deve contenere tutti i metadati previsti dall'ontologia [ADMS-AP_IT](#)⁶⁵. Questo profilo di metadattazione si basa su [DCAT-AP_IT](#). È l'adozione di questo modello che consente l'harvesting anche degli schemi dati.

Mentre il file che riporta lo schema del servizio deve avere l'estensione `.yaml` (se viene utilizzata la versione 3.0 di *OpenAPI* si usa `oas3.yaml`).

Il file OpenAPI deve essere compatibile con il [nuovo modello d'Interoperabilità \(ModI\)](#)⁶⁶.

Per descrivere le connessioni fra gli schemi dati e il loro equivalente nel paradigma *linked-data*, si utilizzano alcuni attributi custom, che vengono derivati dal formato JSON-LD:

- `x-jsonld-type` permette di indicare il tipo di oggetto che si sta descrivendo
- `x-jsonld-context` permette di definire il `@context` di JSON-LD. Notare che è possibile definire `@context` annidati, via via che si esprimono i concetti

In particolare, le sezioni principali e obbligatorie all'interno del file che riporta lo schema del servizio sono le seguenti:

- `info`: le informazioni iniziali riguardanti il titolo (`title`) e la descrizione (`description`) dello schema dati del servizio;
- `components`:

⁶¹ <https://app.diagrams.net/>

⁶² <https://www.diag.uniroma1.it/degiacom/papers/2022/ft2022lssd.pdf>

⁶³ <https://github.com/obdasystems/eddy>

⁶⁴ <https://json-ld.org/>

⁶⁵ <https://www.schema.gov.it/semantic-assets/details?uri=https%3A%2F%2Fw3id.org%2Fitalia%2Ffonto%2FADMS>

⁶⁶ <https://www.agid.gov.it/it/infrastrutture/sistema-pubblico-connettivita/il-nuovo-modello-interoperabilita>

- schemas: vengono descritti i concetti all'interno del servizio, definendo quali sono i concetti di input obbligatori (required). Per ogni concetto sono dichiarate le seguenti voci:
- type: il tipo di dato (object, string, integer);
- description: si riporta la URI del concetto di riferimento;
- per i concetti di tipo object è necessario elencare le properties. Per le properties è necessario definire il tipo di dato (type): se si tratta di un object si riporta il riferimento al concetto, altrimenti è necessario riportare il tipo di format e quando richiesto il pattern;
- x-jsonld-type: si riporta la URI del concetto di riferimento;
- x-jsonld-context:
 - * @vocab: si riporta la radice della URI dell'ontologia maggiormente referenziata all'interno dello schema dati del servizio.

Si riporta un breve esempio di seguito:

```
components:
  schemas:
    TaxCode:
      type: string
      description: https://w3id.org/italia/onto/CPV/taxCode.
      example: RSSMRA75L01H501A
      maxLength: 16
      minLength: 11
    Person:
      type: object
      description: https://w3id.org/italia/onto/CPV/Person
      x-jsonld-type: https://w3id.org/italia/onto/CPV/Person
      x-jsonld-context:
        "@vocab": https://w3id.org/italia/onto/CPV/
      tax_code:
        "@id": taxCode
      date_of_birth: dateOfBirth
      family_name: familyName
      given_name: givenName
      properties:
        tax_code:
          $ref: "#/components/schemas/TaxCode"
        date_of_birth:
          format: date
          type: string
          pattern: ([0-9]{4})-([0-1][0-9])-([0-3][0-9])
        family_name:
          type: string
        given_name:
          type: string
```

5.3 Indicazioni su aggiornamento di asset semantici esistenti

Di seguito sono riportate le raccomandazioni in presenza di proposte, da parte degli utenti, che richiedono modifiche ad asset semantici già esistenti. Invece, per la modifica delle proprie risorse da parte dei titolari delle stesse, si rinvia al [paragrafo «Contribuzione continua al catalogo»](#)

5.3.1 Modifiche a vocabolari controllati esistenti

Qualora si propongano modifiche a vocabolari esistenti è necessario specificare nella proposta quanto segue:

- se un concetto già esistente è da considerarsi deprecato. Nel qual caso si dovrà aggiungere al concetto la proprietà `owl:deprecated` per indicare che non è più valido;
- aggiungere la definizione del/dei nuovo/nuovi concetto/i seguendo esattamente le definizioni dei concetti già esistenti nel vocabolario (anche in termini di URI);
- qualora si voglia aggiungere un nuovo livello a una tassonomia o a un tesoro, è necessario motivare la richiesta e predisporre il nuovo livello considerando l'aggiunta delle proprietà `skos:narrower` e `skos:broader` prima menzionate ove applicabili nei livelli precedenti già presenti nel vocabolario controllato.

5.3.2 Modifiche a ontologie esistenti

Le modifiche vengono esplicitate evidenziando i requisiti (domande di competenza) che portano a richiedere la modifica di concetti e/o proprietà già esistenti o cambiamenti nelle restrizioni *OWL* attualmente inserite nelle ontologie esistenti. Le domande di competenza sono di fatto interrogazioni sui dati che la modellazione deve poter supportare.

Nuovi concetti e/o proprietà rispetto a quelle esistenti, sempre motivate da specifici requisiti da evidenziare, dovranno essere definiti seguendo esattamente le definizioni già fornite nell'ambito delle ontologie esistenti (i.e., specificando tutte le annotazioni che già vengono utilizzate sia in lingua italiana che in lingua inglese, specificando eventuali restrizioni *OWL* se necessarie).

Si raccomanda di non effettuare cambiamenti negli URI per concetti e/o proprietà già definite, per via della loro natura di identificativi univoci e soprattutto persistenti nel tempo, eccetto il caso in cui i cambiamenti siano indispensabili per garantire un funzionamento complessivo dell'asset (anche in termini di coerenza semantica) e del Catalogo, o quando gli URI contengono evidenti errori di battitura che ne possano compromettere l'usabilità e la leggibilità.

5.3.3 Modifiche a schemi dati

Le modifiche da proporre a schemi dati saranno valutate nel caso in cui saranno evidenziati nuovi casi d'uso. Sulla base di queste nuove necessità si decide se modificare uno schema dati già esistente o crearne uno nuovo. Le modifiche proposte per gli schemi dati dovranno seguire tutte le regole sintattiche e semantiche definite nel contesto del Catalogo.

5.4 Istruzioni su come predisporre il repository in cui pubblicare le risorse semantiche

La sezione descrive le caratteristiche principali che è bene che i repository gestiti dai Contributori abbiano affinché vengano correttamente acquisiti dal Catalogo, nel caso gli stessi vogliano contribuire offrendo risorse semantiche gestite in un repository di propria titolarità.

Allo scopo di semplificare il controllo di qualità che i Contributori devono effettuare sui propri repository, gli Amministratori del Catalogo hanno reso disponibile presso il repository <https://github.com/teamdigitale/dati-semantic-cookiecutter> un template di controlli che possono essere integrati sul proprio repository, **disattivando eventualmente alcuni hook non applicabili**, a supporto dell'esecuzione dei controlli di qualità propedeutici alla richiesta di iscrizione al Catalogo. Per ulteriori approfondimenti, è possibile far riferimento al **file readme del cookiecutter**, nel paragrafo **Controlli Automatici e Test**.

I principali controlli implementati nel template sono descritti nelle successive sotto-sezioni.

5.4.1 Struttura di base del repository

La struttura di base del repository deve rispettare il seguente template:

- Ontologie: in `assets/ontologies`;
- Vocabolari controllati: in `assets/controlled-vocabularies`;
- Schemi: in `assets/schemas`.

5.4.2 Contenuto del repository

I file devono essere codificati in UTF-8.

Le risorse semantiche devono essere presenti almeno nel seguente formato:

- Ontologie: RDF/Turtle (*media-type* `text/turtle`) con estensione del file `.ttl`;
- Vocabolari controllati: RDF/Turtle (*media-type* `text/turtle`) con estensione del file `.ttl`;
- Schemi dati: OAS3 (`.oas3.yaml`) e turtle dei metadati (file `index.ttl`).

Possono essere presenti ulteriori serializzazioni (es. RDF/XML, JSON-LD, CSV, ...) che verranno ignorate dal processo di harvesting del Catalogo.

Possono essere presenti eventuali file di documentazione, ma solo in formato *Markdown* (`.md`).

5.4.3 Nomi delle cartelle e dei file

I nomi di file e directory devono rispettare il seguente pattern `\` [A-Za-z0-9_-.] { , 64 }`.

Il nome di ciascun file deve corrispondere al nome della relativa risorsa nell'URI utilizzato per referenziarla.

I nomi dei file di una directory devono corrispondere al nome della directory che li contiene, a meno dell'estensione degli stessi.

Il controllo può essere disattivato nel caso in cui il Contributore abbia implementato una propria soluzione per le URI stabili, oppure se ha configurato una soluzione basata sul W3id ([Identificativi univoci delle risorse](#)) che tratti opportunamente eventuali discrepanze tra nomi dei file, nomi delle cartelle che li contengono e nomi delle risorse nelle relative URI.

5.4.4 Proiezioni in CSV dei vocabolari controllati

Le directory del vocabolario controllato possono contenere una proiezione in formato `csv` del vocabolario insieme ai metadati necessari per mappare i campi del `csv` alle risorse presenti nel file `rdf`.

Nel caso in cui si voglia pubblicare anche la proiezione `csv` per un vocabolario controllato, questo deve avere la struttura di file piatto, ossia riportare tutte le voci delle classificazioni dal primo livello al livello di foglia. Occorre rispettare le seguenti regole:

- generare la proiezione possibilmente mediante strumenti come *JSON-LD framing*;
- eventuali commenti nel `csv` devono iniziare col carattere `#`;
- l'header del `csv` ovvero la prima riga contenente i nomi delle colonne, deve rispettare le seguenti regole sulla base della presenza o meno, nella stessa directory del `csv`, del file di metadati in un *frictionless data package*, annotato secondo le specifiche indicate in *REST API Linked Data Keywords*:
 - se presente, allora i nomi delle colonne devono essere conformi a quanto indicato nei metadati.

- se non presente, i nomi delle colonne devono solo rispettare il seguente pattern: `^[a-zA-Z0-9_]{2,64}$`.
- si suggerisce di creare la riga di header di vocabolari controllati alberati introducendo tante colonne quanti sono i livelli secondo la nomenclatura: `codice_1livello`, `label_ITA_1livello`, `label_ENG_1livello`, `definizione`, ... `codice_nlivello`, `label_ITA_nlivello`, `label_ENG_nlivello`, `definizione`, `NOTE` (Dove `n` rappresenta il numero di livelli della classificazione);
- le righe successive del `csv` devono contenere i valori delle colonne separati da `,` (virgola) e, se valori relativi a campi di tipo stringa, racchiusi in `"` (doppie virgolette).

I metadati di cui al precedente elenco devono essere espressi tramite un file `datapackage` con estensione `.json`, `.yaml` o `.yml`.

Nel caso in cui le voci di un vocabolario siano già presenti in un *repository triple store* è possibile estrarne il contenuto con una *query SPARQL* secondo il seguente formato:

```
SELECT DISTINCT ?codice_1livello ?label_1livello_it ?label_1livello_en ?label_1_
↳livello_fr ?label_1livello_de

WHERE{

<https://w3id.org/italia/controlled-vocabulary/territorial-classifications/countries/
↳italy/ITA> a skos:Concept;

skos:notation ?codice_1livello;

skos:prefLabel ?label_1livello_it;

skos:prefLabel ?label_1livello_en;

skos:prefLabel ?label_1livello_fr;

skos:prefLabel ?label_1livello_de.

FILTER (LANG(?label_1livello_it) = 'it')

FILTER (LANG(?label_1livello_en) = 'en')

FILTER (LANG(?label_1livello_fr) = 'fr')

FILTER (LANG(?label_1livello_de) = 'de')

}
```

5.4.5 Versionamento

Le directory degli asset possono avere sub-directory per supportare il versionamento. Il nome delle sub-directory rispetta il pattern: `(latest|v?[0-9]+(\.[0-9]+){0,2})`.

Una directory contenente asset non contiene contemporaneamente sub-directory versionate con e senza il prefisso `v` perché questo rende impossibile ordinare le versioni.

In Istruzioni su come predisporre il repository sono contenuti alcuni esempi di versionamento delle risorse semantiche.

5.4.6 Approfondimenti sugli schemi dati

Gli schemi utilizzano delle directory versionate come descritto nel corso del documento.

Gli schemi per le API vengono pubblicati in formato *OpenAPI*, corrispondenti ad una estensione di *JSON Schema Draft 4*⁶⁷, incorporato nella sezione `#/components/schema` del file OAS compatibilmente con le Linee Guida per l'interoperabilità tecnica. L'estensione del file è `.oas3.yaml`.

È opportuno che il file YAML contenga i riferimenti semantici descritti nel documento *I-D-polli-restapi-ld-keywords*⁶⁸ attraverso:

- il campo custom `x-jsonld-context` contenente un `@context JSON-LD` conforme alle indicazioni contenute in *JSON-LD 1.1*;
- il campo custom `x-jsonld-type` contenente il riferimento ad un `rdf:type`.

I metadati associati sono pubblicati solo in formato RDF/Turtle (*media type text/turtle*) in un apposito file `index.ttl`, uno per ciascuno schema dati. È opportuno che questo file sia generato automaticamente dal documento *OpenAPI*.

È possibile verificare sintatticamente gli schemi forniti utilizzando l'*OpenAPI Checker*⁶⁹.

Schema bundling

Quando si pubblica un documento OAS contenente la specifica di un'API, è utile de-referenziare ed accorpare in un unico file tutti i riferimenti a schemi ed operazioni.

Questo processo viene detto *bundling*⁷⁰.

Il prodotto sarà un singolo OAS document (es. un file YAML) utile alla validazione sintattica e semantica dell'API.

Questo meccanismo permette di inserire nell'IDL tutte le informazioni semantiche necessarie a descrivere l'API in base sia ai riferimenti ontologici che agli schemi utilizzati.

In *Istruzioni su come predisporre il repository* verrà fornito un caso specifico per illustrare in dettaglio il processo di *bundling*.

Schemi XSD

Attualmente il materiale semantico pubblicato dalla UE si basa sui formati RDF ed XSD.

Il Catalogo non supporta il processamento di file `.xsd`. Questi potranno essere supportati in un secondo momento.

In *Istruzioni su come predisporre il repository* verrà fornito un caso specifico per illustrare in dettaglio uno *schema XSD*.

5.4.7 Esempi

Repository

Ad esempio, analizziamo un repository strutturato come segue:

⁶⁷ <https://spec.openapis.org/oas/v3.0.3#data-types>

⁶⁸ <https://datatracker.ietf.org/doc/draft-polli-restapi-ld-keywords/>

⁶⁹ <https://github.com/italia/api-oas-checker>

⁷⁰ <https://json-schema.org/understanding-json-schema/structuring.html#bundling>

```

bash
├── README.md
├── publiccode.yaml
├── assets/ontologies/
│   ├── Onto1/
│   │   ├── onto1.ttl
│   │   └── onto1.rdf
│   ├── Onto2/
│   │   └── README.md
│   ├── Onto3/
│   │   ├── Other/
│   │   │   └── temp.md
│   │   └── onto3.ttl
│   ├── Onto4/
│   │   └── latest/
│   │       ├── onto1.ttl
│   │       └── onto1.rdf
│   └── notes.md
├── assets/controlled-vocabularies/
└── ...
    
```

Il repository non contiene schemi, quindi il Catalogo non aggiungerà schemi al catalogo durante l'harvesting. Questo non rappresenta un problema e non è considerato un errore.

I file informativi (es. `README.md`, `notes.md`) presenti sia nella radice che nelle sottodirectory vengono ignorati durante l'harvesting.

Per quanto riguarda la directory `Onto1/`:

- essa non contiene sotto-directory né altre directory al suo interno ed è quindi una cartella foglia. Quindi viene processata come potenzialmente contenente un'ontologia;
- contiene un file RDF/Turtle che verrà processato;
- contiene un altro file RDF, plausibilmente una serializzazione diversa degli stessi contenuti del file `.ttl` in RDF/XML. Poiché il processo di harvesting di schema utilizza solo i file di tipo `text/turtle` con estensione `.ttl`, questo file non è usato nel processo stesso.

La directory `Onto2/` non contiene file `.ttl`: questo viene segnalato solamente come **WARNING**.

La directory `Onto3/` ha una sottodirectory, quindi non è considerata come contenitore di ontologia, ma come directory intermedia nel cammino per altre directory foglia: il file `onto3.ttl` è ignorato e non processato.

La directory `Onto4/` contiene una sottodirectory `latest/` che contiene un file `.ttl`, quindi viene processata come potenzialmente contenente un'ontologia.

Versionamento directory degli asset semantici

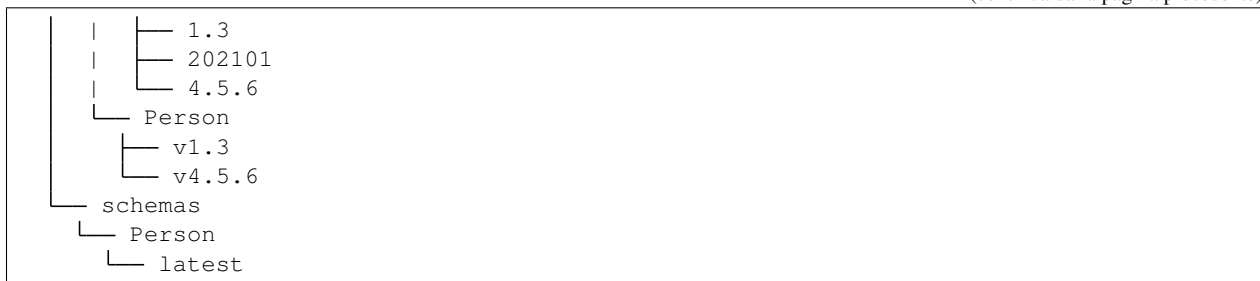
A titolo di esempio, di seguito è fornita una possibile organizzazione delle directory sfruttando il versionamento. È importante notare che le versioni dell'ontologia **Car** non sono prefissate da `v` mentre quelle di **Person** sono tutte prefissate da `v`.

```

bash
├── assets
│   └── ontologies
│       └── Car
    
```

(continues on next page)

(continua dalla pagina precedente)



Nell'esempio di seguito, invece, sono presenti sei esempi di percorsi non validi, anche perché le directory contengono contemporaneamente versioni prefissate da v che senza prefisso.



È possibile che un repository contenga versioni precedenti delle risorse semantiche per fini storici, al di là del versionamento supportato da git.

L'harvesting delle ontologie considera che le directory che contengono ontologie possano essere versionate, non i singoli file. Questo vale anche per le sotto-directory.

Attualmente, il Catalogo non prende in considerazione il versionamento delle cartelle per schemi dati e vocabolari controllati, ma per le ontologie prende in considerazione:

- latest/ se presente;
- quella maggiore secondo il seguente ordinamento:
 - tra due versioni espresse come forme numeriche (con punti), si segue l'ordinamento comunemente condiviso per cui i numeri a sinistra sono i più significativi;
 - qualora due versioni abbiano lunghezza diversa ma una sia prefisso dell'altra, la più lunga viene considerata più recente; ad esempio, v4 . 5 è considerata obsoleta in presenza di v4 . 5 . 2.

Focus su alberatura per le ontologie

Di seguito viene fornito un esempio di alberatura, comprensiva di versionamento, contenente i file che definiscono un'ontologia. In questo caso viene processata solo la directory "latest". Nell'esempio, l'alberatura contiene una serie di file di documentazione opzionali che non vengono processati.



(continues on next page)

(continua dalla pagina precedente)

```
v1.1/
  MyOntology.ttl
latest/
  MyOntology.ttl
  MyOntology.rdf
  LATEST.md
```

Focus su alberatura per i vocabolari controllati

Di seguito l'esempio di un'alberatura contenente un vocabolario controllato e la sua proiezione in formato `csv` generata utilizzando le informazioni di *framing* indicate in `framing.yamllld`.

Il file `datapackage.yaml` contiene i metadati del `csv`.

```
bash
assets/
  controlled-vocabulary/
    my-codelist/
      CHANGELOG.md
      README.md
      my-codelist.ttl
      my-codelist.csv
      datapackage.yaml
      framing.yamllld
```

Il file `my-codelist.ttl` contiene il vocabolario controllato in formato RDF/Turtle.

```
turtle

@prefix skos: <http://www.w3.org/2004/02/skos/core#> .

@prefix at: <http://publications.europa.eu/ontology/authority/> .

@prefix atold: <http://publications.europa.eu/resource/authority/> .

@prefix dc: <http://purl.org/dc/elements/1.1/> .

@prefix owl: <http://www.w3.org/2002/07/owl#> .

@prefix c: <http://publications.europa.eu/resource/authority/country/> .

c:ITA a skos:Concept;

  dc:identifier "ITA";

  skos:prefLabel "Italy"@en, "Italia"@it, "Italie"@fr;

  skos:inScheme atold:country

.

c:DEU a skos:Concept;

  dc:identifier "DEU";
```

(continues on next page)

(continua dalla pagina precedente)

```
skos:prefLabel "Germany"@en, "Germania"@it, "Allemagne"@fr;
skos:inScheme atold:country
.
c:ESP a skos:Concept;
dc:identifier "ESP";
skos:prefLabel "Spain"@en, "Spagna"@it, "Espagne"@fr;
skos:inScheme atold:country
.
```

Il `my-codelist.csv` contiene le proiezioni del vocabolario controllato in formato `csv`.

```
csv
# It is possible to add comments
# metadata is into datapackage.yaml
"id","label_en","label_it","label_fr"
"ITA","Italy","Italia","Italie"
"DEU","Germany","Germania","Allemagne"
"ESP","Spain","Spagna","Espagne"
```

Il file `Datapackage.yaml` contiene tutte le informazioni sui metadati del file `csv`.

```
yaml
# Datapackage.yaml
profile: data-package
resources:
- name: my-codelist
  path: my-codelist.csv
  profile: tabular-data-resource
  dialect:
    delimiter: ","
    doubleQuote: true
    lineTerminator: ""
  schema:
```

(continues on next page)

(continua dalla pagina precedente)

```

x-jsonld-type: skos:Concept
x-jsonld-context:
  "@context":
    skos: http://www.w3.org/2004/02/skos/core#
    dc: http://purl.org/dc/elements/1.1/
    at: http://publications.europa.eu/ontology/authority/
    atold: http://publications.europa.eu/resource/authority/
    c: http://publications.europa.eu/resource/authority/country/
  id: dc:identifier
  label_it:
    "@id": skos:prefLabel
    "@language": it
  label_en:
    "@id": skos:prefLabel
    "@language": en
  label_fr:
    "@id": skos:prefLabel
    "@language": fr
  fields:
    - name: id
      type: string
    - name: label_en
      type: string
    - name: label_it
      type: string
    - name: label_fr
      type: string

```

Schema bundling

Un esempio di file OAS3 metadatato con i campi `x-jsonld-context` e `x-jsonld-type`:

```
yaml
openapi: 3.0.1
...
components:
  schemas:
    Person:
      type: object
      x-jsonld-type: "https://w3id.org/italia/onto/CPV/Person"
      x-jsonld-context:
        "@vocab": "https://w3id.org/italia/onto/CPV/"
        nome_proprio: givenName
        cognome: familyName
      properties:
        nome_proprio: {type: string, ..}
        cognome: {type: string, ..}
      ...
```

Di seguito l'esempio di un'alberatura contenente uno schema.

```
bash
assets/
  schemas/
    Person/
      CHANGELOG.md
      README.md
      person.oas3.yaml
      index.ttl
```

Schemi XSD

L'esempio di seguito fa riferimento al **Countries Authority Table**.

L'*authority table* dei paesi *Countries* viene pubblicata a partire dall'URL su <http://publications.europa.eu/resource/dataset/country> contenente i link a tutti i dataset associati e corrispondente al suo URI.

All'indirizzo <https://publications.europa.eu/resource/authority/country> si trova l'elenco dei paesi in formato RDF; sotto quell'URL ci sono i riferimenti ai singoli paesi, eg. <https://publications.europa.eu/resource/authority/country/ITA>.

Il versionamento è contenuto all'interno degli RDF e l'URL viene dereferenziato all'ultima versione.

Gli URI sono versionati, ad esempio <http://publications.europa.eu/resource/expression/country/20170920-0>.

Da lì è possibile individuare una lista di dataset associati ed eventualmente localizzati: qui https://publications.europa.eu/resource/cellar/07ed8d46-2b56-11e7-9412-01aa75ed71a1.0001.12/DOC_14 la lista delle coppie codice/paese in italiano in formato XML (ATTO table, usate per le traduzioni).

```
xml
<TABLE VL="IT" NAME="countries">
  <LIBELLE CODE="1A0">Kosovo</LIBELLE>
  <LIBELLE CODE="ABW">Aruba</LIBELLE>
  <LIBELLE CODE="AFG">Afghanistan</LIBELLE>
  ...
</TABLE>
```

Qui una codelist (estensione .gc) <http://publications.europa.eu/resource/distribution/country/20210616-0/xml/gc/Country.gc> contenente tutti i dati in un formato xml analogo a quello tabellare.

```
xml
<?xml version="1.0" encoding="UTF-8"?>
<gc:CodeList xmlns:gc="http://docs.oasis-open.org/codelist/ns/genericcode/1.0/">
  <Identification>
    <CanonicalUri>http://publications.europa.eu/resource/dataset/country</CanonicalUri>
    <CanonicalVersionUri>http://publications.europa.eu/resource/dataset/country/20210616-0
    ↪</CanonicalVersionUri>
    <LocationUri>http://publications.europa.eu/resource/distribution/country/20210616-0/
    ↪xml/gc/Country.gc</LocationUri>
  </Identification>
  <ColumnSet>
    <Column Id="code" Use="required">
      <ShortName>Code</ShortName>
      <Data Type="normalizedString" Lang="eng"/>
    </Column>
    <Column Id="Name" Use="optional">
      <ShortName>Name</ShortName>
      <Data Type="string" Lang="eng"/>
    </Column>
```

(continues on next page)

(continua dalla pagina precedente)

```
<Column Use="optional" Id="ita_label">
  <ShortName>engLabel</ShortName>
  <Data Type="string" Lang="ita"/>
</Column>
...
</ColumnSet>
<SimpleCodeList>
  <Row>
    <Value ColumnRef="code">
      <SimpleValue>ITA</SimpleValue>
    </Value>
    <Value ColumnRef="Name">
      <SimpleValue>Italy</SimpleValue>
    </Value>
    <Value ColumnRef="ita_label">
      <SimpleValue>Italy</SimpleValue>
    </Value>
    ...
  </SimpleCodeList>
```

Gli stessi dati possono essere recuperati a partire da <https://data.europa.eu/data/datasets/country>.

La Guida al Catalogo è stata realizzata grazie alla collaborazione tra:

- Dipartimento per la trasformazione digitale - Presidenza del Consiglio dei Ministri
- Consiglio dei Ministri, ISTAT - Istituto Nazionale di Statistica
- CNR - Consiglio Nazionale delle Ricerche

Hanno collaborato alla stesura della presente guida: Fabio Bonelli, Fabrizio Davide, Matteo Fortini, Giorgia Lodi, Claudia Pollina, Mauro Pratesi, Roberto Puglisi, Agostino Purificato, Roberta Radini.

La guida è aggiornata a Giugno 2024