

CATALOGO NAZIONALE DATI

Metodologia per l'interoperabilità semantica



CATALOGO NAZIONALE DATI

Metodologia per l'interoperabilità semantica.

A cura di: Giovanna Bellitti e Massimo Fedeli.

Revisione redazionale: Lara Parisella.

Attività editoriali: Manuela Marrone.

Grafica: Sofia Barletta.

© 2025

Istituto Nazionale di Statistica

Via Cesare Balbo, 16 - Roma

ISBN: 978-88-458-2179-0



Salvo diversa indicazione, tutti i contenuti
pubblicati sono soggetti alla licenza

Creative Commons - Attribuzione - versione 4.0.

<https://creativecommons.org/licenses/by/4.0/deed.it>

È dunque possibile riprodurre, distribuire,
trasmettere e adattare liberamente dati e analisi
dell'Istituto nazionale di statistica, anche a scopi
commerciali, a condizione che venga citata la fonte.

Immagini, loghi (compreso il logo dell'Istat),
marchi registrati e altri contenuti di proprietà di terzi
appartengono ai rispettivi proprietari e non
possono essere riprodotti senza il loro consenso.

INDICE

	Pag.
Introduzione	5
Capitolo 1 - Strategia per l'interoperabilità	7
1.1 Progetti del Piano Nazionale di Ripresa e Resilienza	7
1.1.1 Prospettive future	10
1.2 Metodologia per lo sviluppo del Catalogo Nazionale Dati	10
1.2.1 Analisi concettuale dei domini da modellare	12
1.2.2 Modellazione delle risorse semantiche	14
1.2.2.1 Descrizione del metodo di modellazione e l'ingegnerizzazione dell'ontologia	15
1.2.2.2 Ciclo di vita delle risorse semantiche	19
1.3 Modalità di contribuzione al Catalogo	21
Capitolo 2 - Casi di contribuzione al Catalogo	23
2.1 Casi esemplificativi di contributo al Catalogo	23
2.2 Risorse semantiche <i>core</i> dell'Istat	23
2.3 Risorse semantiche del Ministero dell'Interno: <i>asset</i> semantici dell'Anagrafe Nazionale della Popolazione Residente	24
2.4 Risorse semantiche del Parlamento	25
2.5 Risorse semantiche di Inps e Inail	25
2.6 Risorse semantiche dell'Agenzia del Demanio	26
Conclusioni	27
Riferimenti bibliografici	29

INTRODUZIONE¹

Il Catalogo Nazionale Dati (NDC) è una piattaforma digitale progettata e realizzata dall'Istat in qualità di soggetto attuatore del progetto PNRR (Piano Nazionale di Ripresa e Resilienza) di titolarità del Dipartimento per la trasformazione digitale. L'obiettivo del Progetto è il perseguimento dell'interoperabilità semantica tra i sistemi informativi al fine di garantire che il significato dei dati scambiati sia specifico e uguale per tutti i partecipanti allo scambio informativo. L'interoperabilità semantica è fondamentale per una strategia dati nazionale nella misura in cui garantisce che i dati provenienti da fonti diverse siano compresi e utilizzati in modo coerente, permettendo l'integrazione efficace tra sistemi informativi pubblici e privati e migliorando la qualità delle decisioni. Si favorisce, inoltre, la condivisione sicura e trasparente dei dati, riducendo ambiguità e duplicazioni e si supporta l'automazione dei processi e l'innovazione nei servizi digitali. L'interoperabilità semantica è essenziale per costruire un ecosistema digitale basato su fiducia, efficienza e collaborazione.

In tale contesto, la funzione del Catalogo nazionale della semantica dei dati (*schema.gov.it*) è rendere liberamente accessibili le risorse semantiche (ontologie, schemi dati e vocabolari controllati) relative ai dati delle pubbliche amministrazioni e alle API (*Application Programming Interface*) resi disponibili tramite la Piattaforma Digitale Nazionale Dati (PDND). Sotto tale profilo, emerge la stretta connessione tra il Catalogo Nazionale Dati e la PDND.

Occorre assicurare un'armonizzazione del modo di rappresentare ed esporre il significato dei dati destinati all'interscambio digitale tra pubbliche amministrazioni. Per soddisfare tale esigenza, già affrontata nell'ambito del trattamento statistico dei dati, l'Istituto nazionale di statistica ha messo in campo iniziative per valorizzare le proprie competenze ed esportarle.

Il presente contributo descrive la metodologia elaborata dall'Istat per l'attuazione del Progetto del Catalogo, tramite le proprie strutture competenti per gli aspetti metodologici, informatici e giuridici. Il Progetto rappresenta un'occasione per indirizzare la digitalizzazione della Pubblica amministrazione verso il traguardo del miglioramento dei servizi resi ai cittadini, a cominciare dall'attuazione del principio *una tantum*, in virtù del quale una Pubblica amministrazione non deve richiedere al cittadino un dato dallo stesso già fornito alla Pubblica amministrazione.

Giovanna Bellitti
Direttrice centrale affari giuridici e legali

Massimo Fedeli
Direttore del Dipartimento per lo sviluppo di metodi e tecnologie
per la produzione e diffusione dell'informazione statistica

¹ L'ebook è stato curato da Giovanna Bellitti e Massimo Fedeli. La metodologia per l'attuazione del Catalogo Nazionale Dati (NDC) è stata sviluppata dall'Istat tramite le proprie strutture competenti per l'attuazione del Progetto PNRR "Catalogo nazionale della semantica dei dati". Hanno collaborato il Dipartimento per lo sviluppo di metodi e tecnologie per la produzione e diffusione dell'informazione statistica, diretto da Massimo Fedeli, la Direzione centrale affari giuridici e legali, diretta da Giovanna Bellitti, la Direzione centrale per la metodologia e il disegno dei processi statistici, diretta da Orietta Luzi, e la Direzione centrale per le tecnologie informatiche, diretta da Cecilia Colasanti.

CAPITOLO 1

STRATEGIE PER L'INTEROPERABILITÀ¹

1.1 Progetti del Piano Nazionale di Ripresa e Resilienza

Una delle leve strategiche dell'Unione europea per la realizzazione del mercato unico digitale UE è la trasformazione digitale della Pubblica amministrazione, obiettivo fondato, innanzitutto, sulla valorizzazione del patrimonio informativo pubblico. Le iniziative messe in campo a livello nazionale sono basate, da una parte, sulla convinzione di digitalizzare gli scambi informativi tramite un'infrastruttura tecnologica per l'erogazione di *e-service*, dall'altra, sulla consapevolezza circa la necessità di prevedere strumenti che agevolino la messa a terra dell'infrastruttura stessa.

Due progetti del Piano Nazionale di Ripresa e Resilienza (PNRR)² sono strategici per realizzare tali intenti: la Piattaforma Digitale Nazionale Dati (PDND) e il Catalogo Nazionale Dati (*National Data Catalog* - NDC), che hanno la funzione di mettere a disposizione, rispettivamente, l'accesso a un catalogo di *e-service* erogati dalle pubbliche amministrazioni e una raccolta di risorse semantiche idonee a garantire l'interoperabilità semantica tra i sistemi informativi³.

La PDND pubblica il Catalogo API (*Application Programming Interface*) che espone tutti gli *e-service* pubblicati dagli enti erogatori e tramite il quale è possibile effettuare la richiesta di fruizione dati e, successivamente, integrare i propri servizi per i cittadini.

Al fine di agevolare lo sviluppo della PDND, all'Istituto Nazionale di Statistica - Istat è stata affidata la progettazione, creazione e attuazione di un catalogo centrale in *open data* comprendente schemi di dati, ontologie e vocabolari di base a sostegno della creazione e della progettazione di servizi digitali interoperabili: il Catalogo Nazionale Dati (NDC). Il Progetto è attuato secondo i principi dell'*European Interoperability Framework* (EIF)⁴ e dell'*Interoperable Europe Act*⁵, entrambi ispirati dall'obiettivo di perseguire l'interoperabilità tra i sistemi informativi a tutti i livelli: giuridico, semantico, tecnico e organizzativo⁶.

1 Il Capitolo è stato redatto da Roberto Puglisi (paragrafi 1.1, 1.2, 1.2.1, 1.3) e Roberta Radini (paragrafi 1.1.1, 1.2.2, 1.2.2.1, 1.2.2.2, 1.3); Figure 1.1, 1.2 e 1.3 di Paolo Nicolai. Viene, qui, descritta la metodologia per lo sviluppo del Catalogo elaborata dalle strutture competenti dell'Istat (cfr. Introduzione, nota 1) e la conseguente attuazione operativa svolta dal Gruppo di lavoro costituito con deliberazione del Presidente dell'Istat (DOP/870/2022 del 19 ottobre 2022) e organizzato, riguardo allo sviluppo delle risorse semantiche, in due *Work Package* dedicati rispettivamente alla modellazione delle risorse semantiche (WP2, coordinato dalla dott.ssa Roberta Radini) e all'analisi concettuale dei domini di interesse (WP3, coordinato dal dott. Roberto Puglisi).

2 Il PNRR, approvato il 13 luglio 2021 con Decisione di esecuzione del Consiglio (CID) contenente in un allegato gli obiettivi e i traguardi per ciascuna missione, richiama alla Missione 1 – componente 1 “Digitalizzazione della Pubblica amministrazione” le misure volte a migliorare la digitalizzazione della PA, che si articolano in sette investimenti e tre riforme. Gli investimenti sono diretti in particolare a: 1) razionalizzare e consolidare le infrastrutture digitali esistenti della PA; 2) promuovere la diffusione del *cloud computing*; 3) con particolare attenzione all'armonizzazione e all'interoperabilità delle piattaforme e dei servizi di dati, la realizzazione del principio *once only* e l'accessibilità dei dati tramite un catalogo centrale di “connettori automatici” (cosiddette API - *Application Programming Interface*).

3 Cfr. Figura 1.1.

4 *Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions* (COM/2017/0134 final), *European Interoperability Framework - Implementation Strategy* (cfr. <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex:52017DC0134>).

5 Regolamento (UE) 2024/903 del Parlamento europeo e del Consiglio del 13 marzo 2024, che stabilisce misure per un livello elevato di interoperabilità del settore pubblico nell'Unione (Regolamento su un'Europa interoperabile, cfr. <https://eur-lex.europa.eu/legal-content/IT/ALL/?uri=CELEX:32024R0903>).

6 Cfr. paragrafo 1.2.

Il NDC ha l'obiettivo di fornire un modello e uno standard comune per lo scambio, l'armonizzazione e la comprensione dei dati e dei metadati gestiti dalle pubbliche amministrazioni, rendendoli aperti, strutturati e interoperabili. A tale fine, l'Istat ha realizzato il portale *schema.gov.it*, l'infrastruttura informatica del Catalogo la cui funzione è rendere disponibili le descrizioni semantiche dei dati detenuti dalle amministrazioni pubbliche. Il portale mette a disposizione numerose sezioni e funzionalità; in tale modo, l'utente può ricercare gli strumenti del Catalogo per tipologia (schemi dati, ontologie e vocabolari controllati), per categorie tematiche o per parole chiave. Il funzionamento di *schema.gov.it* è fondato sulla raccolta delle risorse semantiche (*harvesting*) allocate dai rispettivi titolari su propri *repository* e dotate di un *Uniform Resource Identifier* (URI), vale a dire un identificativo univoco che risulti stabile e permanente nel tempo⁷.

Le applicazioni principali della modellazione e della standardizzazione delle informazioni dei dati condivisi dalle pubbliche amministrazioni sono espresse attraverso l'interoperabilità semantica dei dati implementata da API e *Linked Open Data* (LOD). I LOD costituiscono una tecnica per pubblicare sul web i *Linked Data* che sono dati identificati univocamente che non solo sono collegati tra loro in una forma che costituisce una rete di informazioni, ma sono soprattutto aperti.

Alla luce delle funzioni e delle attività progettuali necessarie, l'Istat e il Dipartimento per la trasformazione digitale (DTD) hanno stipulato un accordo ai "sensi dell'art. 5 comma 6, del D.lgs. n. 50/2016, per la realizzazione della misura 1.3.1 Catalogo Nazionale Dati della Missione M1 – Componente C.1 – Asse 1 del PNRR", con il quale l'Amministrazione titolare ha affidato al Soggetto attuatore "l'attuazione del suddetto intervento alle condizioni di cui al presente Accordo"⁸.

L'accordo prevede, altresì, la costituzione di un Comitato di Attuazione composto da cinque componenti, di cui tre designati dall'Amministrazione titolare e due dall'Istat⁹. Tale Comitato ha la funzione di supportare le parti nell'esercizio delle funzioni di coordinamento tecnico operativo delle attività previste e, in particolare: i) garantisce il costante monitoraggio delle attività, anche al fine di proporre adeguate soluzioni a eventuali criticità emergenti in corso di attuazione; ii) esamina i contenuti dei report prodotti nel corso di attuazione ed evidenzia eventuali scostamenti rispetto alle attività programmate e ai tempi di attuazione previsti; iii) assicura che non siano effettuate attività in sovrapposizione con altri interventi del PNRR. L'intero impianto organizzativo mira a garantire un presidio completo del progetto sotto il profilo tecnico, giuridico, metodologico e strategico, in coerenza con gli indirizzi del Piano triennale IT 2021-2023 dell'Istat e con le finalità del PNRR in materia di innovazione digitale, trasparenza e interoperabilità.

La logica funzionale che guida la realizzazione del Progetto NDC si fonda su una divisione di competenze tra il DTD¹⁰, quale soggetto titolare dell'intervento, l'Istat, quale soggetto attuatore, e le altre pubbliche amministrazioni, quali soggetti responsabili delle risorse semantiche (ontologie, vocabolari controllati e schemi dati) relative ai propri dati e raccolte nel Catalogo.

Per il conseguimento dell'obiettivo di agevolare l'interoperabilità semantica fornendo una descrizione dei dati pubblici, l'Istat ha coniugato le esigenze multidisciplinari tipiche della digitalizzazione impiegando competenze giuridiche, metodologiche e informatiche per rendere disponibile una raccolta di risorse semantiche dei dati della PA idonee ad abilitare l'interscambio digitale delle informazioni tra le pubbliche amministrazioni. È stata, così, predisposta una

7 Per un approfondimento sul funzionamento dell'infrastruttura tecnologica del portale <https://schema.gov.it/>, cfr. *Guida al Catalogo nazionale della semantica dei dati* (<https://teamdigitale.github.io/dati-semantic-guida-ndc-docs/index.html>).

8 Cfr. https://presidenza.governo.it/AmministrazioneTrasparente/BandiContratti/AccordiTraAmministrazioni/allegati/220303_Accordo_ISTAT.pdf.

9 Per l'Istat sono stati individuati, quali membri effettivi del Comitato di Attuazione Istat-DTD, la Dott.ssa Giovanna Bellitti e il Dott. Massimo Fedeli; quali membri supplenti, la Dott.ssa Cristina Freguja, la Dott.ssa Orietta Luzi e il Dott. Fabio Massimo Rapioti.

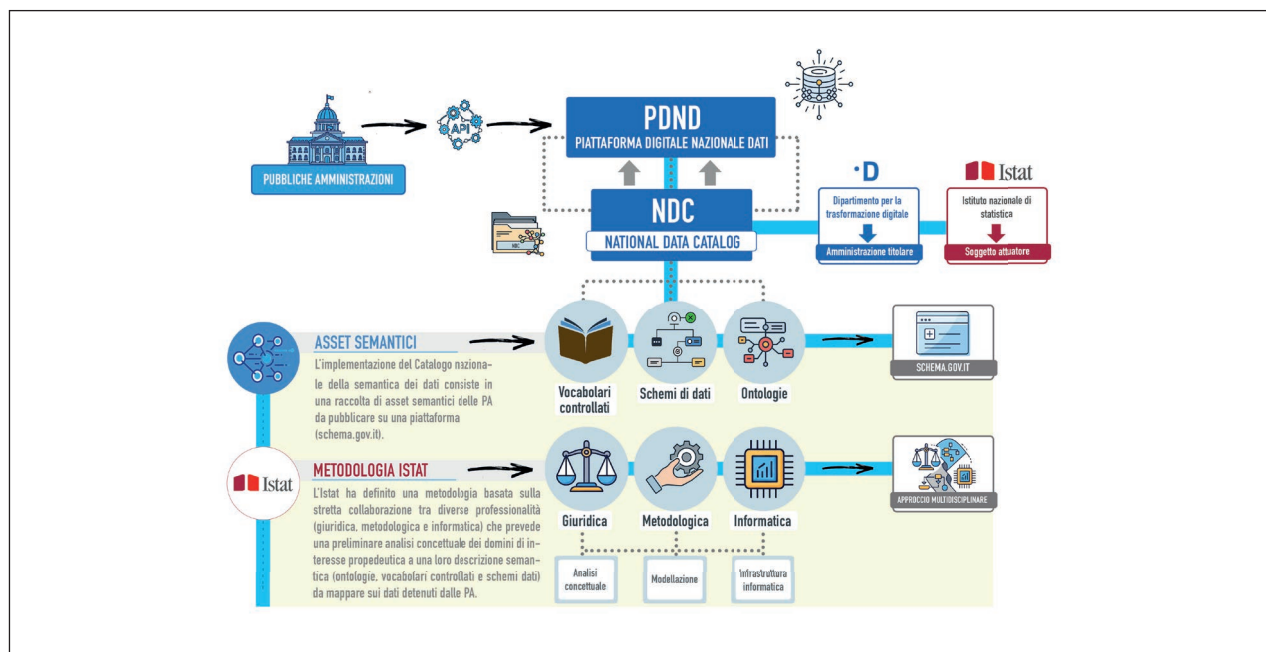
10 L'articolo 24-ter del DPCM 1° ottobre 2012 individua il Dipartimento per la trasformazione digitale quale soggetto preposto alla promozione e al coordinamento delle azioni del Governo finalizzate alla definizione di una strategia unitaria in materia di "trasformazione digitale e di modernizzazione del Paese attraverso le tecnologie digitali" e, a tale fine, "dà attuazione alle direttive del Presidente in materia e assicura il coordinamento e l'esecuzione dei programmi di trasformazione digitale".

metodologia di sviluppo del Catalogo fondata sulla stretta collaborazione tra le diverse professionalità coinvolte al fine di valorizzarne i rispettivi contributi.

Con l'elaborazione e l'applicazione di tale metodo, l'Istat ha valorizzato l'esperienza nello svolgimento dei propri compiti istituzionali¹¹ sviluppando nuove competenze nel trattamento dei dati da mettere al servizio dell'interoperabilità dei sistemi informativi¹².

Il Catalogo Nazionale Dati, conclusa la fase della sua progettazione e realizzazione da parte dell'Istat¹³ e del conseguente suo primo popolamento con risorse di diversi contributori¹⁴ supportati dall'Istat¹⁵, dovrà essere messo a regime tenendo conto delle esigenze soddisfatte con la metodologia qui descritta e avendo cura di stabilizzare, dunque, i requisiti che consentono la disponibilità di risorse semantiche relative ai dati della Pubblica amministrazione. L'attuazione del Progetto del Catalogo apre la strada a un nuovo metodo di trattamento dei dati progettato e attuato in modo interoperabile tra sistemi informativi, sfruttando API aperte, standard comuni, metadati condivisi e governance organizzativa coordinata. Porre l'interoperabilità semantica quale fulcro del trattamento dei dati consente di affrontare in maniera adeguata le sfide sollevate dalla digitalizzazione.

Figura 1.1 Metodologia e governance del Catalogo Nazionale Dati



Fonte: Istat

11 Il D.lgs. 6 settembre 1989, n. 322, recante "Norme sul Sistema statistico nazionale e sulla riorganizzazione dell'Istituto Nazionale di Statistica, ai sensi dell'art. 24 della legge 23 agosto 1988, n. 400", in particolare l'art. 15, comma 1, attribuisce all'Istat, tra gli altri, il compito di provvedere "alla predisposizione delle nomenclature e metodologie di base per la classificazione e la rilevazione dei fenomeni di carattere demografico, economico e sociale" - lett. e); e di provvedere "alla promozione e allo sviluppo informatico a fini statistici degli archivi gestionali e delle raccolte di dati amministrativi".

12 Con deliberazione del Presidente dell'Istat (DOP/870/2022 del 19 ottobre 2022), è stato istituito un Comitato interdipartimentale a supporto del quale sono stati attivati quattro Gruppi di lavoro, ognuno dei quali dedicato a un'area specifica: giuridica, metodologica, informatica e amministrativo-contabile.

13 La durata dell'Accordo tra il Dipartimento per la trasformazione digitale (DTD) e l'Istat è prevista per il 30 giugno 2026.

14 Cfr. paragrafo 1.4 e successivi.

15 Cfr. paragrafo 1.3.

1.1.1 Prospettive future

L'interoperabilità semantica per gli *e-service* così sviluppata deve essere gestita nel tempo prevedendo attività di aggiornamento delle risorse semantiche che debbono essere coerenti con il variare delle procedure amministrative e delle relative discipline, ma anche con il crescente fabbisogno informativo correlato a un ricorso su vasta scala all'interoperabilità semantica per i processi amministrativi.

L'integrazione del linguaggio ontologico per l'espressione della semantica dei dati rappresenta, oggi, una componente strategica anche per lo sviluppo di sistemi di intelligenza artificiale (IA) avanzati che, nella loro evoluzione, avranno bisogno di dati sempre maggiori di alta qualità e affidabili (il cui significato è reso esplicito, condiviso e interpretato e manipolato anche da macchine). Quindi, l'impiego di dati con tale requisiti all'interno di architetture intelligenti permette di costruire un'infrastruttura conoscitiva in grado di supportare processi decisionali più robusti, coerenti e verificabili.

Gli agenti intelligenti, alimentati tramite *e-service*, che forniscono conoscenza strutturata e semanticamente arricchita, sono così in grado di operare in modo più contestualizzato, eseguendo inferenze che rispettino vincoli ontologici e logiche definite a monte.

Questa modalità operativa non solo accresce l'affidabilità delle risposte fornite dagli agenti IA, ma le rende anche certificabili e abilita la tracciabilità delle fonti informative e dei processi logici che hanno portato a una determinata conclusione o suggerimento. La presenza di un substrato semantico consente infatti di documentare ogni passaggio decisionale, evidenziando i concetti coinvolti, le relazioni attivate e i dati utilizzati, secondo modelli trasparenti e formalizzati.

In tale contesto assume un ruolo cruciale il principio di *explainability*, ovvero la capacità di un sistema di intelligenza artificiale di fornire spiegazioni comprensibili e strutturate circa il proprio comportamento. L'adozione di modelli ontologici e dati semanticamente annotati rende possibile implementare meccanismi di spiegazione che non si limitino alla descrizione *ex post* delle decisioni, ma che possono motivare in modo preciso e coerente le ragioni che hanno condotto l'agente a una certa inferenza, indicando i fattori che hanno influenzato il processo e le regole semantiche attivate.

Ciò rappresenta un elemento di fondamentale importanza soprattutto in contesti nei quali è richiesta non solo un'elevata accuratezza nelle risposte fornite dai sistemi intelligenti, ma anche la possibilità di validare tali risposte in termini chiari, normativamente conformi e cognitivamente accessibili all'essere umano.

1.2 Metodologia per lo sviluppo del Catalogo Nazionale Dati

L'implementazione del Catalogo nazionale della semantica dei dati (*National Data Catalog* - NDC) consiste, dunque, in una raccolta di *asset* semantici (ontologie, vocabolari controllati e schemi dati) delle pubbliche amministrazioni da pubblicare su una piattaforma (*schema.gov.it*). Tali *asset* sono funzionali alla Piattaforma Digitale Nazionale Dati (PDND), il cui obiettivo è l'interoperabilità dei sistemi informativi della Pubblica amministrazione e dei gestori di servizi pubblici. NDC, infatti, punta ad abilitare lo sviluppo, da parte di tutti gli enti pubblici erogatori di API ed *e-service* propedeutici all'interoperabilità, dei sistemi informativi degli stessi enti.

Interoperabilità dei flussi informativi tra le pubbliche amministrazioni significa accesso diretto di una PA ai dati di un'altra PA; l'immediata conseguenza favorevole è, da una parte, disporre tempestivamente delle informazioni utili alla conclusione di un determinato procedimento amministrativo e, dall'altra, ridurre il *burden*

amministrativo e burocratico sui cittadini: maggiore efficienza e minore burocrazia dell'azione amministrativa e, dunque, un miglioramento dei servizi per i cittadini. L'interoperabilità si fonda sull'idea di fare dialogare i sistemi informativi pubblici originariamente progettati, da ciascuna PA, per soddisfare esigenze informative in relazione solo ai propri fini istituzionali; il superamento di un assetto monadistico dipende dalla disponibilità di un catalogo di asset semantici (NDC, appunto) comune a tutti i titolari di basi di dati e sistemi informativi pubblici.

Nell'ambito dell'UE, secondo il quadro europeo di interoperabilità (*European Interoperability Framework* - EIF) e l'*Interoperable Europe Act*, lo sviluppo dell'interoperabilità deve riguardare quattro livelli orizzontali: giuridico, organizzativo, semantico e tecnico (gli ultimi tre sono direttamente connessi alle accennate componenti necessarie di un sistema informativo – rispettivamente, processi, dati e strumenti informatici)¹⁶. A tale riguardo, vanno considerate le differenze tra il livello europeo e il livello nazionale di interoperabilità legate alle rispettive diverse potenzialità di implementazione giuridica e semantica.

L'interoperabilità giuridica, ad esempio, consiste nella possibilità di instaurare rapporti tra le pubbliche amministrazioni senza ostacoli derivanti dalle diversità relative alla disciplina di riferimento; l'esigenza di un'interoperabilità giuridica si manifesta, dunque, solo con riferimento alla dimensione europea in considerazione della frammentarietà dovuta alla perdurante esistenza di ordinamenti giuridici nazionali che, unitamente al fattore multilinguistico e multiculturale, determina un alto tasso di complessità nel perseguimento dell'interoperabilità semantica. Si pensi, infatti, all'obiettivo di rendere interoperabile una qualunque procedura amministrativa che può ricevere, in via di principio, per ogni Stato membro una diversa regolamentazione che, a cascata, può comportare diverse concettualizzazioni giuridiche influenzate, per di più, dalle rispettive matrici linguistiche.

L'interoperabilità semantica assume, invece, un carattere essenziale a prescindere dall'ambito nazionale o sovranazionale di riferimento, attenendo alla necessità di garantire che il significato del dato sia specifico e uguale per tutti i partecipanti allo scambio informativo. È importante, a tale proposito, distinguere due diversi aspetti dell'interoperabilità semantica: il significato dei dati e le tecniche informatiche di rappresentazione della conoscenza (ontologie); sono le due componenti delle risorse semantiche raccolte nel Catalogo (*schema.gov.it*). Così, da una parte, è necessario acquisire la conoscenza mediante un processo di concettualizzazione dei dati in virtù del quale gli esperti giuridici del dominio di interesse relativo alla procedura amministrativa analizzano l'informazione (ad esempio, l'indirizzo dove una persona dimora abitualmente) per giungere *bottom-up* a un livello di astrazione (ad esempio, concetto di residenza) tale da consentire il riferimento *top-down* ad altri dati concreti classificabili nei medesimi concetti. D'altro canto, le tecniche informatiche di rappresentazione della conoscenza si sviluppano in un processo che, acquisita la conoscenza di un dominio di interesse mediante la relativa concettualizzazione, la descrivono impiegando modelli di dati (ontologie) che possano essere riutilizzati per lo sviluppo di *e-service* aventi a oggetto lo scambio di dati. Ciò che appare evidente è che, se il sostrato concettuale che cattura la conoscenza deve ricorrere alle tecniche informatiche di rappresentazione della conoscenza (ontologie) per evolversi digitalmente, quest'ultime non possono prescindere da una rigorosa analisi concettuale compiuta da esperti del dominio di interesse, pena un'inconsistenza tale da renderle del tutto inadeguate allo scopo. A tale proposito, l'Istat assegna pari rilievo ai due processi, considerandoli entrambi essenziali e da svolgersi in stretta collaborazione per la realizzazione degli obiettivi del Progetto.

16 Cfr. *European Interoperability Framework*, COM/2017/0134 final, All. 2, par. 3. Precedentemente, nel 2004, con l'*European Interoperability Framework for Pan-European eGovernment Services* non si dava autonoma rilevanza al livello giuridico, emerso invece più chiaramente dopo, alla luce della elevata frammentarietà del panorama normativo dei singoli Stati membri dell'UE.

Per il perseguimento di tale obiettivo, l'Istat ha definito una metodologia in cui ciascuna delle accennate professionalità necessarie per lo sviluppo di risorse semantiche interoperabili offre il proprio contributo strettamente correlato con quello degli altri. Si prevede, così, una preliminare analisi concettuale dei domini di interesse propedeutica a una loro descrizione semantica (ontologie, vocabolari controllati e schemi dati) da mappare sui dati detenuti dalle pubbliche amministrazioni.

La metodologia elaborata per la realizzazione del Progetto NDC prevede, in prima istanza, l'analisi del dominio e dei suoi concetti, propedeutica all'attività definitoria degli stessi. Il prodotto di tale analisi concettuale è la definizione del dato destinato a essere oggetto di interscambio informativo tra i soggetti aderenti alla PDND¹⁷. Naturalmente, vista la natura amministrativa delle procedure prese in considerazione, le norme giuridiche assurgono a fonte primaria da cui trarre le definizioni e il riferimento principale da cui trarre il termine preferito e, conseguentemente, occorre un'adeguata specializzazione giuridica per procedere sia all'analisi concettuale del dominio, sia alla sua descrizione. L'assunzione di tale consapevolezza appare il vero valore aggiunto della metodologia in esame nella misura in cui implica un approccio necessariamente multidisciplinare, evitando che in un'unica professionalità si possano consumare tutte le fasi attuative del Progetto. In altre parole, in ambito di *e-service*, esperti di classificazioni giuridiche dei processi amministrativi possono procedere all'analisi concettuale dei domini di interesse e alla relativa descrizione semantica, ed esperti di analisi metodologica possono provvedere alla modellazione delle relative risorse semantiche in una logica di integrazione. Il rispetto di tale metodo garantisce standard di qualità necessari per l'interscambio informativo di dati amministrativi; la creazione delle risorse semantiche si svolge in stretta collaborazione con il soggetto titolare del *dataset* sia nella fase di analisi concettuale sia in quella di modellazione ontologica.

1.2.1 Analisi concettuale dei domini da modellare

L'interoperabilità semantica dei sistemi informativi della PA si deve fondare sulla disponibilità di un linguaggio comune a tutti gli utenti abilitati a scambiarsi i dati tramite PDND. Sotto questa prospettiva, occorre garantire l'unicità del significato dei dati scambiati in modo che l'informazione trasmessa da un sistema corrisponda a quella effettivamente richiesta (e necessaria per un determinato procedimento amministrativo) da un altro; è necessario, dunque, che i concetti alla base delle informazioni scambiate vengano descritti univocamente evitando, così, la possibilità di errori semantici (associare al termine un significato diverso dal "suo"). L'interoperabilità tra i sistemi informativi di due o più banche dati comporta, così, una necessaria uniformità concettuale rispetto ai dati stessi; il significato di un dato deve, in altre parole, essere lo stesso per i soggetti che rispettivamente lo richiedono e lo rilasciano. I titolari delle banche dati interessate devono, dunque, condividere lemma e definizione di un determinato concetto.

L'interoperabilità semantica si traduce, perciò, nella condivisione concettuale dei termini e dei dati da parte di tutti gli utilizzatori degli stessi. Riferito ai sistemi informativi pubblici, ciò comporta la loro armonizzazione sotto il profilo semantico-giuridico attraverso un processo definitorio da cui fare scaturire la "corretta" relazione tra il lemma impiegato per l'erogazione di un *e-service* e la relativa definizione. La correttezza dei termini, espressa dal binomio lemma-definizione, va perseguita, infatti, prendendo come punto di riferimento l'ordinamento giuridico (ad esempio, impiegare il lemma "azienda" solo per quei dati su cui può mapparsi il concetto

17 Cfr. Figura 1.1.

di “azienda” definito dal codice civile). Per la realizzazione dell'interoperabilità semantica dei dati è necessaria, così, una descrizione concettuale dei dati pubblici armonizzata a livello nazionale per disporre di una chiave di accesso univoca, per titolare del dato e fruitore, ai concetti coinvolti nelle diverse procedure amministrative¹⁸.

La caratteristica principale dei *data asset* pubblici è quella di costituire patrimoni informativi derivanti dall'esercizio di funzioni amministrative pubbliche che, per la Costituzione, la legislazione ordinaria e la normativa regolamentare, rispondono al principio di legalità. Ciò significa, tra l'altro, che i provvedimenti amministrativi devono avere il carattere della tipicità e della tassatività. Una diretta conseguenza di ciò sull'interoperabilità semantica delle PA (e, quindi, sull'implementazione dei contenuti di NDC) è la necessità di utilizzare, laddove si tratti di concetti già definiti dal legislatore, definizioni ricavabili dall'ordinamento giuridico. Tale approccio trova riscontro nei regolamenti europei in cui si dà atto di come l'interoperabilità debba essere intesa in senso ampio, spaziando dagli aspetti tecnici a quelli giuridici e comprendendo elementi strategici del settore. La gamma di attività dovrebbe dunque spingersi al di là del ciclo di vita abituale delle soluzioni per includere tutti gli elementi degli interventi in grado di promuovere le condizioni necessarie per un'interoperabilità costante a livello generale.

Tale contesto non consente, con riferimento ai concetti già definiti in norme, la possibilità di procedere con definizioni concettuali avulse dall'ordinamento giuridico ovvero approssimative, nella misura in cui le stesse non garantirebbero un requisito fondamentale per l'interoperabilità della PA: la condivisione unanime di un concetto e del suo significato da parte di tutti i soggetti istituzionali coinvolti nell'interscambio.

Per il perseguimento di tale obiettivo, dunque, l'unica opzione idonea è agganciare i contenuti semantici da mappare sui dati prodotti da procedure amministrative a un unico fattore comune: le fonti normative. La conoscenza prodotta dall'analisi concettuale delle normative di settore relative ai procedimenti amministrativi (in forza dei quali vengono prodotti i dati scambiati con gli *e-service*) è destinata a essere modellata in ontologie¹⁹ che, con l'utilizzo di un linguaggio formale standardizzato e *machine readable*²⁰, abilitano l'interoperabilità semantica tra i sistemi informativi. A tale riguardo, si richiama la funzionalità di NDC rispetto alla PDND espressa dall'Accordo di collaborazione Istat-DTD sopra richiamato.

Un ulteriore aspetto derivante da una calibrazione dei concetti esposti nei contenuti semantici del Catalogo su parametri normativi (laddove, ovviamente, gli stessi concetti abbiano una derivazione giuridica) è garantire la legittimità dei procedimenti amministrativi che verranno istruiti sulla base di un interscambio di dati mediante la Piattaforma Digitale Nazionale Dati; solo l'accesso a un dato amministrativo con efficacia giuridica consentirà, alla PA richiedente il dato, di concludere legittimamente la procedura amministrativa che ha dato impulso alla richiesta del dato stesso (si pensi, ad esempio, al concetto di residenza con riferimento al quale è possibile enucleare i concetti di residenza anagrafica di cui al DPR 223/1989, residenza civilistica di cui all'art. 43 c.c. e residenza fiscale di cui all'art. 2, comma 2, Testo unico delle imposte sui redditi, non necessariamente sempre coincidenti considerato rispettivamente il valore presuntivo della prima ed effettivo delle seconde).

Sotto il profilo semantico-giuridico occorre, dunque, individuare i concetti fondamentali coinvolti in un determinato procedimento amministrativo attraverso un'analisi giuridica del flusso di dati coinvolti (ad esempio, emissione di un certificato residenza), nonché le definizioni da assegnare ai concetti fondamentali attraverso

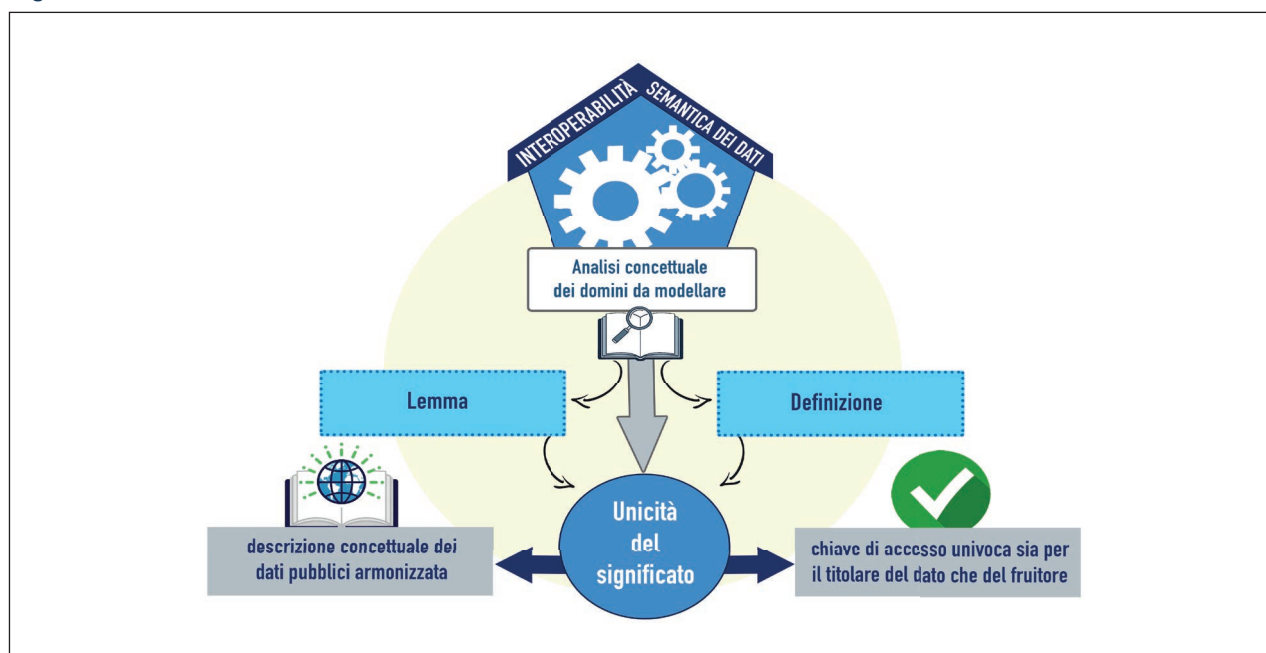
18 Cfr. Figura 1.2.

19 Cfr. Figura 1.3.

20 Cfr. paragrafo 1.2.2.

un'analisi giuridica della disciplina normativa di riferimento che conduca a un significato unico e comune per ciascun concetto. Tale operazione deve essere svolta da esperti giuridici che devono assicurare il riferimento corretto al dato normativo e garantire, così, a ciascuna definizione una solida condivisione da parte di tutti i soggetti coinvolti; al tempo stesso, tale prima fase deve essere svolta in funzione della successiva modellazione ontologica condotta da esperti metodologi. L'analisi concettuale non può, quindi, consistere esclusivamente in una trattazione giuridica dei concetti da modellare. Le definizioni dei concetti hanno l'obiettivo di catturare sinteticamente il significato dei dati amministrativi univocamente stabilito dall'ordinamento giuridico e, perciò, devono consistere in una descrizione essenziale corredata da puntuali riferimenti normativi e contenente informazioni esclusivamente di derivazione normativa (attribuendo altro senso che quello fatto palese dal significato proprio delle parole secondo la connessione di esse, e dalla intenzione del legislatore); solo laddove manchi una precisa disposizione, si farà ricorso alle disposizioni che regolano casi simili o materie analoghe; se il caso rimane ancora dubbio, si decide secondo i principi generali dell'ordinamento giuridico dello Stato.

Figura 1.2 Analisi concettuale dei domini di interesse



Fonte: Istat

1.2.2 Modellazione delle risorse semantiche

L'obiettivo finale, l'interoperabilità dei dati, può rappresentare oggi il motore principale della crescita e dello sviluppo della digitalizzazione della PA, fenomeno supportato e favorito dall'incredibile capacità delle macchine di elaborare informazioni e di metterle a disposizione sia di altre macchine sia di utenti finali, che nel caso dei dati amministrativi possono essere altre pubbliche amministrazioni e/o cittadini. Si evidenzia come la caratteristica fondamentale di questo scambio di dati preveda sia meccanismi di autoalimentazione, attraverso il riutilizzo dei dati condivisi e impiegati da soggetti diversi dal titolare originario, sia la derivazione di nuove informazioni da quelle disponibili, favorendo così un sistema che si arricchisce e si integra costantemente.

Per realizzare queste funzioni è necessario sviluppare sistemi interoperabili, ovvero capaci di comunicare e scambiarsi dati in modo automatico sfruttando un linguaggio formale e standardizzato leggibile e interpretabile da macchine, ma anche una semantica informativa comune definita attraverso il processo sopra descritto.

Se alle esigenze dell'interoperabilità tecnica dei dati risponde la PDND, lo scopo principale del Catalogo è quello di fornire un vocabolario di concetti la cui definizione sia univoca, l'informazione sia categorizzata e i concetti siano interconnessi da relazioni esplicite.

Per implementare l'interoperabilità semantica dei dati si deve creare un connubio tra:

- una semantica comune, realizzata da analisi tematiche del dominio specifico;
- un modello dell'informazione implementato attraverso un linguaggio formale standardizzato;
- una soluzione tecnica che, sfruttando la semantica univoca e la modellazione condivisa dell'informazione, implementi soluzioni software come i LOD, le API e gli e-service.

1.2.2.1 Descrizione del metodo di modellazione e l'ingegnerizzazione dell'ontologia

A tal fine è stata messa a punto una metodologia specifica che, all'analisi dei processi amministrativi e alla definizione univoca della semantica dei concetti espressi attraverso lemma e definizione, descritti nel paragrafo precedente, consenta di modellare una risorsa semantica distinta in ontologia, vocabolario controllato e schema dati.

Le risorse semantiche sono espresse attraverso un linguaggio rappresentativo della conoscenza, come quello utilizzato per l'ingegnerizzazione delle ontologie: OWL (*Web ontology language*) e RDFS (*Resource description framework schema*).

Il linguaggio ontologico consente di formalizzare e standardizzare le rappresentazioni della realtà e costruire una rete di concetti e proprietà interconnessi tra di loro con relazioni più o meno forti tanto da consentire di parlare di un grafo della conoscenza: in altri termini, di una rete semantica. La concettualizzazione della rete semantica applicata ai dati li rende compatibili e facilmente scambiabili tra diversi sistemi informativi.

L'integrazione degli aspetti concettuali e tecnici richiede una stretta collaborazione tra esperti di dominio tematico e di modellazione dell'informazione. Questa sinergia consente di ottenere una modellazione dell'informazione il più possibile avulsa dall'uso specifico e di mantenere un livello di generalizzazione che consente di utilizzare l'informazione per diversi scopi, pur mantenendo integra la consistenza dell'informazione. Per portare un esempio, l'ontologia dei certificati anagrafici consente di modellare le evidenze di tutti i certificati anagrafici e, nello stesso tempo, consente di mantenere l'informazione connessa sia al concetto di "Persona residente in un comune", sia al concetto più generale di "Persona".

Quindi la modellazione ontologica delle risorse semantiche si deve muovere secondo due specifiche direttrici: la *modellazione di concetti interdisciplinari* che consentono la creazione di interconnessioni tra diversi ambiti conoscitivi, sfruttando concetti con una semantica univoca e condivisa, e la *modellazione di concetti specifici di un dominio* che ne rappresentino le peculiarità, pur mantenendo le interconnessioni con i concetti più generali e interdisciplinari.

Le risorse semantiche del Catalogo sono rappresentate da ontologie, da vocabolari controllati e da schemi dati.

Per tutte le risorse semantiche del Catalogo è stata impostata una modellazione conforme a quanto specificato nella documentazione pubblicata dal team per la trasformazione digitale congiuntamente con AgID (Agenzia per l'Italia Digitale) e secondo i principi e i profili applicativi di modellazione descritti in DCAT-AP (*Data Catalog Vocabulary Application Profile*, definito nell'ambito del Programma ISA della Commissione europea) e DCAT-AP_IT (profilo italiano dei metadati utili per descrivere i dati delle pubbliche amministrazioni) nel caso di

vocabolari controllati, mentre per le ontologie e gli schemi dati sono state adottate le regole di metadatazione dell'ontologia ADMS-AP_IT (*Semantic Asset Description Metadata Schema Application Profile*, profilo italiano dei metadati utili per descrivere le ontologie delle pubbliche amministrazioni).

Come già introdotto, la modellazione ontologica consente non solo di definire la semantica di ogni concetto e le relazioni dei vari concetti, il tutto espresso in un linguaggio standard interpretabile dalle macchine (*machine readable*), ma soprattutto di attribuire un identificativo univoco e persistente nel tempo sul web a ogni concetto.

Questo implica che ogni dato riferito al concetto "Persona anagraficamente residente" è identificato da una IRI (*Internationalised Resource Identifier*)²¹ e il suo significato è "Persona fisica avente dimora abituale in un determinato Comune come risultante nell'anagrafe della popolazione residente (artt. 3 e 7 del DPR n. 223/1989; art. 43 del Codice Civile)".

Anche la modellazione di vocabolario controllato produce un elenco standardizzato di concetti caratterizzati da termini di riferimento (lemmi + definizioni), da codici e descrizioni e, nel caso di elenchi gerarchici, anche da livelli e relazioni di dipendenza tra le voci. L'elenco è utile per organizzare, descrivere, predefinire e indicizzare la conoscenza di un dominio (ad esempio, vocabolario controllato dei motivi di cessazione di un rapporto di lavoro) e, in particolare, per codificare in modo univoco i dati (anche in questo caso le singole modalità dei valori che può assumere un vocabolario sono identificate da una IRI).

I vocabolari controllati si distinguono in:

- Lista di termini (*List/Pick List*): "un insieme limitato di termini disposti come semplice elenco alfabetico o in qualche altro modo logicamente evidente". Ad esempio, la *Code List* è una lista di oggetti identificati da un codice e denotati da un termine che copre le diverse modalità descrittive di un concetto, senza pretesa di esaustività e mutua esclusività.
- Tassonomia (*Taxonomy*): "vocabolario controllato composto da termini preferenziali, tutti collegati in una gerarchia o poli-gerarchia". Ad esempio, l'Ateco 2025 è organizzato in una struttura gerarchica (Sezioni, Divisioni, Gruppi, Classi, Categorie e Sottocategorie) dove esistono livelli più ampi (ad esempio le Divisioni) che comprendono sottolivelli di dettaglio (ad esempio i Gruppi), ognuno con relativi termini e codici.
- Tesaurus (*Thesaurus*): "vocabolario controllato disposto in un ordine noto e strutturato in modo che le varie relazioni tra i termini siano chiaramente visualizzate e identificate da indicatori di relazione standardizzati che devono essere impiegati in maniera reciproca". Nel tesaurus le relazioni semantiche possono essere sia in forma gerarchica sia di equivalenza, come le sinonimie, o in forma associativa, e si possono applicare più criteri di classificazione allo stesso concetto: ad esempio, un tesaurus geografico può prendere in considerazione il territorio parallelamente sia per caratteristiche fisiche (ad esempio i confini geografici) sia per entità politico-amministrative (ad esempio le cittadinanze).

I vocabolari controllati vengono modellati secondo l'ontologia SKOS (*Simple Knowledge Organisation System Reference*)²², che è uno standard del web, e, in alcuni casi, secondo la XKOS (*Extended Knowledge Organisation System*)²³, che estende SKOS per le esigenze delle classificazioni statistiche.

La classificazione dei concetti attraverso i vocabolari viene riportata nelle ontologie, associandola ai concetti, che possono essere enumerati da un vocabolario in modo che anche altri enti possano utilizzare la stessa codifica. Questo consente sia di standardizzare i dati ma anche di renderli automaticamente integrabili.

21 Schema.gov.it, "Persona anagraficamente residente" (<https://schema.gov.it/ldview/onto/RPO/RegisteredResidentPerson>).

22 Cfr. <https://www.w3.org/TR/skos-reference/>.

23 Cfr. <https://ddialliance.org/xkos#documentation>.

In ultimo, le risorse semantiche degli schemi dati descrivono, utilizzando le IRI dei concetti ontologici o dei valori che possono assumere attraverso i vocabolari controllati, lo schema dei dati di input e di output di una API. Questo consente non solo di standardizzare le interfacce dati delle API che possono essere pubblicate come *e-service* su PDND, ma anche standardizzare la nomenclatura di trasmissione dei dati da una Pubblica amministrazione a un'altra.

Nell'esperienza maturata nella modellazione delle risorse semantiche del Catalogo, spesso l'esigenza di modellare schemi dati usati per la modellazione di *e-service* ha fatto emergere come alcuni domini tematici fossero assenti dal Catalogo o che ci fosse la necessità di arricchirlo con nuovi concetti.

Si può dire che, come NDC è servente PDND, vale anche viceversa, ossia le esigenze semantiche espresse dagli *e-service* hanno consentito di ampliare la rete semantica di NDC.

La modellazione dei prodotti semantici si può suddividere in attività specifiche:

- a. Specificazione del dominio: avvio di un'istruttoria condotta da esperti giuridici-amministrativi dello specifico tematismo, che delineano l'ambito informativo della modellazione. Questa attività è svolta insieme al titolare del dominio tematico e/o dal fruitore dei concetti che verranno utilizzati nella modellazione di schemi dati.
- b. Acquisizione di conoscenza di dominio: da questo lavoro scaturisce un documento descrittivo del quadro di insieme di una procedura amministrativa o del quadro normativo di riferimento che ne costituisce l'ambito.
- c. Concettualizzazione: da questo prodotto, il team di modellazione semantica e di esperti di dominio individuano i concetti principali, le proprietà e le relazioni. Viene stilata una prima lista di concetti principali e il loro significato. In questo contesto, si potrebbero utilizzare analisi formali e informali attraverso lo studio di testi e manuali, in combinazione con interviste strutturate e non strutturate con esperti, per definire un glossario e successivamente includere nuovi termini (o rimuoverne, se non pertinenti) o definire sinonimie.

Dalla concettualizzazione del dominio, il team di esperti di dominio e di semantica provvede a reperire le informazioni tecnico-giuridiche che ne descrivono l'ambito giuridico amministrativo e compongono la definizione, che viene poi revisionata per essere resa in un linguaggio fruibile anche dai non tecnici. Produce un testo descrittivo del dominio e un glossario dei termini principali che lo caratterizzano.

- d. Integrazione: il team di modellazione ontologica analizza le ontologie già disponibili nella rete semantica del Catalogo per definire il riutilizzo e le interconnessioni con la concettualizzazione già modellata ontologicamente dalla rete di conoscenza del Catalogo, perché di competenza di altri domini. In questa fase si possono fare analisi di completezza di modellazione di alcuni domini ed effettuare la richiesta di espansione del dominio con i concetti mancanti o da ridefinire.
- e. Implementazione e ingegnerizzazione dell'ontologia: il team di modellazione ontologica delinea la modellazione collegando eventuali domini già modellati oppure sotto-domini che possono includere eventuali "sfumature" definitorie o nuove appartenenze concettuali. Questo lavoro è condiviso con il team di esperti di dominio.

Premesso che un'ontologia si compone essenzialmente di classi, proprietà e ruoli (ossia le relazioni), i passi operativi per la modellazione di un'ontologia si compongono di:

- individuare i concetti principali che diventeranno le classi, rilevando, se presenti, delle relazioni gerarchiche, e quindi organizzare le classi in sottoclassi o superclassi;

- definirme le proprietà e descriverne i valori leciti (dominio e *range*), attribuire i valori alle proprietà per tutte le istanze create e i vincoli sulle proprietà in termini di cardinalità e funzionalità;
 - individuare le relazioni tra classi che diventeranno i “ruoli” e i rispettivi domini e *range* o, nel caso di sotto-classi, rappresentare relazioni di appartenenza, composizione, unione e disgiunzione.
- f. Documentazione: i risultati delle lavorazioni dei due team sono formalizzati in due prodotti distinti che vengono integrati nella modellazione a grafo dell'ontologia. Questo rappresenta il primo prodotto di modellazione che viene affinato, ampliato ripercorrendo i passi descritti in modo ciclico.
- g. Valutazione: i prodotti di modellazione vengono condivisi con il titolare del dominio, che valida la modellazione e la semantica dei concetti. La modellazione viene condivisa in forma grafica per renderla più fruibile anche a chi non è esperto di linguaggi di modellazione ontologica. Alla valutazione semantica viene affiancata anche una valutazione sintattica attraverso ragionatori automatici che effettuano una valutazione formale del modello ontologico.

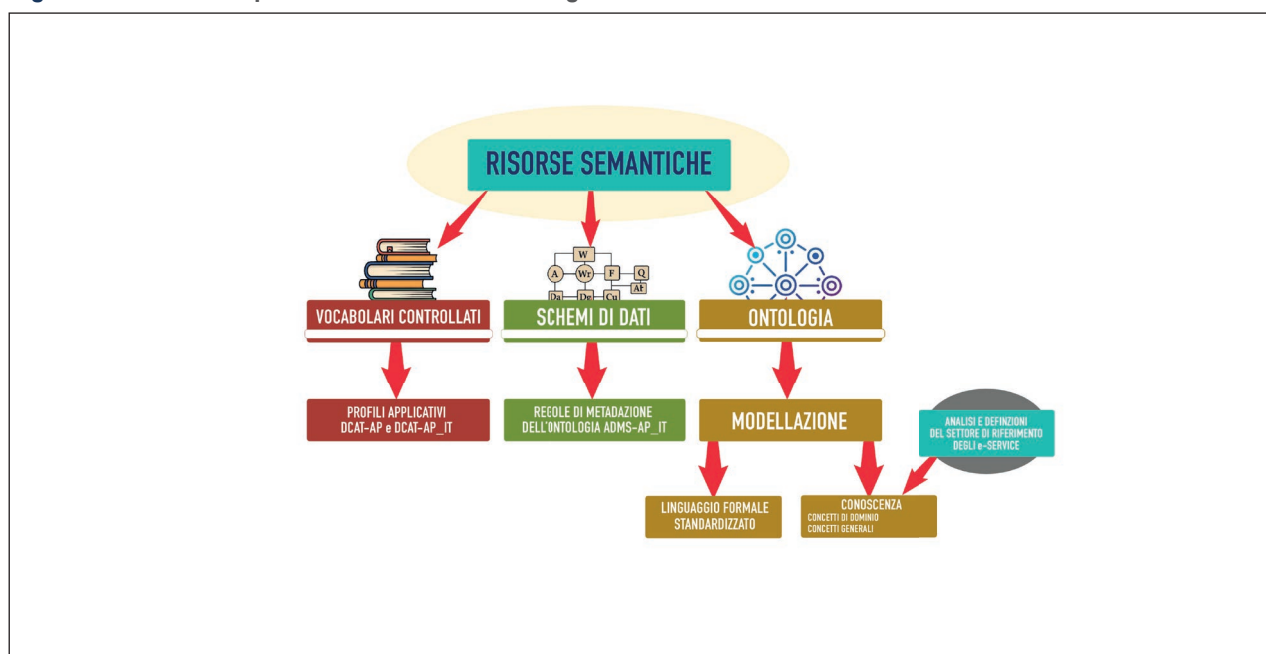
In particolare, le attività di identificazione delle principali componenti semantiche vengono affrontate secondo tre diversi approcci:

1. Approccio *bottom-up*: si individuano prima tutti i concetti più specifici, poi quelli più generali. Tale procedimento va iterato fino al raggiungimento di un livello di astrazione (generalizzazione) sufficiente per gli scopi per cui viene progettata l'ontologia. I concetti saranno organizzati secondo un ordine che rappresenta la gerarchia degli stessi.
2. Approccio *top-down*: si individuano prima i concetti più generali, e poi, via via, quelli più specifici, arrivando a un livello di specializzazione adeguato agli scopi dell'ontologia. Anche qui i concetti saranno organizzati in un ordine che rispetta la gerarchia degli stessi.
3. Approccio *middle-out*: si procede individuando prima tutti i concetti che si ritengono importanti per il dominio da rappresentare, poi i legami gerarchici che legano i concetti, per organizzare gli stessi attraverso la loro specializzazione o la generalizzazione secondo le esigenze specifiche.

Quindi, lo sviluppo dell'ontologia prevede come ultimo passo la creazione del grafo dell'ontologia di dominio integrata con la rete semantica già modellata e pubblicata sul Catalogo. Infatti, sul Catalogo, dovendo garantire la coerenza con quanto già definito nel corpus delle ontologie già pubblicato, si preferisce applicare il metodo misto.

In generale, la modellazione ontologica parte sempre da una ricognizione sulle ontologie già pubblicate, in particolare sul Catalogo, per verificare se esistano concetti comuni già definiti e per prevederne l'integrazione attraverso assiomi che esprimano la coincidenza espressa con lemmi diversi (*same_as*) o la definizione di sottoinsiemi concettuali, oppure, eventualmente, nuove relazioni. Queste operazioni consentono l'integrazione di dati esposti da diverse fonti delle pubbliche amministrazioni che forniscono dati rispetto allo stesso concetto o dominio. Ogni sviluppo ontologico di un dominio garantisce la completezza rispetto all'uso per cui è stata effettuata la modellazione, ma questa può rappresentare una copertura parziale rispetto all'intero dominio di pertinenza, che viene via via completato con interventi successivi, a volte dettati anche da esigenze di altri domini.

Figura 1.3 Metodo per la modellazione ontologica



Fonte: Istat

1.2.2.2 Ciclo di vita delle risorse semantiche

Le attività di predisposizione degli asset semantici prevedono due distinte macro-fasi:

1. la fase di “gestione del ciclo di vita dei prodotti semantici” che parte dalla loro prima modellazione ai successivi aggiornamenti, che spesso consistono nell’ampliamento concettuale o nella creazione di nuovi link concettuali con prodotti già pubblicati;
2. la fase di “pubblicazione” sulla piattaforma *schema.gov.it*.

Come descritto nel paragrafo sulla descrizione del metodo di modellazione e ingegnerizzazione delle ontologie, queste possono subire degli aggiornamenti di diverso carattere, che possono sintetizzarsi in:

- a. *Nuove esigenze conoscitive* della rete semantica del Catalogo emerse nella modellazione di altre procedure amministrative dello stesso dominio oppure di altri domini che vi si riferiscono. Ad esempio, modellando il dominio dell'erogazione pensionistica, gestito da Inps, è scaturita l'esigenza di specificare la caratteristica dell'identificativo univoco delle persone residenti registrate nell'Anagrafe Nazionale della Popolazione Residente (ANPR) appartenente al dominio delle persone residenti gestito dal Ministero dell'Interno. In questo caso è stata ampliata la modellazione dell'ontologia RPO (*Registered Resident Person Ontology*) aggiungendo l'attributo che si riferisce all' ID_ANPR.
- b. *Aggiornamento della normativa o del processo amministrativo* a cui fa riferimento il dominio specifico. Questo aggiornamento può impattare soltanto sulla parte semantica perché modificata la normativa di riferimento, oppure anche sulla modellazione ontologica perché modifica i concetti, le proprietà o le relazioni tra i concetti.
- c. *Aggiornamento descrittivo di classificazioni*. Ad esempio, si è passati dal vocabolario controllato dell'Ateco 2007, aggiornamento 2022, alla nuova struttura dell'Ateco 2025. In questo caso si aggiornano i prodotti semantici aggiungendo un nuovo vocabolario controllato e aggiornando il concetto ontologico Ateco alla nuova classificazione.

In particolare, in questo ultimo anno ha ricoperto una rilevanza particolare l'esigenza da parte dei contributori di PDND, che si apprestano a realizzare e pubblicare i propri *e-service*, di creare gli schemi dati degli *e-service* con l'annotazione semantica della rete di concetti pubblicate dal Catalogo.

Infatti, per realizzare l'interoperabilità semantica, come previsto nelle linee guida dell'interoperabilità (AgID 2023), alla soluzione tecnica che implementa "l'interoperabilità tecnica" attraverso API e la gestione della sicurezza con gli *e-service*, si affianca l'interoperabilità semantica esplicita con le raccomandazioni REG_SEMANTICA_001 e REG_SEMANTICA_002.

In questo caso è possibile avere diverse casistiche che variano da "tutte le entità e le relative proprietà trovano copertura" tra le risorse del Catalogo a "nessuna o almeno una delle entità o proprietà è compresa nelle rappresentazioni". In questo secondo caso si è dato luogo all'aggiornamento delle risorse semantiche del Catalogo (caso a) anche con la creazione di nuove ontologie o vocabolari. Lo sviluppo di nuovi *e-service* ha dato luogo spesso anche allo sviluppo di nuovi schemi dati pubblicati sul Catalogo (cfr. i casi di Inps e dell'Agenzia del Demanio).

Altro *driver* di sviluppo del Catalogo in questo ultimo anno è stato quello di definire specifiche ontologie per la pubblicazione di LOD (cfr. il caso del progetto *OpenCUP* e le nuove ontologie di Camera e Senato).

Il modello di gestione del ciclo di vita delle risorse semantiche del Catalogo più idoneo è quello *incrementale*, che procede per affinamenti successivi, come già descritto, mentre quello *a cascata* non è adeguato per l'assenza di una specifica completa dei requisiti nelle prime fasi del processo di sviluppo e per l'evoluzione delle definizioni delle ontologie nel tempo.

Inoltre, l'operazione di pubblicazione delle risorse semantiche sulla piattaforma *schema.gov.it* richiede il rispetto di vincoli tecnici secondo quanto previsto dal corretto funzionamento dell'*harvester*, che prevede dei passi aggiuntivi, rispetto alla modellazione ontologica e alla conseguente sua formalizzazione in linguaggio OWL, ossia:

- la registrazione degli identificativi unici delle risorse semantiche;
- la predisposizione dei metadati descrittivi dei prodotti semantici, secondo quanto stabilito dal processo di *harvesting* del portale e delle ontologie standard DCAT-AP_IT e ADMS-AP_IT, che contribuiscono all'indicizzazione dei prodotti semantici e quindi alla loro possibilità di recupero attraverso la ricerca;
- la predisposizione degli archivi degli *asset* semantici (tipicamente su GitHub) in una schematizzazione standard, secondo quanto indicato nelle linee guida del Catalogo.

L'interoperabilità dei sistemi di dati è uno dei principi base su cui sono stati definiti gli *indicatori di misurazione del livello di digitalizzazione della PA* predisposto da AgID in collaborazione con Istat e DTD.

Tra gli indicatori che caratterizzano il livello di digitalizzazione raggiunto da una PA sono stati considerati sia la presenza di risorse semantiche pubblicate sul Catalogo da ogni PA, sia di *e-service* che sfruttano, attraverso l'annotazione semantica, le risorse della rete ontologica.

Il raggiungimento dell'interoperabilità semantica, quindi, contribuisce alla realizzazione sia della digitalizzazione della Pubblica amministrazione, sia allo sviluppo dell'uso della IA nei processi amministrativi della PA.

Infatti, l'uso integrato del linguaggio ontologico, dell'espressione della semantica dei dati e dei dati annotati semanticamente, come i LOD o quelli acquisiti attraverso *e-service*, consente di ottenere applicazioni di IA che restituiscono risposte automatiche più affidabili e certificabili. Implementando, così, la funzione dell'*explainability* nell'IA, ossia essere in grado di spiegare perché *AI Agent* ha preso determinate decisioni, quali fattori hanno influenzato i suoi risultati o come funziona il suo comportamento.

Inoltre, l'applicazione delle ontologie ai dati consente di implementare il paradigma del *Ontology-Based Data Management* - OBDM (Kontchakov *et al.* 2013), ovvero la possibilità di applicare algoritmi di IA di ragionamento

automatico (*reasoner*) in grado di evidenziare incoerenze sui dati o estrarre pattern conoscitivi non noti, ossia di fare inferenza sui dati.

In definitiva, la digitalizzazione della Pubblica amministrazione e l'interoperabilità tra i vari sistemi sono aspetti essenziali per favorire un progressivo miglioramento nell'efficienza, nella trasparenza e nella qualità dei servizi pubblici. Superare il modello dati a silos, promuovere uno scambio di dati automatico e sicuro e creare un quadro normativo e semantico condiviso sono passaggi essenziali per costruire un sistema amministrativo più aperto, efficace e orientato al cittadino. La cooperazione tra esperti di modellazione ontologica e semantica, istituzioni e tecnologie rappresenta il motore di questa evoluzione. Richiede investimento in competenze e risorse in grado di lavorare ad ampio spettro.

In particolare, le competenze per realizzare modellazione semantica e ontologica di un dominio sono interdisciplinari e Istat le ha costruite creando team che consentono di lavorare trasversalmente su tematiche metodologiche e giuridiche.

1.3 Modalità di contribuzione al Catalogo

Sono tre le modalità tramite le quali, in funzione del proprio grado di autonomia nella modellazione semantica, i soggetti titolari di *asset* di dati pubblici possono contribuire allo sviluppo del Catalogo con proprie risorse semantiche: sviluppo delle risorse in autonomia e supporto dell'Istat per la pubblicazione nel Catalogo, sviluppo e pubblicazione delle risorse da parte dell'Istat (*semantic stewardship*) e/o formazione dell'Istat per la modellazione degli schemi dati e verifica degli schemi dati da parte dell'Istat. Le diverse modalità si possono sintetizzare in:

- a. *Semantic stewardship dell'Istat*. In qualità di Soggetto attuatore del Progetto, l'Istat offre alle pubbliche amministrazioni interessate a pubblicare risorse semantiche nel Catalogo una *semantic stewardship*, ossia un servizio di supporto finalizzato 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 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 scelta nel caso in cui il Contributore non abbia competenze interne sufficienti per la modellazione semantica e ontologica e necessita 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 dall'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 i dati della classificazione ed eventualmente la loro gerarchia nel caso di vocabolari articolati in un formato strutturato, che riporti: il livello di gerarchia, la relazione gerarchica padre-figlio, la codifica, il lemma e la definizione 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.
- b. *Sviluppo autonomo delle risorse*. Partendo dal presupposto per il quale ogni soggetto titolare di dati pubblici è anche il titolare delle risorse semantiche riferite agli stessi dati, lo stesso può scegliere di produrre le risorse semantiche in autonomia, qualora sia in possesso di sufficienti risorse e competenze per farlo. In questo caso, l'Istat fornisce supporto per consentire la pubblicazione delle risorse nel Catalogo e nel controllo del rispetto

dei vincoli di metadatazione necessari al corretto funzionamento del processo di indicizzazione delle risorse attraverso il processo di *harvesting*.

- c. *Formazione da parte dell'Istat per sviluppo schemi dati*. Come descritto, gli schemi dati sono una risorsa semantica che rappresenta lo schema dati di una o di più API che possono essere implementate negli *e-service* della PDND. A tal fine è stato messo a disposizione uno strumento specifico per l'implementazione di questi schemi dati, detto *Schema Editor*. Per lo sviluppo degli schemi dati sono necessarie competenze tecniche di sviluppo degli schemi secondo lo standard *OpenAPI* e competenze semantiche per associare la corretta annotazione semantica allo schema dati. In questo caso, l'Istat mette a disposizione un supporto su entrambe le competenze e, in aggiunta, una formazione specifica nell'uso dell'applicativo *Schema Editor*.

In alcuni casi il supporto o la validazione della modellazione degli schemi dati svolta dall'Istat ha dato luogo all'attivazione delle altre due modalità di contribuzione, perché ha richiesto di revisionare o aggiungere ontologie o vocabolari controllati, come nel caso di Agenzia del Demanio o di Inps.

In tutte le modalità di contribuzione, specialmente la prima volta in cui la PA contribuisce al Catalogo, l'Istat controlla la configurazione dei *repository* GitHub delle risorse semantiche e la corretta metadatazione delle risorse necessarie all'*harvesting*. Tutte le pubblicazioni passano prima per un ambiente di test che si aggiorna automaticamente ogni due ore (nelle ore pari della giornata), che consente sia all'Istat sia al Contributore di validare la corretta procedura di *harvesting* e di visualizzazione sul Catalogo, per poi passare alle 22:00 nell'ambiente di produzione, se è stato dato il via libera dell'Istat e del Contributore.

CAPITOLO 2

CASI DI CONTRIBUZIONE AL CATALOGO¹

2.1 Casi esemplificativi di contributo al Catalogo

Di seguito vengono riportati, a scopo esemplificativo, soltanto cinque delle 18 pubbliche amministrazioni che contribuiscono al Catalogo, per descrivere le diverse modalità di contribuzione, il diverso grado di autonomia nella realizzazione delle risorse semantiche e il supporto garantito dall'Istat, ma, soprattutto, per illustrare come i loro contributi abbiano arricchito la rete semantica del Catalogo, mettendo a disposizione anche di altre pubbliche amministrazioni le loro risorse semantiche.

Attualmente, le pubbliche amministrazioni hanno contribuito al Catalogo con 302 risorse semantiche, suddivise in: 73 ontologie, 155 vocabolari controllati e 74 schemi dati. Molte altre risorse sono in lavorazione e saranno disponibili a breve.

2.2 Risorse semantiche core dell'Istat

Nella progettazione del Catalogo, accanto allo sviluppo di risorse semantiche relative ai diversi domini di interesse coincidenti con le aree di competenza delle pubbliche amministrazioni che contribuiscono, è prevista la disponibilità di risorse semantiche relative a concetti di carattere generale e trasversale rispetto ai singoli domini: le ontologie *core*. La funzione delle ontologie *core* è quella di modellare concetti in risorse destinate a essere riusate dalle ontologie di dominio; i contributori interessati a descrivere il proprio dominio possono, così, fare diretto riferimento alle ontologie *core* anziché procedere ad autonoma modellazione di concetti come, ad esempio, le persone e i luoghi. La funzionalità delle ontologie *core* si valuta, dunque, nella misura in cui due diversi domini, collegandosi alla stessa ontologia *core* possono, a loro volta, collegarsi tra di loro (*Linked Open Data* - LOD). Così, le ontologie *core* svolgono un ruolo fondamentale nell'ecosistema dei LOD, in quanto offrono strutture concettuali standardizzate che permettono l'interoperabilità, la comprensione e il riutilizzo dei dati tra diversi *dataset* e domini.

Alcune risorse semantiche come le ontologie *core* della Persona (*Core Person Vocabulary* - CPV) e del Territorio (*Core Location Vocabulary* - CLV) sono state sviluppate dall'Istat, la prima per rappresentare i LOD del Censimento della popolazione del 2011, mentre la seconda, in collaborazione con Agenzia delle Entrate, è stata ampliata per comprendere la modellazione degli indirizzi in vista della pubblicazione dei dati dell'Archivio Nazionale dei Numeri Civici delle Strade Urbane (ANNCSU). Per la CLV, già nel 2017 è stata adottata un'assistenza da parte dell'Istat alla modellazione ontologica di un dominio in condivisione con Agenzia delle Entrate. Le due ontologie messe a disposizione nel progetto DAF (*Data and Analytics Framework*) si sono evolute nel tempo, arricchendosi di informazioni in funzione di esigenze di altri fruitori.

¹ Il Capitolo è stato redatto da Roberto Puglisi (paragrafi 2.3 e 2.4) e Roberta Radini (paragrafi 2.1, 2.2, 2.5 e 2.6).

Il caso dell'ontologia CLV per Agenzia delle Entrate è la dimostrazione di come l'Istat, da tempo, offre un servizio di affiancamento alla modellazione ontologica (un approccio di “*stewardship completa*” *ante litteram*), ma anche di come le ontologie condivise e messe a disposizione nella prima rete semantica, che era quella della rete *OntoPiA*, si evolvono con lo scopo di coprire esigenze anche di altri fruitori. L'aver creato concetti cardine come quello di *Person* presente nell'ontologia CPV, riusato in molte ontologie, rende i concetti di altri domini collegabili tra loro.

2.3 Risorse semantiche del Ministero dell'Interno: asset semantici dell'Anagrafe Nazionale della Popolazione Residente

Il supporto in *semantic stewardship* dell'Istat al Ministero dell'Interno nasce dalla stretta collaborazione già esistente tra i due enti nell'ambito dell'Anagrafe Nazionale della Popolazione Residente (ANPR), rispetto alla quale l'Istat e il Ministero dell'Interno condividono competenze in materia di vigilanza anagrafica, ciascuna per le proprie competenze, assegnate dal DPR n. 223 del 1989. Con particolare riferimento al sistema informativo dell'ANPR, dunque, la collaborazione si è sviluppata su tematiche ben note all'Istat e al Ministero dell'Interno. L'Istat, in questo caso, si è occupato dell'analisi concettuale prodromica alla modellazione e della modellazione stessa.

Il Ministero ha mostrato condivisione della metodologia qui descritta, chiedendo il supporto dell'Istat. Per quanto concerne l'analisi concettuale, si è trattato di ricostruire la semantica della Pubblica amministrazione relativa ai flussi informativi coinvolti nel dominio di interesse “anagrafico” del Ministero dell'Interno e di una modellazione ontologica dei concetti stessi; l'Istat ha applicato la metodologia descritta nel Capitolo 1, affidando al WP3 - Analisi giuridiche (Direzione centrale affari giuridici e legali) l'analisi dei concetti rilevanti e della relativa semantica della Pubblica amministrazione per consentire, poi, al WP2 - Definizione e sviluppo metodologie (Direzione centrale per la metodologia e il disegno dei processi statistici) la modellazione ontologica degli stessi, il tutto in collaborazione con gli esperti tematici (Direzione centrale delle statistiche demografiche e del censimento della popolazione) con esperienza nel campo di interesse anagrafico.

Con particolare riferimento all'ANPR, si è tenuto conto delle connessioni col Progetto UE del *Single Digital Gateway* (SDG)², il cui scopo è rendere disponibile un unico punto di accesso a informazioni e procedure online di diverse amministrazioni pubbliche europee, eliminando la necessità per gli utenti di visitare singolarmente ogni sito web nazionale; il SDG è, quindi, un esperimento che, attraverso la costituzione del *Once Only Technical System* (OOTS), punta direttamente all'attuazione del principio *una tantum* nella fruizione dei servizi amministrativi pubblici, una tra le tante utilità dell'interoperabilità tra i sistemi informativi. Il collegamento con lo sviluppo di risorse semantiche relative ai dati da interscambiare tramite *e-service* della PDND è costituito da due delle ventuno procedure amministrative coinvolte nell'implementazione del SDG.

Nel raccogliere la semantica dei dati di ANPR in *schema.gov.it* si è, dunque, riconosciuta una specifica rilevanza, inizialmente, ai dati contenuti nei certificati anagrafici di residenza e di nascita.

² Regolamento sull'interoperabilità semantica tra i sistemi informativi degli Stati membri su 21 procedure, tra cui residenza e nascita (Regolamento (UE) 2018/1724 del Parlamento europeo e del Consiglio, del 2 ottobre 2018, che istituisce uno sportello digitale unico per l'accesso a informazioni, procedure e servizi di assistenza e di risoluzione dei problemi e che modifica il regolamento (UE) n. 1024/2012: <https://eur-lex.europa.eu/eli/reg/2018/1724/oj?locale=IT>).

2.4 Risorse semantiche del Parlamento

In tema di sviluppo di risorse semantiche, Camera e Senato già da tempo usano proprie ontologie per l'accesso ai propri dati aperti, rappresentando, dunque, due realtà già esperte in tema di interoperabilità semantica. L'interesse di Camera e Senato rispetto alle funzionalità offerte dal Progetto NDC si appunta sulla capacità di armonizzazione delle risorse; infatti, primaria esigenza dei due contributori è quella di pervenire alla modellazione di un'unica ontologia del Parlamento al fine di annullare le disomogeneità tra l'ontologia del Senato e l'ontologia della Camera dettate dal loro sviluppo in autonomia. La *stewardship* dell'Istat è ispirata, dunque, non dall'assenza di competenze in tema di produzione di risorse semantiche, quanto piuttosto dalle possibilità di armonizzazione connesse alla progettazione di un catalogo unico delle risorse. L'ontologia del Parlamento mette in risalto la diversa utilità marginale che può produrre il Catalogo: interoperabilità semantica non solo per gli *e-service* della PDND, ma anche per gli *open data* che, integrati con le risorse raccolte in *schema.gov.it*, possono evolversi in *Linked Open Data*.

L'analisi del dominio di interesse da modellare comporta, dunque, lo studio delle ontologie attualmente pubblicate dalla Camera e dal Senato al fine di progettare un'unica ontologia di raccordo per tutte le risorse parlamentari. Nell'ambito della definizione concettuale prodromica alla modellazione dell'ontologia del Parlamento, l'applicazione della metodologia di analisi semantica descritta nel Capitolo 1 da parte del WP3 - Analisi giuridiche ha richiesto una preliminare ricostruzione del contesto normativo di riferimento. Le definizioni dei concetti sono, perciò, sviluppate in coerenza con le previsioni normative contenenti l'organizzazione dei due rami parlamentari, in particolare la Costituzione della Repubblica Italiana, il Regolamento Camera approvato il 18 febbraio 1971 con s.m.i. (ultimo aggiornamento deliberato il 30 novembre 2022) e il Regolamento del Senato approvato il 17 febbraio 1971 con s.m.i. (ultimo aggiornamento deliberato il 27 luglio 2022).

L'attuazione del Progetto si è svolta con sessioni di lavoro tra i referenti dell'Istat e quelli di Camera e Senato, attraverso un'interlocazione tecnica e metodologica costante. Tale collaborazione ha permesso lo sviluppo di risorse semantiche idonee ad armonizzare la modellizzazione dei dati esposti dal Parlamento con riferimento alle strutture e attività delle due Camere.

Il Senato e la Camera dei deputati hanno due diverse ontologie che utilizzano separatamente, attraverso soluzioni articolate di *quering* sui dati, per esporre i loro *Linked Open Data*. La nuova ontologia unica per l'intero Parlamento richiede, invece, di rieseguire il processo di *mapping* sui dati e di conseguenza l'estrazione di questi, con l'obiettivo di pervenire a un unico ambiente integrato e condiviso, guidato da una semantica comune.

Il percorso intrapreso è approdato alla modellazione, da parte del WP2 - Definizione e sviluppo metodologie, della struttura degli organi di Camera e Senato e, successivamente, alla illustrazione relativa al procedimento legislativo parlamentare, utilizzando l'approccio metodologico guidato dalla stretta collaborazione tra esperti giuridici e di modellazione ontologica.

2.5 Risorse semantiche di Inps e Inail

Il Catalogo si è arricchito con le risorse semantiche relative al dominio di interesse della previdenza sociale. L'Inps e l'Inail, avendo proceduto in autonomia alla modellazione di proprie risorse, hanno chiesto all'Istat supporto solo per il processo di *onboarding*, al fine di garantire l'interoperabilità tecnica utile all'interscambio informativo.

In particolare, viste le elevate competenze di modellazione ontologica dei due contributori, questi sono stati i primi a pubblicare sul Catalogo gli schemi dati. Da quest'anno Inps ha sfruttato la modellazione ontologica pubblicata sul Catalogo per aggiungere le annotazioni ontologiche anche agli schemi di API pubblicate su PDND.

La collaborazione con Inps, che ha fatto da apripista nelle attività di predisposizione delle informazioni necessarie a realizzare l'interoperabilità semantica degli *e-service*, ha consentito all'Istat di avere un caso d'uso concreto che è stato sfruttato per perfezionare il processo di definizione dell'annotazione semantica degli *e-service*, di predisposizione di indicatori di misurazione dell'interoperabilità semantica e di messa a punto di strumenti software per lo sviluppo degli schemi dati e degli schemi di API.

2.6 Risorse semantiche dell'Agenzia del Demanio

L'Agenzia del Demanio ha avviato un progetto di pubblicazione dei propri *e-service* sul portale PDND e, volendo realizzare l'interoperabilità semantica dei propri dati, ha richiesto il servizio di *stewardship* per la modellazione degli schemi delle API degli *e-service*.

Il modello adottato è quello di incontri con sessioni di lavoro tra i referenti dell'Istat e quelli dell'Agenzia del Demanio, prima per individuare tutti i concetti già disponibili sul Catalogo adeguati a modellare lo scambio di dati implementato con gli *e-service*, poi per la modellazione degli schemi dati e, successivamente, per la verifica di questi. Le sessioni di lavoro sono state condotte con modalità di formazione *on the job*, tanto che dopo poche riunioni gli esperti di Demanio sono diventati autonomi nell'annotazione ontologica degli schemi di API.

Questo lavoro ha dato avvio anche alla creazione di nuove risorse semantiche, come vocabolari controllati e un'ontologia per andare a coprire gli ambiti semantici specifici del loro dominio che non trovano modellazione in nessuna ontologia della rete semantica del Catalogo al momento disponibile.

In questo caso si è adottato un modello di "*stewardship* completa" simile a quella adottata per il progetto ANPR e per Camera e Senato (Parlamento).

CONCLUSIONI

Il Progetto Catalogo Nazionale Dati e le sinergie instaurate con il processo di produzione della Piattaforma Digitale Nazionale Dati (PDND) hanno avviato un percorso per la realizzazione dell'interoperabilità semantica dei dati, tassello di estrema rilevanza nel programma di digitalizzazione della Pubblica amministrazione che è stato finanziato con i fondi PNRR nell'ambito di quanto previsto dalla Missione M1. Come descritto, sono state sviluppate soluzioni software, come il portale *schema.gov.it*, a supporto della pubblicazione e della realizzazione dei prodotti semantici. In particolare, per realizzare gli schemi dati di *e-service* della PDND, che chiudono il cerchio tra la creazione della metadattazione della semantica e la realizzazione dell'interoperabilità semantica, ossia l'automazione dello scambio di dati, della lettura e dell'interpretazione dei dati da parte delle macchine, anche ai fini di un utilizzo di questi per alimentare i processi di intelligenza artificiale.

Soprattutto, possiamo dire che è stato definito e messo in esercizio un metodo e un processo di realizzazione dei prodotti semantici che rappresentano l'elemento chiave e innovativo dell'interoperabilità semantica.

Al di là degli obiettivi primari dell'interoperabilità semantica, la realizzazione delle risorse semantiche del Catalogo, poiché sono direttamente accessibili dalle macchine, possono essere sfruttate per alimentare una IA che può ragionare su una conoscenza certificata e su tematiche prettamente legate all'ordinamento amministrativo interno (*e-service* e decisioni dei *policy maker*), rispondendo quindi all'esigenza per l'IA generativa di poter essere "affidabile", consentire la "tracciabilità" delle fonti che hanno contribuito alle risposte, ma, soprattutto, "personalizzabile" su tematiche specifiche e garantire l'"indipendenza digitale".

In altri termini, mettere a disposizione dell'IA dati e metadati che rappresentano un "linguaggio giuridico amministrativo" caratterizzato dal sistema giuridico italiano, che è molto diverso dal *common law* anglo-americano. La tematica non riguarda soltanto problemi linguistici e interpretazione dei termini utilizzati, ma soprattutto la rappresentazione della specificità degli istituti giuridici che quei termini e dati descrivono.

La sfida di mettere a regime e alimentare la rete semantica che modella tutti i processi amministrativi e, soprattutto, a mantenerla e renderla sempre in linea con gli aggiornamenti normativi e procedurali amministrativi, deve essere gestita nel tempo con l'adozione di norme, l'individuazione di strutture competenti e l'aggiornamento continuo delle linee guida per i processi.

Questo richiede l'individuazione di strutture pubbliche competenti e la collaborazione di tutte le pubbliche amministrazioni che in una forma federata mettano a disposizione di tutti i loro contenuti. Quindi una gestione centralizzata e *super partes* che consenta la mediazione e l'armonizzazione dei contenuti secondo regole, tecniche standard e trasparenti. Ciò implica la necessità di un'adeguata individuazione delle figure professionali necessarie al perseguimento di tale obiettivo; ad esempio, per la produzione di risorse semantiche, sono richiesti esperti di modellazione ontologica e esperti analisti di definizione giuridica che, forse, non tutte le pubbliche amministrazioni sono al momento pronte a mettere in campo.

Il Progetto Catalogo Nazionale Dati ha un mandato che termina a giugno del 2026, ma la strada disegnata è ben delineata; i prodotti realizzati coprono solo una parte dei processi amministrativi, resta quindi da creare la giusta sinergia e organizzazione per proseguire al completamento e non solo alla gestione nel tempo di quanto introdotto dal Catalogo.

Giovanna Bellitti

Direttrice centrale affari giuridici e legali

Massimo Fedeli

Direttore del Dipartimento per lo sviluppo di metodi e tecnologie
per la produzione e diffusione dell'informazione statistica

RIFERIMENTI BIBLIOGRAFICI

- Agenzia per l'Italia Digitale - AgID. 2023. *Linee Guida sull'interoperabilità tecnica delle Pubbliche Amministrazioni*. Roma, Italia: AgID. https://www.agid.gov.it/sites/agid/files/2024-05/linee_guida_interoperabilit_tecnica_pa.pdf.
- American National Standards Institute - ANSI, and National Information Standards Organization - NISO. 2010. ANSI/NISO Z39-19:2005 (R2010). *Guidelines for the Construction, Format, and Management of Monolingual Controlled Vocabularies*. Baltimore, MD, U.S.: NISO.
- Batini, C. 2013. "Un'introduzione ai servizi di eGovernment". *Amministrare*, N. 1: 37-58.
- Bellitti, G., e C. Colasanti (a cura di). 2021. *Manuale sui principali adempimenti in materia di trattamento di dati personali: il caso dell'Istat*. Letture Statistiche - Metodi. Roma, Italia: Istat. <https://www.istat.it/produzione-editoriale/manuale-sui-principali-adempimenti-in-materia-di-trattamento-di-dati-personali-il-caso-dellistat/>.
- Bellitti, G., C. Colasanti, e M. Fedeli. 2023. "Piattaforma Digitale Nazionale Dati: come l'Istat realizzerà il National Data Catalog". *Agenda Digitale* (25 gennaio 2023). Milano, Italia: Network Digital 360. <https://www.agendadigitale.eu/cittadinanza-digitale/data-management/piattaforma-digitale-nazionale-dati-come-listat-realizzera-il-national-data-catalog/>.
- Bellitti, G., e M. Fedeli (a cura di). 2022. *Regole e strategie nel trattamento digitale e nella produzione dei dati quantitativi e qualitativi*. Letture Statistiche - Metodi. Roma, Italia: Istat. <https://www.istat.it/produzione-editoriale/regole-e-strategie-nel-trattamento-digitale-e-nella-produzione-dei-dati-quantitativi-e-qualitativi/>.
- Bertaccini, F., M. Prandi, S. Sintuzzi, e S. Togni. 2006. "Tra lessico naturale e lessici di specialità: la sinonimia". In Rombi, R., G. Cifoletti, F. Fusco, L. Innocente, e V. Orioles (a cura di). *Studi linguistici in onore di Roberto Gusmani*: 171-192. Alessandria, Italia: Edizioni dell'Orso.
- Copenhagen Economics. 2010. "The Economic Impact of a European Digital Single Market". *Informed Decisions* (Study commissioned by the European Policy Centre – EPC). Copenhagen, Denmark: Copenhagen Economics.
- Cordella, A., and A. Paletti. 2019. "Government as a platform, orchestration, and public value creation: The Italian case". *Government Information Quarterly*, Volume 36, Issue 4: Article 101409. <https://doi.org/10.1016/j.giq.2019.101409>.
- Cortellazzo, M. 1994. *Lingue speciali. La dimensione verticale*. Padova, Italia: Unipress.
- DigitalEurope. 2009. "A Transformational Agenda for the Digital Age". *White Paper*. Brussels, Belgium: DigitalEurope.
- Folino, A. 2013. "Tassonomie e thesauri". In Guarasci, R., e A. Folino (a cura di). *Documenti Digitali*: 387-444. Milano, Italia: ITER.
- Geisler, S., M.-E. Vidal, C. Cappiello, B. Farias Lóscio, A. Gal, M. Jarke, M. Lenzerini, P. Missier, B. Otto, E. Paja, B. Pernici, and J. Rehof. 2021. "Knowledge-Driven Data Ecosystems Toward Data Transparency". *Journal of Data and Information Quality*, Volume 14, Issue 1: 1-12.
- Haas, E.B. 2004. *Uniting of Europe. Political, Social, and Economic Forces, 1950-1957*. Notre Dame, IN, U.S.: University of Notre Dame Press.

- International Organization for Standardization - ISO. 2019. ISO 1087:2019. *Terminology work and terminology science - Vocabulary*. Geneva, Switzerland: ISO.
- International Organization for Standardization - ISO. 2013. ISO 25964-2:2013. *Information and documentation. Thesauri and interoperability with other vocabularies – Part 2: Interoperability with other vocabularies*. Geneva, Switzerland: ISO.
- Kontchakov, R., M. Rodríguez-Muro, and M. Zakharyashev. 2013. "Ontology-Based Data Access with Databases: A Short Course". In Rudolph, S., G. Gottlob, I. Horrocks, and F. van Harmelen (Eds.). *Reasoning Web. Semantic Technologies for Intelligent Data Access. 9th International Summer School 2013, Mannheim, Germany, July 30 - August 2, 2013. Proceedings*: 194-299. Berlin, Germany: Springer.
- Lindberg, L.N. 1963. *The Political Dynamics of European Economic Integration*. Redwood City, CA, U.S.: Stanford University Press.
- Lenzerini, M. 2023. "Data Modeling and AI: From Semantic Networks to Knowledge Graphs". Keynote presented at DATA 2023 – 12th International Conference on Data Science, Technology and Applications. Roma, Italy, 11-13 July 2023.
- O'Brien, R., and the Economist Intelligence Unit. 2004. "E-government in Central Europe. Rethinking Public Administration". *White Paper*. London, UK: Economist Intelligence Unit. https://graphics.eiu.com/files/ad_pdfs/central_europe_egov.pdf.
- Palfrey, J.G., and U. Gasser. 2012. *Interop: The Promise and Perils of Highly Interconnected Systems*. New York, NY, U.S.: Basic Books.
- Polizzi, M., e S. Letardi. 2023. "Il Catalogo Nazionale Dati: far parlare i dati per innovare i servizi". Paper presentato alla XLIV Conferenza Scientifica Annuale dell'Associazione Italiana di Scienze Regionali - AISRe. *Europa e Mediterraneo tra transizioni e conflitti. Opportunità e rischi per regioni e territori*. Napoli, Italia, 6-8 settembre 2023.
- Puglisi, R. 2022. "Profili normativi del trattamento dei dati e del processo di produzione quantitativa e qualitativa statistica. Il rapporto tra norma e tecnologia. Le regole giuridiche, tecniche e metodologiche". In Bellitti, G., e M. Fedeli (a cura di). *Regole e strategie nel trattamento digitale e nella produzione dei dati quantitativi e qualitativi*: 27-43. *Lecture Statistiche - Metodi*. Roma, Italia: Istat. <https://www.istat.it/produzione-editoriale/regole-e-strategie-nel-trattamento-digitale-e-nella-produzione-dei-dati-quantitativi-e-qualitativi/>.
- Rosamond, B. 2003. "(European) Integration Theory, EU Studies and the Sociology of Knowledge". Paper presented at the 8th Biennial Conference of the European Union Studies Association. Nashville, TN, U.S., 27-29 March 2003. <https://aei.pitt.edu/2915/1/152.pdf>.
- Scholl, H.J., H. Kubicek, and R. Cimander. 2011. "Interoperability, Enterprise Architectures, and IT Governance in Government". In Janssen, M., H.J. Scholl, M.A. Wimmer, and Yao-hua Tan (Eds.). *Electronic Government. 10th International Conference, EGOV 2011, Delft, The Netherlands, August 29 - September 1, 2011, Proceedings*: 345-354. Berlin, Germany: Springer.
- Sorrentino, F. 2008. "Le fonti del diritto amministrativo". *Trattato di diritto amministrativo* (diretto da Giuseppe Santaniero), Volume XXXV. Padova, Italia: CEDAM.

- United Nations Economic Commission for Europe - UNECE (on behalf of the international statistical community). 2019. *Generic Statistical Business Process Model – GSBPM (Version 5.1, January 2019)*. Geneva, Switzerland: UNECE. <https://unece.org/statistics/documents/2019/01/standards/gsbpm-v51>.
- United Nations Statistical Commission and Economic Commission for Europe. 2000. "Terminology on Statistical Metadata". Conference of European Statisticians, *Statistical Standards and Studies*, N. 53. Geneva, Switzerland: United Nations Publication. <https://digitallibrary.un.org/record/442455?v=pdf>.
- van Staden, S., and J. Mbale. 2012. "The Information Systems Interoperability Maturity Model (ISIMM): Towards Standardizing Technical Interoperability and Assessment within Government". *International Journal of Information Engineering and Electronic Business*, Volume 4, N. 5: 36-41.
- Vicente, S., M. Pérez, X. García, A. Gimeno, and J. Naval. 2005. "eGovernment interoperability on a semantically driven world". In Benamou, N. (Ed.). *Pre Proceedings – Draft of selected papers of the eGov-Interop'05 Conference*. Geneva, Switzerland, 23 - 24 February 2005. <https://interop-esa05.unige.ch/INTEROP/Proceedings/eGovScientific/eGovInterop%20pre%20proceedings.pdf>.

