

# TRASFORMAZIONE DIGITALE DELLA PUBBLICA AMMINISTRAZIONE

METODI PER L'INTEROPERABILITÀ  
PER LO SVILUPPO DI *E-SERVICE*





## **TRASFORMAZIONE DIGITALE DELLA PUBBLICA AMMINISTRAZIONE** METODI PER L'INTEROPERABILITÀ PER LO SVILUPPO DI *E-SERVICE*

A cura di Giovanna Bellitti e Massimo Fedeli.

Coordinamento delle attività tecnico scientifiche a supporto della realizzazione dei contenuti del volume: Daniela Lauriello.

Supporto alla cura redazionale: Lara Parisella.

Attività editoriali: Nadia Mignolli (coordinamento), Alfredina Della Branca, Marco Farinacci, Alessandro Franzò e Manuela Marrone.

Responsabile per la grafica: Sofia Barletta.

ISBN 978-88-458-2125-7

© 2023

Istituto nazionale di statistica  
Via Cesare Balbo, 16 - Roma



Salvo diversa indicazione, tutti i contenuti pubblicati sono soggetti alla licenza Creative Commons - Attribuzione - versione 3.0.  
<https://creativecommons.org/licenses/by/3.0/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.
<b>Introduzione</b>	5
<b>1. Il metodo per l'armonizzazione semantica dei dati</b>	7
1.1 Interoperabilità della Pubblica Amministrazione e competenze nel trattamento dei dati	7
1.2 Il Progetto per l'interoperabilità semantica	9
1.3 L'attuazione del Catalogo	10
1.3.1 Il censimento delle banche dati pubbliche	10
1.3.2 Risultati del Censimento permanente delle Istituzioni pubbliche	11
1.3.3 Formazione e comunicazione	13
<b>2. La strategia dati per la Pubblica Amministrazione</b>	15
2.1 Digitalizzazione e dati	15
2.2 <i>Government as a Platform</i> (GaaP)	16
2.3 L'interscambio dei dati amministrativi	17
2.4 Classificazioni dei dati dell'azione amministrativa ed <i>e-service</i>	19
2.5 Il futuribile: AI generativa al servizio della Pubblica Amministrazione	20
<b>3. L'interoperabilità semantica della Pubblica Amministrazione</b>	23
3.1 I livelli di interoperabilità	23
3.2 Il livello europeo di interoperabilità	25
3.3 Il livello nazionale di interoperabilità	28
3.4 La Piattaforma Digitale Nazionale Dati e il Catalogo Nazionale dei Dati della Pubblica Amministrazione	29
3.5 La semantica della Pubblica Amministrazione	31
<b>4. La semantica e le ontologie per il Catalogo Nazionale dei Dati pubblici: aspetti metodologici</b>	35
4.1 Modellazione Semantica: elementi di base	36
4.1.1 Il ruolo dei termini nel Catalogo Nazionale della Semantica dei Dati	36
4.1.2 Il termine come unità minima di una raccolta terminologica	37
4.1.3 Relazioni semantiche tra termini	38
4.2 Predisposizione degli <i>asset</i> semantici	42
4.2.1 Il ciclo di vita dei prodotti semantici	43
4.2.2 La predisposizione per la pubblicazione dei prodotti semantici	44



	Pag.
<b>5. La piattaforma del Catalogo</b>	47
5.1 Obiettivi dell'infrastruttura	47
5.2 Schema.gov.it	47
5.3 Il processo di <i>harvesting</i>	50
5.4 La qualità dei dati	51
5.5 Profili di sicurezza	52
<b>Riferimenti bibliografici</b>	53

---

## INTRODUZIONE

Lo sviluppo dei metodi di trattamento dei dati segue un percorso inevitabilmente condizionato dall'evoluzione delle tecnologie dell'informazione e della comunicazione, strumenti in grado di imprimere eccezionali accelerazioni nell'elaborazione delle informazioni.

Se, in una fase iniziale, le potenzialità delle nuove tecnologie digitali si esprimono sotto un profilo prevalentemente quantitativo (tramite l'accrescimento esponenziale dei dati elaborati), successivamente, la loro mera capacità di registrare e trasmettere i dati si evolve nella possibilità di processarli estraendone la relativa informazione. Progressivamente, quindi, diviene sempre più chiaro come la digitalizzazione possa influire, se adeguatamente sfruttata, anche sulla qualità del trattamento. Si espande, così, la rilevanza del ricorso alle macchine nel processo di trattamento del dato che, partendo dai dati "grezzi" relativi ai diversi fenomeni sociali, demografici ed economici, conduce all'elaborazione dell'informazione.

Le stesse dinamiche si rispecchiano nell'evoluzione dell'informazione statistica ufficiale caratterizzata da un costante interesse per lo sviluppo di tecniche di trattamento finalizzate al miglioramento della qualità e della quantità della produzione informativa. In tale prospettiva, del resto, si pone l'obiettivo perseguito dall'Istituto Nazionale di Statistica – Istat di realizzare l'unità di indirizzo, l'omogeneità organizzativa e la razionalizzazione dei flussi informativi. Se ciò è avvenuto, in un primo momento, attraverso la standardizzazione della modulistica amministrativa finalizzata al riutilizzo per fini statistici dei dati amministrativi, gli ultimi approdi sono rappresentati dal coinvolgimento dell'Istat nel processo di armonizzazione classificatoria utile all'interoperabilità semantica dei sistemi informativi pubblici. È, così, che le attività istituzionali per la produzione statistica (rimodellate, nel 2010, dal legislatore con l'affidamento all'Istat del compito di "*definire i metodi e i formati*" per lo scambio dell'informazione statistica e finanziaria, nonché del coordinamento dei sistemi informativi delle pubbliche amministrazioni per la raccolta di dati per fini statistici) sono considerate anche nella Strategia Nazionale Dati; per l'implementazione della Piattaforma Digitale Nazionale Dati (PDND) – Progetto ispirato al principio *once only* grazie all'interoperabilità dei sistemi informativi e delle basi di dati pubbliche – è prevista la realizzazione di un Catalogo che raccolga le risorse semantiche relative ai dati pubblici. L'armonizzazione compiuta dall'Istat nel trattamento dei dati amministrativi mediante un'adeguata attività classificatoria costituisce, in questa prospettiva, un prodromo delle moderne tecniche di interoperabilità semantica tra sistemi informativi. Grazie all'interoperabilità tra i sistemi informativi, è possibile ottenere un miglioramento dei servizi offerti dalle pubbliche amministrazioni ai cittadini in funzione della maggiore efficienza dell'azione amministrativa e del corrispettivo minore carico burocratico nei confronti degli interessati.

In tale contesto, la Presidenza del Consiglio dei Ministri – Dipartimento per la trasformazione digitale ha individuato l'Istat quale soggetto attuatore del Progetto PNRR "Catalogo Nazionale della Semantica dei Dati" (*National Data Catalog* – NDC), finalizzato all'agevolazione dell'interoperabilità semantica della Pubblica Amministrazione fornendo una descrizione dei dati pubblici mediante modelli di dati, vocabolari e schemi di dati. Per il conseguimento dell'obiettivo, l'Istat ha coniugato le esigenze multidisciplinari tipiche della digitalizzazione impiegando competenze informatiche, metodologiche, tematiche e giuridiche per rendere disponibile una raccolta di risorse

semantiche idonee ad abilitare l'interscambio digitale delle informazioni tra le pubbliche amministrazioni. Così, per assicurare all'interoperabilità semantica della P.A. lo stesso rigore e gli stessi standard qualitativi della produzione statistica, è stata predisposta una metodologia di sviluppo del Catalogo fondata sulla stretta collaborazione tra le diverse professionalità coinvolte al fine di valorizzarne i rispettivi contributi.

Il presente lavoro descrive la metodologia applicata dall'Istat per la progettazione e l'attuazione del Catalogo Nazionale della Semantica dei Dati offrendo, al contempo, una panoramica sul contesto e sui principi di riferimento per la digitalizzazione della P.A.; si completa, così, l'analisi iniziata con il *“Manuale sui principali adempimenti in materia di trattamento di dati personali: il caso dell'Istat”* (Bellitti, G., e C. Colasanti, 2021) e proseguita con *“Regole e strategie nel trattamento digitale e nella produzione dei dati quantitativi e qualitativi”* (Bellitti, G., e M. Fedeli, 2022).

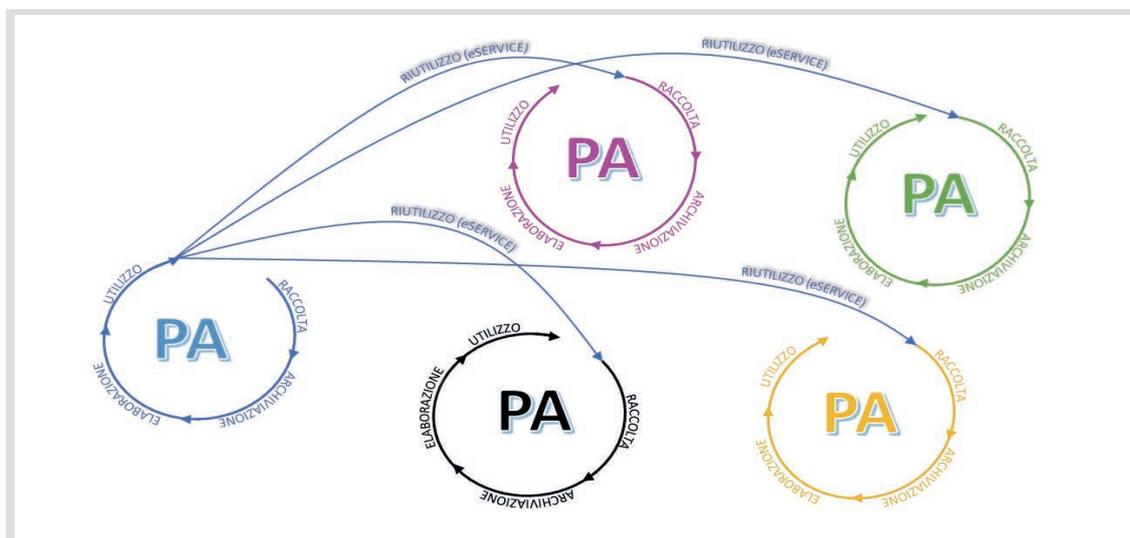
Il Presidente dell'Istat  
Francesco Maria Chelli

## 1. IL METODO PER L'ARMONIZZAZIONE SEMANTICA DEI DATI<sup>1</sup>

### 1.1 Interoperabilità della Pubblica Amministrazione e competenze nel trattamento dei dati

Il progresso della trasformazione digitale della società è ormai assestato sull'esigenza di fare evolvere il trattamento dei dati puntando a un ciclo di vita dell'informazione arricchito dell'ulteriore fase del riutilizzo; non più, dunque, dati destinati a vivere esclusivamente nel patrimonio informativo di chi li ha prodotti, ma deputati a entrare nel trattamento da parte di ulteriori utilizzatori, innescando, così, un interscambio virtuoso in grado di aumentare le capacità di approvvigionamento informativo da parte dei singoli partecipanti. Nell'ambito delle pubbliche amministrazioni, una simile prospettiva si concretizza con la programmazione di *e-service* finalizzati, appunto, al riutilizzo dei dati amministrativi; *e-service* che, dunque, assumono la veste di strumento di collegamento informativo tra i diversi enti. La predisposizione di piattaforme appositamente dedicate agli *e-service* è, così, la strada da percorrere per integrare le fasi della raccolta, dell'archiviazione, dell'elaborazione e dell'utilizzo dei dati con quella del loro riutilizzo.

Figura 1.1 - Ciclo di vita dell'informazione nell'era digitale



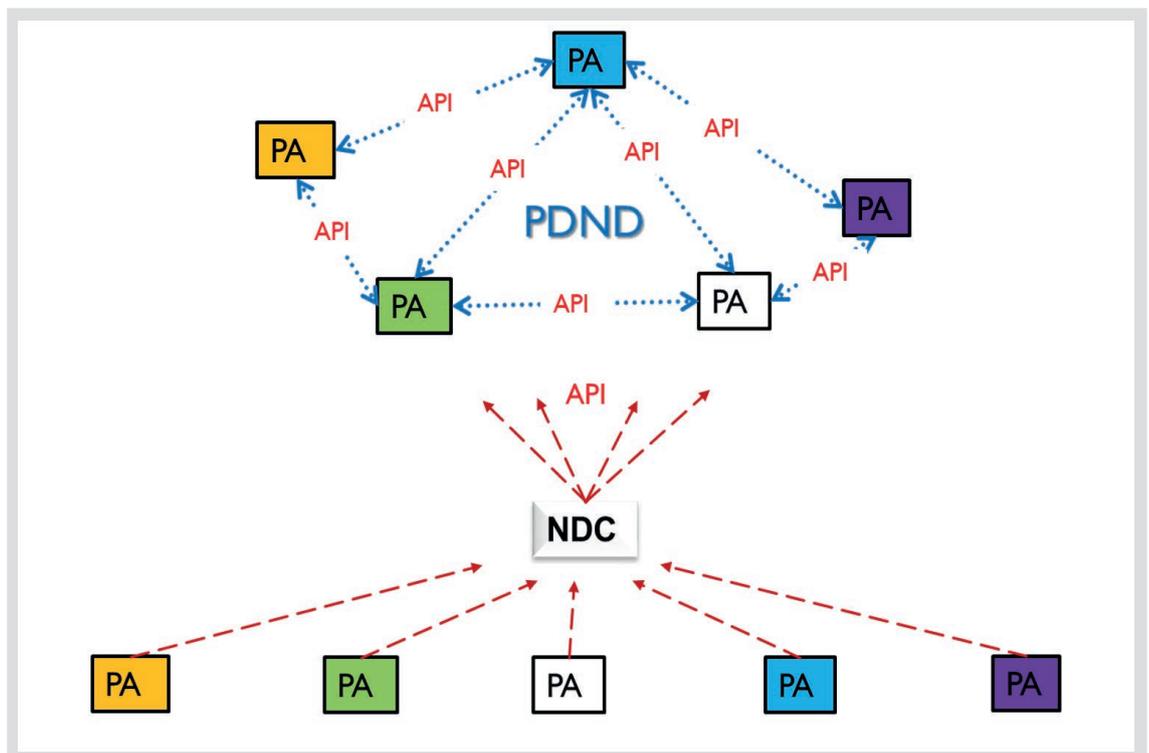
L'obiettivo è l'interoperabilità tra le banche dati pubbliche da perseguire nel contesto di un modello fondato sullo sviluppo di pubblici servizi erogati digitalmente (*e-service*), con lo scopo di creare ecosistemi in grado di trarre elementi di miglioramento dall'interazione con i propri utenti (cfr. paragrafo 2.2). Infatti, se simili forme di *eGovernment* si fondano sulla circolazione digitale delle informazioni tra diversi soggetti, requisito essenziale per la sua attivazione è la possibilità per i sistemi informativi pubblici di comunicare automaticamente per lo scambio di informazioni.

<sup>1</sup> Di Giovanna Bellitti e Massimo Fedeli.

È per tali ragioni che, nei tentativi di incremento della disponibilità dei dati, di rafforzamento della fiducia nella condivisione dei dati e di superamento degli ostacoli tecnici al riutilizzo dei dati, emerge l'esigenza di strutturare un sistema di *data governance* idoneo a guidare, a livello europeo, l'innovazione in materia di dati. Alla luce di ciò, la necessaria considerazione degli organismi con competenze specifiche come gli istituti nazionali di statistica, indicata dal legislatore europeo (*considerando 53, Data Governance Act*), sintetizza l'importanza del coinvolgimento, nel percorso verso forme sempre più evolute di trattamento digitale dei dati, di attori esperti nel trattamento dei dati per finalità di produzione statistica.

In tale contesto, si esprime l'individuazione dell'Istat quale soggetto attuatore del Progetto del Catalogo Nazionale della Semantica dei Dati (*National Data Catalog – NDC*) finalizzato alla raccolta e all'armonizzazione di schemi di dati, ontologie e vocabolari di base a sostegno della creazione e della progettazione di *e-service*. Il NDC punta ad agevolare lo sviluppo di API (*Application Programming Interface*) per l'interscambio dei dati nella Piattaforma Digitale Nazionale Dati (PDND), rendendo trasparente la struttura dei dati delle pubbliche amministrazioni in funzione dell'interoperabilità dei rispettivi sistemi informativi. Se, infatti, la PDND è lo strumento individuato per l'attuazione del modello GaaP – *Government as a Platform* (cfr. Capitolo 2), ispirato all'implementazione di piattaforme digitali generative di ecosistemi complessi in grado di trarre le proprie dinamiche evolutive dall'interazione degli utenti (Bellitti, G., *et al.*, 2023), il Catalogo progettato dall'Istat punta a rafforzare la consapevolezza su consistenza e struttura del patrimonio informativo pubblico in modo da abilitare il pieno funzionamento della PDND.

Figura 1.2 - Catalogo Nazionale della Semantica dei Dati (NDC) e Piattaforma Digitale Nazionale Dati (PDND)



## 1. Il metodo per l'armonizzazione semantica dei dati

---

### 1.2 Il Progetto per l'interoperabilità semantica

“*Interoperability enables sharing information and knowledge and exchanging data. Different departments and remote institutions involved in the eGovernment will be able to work in a transparent manner*” (Vicente, S., et al., 2005). L'interscambio dei dati tra le pubbliche amministrazioni comporta, dunque, il perseguimento dell'interoperabilità semantica dei sistemi informativi della Pubblica Amministrazione (P.A.) in modo da assicurare che tutti i soggetti partecipanti allo scambio informativo concordino sull'attribuire a un determinato dato un unico significato.

In primo luogo, per la progettazione del Catalogo l'Istat ha elaborato il metodo qui descritto, che si sviluppa su tre diversi piani strettamente connessi: l'armonizzazione semantica (cfr. Capitolo 3); l'applicazione di tecniche di rappresentazione concettuale mediante ontologie, schemi di dati e vocabolari controllati (cfr. Capitolo 4); la progettazione dell'infrastruttura informatica per raccogliere le risorse semantiche relative ai domini delle singole amministrazioni ([schema.gov.it](http://schema.gov.it), cfr. Capitolo 5).

La realizzazione del Catalogo comporta, dunque, un approccio multidisciplinare basato sulla permanente collaborazione tra le professionalità di carattere informatico, metodologico, giuridico e tematico.

L'armonizzazione semantica (cfr. Capitolo 3) è un requisito fondamentale individuato dall'Istat per assicurare alla descrizione dei dati della P.A. la consistenza e la qualità necessarie per il perseguimento dell'interoperabilità; viceversa, una raccolta di risorse semantiche priva di uniformità concettuale (con  $n$  amministrazioni che, potenzialmente, possono attribuire  $n$  significati a un determinato concetto) è esposta al rischio di riproporre i limiti del modello a *silos* da cui cerca di rifuggire la digitalizzazione (cfr. paragrafo 2.2) e di mostrarsi, dunque, alla prova dei fatti inadeguata allo scopo di impiegare applicativi in grado di elaborare e “interpretare” automaticamente i dati. A tale fine, occorre procedere con l'analisi concettuale dei domini di interesse alla luce delle competenze classificatorie dell'Istat (art. 15, lett. e), D.Lgs. n. 322/1989). Considerato l'ambito semantico da prendere in considerazione (settori in cui si esplica l'attività delle pubbliche amministrazioni), l'Istat ha assunto come parametro i concetti e le classificazioni giuridiche ogniqualvolta i dati da descrivere abbiano un riferimento normativo; in questi casi, infatti, è il legislatore a fornire una definizione concettuale a valenza generale. In tale scenario, l'Istat accompagna l'analisi giuridica con le competenze tematiche acquisite nel trattamento statistico dei dati amministrativi per l'analisi dei flussi informativi scaturenti dalle procedure amministrative e utili allo svolgimento di ulteriori procedure amministrative.

I risultati dell'analisi concettuale danno impulso alla modellazione ontologica e al disegno di schemi di dati (cfr. Capitolo 4) con l'impiego di tecniche in grado di produrre risorse semantiche *machine readable*, che fungono da strumento agevolatore dell'interoperabilità tra i sistemi informativi; tali metodi evoluti di rappresentazione concettuale consentono, infatti, una via d'accesso allo sviluppo di API.

Le fasi dell'analisi concettuale, effettuata da parte di esperti tematici (nel caso del Progetto del Catalogo sono necessarie professionalità con specifiche competenze giuridiche nell'analisi delle procedure amministrative; cfr. paragrafo 3.5), e della modellazione ontologica dei relativi risultati, eseguita da parte di esperti informatici specializzati in modelli di dati *machine readable*, si svolgono all'unisono onde consentire che la conoscenza fornita dai primi venga adeguatamente recepita nella produzione delle risorse semantiche.

L'infrastruttura informatica ([schema.gov.it](https://schema.gov.it), cfr. Capitolo 5) è caratterizzata da un'architettura funzionale deputata, come specificato meglio nel prosieguo, all'*harvesting* delle risorse semantiche pubblicate dalle pubbliche amministrazioni su propri *repository* esterni rispetto al portale [schema.gov.it](https://schema.gov.it); tale tecnica consente di esporre contenuti aggiornati quotidianamente, evitando, così, il rischio di fare riferimento a strutture di dati e metadati non più mappabili sui dati effettivamente detenuti dall'ente titolare di volta in volta coinvolto.

### 1.3 L'attuazione del Catalogo

L'Istat provvede all'implementazione del Catalogo fornendo, laddove richiesto, supporto tecnico e metodologico alle pubbliche amministrazioni nella definizione di ontologie, vocabolari controllati e schemi di dati relativi ai propri *dataset*. Tale supporto si concretizza in due diverse modalità di intervento a seconda delle esigenze manifestate dall'amministrazione interessata a pubblicare sul Catalogo. La prima si esplica attraverso l'attivazione di una *stewardship* per l'analisi concettuale del dominio di interesse e la relativa modellazione ontologica, oltre che la raccolta su [schema.gov.it](https://schema.gov.it) delle risorse semantiche sviluppate; la seconda modalità prevede la raccolta delle risorse semantiche sviluppate autonomamente dall'ente titolare del dominio di interesse garantendo, comunque, il supporto necessario per il rispetto dei requisiti tecnici di pubblicazione tramite *harvesting*.

Oltre agli aspetti realizzativi, il perseguimento degli obiettivi di attuazione del Catalogo previsti dall'accordo tra il Dipartimento della trasformazione digitale della Presidenza del Consiglio dei Ministri e l'Istat comporta ulteriori attività: il censimento delle banche dati delle pubbliche amministrazioni (PP.AA.), la formazione e la comunicazione.

#### 1.3.1 Il censimento delle banche dati pubbliche<sup>2</sup>

Al fine di acquisire una visione sul patrimonio informativo pubblico potenzialmente coinvolto nel Progetto NDC, l'Istat ha condotto un censimento delle banche dati destinate a divenire interoperabili per l'erogazione di *e-service*. A tale riguardo, nell'ultima edizione del Censimento permanente delle Istituzioni pubbliche (2023) è stata inserita una specifica sezione con un quesito per acquisire informazioni sul grado di interoperabilità dei rispettivi sistemi informativi. Obiettivo di tale Censimento è conoscere lo stato dell'arte dell'interoperabilità delle banche dati, in modo da organizzare azioni per l'estensione dei contenuti del Catalogo Nazionale della Semantica dei Dati. Con i risultati dei dati censuari si tenterà, dunque, di comprendere meglio la situazione attuale, favorendo l'intervento di politiche di innovazione finalizzate alla semplificazione delle procedure di accesso ai servizi delle pubbliche amministrazioni, in ossequio al principio *once only*.

Nel questionario viene chiesto se l'unità istituzionale rispondente detenga banche dati già rese interoperabili o che potrebbero essere rese interoperabili in quanto utili per lo svolgimento di compiti istituzionali di altre pubbliche amministrazioni.

Per quanto qui di interesse, i risultati devono essere letti, evidentemente, alla luce delle dinamiche che (come accennato nel paragrafo 1.1) guidano il perseguimento dell'interoperabilità: l'implementazione di una piattaforma di interscambio (la PDND) alla quale le pubbliche amministrazioni possono aderire sia come soggetti "erogatori" di servizi, sia come "fruitori".

<sup>2</sup> Con studi e ricerche e contributo di Paolo Nicolai.

## 1. Il metodo per l'armonizzazione semantica dei dati

Figura 1.3 - Quesito inserito nel questionario del Censimento permanente delle Istituzioni pubbliche (ed. 2023)

● **INTEROPERABILITÀ E PIATTAFORMA DIGITALE NAZIONALE DATI (PDND)**

**6.7 Ad oggi l'unità istituzionale detiene banche dati che sono state rese interoperabili <sup>a</sup> ?**

1  Sì  
2  No

**6.8 Ad oggi l'unità istituzionale detiene banche dati che potrebbero essere rese interoperabili in quanto utili per lo svolgimento di compiti istituzionali di altre Pubbliche Amministrazioni?**

1  Sì  
2  No

*Se l'unità istituzionale detiene banche dati (dom. 6.7 o 6.8 = 1)*

**6.9 Riportare l'elenco delle banche dati dell'unità istituzionale già rese o da rendere interoperabili**

Denominazione estesa della banca dati	Adesione alla PDND	Ufficio responsabile
<input type="text"/>	1 <input type="checkbox"/> Sì 2 <input type="checkbox"/> No, ma è in previsione 3 <input type="checkbox"/> No e non è in previsione	<input type="text"/>
<input type="text"/>	1 <input type="checkbox"/> Sì 2 <input type="checkbox"/> No, ma è in previsione 3 <input type="checkbox"/> No e non è in previsione	<input type="text"/>
<input type="text"/>	1 <input type="checkbox"/> Sì 2 <input type="checkbox"/> No, ma è in previsione 3 <input type="checkbox"/> No e non è in previsione	<input type="text"/>

<sup>a</sup> L'interoperabilità dei sistemi informativi e delle basi di dati delle Pubbliche Amministrazioni, da realizzarsi anche con l'adesione di queste ultime alla Piattaforma Digitale Nazionale Dati (PDND), ha come obiettivo lo sfruttamento del patrimonio informativo pubblico grazie a un interscambio di dati più efficiente e veloce.

La lettura dei dati del Censimento va effettuata, quindi, tenendo conto di ciò e delle categorie di appartenenza delle unità istituzionali rispondenti; con riferimento a quest'ultimo aspetto, ad esempio, se l'elevato numero dei comuni comporta un'alta incidenza quantitativa degli stessi sul tasso di interoperabilità, da un punto di vista meramente semantico, le relative banche dati sono riferibili (presuntivamente) al medesimo dominio di interesse (anagrafe e stato civile) che confluiscono nell'Anagrafe Nazionale della Popolazione Residente (ANPR). Stesso discorso per gli altri gruppi di enti che possono essere ricondotti sotto il medesimo dominio di interesse. Ciò considerato, il Censimento fornisce comunque dati che consentono un'analisi statistica a più ampio raggio, estesi anche ad aspetti diversi dalla semantica.

### 1.3.2 Risultati del Censimento permanente delle Istituzioni pubbliche<sup>3</sup>

In base ai primi risultati provvisori dell'edizione 2023 del Censimento permanente delle Istituzioni pubbliche, le istituzioni che detengono banche dati che sono state rese interoperabili o che potrebbero essere rese interoperabili sono 3.407, pari al 28,5 per cento delle

3 Fonte dei dati: "Censimento permanente delle istituzioni pubbliche" (<https://www.istat.it/it/censimenti-permanenti/istituzioni-pubbliche>). Istat - Dipartimento per la produzione statistica (Dips), Direttrice: Monica Pratesi; Responsabile, Direttore centrale per le statistiche economiche: Fabio Massimo Rapiti; Responsabile, Direttore centrale per la raccolta dati: Claudio Ceccarelli. Analisi a cura di Elisa Berntsen.

unità rispondenti. Una amministrazione su sei detiene banche dati che sono state rese interoperabili e quasi una su quattro detiene banche dati che potrebbero essere rese interoperabili. La maggiore propensione all'interazione digitale tra pubbliche amministrazioni, cittadini e imprese si registra presso le amministrazioni dello Stato e organi costituzionali o a rilevanza costituzionale, con il 40 per cento di istituzioni che ha banche dati già rese interoperabili e il 48,6 per cento che ha banche dati che potrebbero essere rese interoperabili in quanto utili per lo svolgimento di compiti istituzionali di altre pubbliche amministrazioni. A seguire Regioni (29,3 per cento e 39 per cento) e Università (28,2 per cento e 63,4 per cento). Di contro, la quota più elevata di istituzioni che non detengono banche dati già rese o da rendere interoperabili si registra presso le Comunità montane e unioni di comuni (87,2 per cento), presso gli Enti pubblici non economici (84,9 per cento) e presso istituzioni appartenenti ad Altra forma giuridica (90,3 per cento).

Le banche dati delle unità istituzionali rispondenti oggetto di analisi del Censimento sono oltre 5 mila. La maggiore adesione all'infrastruttura tecnologica che abilita l'interoperabilità dei sistemi informativi e delle basi di dati delle pubbliche amministrazioni riguarda le banche dati dei Comuni e a seguire quelle degli Enti pubblici non economici e delle Aziende o Enti del Servizio Sanitario Nazionale.

**Tavola 1.1 - Primi risultati provvisori dell'edizione 2023 del Censimento permanente delle Istituzioni pubbliche**  
Istituzioni che detengono banche dati che sono state rese o potrebbero essere rese interoperabili per forma giuridica e classe dimensionale. Anno 2022 (valori percentuali)

FORMA GIURIDICA CLASSE DIMENSIONALE DIP	Totale Istituzioni rispondenti		Istituzioni che detengono banche dati che sono state rese interoperabili		Istituzioni che detengono banche dati che potrebbero essere rese interoperabili		Istituzioni che non detengono banche dati		Numero banche dati già rese o da rendere interoperabili	Di cui banche dati che aderiscono alla PDND			Di cui banche dati che non aderiscono alla PDND ma prevedono di aderire		
	v.a.	%	v.a.	%	v.a.	%	v.a.	%		v.a.	%	v.a.	%	v.a.	%
<b>FORMA GIURIDICA</b>															
Amministrazione dello stato e organo costituzionale o a rilevanza costituzionale	35	40,0	14	40,0	17	48,6	15	42,9	63	5	7,9	28	44,4	30	47,6
Regione (Giunta e consiglio regionale)	41	29,3	12	29,3	16	39,0	22	53,7	80	8	10,0	53	66,3	19	23,8
Provincia e città metropolitana	104	11,5	12	11,5	24	23,1	76	73,1	54	6	11,1	22	40,7	26	48,1
Comune	7.408	20,2	1.497	20,2	2.329	31,4	4.732	63,9	4.010	1.704	42,5	1.826	45,5	480	12,0
Comunità montane e unione dei comuni	467	3,6	17	3,6	53	11,3	407	87,2	78	10	12,8	34	43,6	34	43,6
Azienda o ente del servizio sanitario nazionale	198	16,7	33	16,7	40	20,2	149	75,3	79	21	26,6	21	26,6	37	46,8
Università pubblica	71	28,2	20	28,2	45	63,4	22	31,0	85	15	17,6	59	69,4	11	12,9
Enti pubblici non economici	2.833	8,9	252	8,9	361	12,7	2.405	84,9	685	185	27,0	217	31,7	283	41,3
Altra forma giuridica	800	4,6	37	4,6	68	8,5	722	90,3	114	18	15,8	28	24,6	68	59,6
<b>Totale complessivo</b>	<b>11.957</b>	<b>15,8</b>	<b>1.894</b>	<b>15,8</b>	<b>2.953</b>	<b>24,7</b>	<b>8.550</b>	<b>71,5</b>	<b>5.248</b>	<b>1.972</b>	<b>37,6</b>	<b>2.288</b>	<b>43,6</b>	<b>988</b>	<b>18,8</b>
<b>CLASSE DIMENSIONALE (DIPENDENTI CensIP 2020)</b>															
0-9	5.585	12,6	703	12,6	983	17,6	4.401	78,8	1.438	552	38,4	572	39,8	314	21,8
da 10 a 49	4.046	18,5	748	18,5	1.148	28,4	2.729	67,4	1.956	819	41,9	835	42,7	302	15,4
50-249	1.585	17,9	283	17,9	556	35,1	973	61,4	1.127	437	38,8	492	43,7	198	17,6
250-999	311	22,2	69	22,2	133	42,8	166	53,4	323	74	22,9	165	51,1	84	26,0
1000-24999	294	24,8	73	24,8	115	39,1	168	57,1	354	75	21,2	198	55,9	81	22,9
25000 e oltre nuove	5	60,0	3	60,0	4	80,0	1	20,0	16	6	37,5	10	62,5	0,0	
<b>Totale</b>	<b>11.957</b>	<b>15,8</b>	<b>1.894</b>	<b>15,8</b>	<b>2.953</b>	<b>24,7</b>	<b>8.550</b>	<b>71,5</b>	<b>5.248</b>	<b>1.972</b>	<b>37,6</b>	<b>2.288</b>	<b>43,6</b>	<b>988</b>	<b>18,8</b>

## 1. Il metodo per l'armonizzazione semantica dei dati

---

I soggetti che ancora non aderiscono alla PDND ma prevedono di aderirvi sono poco meno della metà (43,6 per cento). La quota sale al 69,4 per cento per le Università e al 66,3 per cento per le Regioni.

### 1.3.3 Formazione e comunicazione

Riguardo ai profili formativi e comunicativi, Istat ha creato e adottato strategie con lo scopo sia di accompagnare e istruire gli altri enti della P.A. interessati a fornire contributi al NDC, sia di diffondere gli obiettivi, l'ambito di applicazione, le modalità di contribuzione e i contenuti semantici del Catalogo.

Gli aspetti legati alla formazione si sono finora concretizzati in attività di condivisione delle linee guida e di supporto alle pubbliche amministrazioni sugli standard e sulle *best practice* adottate per costruire un linguaggio semantico condiviso e per la definizione e l'inserimento dei metadati e delle ontologie di riferimento, anche per favorire la standardizzazione e il riuso della meta-informazione.

Tali attività di accompagnamento si sono realizzate attraverso incontri e scambi periodici tra *team* interni appositamente dedicati e competenti in ambito giuridico, metodologico e informatico e quelle pubbliche amministrazioni che hanno manifestato l'interesse a fornire il loro contributo al Catalogo, tra le quali Inail, Ministero dell'Interno e Inps.

Tra gli sviluppi delle attività di formazione sono in corso di progettazione ulteriori percorsi informativi e di aggiornamento su tematiche specifiche e sulle tecnologie utilizzate, anche attraverso materiali didattici sviluppati *ad hoc* e video *tutorial* a supporto delle pubbliche amministrazioni.

Gli aspetti legati alla comunicazione hanno come principale obiettivo la diffusione della cultura del dato, delle funzionalità del Catalogo e dei suoi contenuti. Le attività di comunicazione finora effettuate si sono tradotte nella partecipazione a eventi e congressi di interesse pubblico già a partire dalla fine del 2022, pochi mesi dopo la prima pubblicazione del Catalogo sul sito web [schema.gov.it](https://schema.gov.it). Sono infatti stati effettuati interventi in occasione di convegni nazionali organizzati da aziende partner<sup>4</sup> e fornitrici<sup>5</sup>, e numerosi *podcast* organizzati da ForumPA e realizzati con la collaborazione dell'Istat<sup>6</sup>.

Nel corso del 2023 il progetto NDC è stato ampiamente illustrato in occasione della Conferenza organizzata dall'Associazione Italiana di Scienze Regionali<sup>7</sup>. Sono stati inoltre realizzati articoli e monografie nel corso del 2022 (Bellitti, G., e M. Fedeli, 2022) e del 2023 (Polizzi, M., e S. Letardi, 2023).

Attualmente è in corso la pubblicazione su ForumPA di un articolo con la collaborazione del Dipartimento per la trasformazione digitale della Presidenza del Consiglio dei Ministri, avente come tematica principale gli attuali contenuti del portale NDC e le modalità di contribuzione al Catalogo.

4 *Analytics Bootcamp 2022* - 13 e 14 ottobre 2022, Milano.

5 *Informatica Data Talks* - 15 novembre 2022, Milano.

6 Cfr. <https://www.forumpa.it/pa-digitale/strategie-europee-interoperabilita-dei-dati-e-collaborazione-tra-pa-questi-temi-della-prima-puntata-di-beyond-data/>; <https://www.forumpa.it/pa-digitale/abilitare-il-single-digital-gateway-per-una-pa-sempre-piu-europea-ascolta-la-seconda-puntata-di-beyond-data/>.

7 XLIV Conferenza scientifica annuale AISRe - 6/8 settembre 2023, Napoli.

Allo scopo di valorizzare sempre più il ruolo del Catalogo ai fini dell'interoperabilità semantica con le piattaforme della P.A. e con la PDND e di condividerne le funzionalità per le pubbliche amministrazioni, per le imprese e per i cittadini interessati al suo utilizzo, è prevista una costante partecipazione a eventi e *meeting* di interesse nazionale e una periodica produzione editoriale.

<b>Quadro delle attività per la progettazione e l'attuazione del Catalogo Nazionale della Semantica dei Dati (NDC)</b>	
<b>Accordo tra l'Istat e la Presidenza del Consiglio dei Ministri – Dipartimento per la trasformazione digitale</b>	
Armonizzazione semantica dei dati	Analisi concettuale e classificazioni dei domini di interesse per lo sviluppo del Catalogo
Modellazione delle risorse semantiche	Modellazione di Ontologie, Schemi di dati e Vocabolari controllati
Realizzazione Infrastruttura informatica	Definizione requisiti, Disegno, Progettazione, Sviluppo, Test e Collaudo, Produzione, Evoluzione
Censimento banche dati della P.A.	Integrazione della Rilevazione censuaria Istat sulle istituzioni pubbliche
Formazione	Accompagnamento delle PP.AA., Sviluppo materiale didattico
Comunicazione/promozione	Diffusione: Cultura del dato, Funzionalità e Contenuti del Catalogo
<b>INTEROPERABILITÀ</b>	

## 2. LA STRATEGIA DATI PER LA PUBBLICA AMMINISTRAZIONE<sup>1</sup>

### 2.1 Digitalizzazione e dati

L'analisi del contesto nell'ambito del quale si inserisce il tema dell'interoperabilità semantica perseguita dal Progetto NDC e dal relativo metodo di sviluppo descritto in precedenza (cfr. Capitolo 1) consente di apprezzare la portata delle iniziative assunte per la sua realizzazione. Abilitando un migliore sfruttamento del patrimonio informativo pubblico, l'interoperabilità semantica si pone, infatti, al centro della digitalizzazione della P.A. e la considerazione del quadro strategico generale in cui la stessa si muove è requisito fondamentale per indirizzarne i progressi attuativi.

La trasformazione digitale della società costituisce per il settore pubblico, ormai da tempo, il principale motore di sviluppo evolutivo in grado di rimettere in gioco non solo le modalità espressive dell'azione amministrativa, ma anche lo stesso ruolo delle istituzioni. Punto di riferimento strategico degli attuali sviluppi è l'Agenda digitale europea del 2010, che si pone l'obiettivo di creare un mercato unico digitale idoneo a garantire il mantenimento degli attuali standard europei in tema di riconoscimento e protezione dei diritti fondamentali<sup>2</sup>; già nell'Agenda digitale europea si parte dal presupposto per il quale la stessa "può essere efficace solo se le sue diverse componenti e applicazioni sono interoperabili e si basano su standard comuni e piattaforme aperte"<sup>3</sup>. Tali propositi vengono perseguiti con una serie di iniziative mirate, da una parte, a puntellare con i valori e principi propri dell'Unione europea (Ue) la protezione dei dati personali (*General Data Protection Regulation* – GDPR), la *data governance*, il riutilizzo dei dati e l'attività di intermediazione di dati (*Data Governance Act*), l'accesso equo ai dati e il loro utilizzo (proposta di *Data Act*), i mercati digitali (*Digital Market Act*), i servizi digitali (*Digital Services Act*), l'intelligenza artificiale (proposta di Regolamento sull'AI – *Artificial Intelligence*) e, dall'altra, a potenziare le possibilità di sviluppo della digitalizzazione sotto il profilo dell'accesso ai dati (*Open data* e riuso del settore informativo pubblico) e dell'interoperabilità dei sistemi informativi (*European Interoperability Framework* – EIF). Con particolare riferimento agli ultimi progressi delle nuove tecnologie, l'Ue si dimostra attenta a garantire uno sviluppo equilibrato dell'intelligenza artificiale, cercando di bilanciare le esigenze di tutela dei diritti fondamentali con le dinamiche proprie dell'innovazione sperimentale. Così, da una parte, si propone un modello di legittimazione di diverse forme di AI basato sul livello di rischio (inaccettabile, elevato, limitato e minimo); dall'altra, si prevedono "spazi di sperimentazione normativa per l'AI" istituiti da una o più autorità competenti degli Stati membri al fine di rendere disponibile un ambiente controllato per lo sviluppo dell'AI.

<sup>1</sup> Di Giovanna Bellitti e Massimo Fedeli.

<sup>2</sup> Cfr. Copenhagen Economics, *The Economic Impact of a European Digital Single Market*. Study commissioned by European Policy Centre, 2010, p. 12: "The DSM can be seen as the fourth wave of economic integration of the EU market. After the 1992 Single Market Programme, EMU and the common currency (Euro) in 1999/2002 and the 2007 Services Directive, the creation of a single market for digital goods and services can be seen as a fourth step of the European Union to foster growth, jobs and innovation".

<sup>3</sup> DigitalEurope, *A Transformational Agenda for the Digital Age*, DIGITALEUROPE's 2020 Vision, Brussels, 2009.

Se la narrazione è sempre ispirata al mercato – una costante dell'approccio Ue fondato sulla perimetrazione dei diritti fondamentali alla loro dimensione economicistica, per poi consentirne un'espansione universale – i riflettori dei decisori europei sono puntati costantemente sui dati: “*the most inviting index of integration – because it can be verified statistically – is the economic one*” (Haas, E.B., 2004; Lindberg, L.N., 1963; Rosamond, B., 2003).

In tale scenario, l'innovazione della P.A. costituisce uno snodo fondamentale per la transizione digitale della società; non solo perché l'azione amministrativa rappresenta uno dei campi principali in cui si misura il riconoscimento dei diritti e dei valori dei cittadini, ma anche in considerazione della rilevanza del bacino informativo pubblico per la sperimentazione di soluzioni digitali per il trattamento dei dati.

## 2.2 *Government as a Platform (GaaP)*

Nel 2021, dallo snodo evolutivo pandemico e dalla conseguente diversa concezione della digitalizzazione scaturisce la Strategia del *Digital compass*. Non solo si è impressa un'improvvisa accelerazione allo sviluppo delle nuove tecnologie digitali, ma si è consolidata la convinzione circa la loro essenzialità per molti aspetti della società (socializzazione, lavoro, salute, apprendimento, intrattenimento, acquisti e, più in generale, fruizione di servizi *online*). Per queste ragioni, l'Ue ha puntato su politiche per un futuro digitale antropocentrico, sostenibile e più prospero, includendo la digitalizzazione dei servizi pubblici tra i quattro punti cardinali della *Bussola digitale 2030* utile per mappare il percorso dell'Ue nella transizione digitale. Viene, così, proposto un modello di governo come piattaforma (*Government as a Platform – GaaP*) in cui servizi pubblici digitali rappresentano la chiave per consentire un accesso globale e agevole ai servizi da parte di cittadini e imprese (O'Brien, R., 2004).

Il modello GaaP è un fattore determinante per perseguire, a livello europeo, la sovranità digitale, elemento necessario per assicurare autonomia e indipendenza dei cittadini e delle imprese. È tale linea di azione a indicare l'indissolubile legame tra interoperabilità e sovranità digitale: un quadro di interoperabilità rafforzato per trarre vantaggio dal valore economico dei dati, creando così un mercato unico per i dati utile a garantire una competitività globale e una sovranità dei dati.

Con le piattaforme si introduce un nuovo modo di costruire i servizi pubblici digitali, utilizzando un modello di sviluppo collaborativo tra amministrazioni pubbliche e cittadini per condividere e migliorare i processi e le capacità digitali nello svolgimento delle procedure amministrative. Le piattaforme vanno considerate quali architetture digitali grazie alle quali è possibile generare degli ecosistemi complessi entro i quali sono le interazioni tra gli utenti a generare – grazie al trattamento (eventualmente algoritmico) dei dati così prodotti – le possibilità evolutive della piattaforma stessa. In tale prospettiva, l'attività di governo e l'erogazione stessa dei servizi potranno, da una parte, migliorare continuamente grazie a capacità avanzate di elaborazione dati e, dall'altra, subiranno modificazioni nel loro modo di essere.

Il modello a piattaforma influisce, dunque, drasticamente sui processi attraverso i quali sono prodotti i servizi pubblici e, soprattutto, sul valore degli stessi, ispirandosi a uno stretto coordinamento tra tutte le amministrazioni; l'obiettivo è affrancarsi da modelli basati su relazioni chiuse, strutturate e formali, per abilitare una maggiore apertura e fluidità operativa nei rapporti tra i soggetti del settore pubblico (Cordella, A., e A. Paletti, 2019). Le piattaforme, da una parte, sono facilmente accessibili, favorendo, così, la semplificazione amministrativa; dall'altra, fondandosi sulla circolazione aperta dei dati e delle risorse, autoa-

limentano lo sviluppo di nuove soluzioni. A tale ultimo proposito, vanno considerate le maggiori potenzialità partecipative liberate dalle piattaforme in favore dei cittadini che, dunque, possono essere più facilmente coinvolti nella modellazione dei processi amministrativi; in una società caratterizzata dalla rapida mutevolezza dei bisogni dei cittadini, appare fondamentale acquisire agili strumenti in grado di intercettare efficacemente i cambiamenti.

Se il ricorso a forme di *eGovernment* appare, dunque, una strada obbligata (anzi, ormai scontata) nel percorso evolutivo delle pubbliche amministrazioni, va rilevato come lo stesso debba essere accompagnato da un approccio culturale in grado di recepire l'interoperabilità abbandonando il modello a *silos* che ha sinora contraddistinto l'organizzazione delle amministrazioni, costringendole a sistemi informativi chiusi (Batini, C., 2013). La transizione digitale della P.A. non può compiersi senza interoperabilità che se, come illustrato nel Capitolo 3, è concetto da riferire essenzialmente all'impostazione dei sistemi informativi, ancor prima deve fungere da principio guida per l'azione amministrativa. È lo stesso CAD (Codice dell'Amministrazione Digitale), del resto, a ricordarci che lo Stato, le Regioni e le autonomie locali devono assicurare la disponibilità, la gestione, l'accesso, la trasmissione, la conservazione e la fruibilità dell'informazione in modalità digitale e organizzarsi e agire a tale fine utilizzando, con le modalità più appropriate e nel modo più adeguato al soddisfacimento degli interessi degli utenti, le tecnologie dell'informazione e della comunicazione; su tali premesse, dunque, viene affermato il principio della disponibilità dei dati della P.A. in virtù del quale gli stessi sono formati, raccolti, conservati, resi disponibili e accessibili con l'uso di tecnologie dell'informazione e della comunicazione che ne consentano la fruizione e riutilizzo da parte delle altre pubbliche amministrazioni e dai privati.

### 2.3 L'interscambio dei dati amministrativi

La circolazione dei dati nell'ambito della P.A. è un'esigenza evidentemente congenita alla stessa azione amministrativa. Se per lo svolgimento di un determinato procedimento amministrativo è necessaria la conoscenza di determinati dati formati e detenuti da altra amministrazione, l'amministrazione competente per il procedimento deve evidentemente stabilire un rapporto di scambio informativo con l'amministrazione titolare del dato amministrativo. Fintantoché la trasformazione digitale non aveva ancora espresso il suo potere dirompente anche nella P.A., l'interscambio analogico dei dati amministrativi è avvenuto con il ricorso a una modulistica adeguatamente predisposta per consentire l'interscambio delle informazioni tra le pubbliche amministrazioni. Il fine è consentire un trattamento quantitativo e qualitativo dei dati utile non solo alla conclusione di un determinato procedimento, ma anche, e soprattutto, al migliore sfruttamento del patrimonio informativo. Lo sviluppo, infatti, di standard (originariamente tramite la modulistica, appunto) per il trattamento di informazioni è funzionale all'armonizzazione del "*processamento*" del dato, vale a dire l'attività che consente di estrarre una determinata informazione dal dato acquisito. L'uniformità dei processamenti dei dati (i trattamenti) è uno strumento per perseguire la condivisione della conoscenza e non solo del dato.

L'adozione di una determinata modulistica standardizzata alla luce di un'adeguata attività classificatoria costituisce, in questa prospettiva, un prodromo delle moderne tecniche di interoperabilità tra sistemi informativi della P.A. Non a caso, nell'ambito anagrafico, caratterizzato da un trattamento dati con l'adozione di "modelli conformi agli standard indicati" dall'Istat (cfr. DPR n. 223 del 1989, Regolamento anagrafico, art. 12,

per le comunicazioni dello stato civile; art. 13, per la dichiarazione; art. 19, per gli accertamenti richiesti dall'ufficiale di anagrafe; art. 22, per le schede di convivenza), si è espresso uno dei più significativi progetti riusciti di digitalizzazione dei dati della P.A.: la costituzione dell'Anagrafe Nazionale della Popolazione Residente (ANPR), base di dati di interesse nazionale ai sensi dell'art. 60 CAD.

Così, il trattamento statistico dei dati della P.A. si presenta quale specifica competenza da mettere al servizio della *data governance* nazionale come, del resto, sottolineato dalle iniziative europee impegnate a segnare il percorso da seguire nella costruzione di un Mercato unico digitale; il riferimento è alla *Data Governance Act* che, al fine di attuare il quadro di *governance* dei dati, ritiene opportuno costituire un apposito organismo (Comitato europeo per l'innovazione in materia di dati) per la cui composizione occorre considerare, altresì, *“organismi con competenze specifiche, quali gli istituti nazionali di statistica”* (considerando 53). Alla luce di ciò, risulta senz'altro utile volgere lo sguardo alle dinamiche della produzione della statistica ufficiale; come è noto, infatti, quest'ultima, in maniera sempre più preponderante con lo sviluppo di nuove Tecnologie dell'Informazione e della Comunicazione (TIC), annovera, tra le sue principali fonti di approvvigionamento informativo, i registri amministrativi; per l'elaborazione di statistiche ufficiali, dunque, la statistica si è da sempre nutrita dei dati prodotti dall'azione amministrativa dello Stato (Puglisi, R., 2022).

La consapevolezza circa l'alto valore del patrimonio informativo pubblico si scorge nell'osmosi sorta tra azione amministrativa e produzione statistica nell'ambito dei dati relativi alla popolazione. Il coordinamento informativo statistico, da parte dell'Istat, dei dati dell'amministrazione statale, regionale e locale risponde all'esigenza di un *“sistematico collegamento e [una] interconnessione di tutte le fonti pubbliche preposte alla raccolta e alla elaborazione dei dati statistici”* mediante una rete di uffici di statistica presso ogni amministrazione e alle dipendenze funzionali dell'Istat, cui vengono attribuiti compiti di indirizzo e coordinamento di tale sistema (art. 24, legge 23 agosto 1988, n. 400), all'insegna dell'obiettivo di *“realizzare l'unità di indirizzo, l'omogeneità organizzativa e la razionalizzazione dei flussi informativi a livello centrale e locale”* (art. 1, D.Lgs. 6 settembre 1989, n. 322). La lungimiranza del legislatore nel predisporre un simile efficace strumento di coordinamento informativo si è rivelata, oltre che nella qualità dei dati statistici prodotti, nel richiamo al Sistan e alle sue regole tecniche operato da parte dei primi tentativi di razionalizzazione dell'impiego dei *“sistemi informativi automatizzati”* con la definizione dei relativi standard, *“al fine di garantire l'interconnessione dei sistemi informatici pubblici”*<sup>4</sup>. Su tali basi è stato previsto che l'utilizzazione dei sistemi informativi automatizzati deve garantire il collegamento con il Sistan<sup>5</sup> e, sotto il profilo organizzativo, che il *“collegamento delle attività degli uffici”* si adegui *“al dovere di comunicazione interna ed esterna, e interconnessione mediante sistemi informatici e statistici pubblici”*<sup>6</sup>.

L'ultimo passo evolutivo della naturale convivenza tra trattamenti statistici dei dati e trasformazione digitale del trattamento dei dati della P.A. si riscontra nella norma del CAD secondo cui le basi di dati di interesse nazionale (vale a dire, sistemi informativi unitari con lo scopo di garantire l'allineamento delle informazioni e l'accesso alle me-

4 Art. 2, comma 1, lett. mm), Legge 23 ottobre 1992, n. 421, *Delega al Governo per la razionalizzazione e la revisione delle discipline in materia di sanità, di pubblico impiego, di previdenza e di finanza territoriale.*

5 D.Lgs. 12 febbraio 1993, n. 39.

6 Art. 2, comma 1, lett. c) del D.Lgs. n.165 del 2001, *Norme generali sull'ordinamento del lavoro alle dipendenze delle amministrazioni pubbliche.*

## 2. La strategia dati per la Pubblica Amministrazione

---

desime da parte delle pubbliche amministrazioni) debbono possedere “*le caratteristiche minime di sicurezza, accessibilità e interoperabilità e sono realizzati e aggiornati*” secondo le Linee guida dell’ AgID (Agenzia per l’Italia Digitale) e secondo le vigenti regole del Sistema Statistico Nazionale (Sistan)<sup>7</sup>.

### 2.4 Classificazioni dei dati dell’azione amministrativa ed e-service

Oltre che a livello organizzativo e tecnico, l’interoperabilità *ante litteram* imbastita con le metodologie per il trattamento dati definite dall’Istat si esplica, altresì, al livello semantico. Tra i compiti dell’Istat, infatti, rientra anche la “*predisposizione delle nomenclature e metodologie di base per la classificazione e la rilevazione dei fenomeni di carattere demografico, economico e sociale*”<sup>8</sup>. Su tali presupposti, il legislatore affida la *data governance* dei Censimenti all’Istat (da ultimo, art. 1, comma 232, Legge 27 dicembre 2017, n. 205).

Così, viene messa in campo l’attività per l’effettuazione del Censimento della Popolazione, del Censimento delle Istituzioni *non profit* o in campo economico. Nell’assolvere agli obblighi di rilevazione stabiliti dal Regolamento (Ce) n. 763/2008 relativo ai Censimenti della Popolazione e delle Abitazioni, l’Istat ha elaborato metodologie di trattamento dei dati per consentire alle informazioni detenute nei registri anagrafici comunali di transitare in banche dati adeguatamente organizzate per il loro riutilizzo per finalità statistiche. I Censimenti della popolazione rappresentano, così, uno degli esempi più significativi di interscambio di dati tra pubbliche amministrazioni per l’erogazione di servizi pubblici; infatti, “*per raggiungere i predetti obiettivi, il Censimento deve raccogliere informazioni e dati riferiti a diverse unità di rilevazione, adottando definizioni e classificazioni armonizzate*”, secondo i criteri previsti in campo internazionale<sup>9</sup>. Strettamente connessa all’esperienza censuaria, è la costruzione dell’Archivio Nazionale dei Numeri Civici delle Strade Urbane (ANNCSU) da parte dell’Agenzia delle Entrate e dell’Istat, la cui realizzazione, prevista dall’art. 3, comma 2, Decreto Legge 18 ottobre 2012, n. 179, è da ricondurre tra gli obiettivi nazionali in tema di *eGovernment*<sup>10</sup>.

I trattamenti dei dati per finalità statistiche sono, dunque, esperienze in cui si può constatare l’adozione di tecniche di trattamento dei dati per garantire, “*fin dalla progettazione e per impostazione predefinita*”, la futura interoperabilità dei sistemi informativi, aprendo la strada al riutilizzo del patrimonio informativo pubblico. Tale possibilità di riutilizzo è un valore aggiunto non solo per i fruitori dei dati, ma anche per lo stesso ente pubblico titolare del dato amministrativo sotto un duplice profilo. Da una parte, si favorisce la trasparenza e la responsabilizzazione dell’ente e, dall’altra, il ritorno dell’informazione da parte degli utilizzatori consente all’amministrazione titolare di migliorare la qualità dei dati raccolti e, quindi, l’adempimento ai propri compiti.

Sotto il profilo del riutilizzo del patrimonio informativo pubblico, si evidenzia come la realizzazione di risultati positivi come ANPR sia propedeutica all’erogazione di *e-service*

7 Art. 60, comma 2, CAD.

8 Art. 15, comma 2, lett. e), del D.Lgs. n. 322 del 1989.

9 Deliberazione Presidente Istat del 18 febbraio 2011, n. 6, *Piano Generale del 15° Censimento Generale della Popolazione e delle Abitazioni*.

10 Piano *eGovernment* 2009-2012. L’ANNCSU fornisce la disponibilità di un unico sistema informativo (costantemente aggiornato dai Comuni) dei dati relativi agli indirizzi del territorio nazionale organizzati sulla base di un unico standard; anche in questo caso, è sempre il legislatore ad affidare all’Istat il compito di fornire indicazioni per la compilazione e l’aggiornamento degli stradari comunali da parte dei comuni stessi (art. 45 Regolamento anagrafico). L’ANNCSU è realizzata secondo criteri di interoperabilità con le altre banche dati di rilevanza nazionale e regionale.

nei confronti di una variegata platea di utenti. Oltre all'accesso digitale ai propri dati anagrafici da parte del cittadino, ANPR eroga servizi di interscambio dati tramite la Piattaforma Digitale Nazionale Dati<sup>11</sup>.

La costruzione di sistemi informativi pubblici progettati nativamente interoperabili grazie, sotto il profilo semantico, a classificazioni standard dei dati apre, così, la strada all'erogazione di *e-service*. Alla luce delle esperienze appena descritte, si può notare come, coltivata l'interoperabilità dei sistemi informativi, gli *e-service* maturino quasi spontaneamente.

## 2.5 Il futuribile: AI generativa al servizio della Pubblica Amministrazione

Solo nell'immediato, l'obiettivo di rendere interoperabili i sistemi informativi della P.A. arretra le proprie implicazioni a un più agevole interscambio di dati tra le amministrazioni propedeutico all'erogazione di *e-service*; gli sviluppi delle nuove tecnologie, infatti, introducono nuove possibilità di impiego delle macchine che puntano in maniera sempre più insistente a un maggiore coinvolgimento delle stesse nello svolgimento delle procedure amministrative. Il riferimento è, evidentemente, alle forme di intelligenza artificiale che, grazie all'impiego di algoritmi, non si limitano a rendere immediatamente disponibili informazioni già note, ma si spingono fino alla creazione di nuovi contenuti. Se le alte potenzialità di simili strumenti possono tradursi nell'erogazione di nuovi servizi ai cittadini, sono molteplici gli aspetti da tenere in considerazione prima di adottare forme di loro impiego nella società in generale e, in particolare, nell'azione amministrativa.

Per tali motivi, l'intelligenza artificiale è oggetto di specifica attenzione da parte dei decisori europei, che mirano a conciliare lo sviluppo delle nuove tecnologie con i valori e i principi democratici fondanti dell'Ue. La proposta di Regolamento sull'intelligenza artificiale<sup>12</sup> prevede, così, un sistema *risk-based* finalizzato a limitare l'accesso nel mercato digitale Ue solamente a quelle forme di AI che risultino "sicure" per la persona.

Sono quattro i livelli in cui è diversamente modulato il rischio: inaccettabile (sistemi AI che rappresentano una chiara minaccia per la sicurezza delle persone); elevato (sistemi AI che, in ragione del loro ambito applicativo, necessitano della previsione di obblighi rigorosi prima di potere essere immessi sul mercato); limitato (sistemi AI per i quali il rischio si riferisce a esigenze di trasparenza affrontabili con la previsioni di specifici obblighi informativi); minimo o nullo (sistemi AI liberamente utilizzabili).

11 Si consideri, inoltre, l'esperienza ANNCSSU. L'ANNCSSU espone servizi differenziati in favore dei Comuni (consultazione puntuale e massiva degli stradari, indirizzari e interni del Comune; trasmissione di documentazione a supporto della verifica sull'ordinamento ecografico; verifica della rispondenza degli indirizzi ai requisiti tecnici stabiliti dall'Istat; inserimento, modifica e aggiornamento dei dati degli stradari, indirizzari e interni, ai fini dell'aggiornamento), dell'Agenzia delle Entrate (consultazione ed estrazione dei dati per lo svolgimento delle attività istituzionali dell'Agenzia, come la corretta individuazione degli indirizzi per la notifica di atti, nonché per allineare gli indirizzi contenuti nell'Anagrafe tributaria agli indirizzi dell'ANNCSSU), dell'Istat (consultazione ed estrazione dei dati; funzioni per il monitoraggio delle attività in qualità di gestore dell'infrastruttura), di tutte le pubbliche amministrazioni e dei gestori di servizi postali (possibilità di accesso in cooperazione applicativa di interrogazione puntuale di stradari e indirizzari) e di tutti gli utenti (effettuazione gratuita di interrogazioni sulla presenza o meno in ANNCSSU di un determinato odonimo o di uno specifico indirizzo). Cfr. Provvedimento interdirigenziale del 10 maggio 2022 adottato ai sensi dell'art. 11, DPCM 12 maggio 2016, Istat – Agenzia delle Entrate per la definizione di "Specifiche tecniche e modalità di accesso ai servizi erogati dall'Archivio Nazionale dei Numeri Civici delle Strade Urbane (ANNCSSU)".

12 COM (2021) 206 final, Proposta di Regolamento del Parlamento europeo e del Consiglio che stabilisce regole armonizzate sull'intelligenza artificiale (Legge sull'intelligenza artificiale).

## 2. La strategia dati per la Pubblica Amministrazione

---

Tali forme di tutela sono bilanciate con strumenti di incentivazione dell'innovazione, veicolandola in forme di sviluppo controllate in ambienti sicuri. Così, sono previsti “*spazi di sperimentazione normativa per l'intelligenza artificiale*” istituiti da una o più autorità competenti degli Stati membri al fine di rendere disponibile un ambiente controllato che faciliti lo sviluppo, le prove e la convalida di sistemi di AI innovativi per un periodo di tempo limitato prima della loro immissione sul mercato.

Infine, si prevede un sistema di *governance* tramite l'istituzione di un Comitato europeo per l'intelligenza artificiale, costituito da rappresentanti degli Stati membri e della Commissione; a livello nazionale, gli Stati membri dovranno designare una o più autorità nazionali competenti e, tra queste, l'autorità nazionale di controllo, al fine di controllare l'applicazione e l'attuazione del regolamento.

L'Ue, dunque, adotta una prospettiva (comune ad altre iniziative in tema di digitalizzazione) in virtù della quale si cerca di bilanciare i progressi tecnologici con i diritti e valori fondamentali europei; è, questa, la strada da seguire per l'impiego, anche nell'azione della P.A., delle nuove tecnologie basate su una progressiva autonomia delle macchine.



### 3. L'INTEROPERABILITÀ SEMANTICA DELLA PUBBLICA AMMINISTRAZIONE<sup>1</sup>

#### 3.1 I livelli di interoperabilità

I dati sono, dunque, il motore di sviluppo della digitalizzazione, fenomeno, quest'ultimo, caratterizzato dalla straordinaria amplificazione delle capacità di elaborazione dei dati da parte delle macchine in grado di condurre a risultati mai perseguibili dall'uomo. Come visto, il metodo di sviluppo del Catalogo (cfr. Capitolo 1) si fonda sul presupposto per il quale il ciclo di vita dell'informazione amministrativa – che attraversa le fasi della nascita, raccolta, archiviazione, elaborazione, utilizzo – proceda ad autoalimentarsi attraverso il riutilizzo dei dati anche da parte di soggetti diversi dal titolare degli stessi.

La chiave d'accesso alle potenzialità del digitale è, dunque, l'interoperabilità tra i sistemi informativi intesi non solo nella loro dimensione informatico-infrastrutturale (reti, risorse hardware e software per l'elaborazione, la memorizzazione e la trasmissione dei dati), ma nel senso di complesso di risorse – quali i dati, le informazioni, le persone, gli strumenti (tra cui, appunto anche i sistemi informatici) e i processi – organizzate per raccogliere, archiviare, elaborare i dati scaturenti da una determinata attività aziendale, al fine di renderli disponibili nei futuri processi che potranno, così, beneficiarne in termini di maggiore efficienza ed efficacia.

L'interoperabilità è la capacità delle amministrazioni di interagire in vista di obiettivi reciprocamente vantaggiosi, ricorrendo alla condivisione di conoscenze e informazioni per mezzo dei processi lavorativi su cui si basano, tramite lo scambio di dati tra i rispettivi sistemi informatici (*European Interoperability Framework - EIF, 2017*). “*One of the key insights offered by interop theory is the degree to which the proper functioning of systems that seem to be predominantly technical in nature [...] depends on how well human beings and institutions can work together*” (Palfrey, J.G., e U. Gasser, 2012).

Per il settore pubblico, quindi, perseguire l'interoperabilità dei relativi sistemi significa fondare i presupposti affinché il relativo patrimonio informativo (dati prodotti dallo svolgimento delle attività istituzionali del settore pubblico) possa divenire oggetto di interscambio automatico tra i sistemi stessi, ossia senza la necessaria intermediazione dell'operatore nel processamento del dato “grezzo” finalizzato all'acquisizione dell'informazione.

L'EIF 2017 distingue quattro livelli orizzontali dell'interoperabilità: 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)<sup>2</sup>. Allo stesso tempo, emergono differenze tra il livello europeo e il livello nazionale di interoperabilità legate alle rispettive diverse potenzialità di implementazione giuridica e semantica.

<sup>1</sup> Di Roberto Puglisi.

<sup>2</sup> Cfr. EIF 2017, COM(2017) 134 *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.

L'interoperabilità giuridica 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.

Emerge, di conseguenza, anche l'essenzialità del livello semantico di interoperabilità che attiene 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, queste ultime, le due componenti delle risorse semantiche raccolte nel catalogo ([schema.gov.it](http://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 del dominio di interesse 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 successivo 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 (cfr. Capitolo 4 per una più ampia trattazione delle ontologie e degli schemi 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, queste 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. Come accennato nell'esposizione del metodo di sviluppo del Catalogo (cfr. paragrafo 1.2), l'Istat assegna pari rilievo ai due processi, considerandoli entrambi essenziali e da svolgersi in stretta collaborazione per la realizzazione degli obiettivi del Progetto.

Con riferimento, invece, al livello organizzativo<sup>3</sup> e tecnico<sup>4</sup>, appare evidente la necessità di garantire la compatibilità comunicativa dei sistemi informativi, oltre che per i dati, anche per le altre componenti di un sistema informativo: processi e strumenti informatici. I sistemi informativi devono potersi integrare nelle modalità attraverso cui si esplicano le procedure e, allo stesso tempo, devono naturalmente adottare soluzioni tecniche per l'archiviazione, l'elaborazione e la trasmissione dei dati che risultino tra loro compatibili.

3 Cfr. Vicente, S., M. Pérez, X. García, A. Gimeno, e J. Naval. 2005. "eGovernment interoperability on a semantically driven world", p. 465: "Interoperability also involves some concerns regarding organisational issues, such as co-ordinating processes that span not only intra-organisational boundaries, but also co-operate with partner organisations that may well have different internal organisation and operations".

4 Con riferimento all'interoperabilità tecnica, cfr. Van Staden, S., e J. Mbale, 2012: "Data interoperability denotes the ability of different software from heterogeneous systems to understand the syntactical and semantic meaning of data from different data models through the use of common data models, mappings and structures".

In assenza di interoperabilità, le amministrazioni rimangono ancorate al modello a *silos* in virtù del quale, al più, è possibile stabilire un collegamento che si limiti a un interscambio di dati “grezzi” (da intendersi quali contenitori delle informazioni) e non anche delle informazioni negli stessi contenuti. Qualora due amministrazioni riescano, al più, a perseguire una trasmissione (spesso massiva) dei dati, l'amministrazione ricevente dovrà, comunque, procedere al loro processamento propedeutico all'estrazione delle informazioni da fare confluire nel proprio sistema informativo.

L'obiettivo è, dunque, superare progressivamente tali schemi propri di una fase della digitalizzazione connotata dall'impiego delle macchine solo come strumenti impiegati sotto la costante supervisione dell'operatore, chiamato ad assegnare un significato ai dati elaborati e memorizzati nei sistemi informativi. “*Despite the advent of modern information and communication technology (ICT), nothing much changed this state of affairs until the Internet made possible the effective connection of diverse information systems*” (Scholl, H.J. et al., 2011). Si vuole pervenire, in altre parole, a una gestione dei dati del settore pubblico per mezzo di sistemi informativi delle pubbliche amministrazioni chiamate a svolgere il ruolo di erogatori e fruitori di *e-service* a seconda che, rispettivamente, detengano o richiedano le informazioni utili allo svolgimento delle procedure amministrative; tale scenario apre la strada a una maggiore efficienza dell'azione amministrativa nella misura in cui, da una parte, il cittadino dovrà fornire i propri dati una sola volta (principio *una tantum*) ad amministrazioni che, dall'altra, potranno dialogare direttamente tramite i propri sistemi informativi interoperabili.

Riguardo al livello giuridico e a quello semantico dell'interoperabilità, si notano differenze tra lo sviluppo europeo e quello nazionale dovute alla circostanza per la quale solo il secondo può contare su un'armonizzazione normativa a monte; ciò determina delle differenti velocità evolutive tra la dimensione europea e nazionale dell'interoperabilità che, comunque, dovranno tendere a incontrarsi per stabilire un unico livello di interoperabilità.

#### 3.2 Il livello europeo di interoperabilità

Consapevole del valore strategico dell'interoperabilità, l'Ue è da tempo impegnata a perseguirla nell'ambito del patrimonio informativo del settore pubblico. Sono molteplici le azioni europee specificamente intraprese a tale proposito: dal Programma IDA (*electronic Interchange of Data between Administrations*)<sup>5</sup> – con lo scopo di sviluppare reti telematiche in particolari settori amministrativi – ai Programmi IDABC e ISA<sup>6</sup> che – promuovendo servizi condivisi – aprono la strada per il Quadro Europeo di Interoperabilità (QEI) del 2010<sup>7</sup> aggiornato, poi, nel 2017<sup>8</sup>.

5 Decisione n. 1719/1999/CE del Parlamento europeo e del Consiglio, del 12 luglio 1999, relativa a una serie di orientamenti, compresa l'individuazione di progetti di interesse comune, per reti transeuropee di trasmissione elettronica di dati tra amministrazioni.

6 Il Programma ISA - *Interoperability solutions for european public administrations* (Decisione n. 922/2009/CE del Parlamento europeo e del Consiglio, del 16 settembre 2009, sulle soluzioni di interoperabilità per le pubbliche amministrazioni europee) sostituisce il Programma IDABC - *Interoperable Delivery of Pan-european eGovernment Services to Public Administrations, Business and Citizens*, previsto dalla Decisione 2004/387/CE del Parlamento europeo e del Consiglio, del 21 aprile 2004, relativa all'erogazione interoperabile di servizi paneuropei di governo elettronico alle amministrazioni pubbliche, alle imprese e ai cittadini.

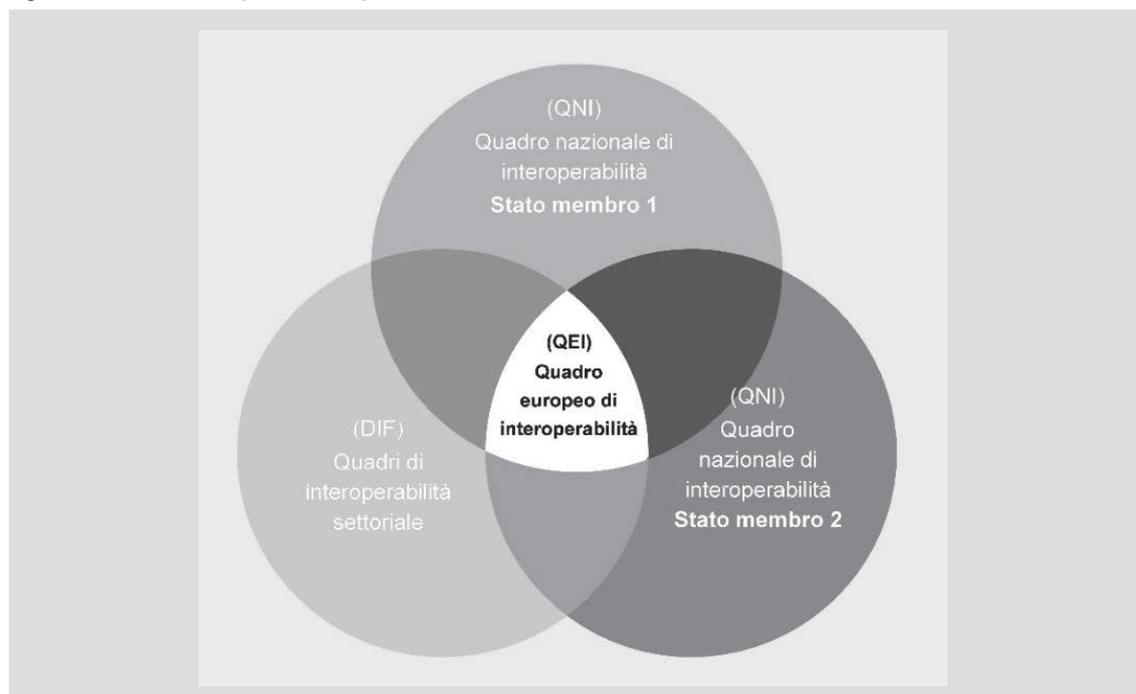
7 COM(2010) 744 del 16 dicembre 2010, Comunicazione della Commissione al Parlamento europeo, al Consiglio, al Comitato economico e sociale europeo e al Comitato delle regioni, *Verso l'interoperabilità dei servizi pubblici europei*, Allegato 2.

8 COM(2017) 134, Comunicazione della Commissione al Parlamento europeo, al Consiglio, al Comitato economico e sociale europeo e al Comitato delle regioni, *Quadro europeo di interoperabilità. Strategia di attuazione*.

Ispirato al principio di sussidiarietà in ottica di superamento dell'eterogeneità dei quadri nazionali di interoperabilità, il QEI rappresenta "il denominatore comune" delle politiche nazionali in tema di interoperabilità (QEI, principio base 1). Così, in una prospettiva evolutiva dei sistemi informativi, in cui lo scambio di informazioni dovrà avvenire *machine to machine*, si individuano, come visto (cfr. paragrafo 2.1), i quattro diversi livelli di interoperabilità che, a ben vedere, corrispondono alle tre componenti funzionali di qualsiasi sistema informativo (organizzativa, semantica e tecnica) cui si aggiunge il livello giuridico dettato dalla peculiare eterogeneità dei sistemi giuridici nazionali.

Alla luce dell'implementazione multilivello descritta nel paragrafo precedente, si è visto come l'interoperabilità giuridica e semantica costituiscano due fattori che, perlomeno allo stato attuale, possono ritardare criticamente l'avanzamento dell'interoperabilità europea rispetto a quella delle singole realtà nazionali; ciò per l'evidente ragione che, nella dimensione locale, si annulla la frammentarietà e, con essa, le differenze relative all'ordinamento giuridico di appartenenza e linguistiche. Sotto il profilo (giuridico), in ambito europeo, l'unica strada percorribile è agire prevalentemente in quei settori che già conoscono forme di convergenza normativa in grado di agevolare un dialogo non solo grazie all'assenza di ostacoli derivanti dalla frammentazione regolamentare, ma anche per la presenza di concetti comuni su cui fondare un'armonizzazione semantica. È in questi termini, perciò, che il QEI va inteso, come accennato, quale "denominatore comune" (minimo) dei quadri nazionali di interoperabilità: *"un QNI [Quadro Nazionale di Interoperabilità] in linea con il QEI per l'attuazione di servizi pubblici a tutti i livelli amministrativi nazionali [...] crea le condizioni di interoperabilità per estendere il campo di applicazione di questi servizi oltre i confini nazionali"*. L'Ue, quantomeno in questa fase "iniziale", non può ambire allo stesso grado di interoperabilità cui, invece, possono (almeno potenzialmente) già accedere i sistemi nazionali scevri da differenze linguistiche e ordinamentali.

Figura 3.1 - Quadro Europeo di Interoperabilità - QEI



Per tali ragioni, da una parte, gli sforzi europei si concentrano, tramite lo Sportello unico digitale (*Single Digital Gateway* - SDG), su determinate procedure amministrative attorno alle quali sviluppare l'erogazione di *e-service* comuni a tutti i cittadini europei; dall'altra, l'interoperabilità semantica viene fondata facendo ricorso a modelli di dati semplificati (*Core vocabularies*).

Per ciò che concerne il SDG, il Regolamento (Ue) 2018/1724<sup>9</sup> lo introduce per fondare uno sportello unico di accesso ai servizi mediante il quale i cittadini e le imprese possano completare le procedure in linea, interagendo con gli enti pubblici a livello dell'Unione e nazionale<sup>10</sup>. Il SDG è composto da un'interfaccia utenti comune gestita dalla Commissione e integrata nel portale "La tua Europa", che fornisce accesso ai pertinenti siti web dell'Unione e nazionali. Con riferimento a 21 procedure, è previsto che ciascuno Stato membro provveda affinché gli utenti possano accedervi ed espletarle interamente in linea, sempre che la procedura sia prevista nell'ordinamento interno. Per le procedure SDG espletabili in linea, gli Stati membri devono garantirne l'accesso transfrontaliero. A tale fine, il Regolamento istituisce il Sistema tecnico per lo scambio transfrontaliero automatizzato di prove (*Once-Only Technical System* - OOTS) in relazione alle 21 procedure SDG. Le specifiche tecniche e operative dell'OOTS sono state definite dal Regolamento Ue 2022/1463 (efficace a decorrere dal 12 dicembre 2023). In particolare, ai fini dell'OOTS, è previsto che *"Se le autorità competenti rilasciano legalmente, nel proprio Stato membro e in un formato elettronico che consente uno scambio automatizzato, prove rilevanti ai fini delle procedure in linea [...], esse mettono tali prove anche a disposizione delle autorità competenti richiedenti di altri Stati membri in un formato elettronico che consenta uno scambio automatizzato"*. Ai sensi del Regolamento Ue 1724/2018, per "prova" si intende *"ogni documento o dato, compresi testi o registrazioni sonore, visive o audiovisive, su qualsiasi supporto, richiesto da un'autorità competente per dimostrare la veridicità dei fatti o il rispetto degli obblighi procedurali di cui all'articolo 2, paragrafo 2, lettera b)"*<sup>11</sup>.

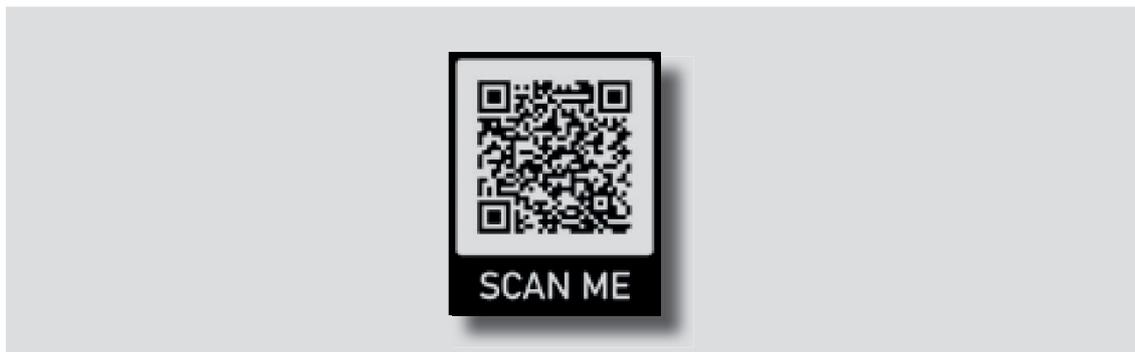
Con riferimento, invece, ai profili semantici, per sostenere lo sviluppo di soluzioni digitali per aprire la strada all'erogazione, da parte delle pubbliche amministrazioni, di servizi pubblici interoperabili, transfrontalieri e intersettoriali, l'Ue (Programma ISA<sup>2</sup>) ha proposto diverse soluzioni tra cui l'elaborazione dei cosiddetti *Core vocabularies*, vale a dire dei "modelli di dati semplificati, riutilizzabili ed estensibili che catturano le caratteristiche fondamentali di un'entità, come una persona o un'organizzazione pubblica, in modo neutrale rispetto al contesto". I *Core vocabularies* individuati dal Programma ISA<sup>2</sup> (*Core person, Core business, Core location, Core criterion and Core evidence, e Core public organization*)<sup>12</sup> sono delle soluzioni che, tenendo conto della frammentarietà dei quadri giuridici nazionali, puntano su un livello alto di astrazione in grado di abbracciare concetti descritti semanticamente in modo neutro rispetto al contesto.

9 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.

10 *"Una procedura dovrebbe essere considerata come interamente in linea quando l'utente può espletare tutte le fasi, dall'accesso al completamento, interagendo con l'autorità competente ("front office") per via elettronica, a distanza e mediante un servizio in linea"* (considerando 24, Regolamento (Ue) 2018/1724).

11 Art. 2, par. 2, lett. b): *"informazioni sulle procedure in linea e non in linea e link alle procedure in linea, incluse le procedure di cui all'allegato II, stabilite a livello dell'Unione o nazionale per consentire agli utenti di esercitare i diritti e rispettare gli obblighi e le norme nell'ambito del mercato interno nei settori di cui all'allegato I"*.

12 Cfr. [https://ec.europa.eu/isa2/solutions/core-vocabularies\\_en/](https://ec.europa.eu/isa2/solutions/core-vocabularies_en/).

Figura 3.2 - ISA<sup>2</sup> Core vocabularies poster

### 3.3 Il livello nazionale di interoperabilità

Si è già detto che la prima fase della trasformazione digitale si contraddistingue per la resistenza delle pubbliche amministrazioni a replicare il modello a *silos* (trapiantato dal mondo analogico), in virtù del quale in ogni organizzazione si crea un sistema informativo non progettato per comunicare con l'esterno e nel quale, sotto il profilo prettamente semantico, ciascuna impiega una propria concettualizzazione delle informazioni trattate (i dati pubblici) non spendibile in altri sistemi informativi. Il risultato negativo è l'impossibilità di connettere i diversi sistemi informativi per uno scambio immediato delle informazioni perché manca l'interoperabilità semantica. Ci si può limitare a una trasmissione massiva di dati da un'amministrazione all'altra e al conseguente ulteriore processamento del dato da parte dell'amministrazione ricevente; le inefficienze di tali dinamiche sono evidenti, costringendo a duplicazioni che impattano negativamente sui servizi resi dalla P.A.

Interoperabilità dei flussi informativi tra le pubbliche amministrazioni significa accessibilità di una amministrazione pubblica ai dati di un'altra amministrazione pubblica per il tramite di API (*Application Programming Interface*, ossia degli strumenti che consentono di superare le differenze e, quindi, abilitano l'interoperabilità tecnica tra due sistemi informativi); 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. Dunque, un miglioramento dei servizi per i cittadini; è, infatti, l'impossibilità di mettere in comunicazione più banche dati a costringere il cittadino, interessato in un determinato procedimento amministrativo, a fungere da punto di raccolta dei dati necessari per il procedimento stesso e detenuti da altre pubbliche amministrazioni.

L'attenzione del legislatore per l'interoperabilità non è nuova e si manifesta anche indirettamente con norme che presuppongono una P.A. interoperabile. Si pensi, ad esempio, all'inutilizzabilità, nei rapporti tra cittadino e un'amministrazione, delle certificazioni rilasciate da una amministrazione pubblica in ordine a stati, qualità o fatti; sono le amministrazioni pubbliche a dovere acquisire d'ufficio le informazioni oggetto delle dichiarazioni sostitutive, nonché tutti i dati e i documenti che siano in possesso delle pubbliche amministrazioni, previa indicazione, da parte dell'interessato, degli elementi indispensabili per il reperimento delle informazioni o dei dati richiesti ovvero ad accettare la dichiarazione sostitutiva prodotta dall'interessato<sup>13</sup>.

<sup>13</sup> Cfr. art. 43 D.P.R. n. 445/2000, TUDA (Testo Unico delle disposizioni legislative e regolamentari in materia di Documentazione

### 3. L'interoperabilità semantica della Pubblica Amministrazione

---

Tale schema, se non presuppone (ancora) necessariamente l'interoperabilità dei sistemi informativi delle pubbliche amministrazioni, quantomeno la immagina fortemente, traducendosi in un'esortazione al cambiamento. Siamo ai primi passi della P.A. nella digitalizzazione, in un momento in cui è sempre il documento a porsi al centro dell'esercizio dell'azione amministrativa e, dunque, dei rapporti tra la P.A. e il cittadino; ancora non sono scattati gli automatismi che, oggi, consentono di pensare in digitale e, così, si cerca, agli albori della digitalizzazione della P.A., di esportare (talvolta goffamente) le dinamiche analogiche negli ambienti informatici. Tuttavia, ciò che appare sorprendente è come l'interoperabilità già agisca sottotraccia, cercando di farsi strada attraverso l'imposizione di un divieto al diretto beneficiario dell'applicazione del principio *once only*; non riuscendo a ottenere l'interoperabilità con iniziative da parte della P.A., si impedisce ai cittadini di continuare a sopperire alle inefficienze della P.A. sostituendosi nella raccolta dei dati già in suo possesso.

Se, come detto, l'interoperabilità si fonda sull'idea di fare dialogare sistemi informativi pubblici originariamente progettati a *silos* (ossia per soddisfare esigenze informative in relazione solo ai propri fini istituzionali), il superamento di un assetto monadistico passa anche (e soprattutto) dalla disponibilità di una raccolta di *asset* semantici comune a tutti i titolari di basi di dati e sistemi informativi pubblici: un Catalogo Nazionale Dati. Per riuscire a stabilire un flusso informativo armonizzato tra sistemi informativi pubblici, occorre, infatti, garantire la disponibilità di una descrizione chiara dei dati detenuti da ogni amministrazione pubblica.

In perfetta aderenza al QEI, in ambito nazionale si distinguono, da una parte, modelli di dati riferiti ai singoli domini di interesse identificabili con le aree di attività delle diverse pubbliche organizzazioni (ad esempio, anagrafe, stato civile, salute, istruzione, fisco) e, dall'altra, a un livello più generale, i modelli di dati relativi ad ampie categorie concettuali (ad esempio persone, organizzazioni, luoghi) in grado di mapparsi trasversalmente sui singoli domini di interesse e creare, così, dei collegamenti tra gli stessi (cfr. Capitolo 4).

Il Progetto NDC ([schema.gov.it](https://schema.gov.it)) è una misura scelta per attuare il quadro nazionale di interoperabilità e accelerare il processo di digitalizzazione della P.A. che, alla luce di quanto sinora considerato e riscontrato nell'attuazione del Progetto stesso, non dipende solamente dalle soluzioni tecniche messe in campo ma soprattutto dalla disponibilità ad accoglierle, superando le resistenze culturali che ostacolano il completamento della transizione digitale della P.A. nel segno, come visto (cfr. Capitolo 2), della progressiva adozione del *platform model*, che vede come caratteristica principale la partecipazione collaborativa abilitata da un'interoperabilità in grado di assicurare una circolazione consapevole dei dati.

#### 3.4 La Piattaforma Digitale Nazionale Dati e il Catalogo Nazionale dei Dati della Pubblica Amministrazione

La Piattaforma Digitale Nazionale Dati (PDND) viene introdotta nel tessuto normativo del CAD per favorire la conoscenza e l'utilizzo del patrimonio informativo pubblico (art. 50 *ter* CAD). La PDND si affaccia in un contesto caratterizzato dal nuovo EIF e dal nuovo modello di interoperabilità ispirato ai suoi principi, prendendo il posto del *Data & Analytics Framework* (DAF), nato come una piattaforma di *Big data* composta da un *data lake* (dove memorizzare dati di potenziale interesse come le basi di dati, i *dati log*

---

Amministrativa) che precisa come l'acquisizione d'ufficio debba avvenire esclusivamente per via telematica.



e i dati di utilizzo delle PP.AA., nonché i dati provenienti del web e dai *social network*), da un sistema di *Big data engine* per armonizzare ed elaborare i dati grezzi del *data lake* implementando modelli di *machine learning* e da strumenti per la comunicazione dei dati (API esponenti dati e funzionalità).

Il Piano triennale per l'informatica nella P.A. 2022/2024 rammenta, dunque, come l'obiettivo della PDND sia quello di permettere di autorizzare e di autenticare le PP.AA. alla comunicazione tra i loro sistemi informativi e alla condivisione dei dati a loro disposizione, realizzando l'interoperabilità attraverso l'esposizione di servizi digitali implementati dalle API; secondo lo stesso Piano triennale, la PDND in futuro dovrà consentire anche l'accesso ai *Big data* prodotti dalle amministrazioni, nonché l'elaborazione di politiche *data-driven*.

La PDND è un'infrastruttura predisposta per consentire l'interscambio informativo tra le pubbliche amministrazioni mediante la disponibilità di *e-service*. Per lo sviluppo della PDND, è necessario fornire un'adeguata descrizione semantica dei dati pubblici con modelli di dati (ontologie), vocabolari controllati e schemi dati che consentano lo sviluppo delle API, vale a dire interfacce software con la funzione, appunto, di intermediare tra diverse applicazioni ognuna con i propri schemi di funzionamento.

Al fine di agevolare lo sviluppo della PDND, all'Istat è stata affidata la "Progettazione, creazione e attuazione di un catalogo centrale (in *open data*, secondo i principi del Quadro europeo di interoperabilità) comprendente schemi di dati, ontologie e vocabolari di base a sostegno della creazione e della progettazione di servizi digitali interoperabili".

L'attuazione del Progetto NDC consiste, da una parte, nell'elaborazione di un metodo idoneo a garantire che la descrizione semantica dei dati risulti adeguata (cfr. Capitolo 1), allo scopo di consentire che il significato assegnato a (o, detto in altri termini, che l'informazione risultante dal processamento di) ciascun dato risulti comune a tutte le pubbliche amministrazioni (enti erogatori e fruitori degli *e-service* della PDND, cfr. Capitoli 3 e 4); dall'altra, nella realizzazione dell'infrastruttura informatica destinata a esporre le risorse semantiche relative ai dati della P.A. (cfr. Capitolo 5).

La soluzione individuata dall'Istat per lo sviluppo delle descrizioni semantiche da pubblicare in [schema.gov.it](https://schema.gov.it) parte, così, dalla considerazione per la quale l'interoperabilità semantica dei sistemi informativi della P.A. si debba fondare sulla disponibilità di una rappresentazione concettuale comune a tutti gli utenti destinati a scambiarsi i dati tramite PDND. Sotto questa prospettiva, occorre garantire l'unicità del significato dei dati scambiati in modo che l'informazione contenuta nel dato trasmesso da un sistema corrisponda a quella effettivamente richiesta (e necessaria per un determinato procedimento amministrativo) da un altro sistema.

Per realizzare tale intento, è necessario che i concetti alla base delle informazioni scambiate vengano descritti univocamente evitando, così, la possibilità di disallineamenti semantici (associare, da parte di ciascuno, un termine a un significato diverso); si deve, dunque, semantizzare i dati detenuti dalle pubbliche amministrazioni in una prospettiva di armonizzazione. L'interoperabilità semantica si traduce, perciò, nella condivisione della descrizione concettuale (lemma + definizione) dei dati da parte di tutti gli utilizzatori degli stessi.

Il perseguimento di tale obiettivo ha richiesto la predisposizione, da parte dell'Istat, di una metodologia basata sulla stretta collaborazione tra diverse professionalità (giuridica, tematica e metodologica) che prevede una preliminare analisi concettuale dei domini di interesse; l'obiettivo di tale fase iniziale è l'individuazione dei concetti da collegare ai dati de-

tenuti da una determinata amministrazione. È un primo livello di astrazione indispensabile per aprire la strada alla loro descrizione semantica con ricorso a modelli di dati (ontologie, vocabolari controllati e schemi di dati) da mappare sui dati detenuti dalle pubbliche amministrazioni. Dal dato contenente l'informazione si passa alla definizione del concetto sotteso alla stessa informazione, collegandolo al termine giuridico usato per richiamare quello stesso concetto. Ad esempio, partendo dai dati di una scheda anagrafica (persona fisica, nome, cognome, residenza) si astrae il concetto di "Persona anagraficamente residente = Persona fisica avente dimora abituale in un determinato Comune come risultante nell'Anagrafe della popolazione residente" costruendo, così, il binomio lemma-definizione concettuale. La stessa operazione replicata nell'ambito dell'anagrafe tributaria porta a differenziare il concetto di "residenza fiscale", tenuto conto della sua non necessaria coincidenza con la residenza anagrafica. Senza tale attività preliminare di analisi semantico-giuridica, cui segue la fase di descrizione semantica tramite la modellazione di ontologie, in un ambiente interoperabile è alto il rischio che a una richiesta di un dato di residenza non meglio specificato si risponda con l'informazione non effettivamente attesa. Si perseguono, così, standard minimi di qualità.

Alla luce del funzionamento di [schema.gov.it](http://schema.gov.it) (cfr. Capitolo 5), che funge da punto di raccolta delle risorse semantiche già pubblicate dalle pubbliche amministrazioni in propri *repository* esterni a [schema.gov.it](http://schema.gov.it), ogni ente titolare di un *dataset* è, altresì, titolare e responsabile delle risorse semantiche che vi si riferiscono; spetta, dunque, a ciascun ente titolare provvedere all'aggiornamento delle proprie risorse.

In tale contesto, l'Istat, oltre alla progettazione e gestione informatica di [schema.gov.it](http://schema.gov.it), può intervenire anche fornendo, laddove richiesta, assistenza all'ente titolare nella descrizione semantica dei propri dati sin dalla fase di analisi concettuale e successiva modellazione ontologica ovvero analizzando e revisionando i contenuti già prodotti autonomamente dall'ente; ciò garantisce la conformità agli standard qualitativi scaturenti dalla accennata metodologia di sviluppo.

#### 3.5 La semantica della Pubblica Amministrazione

Si è detto come l'interoperabilità tra i sistemi informativi di due o più banche dati presupponga 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 termine (lemma) e definizione di un determinato concetto.

Per la realizzazione dell'interoperabilità 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 destinatario, ai concetti coinvolti nelle diverse procedure amministrative.

Tale approccio è conforme al Quadro europeo di interoperabilità 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.

In tale contesto, nel momento in cui si punta, con la progettazione del NDC, a una raccolta di risorse semantiche di supporto alla PDND il cui obiettivo è lo sviluppo di *e-service* per l'interscambio di dati pubblici tra pubbliche amministrazioni, occorre prendere atto del carattere speciale (o settoriale) del linguaggio giuridico (Cortellazzo, M., 1994). A tale proposito va sottolineato, infatti, come una delle caratteristiche principali dei *dataset* pubblici sia quella di costituire patrimoni informativi derivanti dall'esercizio di funzioni amministrative pubbliche che, per Costituzione, legge e regolamento, rispondono al principio di legalità (Sorrentino, F., 2008). 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 PP.AA. (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.

Riferito ai sistemi informativi pubblici, ciò comporta la loro armonizzazione sotto il profilo semantico-giuridico e terminologico attraverso un processo definitorio da cui fare scaturire la "corretta" relazione tra il termine (lemma) che verrà impiegato per l'erogazione di un *e-service* e la relativa definizione. La correttezza del binomio termine-definizione va perseguita prendendo come punto di riferimento l'ordinamento giuridico (solo laddove ricorra un concetto definito da un provvedimento normativo). Ad esempio, va perseguito l'obiettivo di impiegare il termine "azienda" solo per quei dati su cui può mapparsi il concetto di "azienda" così come definito dal codice civile ed evitando il ricorso allo stesso lemma per riferirsi ad altro tipo di informazioni. Tale considerazione si basa sulla legittima aspettativa dell'utente di potere collegare, ai termini già in uso da parte del legislatore, il significato da quest'ultimo assegnatogli; la legittima aspettativa, in altre parole, dell'uso della semantica giuridica nel settore pubblico.

Il 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 P.A.: la condivisione unanime di un concetto e del suo significato da parte di tutti i soggetti istituzionali coinvolti nell'interscambio.

A tale fine, 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. A tale riguardo, si richiama la funzionalità di NDC rispetto alla PDND.

Un ulteriore aspetto derivante da una calibrazione dei concetti esposti nei contenuti semantici del Catalogo su parametri normativi (esclusivamente laddove, come già detto, 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 PDND; solo l'accesso al dato amministrativo atteso consentirà, all'amministrazione 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, di residenza civilistica di cui all'art. 43 c.c., e residenza fiscale di cui all'art. 2, comma 2, TUIR (Testo Unico delle Imposte sui Redditi), non necessariamente sempre coincidenti considerato rispettivamente il valore presuntivo della prima ed effettivo delle seconde. Nel momento in cui una qualsiasi amministrazione, anziché richiedere all'interessato in un determinato procedimento amministrativo la prova della sua residenza, attinge – attraverso un *e-service* appositamente sviluppato tramite la PDND per collegarsi con i dati amministrativi acquisiti da parte del Ministero dell'Interno – direttamente alla base di dati dell'Anagrafe Nazionale della Popolazione Residente (ANPR),

### 3. L'interoperabilità semantica della Pubblica Amministrazione

---

appare evidente la necessità che ente erogatore ed ente fruitore si stiano riferendo allo stesso dato (residenza anagrafica), assegnandogli un comune significato (“Luogo in cui una persona ha fissato la dimora abituale [art. 43 c.c.] come risultante in Anagrafe”) senza possibilità di fraintendimenti con altri concetti come la “residenza effettiva” (da intendersi come “Luogo in cui una persona ha fissato la dimora abituale [art. 43 c.c.]”, a prescindere dalle risultanze anagrafiche) ovvero la “residenza fiscale”.

Per realizzare tale obiettivo, sotto il profilo semantico-giuridico, occorre 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 stessi attraverso un'analisi giuridica della disciplina normativa di riferimento che conduca a un significato unico e comune per ciascun concetto. Tale operazione deve essere guidata da un riferimento al dato normativo in modo da garantire a ciascuna definizione una solida condivisione da parte di tutti i soggetti coinvolti oltre che una legittima base giuridica.

Prima di procedere alla descrizione semantica per mezzo di modelli di dati (ontologie), vocabolari controllati o tassonomie, si analizza, dunque, il dominio di interesse cui si riferiscono i dati detenuti da un'amministrazione, al fine di individuarne, come accennato, i concetti utili a mappare le informazioni che saranno oggetto di interscambio tramite *e-service* esposti nella PDND. Il prodotto di tale analisi interpretativa è il binomio termine-definizione concettuale riferito al dato di rilievo per l'azione amministrativa; considerata la natura amministrativa dei dati oggetto di interscambio tramite PDND, le norme giuridiche assurgono a fonte primaria da cui trarre le definizioni e il riferimento principale da cui trarre il termine preferito.

I concetti, descritti dal binomio termine (o lemma, etichetta, *label*) - definizione concettuale, consentono il passaggio alla seconda fase, ossia quella di descrizione semantica (cfr. Capitolo 4) per mezzo di modelli di dati (ontologie) in grado di rappresentare, in una modalità *machine readable*, la conoscenza acquisita dalla prima fase di analisi concettuale del dominio.

Sono, questi, elementi necessari per un interscambio informativo utile allo svolgimento di un procedimento amministrativo. La correttezza definitoria di quest'ultimo non può, infatti, prescindere dalla circostanza per la quale i dati scambiati tra le PP.AA. corrispondano alla concettualizzazione propria dell'ordinamento giuridico.



## 4. LA SEMANTICA E LE ONTOLOGIE PER IL CATALOGO NAZIONALE DEI DATI PUBBLICI: ASPETTI METODOLOGICI<sup>1</sup>

Nel percorso di sviluppo dell'interoperabilità semantica della P.A. tracciato con la progettazione del Catalogo Nazionale della Semantica dei Dati (*National Data Catalog – NDC*) (cfr. Capitolo 1), emerge come requisito fondamentale l'uso di tecniche digitali di rappresentazione concettuale che rendano possibili nuove applicazioni della digitalizzazione dei dati stabilendo un legame tra la conoscenza e le macchine. La realizzazione di tali obiettivi viene, così, sintetizzata con il disegno delle ontologie, vale a dire modelli di dati *machine readable* in grado di agevolare lo sviluppo di API e, dunque, rendere automaticamente fruibile da parte di un sistema informativo la conoscenza relativa a un determinato dominio di interesse, conoscenza rappresentata da alcuni termini il cui significato è già sistematizzato in norme e regolamenti (la cui rappresentazione semantica è stata definita nel paragrafo 3.5) e da altri termini che invece non trovano riscontro in norme e regolamenti. Questi ultimi termini non sono una parte residuale dell'insieme di concetti utili a definire il contenuto di un dominio di interesse: spesso, infatti, corrispondono a una quota maggioritaria del complesso di termini utili.

In tale prospettiva, la modellazione ontologica compie l'applicazione del metodo elaborato dall'Istat fornendo gli strumenti espressivi idonei a digitalizzare la conoscenza mutuata dall'analisi concettuale dei domini di interesse.

Entrando nel dettaglio dell'organizzazione di questo Capitolo, il paragrafo 4.1 fornisce l'organizzazione concettuale standard internazionale per la terminologia (cfr. paragrafo 4.3.1). Rispetto a questi termini, si descrivono anche le possibili relazioni semantiche (cfr. paragrafo 4.1.3), che aiutano a rappresentare similarità, inclusioni logiche o altri tipi di relazione, modellate in coerenza con le norme ISO (*International Organization for Standardization*). Il patrimonio terminologico complessivo, comprendente anche i termini rappresentati nel paragrafo 3.5, viene riversato nelle ontologie (cfr. paragrafo 4.2). Le ontologie che vengono pubblicate su NDC modellano il dominio dei dati e delle procedure delle PP.AA. e dei gestori di servizi pubblici che si vogliono rendere interoperabili.

Attraverso la modellazione delle ontologie (dette anche *knowledge graph*) si definisce un vocabolario semantico unico delle informazioni contenute nei dati di interesse, espresse in un linguaggio standard secondo un costrutto che consente una interpretazione delle informazioni in modo automatico (*Machine-to-Machine, M2M*). Inoltre, il modello ontologico fornisce una rappresentazione del dominio dei dati concettuale, integrata e astratta dalle strutture fisiche, ossia modella l'informazione in modo trasparente rispetto alle strutture fisiche dei dati che possono essere diverse per ogni singolo contribuente al NDC.

L'adozione da parte degli enti pubblici dei modelli ontologici esposti su NDC consentirà, attraverso l'operazione di *Mapping* delle ontologie sui dati delle (PP.AA.) di erogare servizi di *Ontology-Based Data Access (OBDA)*, ovvero l'accesso ai dati delle proprie basi attraverso l'ontologia o, più genericamente, attraverso degli *e-service* che rendono interconnettibili i dati delle diverse PP.AA. che li utilizzano, come illustrato nel paragrafo 2.4. Infatti, attraverso questo servizio è possibile calcolare le risposte alle interrogazioni, o *query*, poste

<sup>1</sup> Il Capitolo è stato curato da Orietta Luzi ed è stato redatto da: Claudia Brunini (paragrafi 4.1.1, 4.1.2); Emanuela Recchini (paragrafo 4.1.3); Roberta Radini (paragrafi 4.2, 4.2.1); Raffaella Maria Aracri (paragrafo 4.2.2); Mauro Scanu (paragrafo 4.1).

sotto forma dei termini dell'alfabeto dell'ontologia, elaborandole attraverso un processo di ragionamento sull'ontologia e sui *mapping* e accedendo quindi alle sorgenti informative, restituendo le informazioni corrette all'utente richiedente in modo trasparente rispetto alla struttura dei dati.

Questa è una funzionalità avanzata implementata dal Progetto della Piattaforma Digitale Nazionale Dati (PDND) che sfrutta la modellazione ontologica e l'armonizzazione concettuale dei vari domini amministrativi realizzata da NDC. Sin d'ora le PP.AA. possono rendere interoperabili i dati attraverso *e-service* che sfruttano la semantica descritta in NDC per rendere lo scambio dei dati privo di fraintendimenti semantici, mentre la funzione di modellazione concettuale e semantica del contenuto informativo delle fonti della P.A. viene realizzato in prima istanza già da NDC.

## 4.1 Modellazione Semantica: elementi di base

In questa parte si definisce cosa è un termine e quali sono le sue componenti essenziali (lemma e definizione) e accessorie, ossia le relazioni semantiche tra termini diversi. Tali informazioni sono poi caricate come contenuti sulle ontologie.

### 4.1.1 Il ruolo dei termini nel Catalogo Nazionale della Semantica dei Dati

Il flusso di dati, documenti e informazioni tra enti diversi, sia a livello nazionale sia globale, è enorme e la chiarezza nella comunicazione è un'esigenza estremamente urgente. Sul versante dello scambio di informazioni amministrative, sono ricorrenti anche le questioni terminologiche laddove è essenziale allineare tra loro le fonti amministrative originali.

Un insieme di termini, quindi, è lo strumento in grado di rendere univoco e condiviso il significato della lingua utilizzata e garantire qualità nella comunicazione fornita. Consente l'utilizzo di concetti complessi e strettamente correlati anche da parte di non esperti, facilitando così la diffusione e il corretto utilizzo dei dati, della loro raccolta e archiviazione e dello scambio tra fonti eterogenee. Inoltre, in una prospettiva basata sui metadati, una terminologia coerente è certamente un prerequisito essenziale per la costruzione di raccolte terminologiche controllate che favoriscano l'interoperabilità (per ulteriori dettagli si veda ISO 25964, lo standard internazionale per i *thesauri*).

L'utilizzo di raccolte terminologiche gestite in modo non centralizzato è una pratica consolidata in diversi enti. Tali collezioni, gestite per lo più all'interno dei singoli enti, non sono allineate tra loro, hanno riferimenti temporali diversi, modalità di aggiornamento non documentate, sono compilate secondo criteri concettuali non condivisi e solitamente mancano di documentazione delle fonti utilizzate. Poiché vengono implementate in modo autonomo, sono frequenti i casi di duplicazioni improprie, sinonimi non documentati e relazioni semantiche non gestite.

Risulta quindi urgente l'esigenza di unire e armonizzare, a livello centralizzato, i contenuti del patrimonio informativo relativo alle procedure amministrative rese disponibili tramite gli *e-service* della PDND. Il superamento di semantiche locali in favore di un'unica raccolta di termini che raccolga centralmente tutte le risorse terminologiche e semantiche e alla quale i diversi utenti possano attingere in futuro è la risposta più efficace.

Un'unica raccolta terminologica presenta diversi vantaggi:

- dal lato della qualità, mira a favorire l'uso di un linguaggio chiaro e univoco migliorando le informazioni diffuse;
- dal lato del processo, riduce l'utilizzo delle risorse e aumenta l'efficienza produttiva, evitando la gestione locale dei termini e quindi la duplicazione di informazioni e attività;
- in prospettiva, un'unica raccolta terminologica promuove l'interoperabilità tra diversi sistemi, rappresentando il collegamento tra semantica e metadati.

L'unità minima di un'unica raccolta terminologica è il *termine*, la cui formulazione è sottoposta a criteri precisi e standardizzati forniti dalla letteratura terminologica e adattati alle esigenze interne di ciascun ente.

##### 4.1.2 Il termine come unità minima di una raccolta terminologica

Il *termine*, unità minima del linguaggio, è definito come l'associazione di due elementi: il *lemma* e la sua *definizione*. È necessario fare riferimento alla coppia *lemma-definizione* come elemento univocamente determinato perché vi sia la possibilità di utilizzare termini con lo stesso *lemma* ma diversa *definizione* (polisemia) e viceversa, termini con la stessa *definizione* ma diverso *lemma* (sinonimia). Sinonimie e polisemie, spesso "ineliminabili" perché insite nella vitalità stessa del linguaggio (ad esempio sono generate in contesti specifici in cui si usano termini derivati da altre lingue settoriali, tra le più frequenti troviamo quella giuridica ed economica), costituiscono a livello semantico due delle possibili relazioni esistenti tra termini. Maggiori dettagli al riguardo si possono trovare nel paragrafo 4.1.3.

Un'unica raccolta terminologica è quindi un catalogo di *lemmi* con una definizione associata. La coppia *lemma-definizione* è documentata dalle relative meta-informazioni, essenziali per una corretta gestione dei termini. Il rispetto di criteri standard nella formulazione del *lemma* e della *definizione* garantisce l'armonizzazione e la coerenza interna del linguaggio.

##### *Lemma*

Il *lemma* svolge la funzione onomasiologica di nominare il concetto. Può essere costituito da una sola parola, da un gruppo di parole o da una frase più o meno completa. Si tratta quindi di un campo di testo di media lunghezza.

Il *lemma* dovrebbe avere le seguenti caratteristiche:

- essere espresso nella forma base: i sostantivi, quando possibile, vanno espressi al nominativo singolare, i verbi all'infinito;
- essere chiaro rispetto al dominio di riferimento, ad esempio: il *lemma* "tasso di natalità" può riferirsi sia alla popolazione sia alle imprese, allora è opportuno chiamare quello delle imprese "tasso di natalità delle imprese";
- le locuzioni tecniche e i termini complessi dovrebbero essere introdotti nel loro ordine naturale;
- gli acronimi e le abbreviazioni non sono *lemmi*, ma un campo relativo al *lemma* esteso.

##### *Definizione*

La *definizione* ha la funzione di specificare il ruolo assegnato al *lemma* nel dominio di riferimento. Richiede un campo di testo di notevole lunghezza.

Semanticamente si formula collocando innanzitutto il concetto nella classe di oggetti a cui appartiene, rendendo così immediatamente chiaro l'ambito di riferimento del termine. Esprime poi eventuali ulteriori spiegazioni, compresi esempi. A una eventuale *definizione* specialistica, può far seguito una *definizione* generica.

Le *definizioni*, tutte espresse utilizzando lo stesso sistema concettuale, dovrebbero avere le seguenti caratteristiche:

- non ripetere all'inizio della frase il termine stesso, ma l'*incipit* definente, cioè l'elemento lessicale a cui si riferisce il *lemma*;
- essere concise, ma complete;
- contenere termini consolidati e inequivocabili;
- essere prive di abbreviazioni e acronimi;
- essere sempre formulate in senso positivo;
- non essere circolari.

#### 4.1.3 Relazioni semantiche tra termini

Le procedure amministrative rappresentano un dominio complesso ed estremamente articolato, formato da numerosi sottodomini, ciascuno con un proprio linguaggio specifico: si devono anche tenere presenti i numerosi scambi che avvengono con altri universi settoriali, come la galassia dei tanti produttori di dati a livello sia europeo sia internazionale, il mondo della ricerca scientifica pubblica e privata, gli enti con facoltà di erogare servizi o le infrastrutture di varia natura tecnologica o amministrativa, non ultimi i fruitori finali dei servizi medesimi. Tali scambi possono avvenire su registri linguistici che vanno da quello più specializzato a quello più generalista, passando per diversi livelli intermedi. Questa complessa rete di relazioni rende il linguaggio da usare per il NDC estremamente dinamico, soggetto a continui aggiornamenti, revisioni e allineamenti. D'altro canto, i lessici settoriali non vanno considerati mere nomenclature, cioè etichette incollate su oggetti e concetti identificati indipendentemente, ma vere e proprie strategie di denominazione e concettualizzazione che nascono sotto la pressione di specifici compiti funzionali, soggetti a grande mobilità.

La complessità dei compiti, il notevole dinamismo delle funzioni e le molteplici interconnessioni con diversi linguaggi settoriali rendono il linguaggio nel NDC estremamente mobile e l'adozione e l'uso di termini in costante cambiamento ed evoluzione. La duplice esigenza da un lato di rendere trasparente l'uso della lingua, in modo da dare fluidità e chiarezza alla comunicazione tra gli specialisti e tra questi e il mondo esterno, e dall'altro di soddisfare la domanda sempre più pressante di interoperabilità tra diversi sistemi sia a livello nazionale sia internazionale, richiede una *governance* attenta e dettagliata delle risorse terminologiche. La compilazione di un semplice elenco di termini con *lemma* e *definizione* non è più sufficiente, ma diventa essenziale la documentazione analitica delle relazioni semantiche tra i singoli termini nonché dei metadati a essi associati.

Nel predisporre la raccolta di terminologia controllata per il NDC si è fatto riferimento a tre standard internazionali specifici: ANSI/NISO Z39-19:2005 (R2010); ISO 25964-2-2013; ISO 1087:2019.

L'obiettivo di una raccolta terminologica controllata è quello di evidenziare le relazioni semantiche tra i termini. Se un certo concetto è contrassegnato da più forme linguistiche, una raccolta controllata denota la forma con il ruolo di termine preferito e associa a essa le

numerose altre forme utilizzate per indicarla. Ciò consente alle informazioni su un determinato concetto di convergere in un unico punto, vale a dire il termine preferito, attraverso i collegamenti a esso associati. I sinonimi e le forme semantiche equivalenti sono adeguatamente documentati affinché l'utente possa associare al concetto il termine preferito o, in alternativa, uno degli equivalenti. Una raccolta terminologica controllata documenta anche le polisemie, cioè i casi in cui più concetti sono contrassegnati da termini con lo stesso *lemma* e diversa *definizione*. Ciò avviene con riferimento a diversi sottodomini ed è importante che gli utenti dispongano di adeguata documentazione per questi casi.

Secondo i suddetti standard internazionali, è necessario esplicitare tre tipi standardizzati di relazioni semantiche: equivalenza, gerarchia, associazione. La relazione di equivalenza consente la gestione di sinonimie, quasi-sinonimie e varianti linguistiche. Le relazioni gerarchiche si basano su gradi o livelli di superordinazione/subordinazione, dove il termine sovraordinato rappresenta una classe o un insieme, e i termini subordinati si riferiscono ai suoi elementi, parti o individui. La relazione associativa riguarda le associazioni tra termini che non sono né equivalenti né gerarchici, ma che sono semanticamente o concettualmente associati a tal punto che il collegamento tra loro dovrebbe essere reso esplicito in una raccolta terminologica controllata, sulla base del fatto che potrebbe suggerire termini aggiuntivi da utilizzare nell'indicizzazione o nel recupero.

Una proprietà fondamentale delle relazioni semantiche è che sono reciproche; ovvero ogni relazione indicata tra il Termine A e il Termine B deve avere una corrispondente relazione dal Termine B al Termine A: questa regola deve essere rispettata per tutti i tipi di rapporti. Gli indicatori di relazione (collegamento semantico) sono sempre operatori accoppiati. Alcuni indicatori sono simmetrici mentre altri sono asimmetrici. Le relazioni semantiche sono espresse mediante operatori semantici specifici.

##### *Relazioni di equivalenza*

La *relazione di equivalenza* permette la gestione di termini che esprimono lo stesso concetto (o uno simile). A uno di questi termini viene assegnato il ruolo di preferito, mentre i restanti, in quanto non preferiti, sono a esso collegati tramite esplicito riferimento. Questa relazione è esplicitata mediante gli operatori semantici *USE* (U), che rimanda dal/i termine/i non preferito/i a quello preferito, e *USED FOR* (UF), che è definito nel senso opposto.

Esempio: gruppo di equivalenza {morte\*, scomparsa, decesso};  
morte UF scomparsa;  
scomparsa U morte;  
decesso U morte.

La *relazione di equivalenza* comprende:

- *Sinonimi*: termini aventi *lemma* diverso e identica *definizione*.
  - Esempio 1: ospedale\*, istituto di cura, struttura di ricovero.
  - Esempio 2: morte\*, scomparsa, decesso.
- *Varianti lessicali*: termini i cui *lemmi* differiscono per varianti ortografiche e morfologiche (sono incluse abbreviazioni e acronimi).
  - Esempio 1: *online/on-line*.
  - Esempio 2: Pvc/cloruro di polivinile.
- *Quasi sinonimi*: termini semanticamente vicini ma non sinonimi perfetti.
  - Esempio 1 (ripreso dal linguaggio comune): acqua di mare/acqua salata.

- Esempio 2 (ripreso dal Glossario Istat): “*Numero medio di componenti per famiglia*”, “*Dimensione media della famiglia*” e “*Ampiezza della famiglia*”.
- Esempio 3 (ripreso dal Glossario Istat): “*Intestatario del foglio di famiglia*” e “*Persona di riferimento*”.
- Esempio 4 (ripreso dal Glossario Istat): “*Dati di flusso*” e “*Dati longitudinali*”.
- Esempio 5 (ripreso dal Glossario Istat): “*Lavoro a tempo parziale*” e “*Lavoro part-time*”.
- *Collocazione generica*: tecnica in cui il *lemma* di una classe e i *lemmi* dei suoi membri sono trattati come equivalenti, con il *lemma* della classe più ampia che funziona come termine preferito.
  - Esempio:
    - mobile UF scrivania;
    - mobile UF tavolo;
    - mobile UF sedia;
    - scrivania U mobile;
    - tavolo U mobile;
    - sedia U mobile.
- *Rimando a elementi di termini composti*: rimando da elementi lessicali in forma composta ai singoli elementi componenti. Si tratta di un *rapporto di equivalenza* tipico di quei linguaggi documentari che preferiscono rinunciare alla coordinazione, cioè alla combinazione di più elementi lessicali all'interno di un unico elemento lessicale in forma composta.
  - Esempio:
    - indice dei prezzi UF indice dei prezzi al consumo;
    - indice dei prezzi UF indice dei prezzi alla produzione;
    - indice dei prezzi UF indice dei prezzi delle case;
    - indice dei prezzi al consumo U indice dei prezzi;
    - indice dei prezzi alla produzione U indice dei prezzi;
    - indice dei prezzi delle case U indice dei prezzi.

### *Relazioni gerarchiche*

Le *relazioni gerarchiche* sono strutture verticali che mettono in correlazione i termini secondo un rapporto di subordinazione, dove il termine sovraordinato rappresenta una classe o un insieme da cui si diramano i termini subordinati. In questo tipo di relazione viene esplicitato il concetto e il grado di subordinazione (o superordinazione) tra termini appartenenti allo stesso albero gerarchico, dove il termine sovraordinato rappresenta una classe o un insieme, mentre il termine subordinato costituisce un elemento o parte di essa. Queste relazioni sono tipiche delle tassonomie, ma queste ultime, a differenza dei vocabolari controllati, non prevedono relazioni diverse da quelle gerarchiche, né che tali relazioni siano esplicitate e fornite di definizioni rigorose.

#### 4. La semantica e le ontologie per il Catalogo Nazionale dei Dati pubblici: aspetti metodologici

In una *raccolta terminologica controllata* i rapporti gerarchici sono esplicitati, nel senso che sono espressamente indicati il genitore, cioè il termine sovraordinato dettagliato in ciascuno dei suoi elementi, e i figli, cioè i termini subordinati. È possibile che un termine abbia più di un genitore (con riferimento a differenti *fundamenta divisionis*) e quindi appartenga a più gerarchie che possono anche sovrapporsi (poligerarchie). In questi casi, poiché diventa opportuno indicare il *fundamentum divisionis* che genera le diverse relazioni, viene aggiunta una “*etichetta del nodo*”, riportante il concetto di origine.

Generalmente, la *raccolta terminologica controllata* è un *vocabolario controllato*. Nel contesto del NDC, si indica con *vocabolario controllato* solo un sottoinsieme della *raccolta terminologica controllata* che include le relazioni gerarchiche, ossia, usando un termine statistico, le classificazioni. L'operatore semantico utilizzato per identificare il termine sovraordinato è BT (*Broader Term*), al quale può essere aggiunta una cifra che indica la distanza in “*passo*” gerarchico dal termine subordinato.

Esempio 1 (ripreso dalla Classificazione delle attività economiche – Ateco):

Fusione di ghisa

BT1 Fonderie

BT2 Metallurgia

BT3 Attività manifatturiere

L'operatore semantico che identifica la relazione inversa, cioè i subordinati di un dato termine, è NT (*Narrower Term*).

Esempio 2:

Attività manifatturiere

NT1 Metallurgia

NT2 Fonderie

NT3 Fusione di ghisa

Tre sottotipi di relazioni rientrano nella categoria delle relazioni gerarchiche:

- La *relazione generica* documenta relazioni gerarchiche che non rientrano nei seguenti sottotipi e può essere verificata con l'affermazione “*tutti-alcuni*”, dove il termine sovraordinato deve contenere tutti i subordinati. Tale relazione potrà essere identificata dagli operatori BTG (*Broader Term Generic*) e NTG (*Narrower Term Generic*).
- La *relazione di istanza* esprime il legame tra un evento, una categoria generale e le sue componenti o istanze. Può essere testata utilizzando l'istruzione “*IsA...*”. Tale rapporto potrà essere individuato dagli operatori BTI (*Broader Term Instance*) e NTI (*Narrower Term Instance*).

Esempio 3: regioni

NT11 Piemonte

NT11 Lombardia

NT11 Trentino-Alto Adige/Südtirol

NT11 Friuli-Venezia Giulia

La relazione può essere affinata inserendo un primo livello di relazione partitiva e poi al secondo livello una relazione di istanza.

Esempio 4: Regioni

NTP1 Regioni a statuto ordinario

NTI2 Piemonte

NTI2 Lombardia

NTP1 Regioni autonome

*NTI2 Trentino-Alto Adige/Südtirol*

*NTI2 Friuli-Venezia Giulia*

- c. *La relazione parte-tutto* esprime il caso in cui l'elemento subordinato è una parte di un tutto che assume il ruolo di termine sovraordinato. Può essere testata con l'affermazione "*Fa parte di...*". Tale rapporto può essere individuato dagli operatori BTP (*Broader Term Partitive*) e NTP (*Narrower Term Partitive*).

Esempio 5: gerarchie territoriali (rientrano tutte in questo sottotipo di rapporto)

Nazione

*NTP1* Regioni

*NTP2* Comuni

*NTP3* Sezioni di Censimento

#### *Relazioni associative*

Nei casi in cui viene accertato un legame semantico tra termini, ma questa relazione non è né di equivalenza né gerarchica, allora si parla di *relazione associativa*. Si tratta quindi di relazioni di tipo residuo e sono indicate dall'operatore semantico RT (*Related Term*). È il caso, ad esempio, di due termini che appartengono a due gerarchie in parte sovrapposte, come "*Tipo famiglia*" e "*Tipo nucleo familiare*", che sono due classificazioni i cui elementi sono in parte sovrapposti.

## 4.2 Predisposizione degli *asset* semantici

Il progetto NDC prevede la predisposizione di diverse tipologie di *asset* semantici che vanno dalle ontologie ai vocabolari controllati e alla modellazione degli schemi dati. La modellazione semantica di questi prodotti viene condotta in modo conforme a quanto specificato nella documentazione pubblicata dal *team* per la trasformazione digitale congiuntamente con AgID e secondo i principi e i profili applicativi di modellazione descritti in DCAT-AP (definiti nell'ambito del Programma ISA della Commissione Europea) e DCAT-AP\_IT (profilo italiano dei metadati utili per descrivere i dati delle PPA.A.).

In questo contesto 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 *link* concettuali con nuovi prodotti pubblicati;
2. la fase di "*Pubblicazione*".

La modellazione dei prodotti semantici parte dall'analisi e dalla definizione del dominio amministrativo e dell'informazione di interesse, individuando i concetti, le relazioni e le proprietà principali come descritto nel paragrafo 4.2.1. Questi devono essere descritti da definizioni rigorose, anche in un linguaggio tecnico-giuridico, che non consenta interpretazioni ambigue.

Quindi la modellazione dei prodotti semantici parte da una istruttoria condotta da esperti giuridici-amministrativi dello specifico tematismo in esame (cfr. Capitolo 3), mentre le definizioni vengono stilate nel rispetto di standard di scrittura delle definizioni semantiche come descritto nel paragrafo 4.1.

### 4.2.1 Il ciclo di vita dei prodotti semantici

Nella descrizione del ciclo di vita dei prodotti semantici si fa riferimento specifico alle ontologie, perché esemplificative della gestione del ciclo di vita degli altri tipi di prodotti semantici, come i vocabolari controllati e la modellazione degli schemi dati, in particolare degli *e-service*. In generale, tutti questi tipi di prodotti semantici rappresentano il complesso degli *asset* semantici di un dominio (a cui possono contribuire più pubbliche amministrazioni) oppure di una singola amministrazione pubblica nel caso di un dominio specifico.

Il ciclo di vita dei prodotti semantici di seguito illustrato definisce il processo di prima modellazione oppure di aggiornamento di prodotti già pubblicati o di prodotti semantici prodotti per altri scopi o pubblicati su altre piattaforme e non pubblicati su NDC.

La modellazione ontologica parte sempre da una ricognizione sulle ontologie già pubblicate, in particolare sul portale NDC, per verificare se esistano concetti comuni già definiti e per prevederne l'integrazione attraverso assiomi che esprimano la coincidenza (*same\_as*) o la definizione di sottoinsiemi concettuali, oppure, eventualmente, nuove relazioni.

Queste operazioni consentono l'integrazione di dati esposti da diverse fonti delle PP.AA. che forniscono dati rispetto allo stesso concetto o dominio. In alcuni casi si procede alla modellazione dettagliata di sotto domini che possono includere eventuali "*sfumature*" definitorie o nuove appartenenze concettuali.

Premesso che una ontologia si compone essenzialmente di classi, proprietà e ruoli (ossia le relazioni), i passi operativi per la modellazione di una ontologia si possono riassumere come segue:

1. Individuare i concetti principali che diventeranno le classi:
  - se presenti delle relazioni gerarchiche, organizzare le classi in sottoclassi o superclassi;
  - definirne le proprietà e descriverne i valori leciti (Dominio e *Range*);
  - attribuire i valori alle proprietà per tutte le istanze create.
2. Individuare le relazioni tra classi che diventeranno i "*ruoli*":
  - nel caso di sottoclassi, rappresentare relazioni di appartenenza, composizione, unione e disgiunzione;
  - dominio e *range* delle relazioni;
  - vincoli sulle proprietà in termini di cardinalità, funzionalità.

La modellazione della definizione semantica delle classi, degli attributi e delle relazioni tra di essi viene definita in ottemperanza agli standard definiti nel paragrafo 4.1, con la particolare attenzione ai contenuti semantici come sono definiti nel Capitolo 3.

Le attività di definizione semantica lavorano sull'identificazione delle principali componenti semantiche individuate secondo tre diversi approcci, come descritti di seguito.

I metodi per individuare un primo elenco di termini più importanti di un certo *asset* semantico, sui quali si vogliono fare asserzioni, partono dalla scelta di uno di questi tre approcci:

1. Approccio *Bottom-up*: la progettazione avviene dal basso verso l'alto, ovvero 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*: la progettazione dell'ontologia avviene dall'alto verso il basso, ovvero individuando 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 *Misto*: come suggerisce il nome, 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 dal più generale al più specifico.

Quindi lo sviluppo dell'ontologia prevede come ultimo passo la creazione delle istanze di tutti gli oggetti semantici individuati nella procedura sopra descritta.

Nello sviluppo di ontologie per NDC, dovendo garantire la coerenza con quanto già definito nel *corpus* delle ontologie già pubblicato, si preferisce applicare il metodo misto.

Per l'implementazione dei prodotti semantici, in particolare per le ontologie, si possono utilizzare applicativi *open source*, e sono da preferire applicativi con funzioni di *editing* grafico delle ontologie. Questa funzione realizza la modellazione delle ontologie con un linguaggio completamente visivo, che rende il disegno delle ontologie estremamente agevole a chi si occupa di modellazione ontologica e a chi le deve consultare pur non essendo un esperto di linguaggi tecnici, come può essere OWL (*Web Ontology Language*). Vale la pena sottolineare che l'espressione grafica di una ontologia consente un lavoro di revisione in collaborazione con gli esperti tematici che tipicamente hanno già familiarità con i linguaggi diagrammatici.

Inoltre, è demandato al *tool* grafico di garantire la correttezza sintattica e di fornire funzioni di esportare l'ontologia nella sintassi standard di OWL 2 nelle diverse serializzazioni, come TTL, RDF, JSON.

Uno degli strumenti grafici di modellazione delle ontologie usato è *Eddy*, che è stato progettato per creare ontologie in *Graphol*, un linguaggio ontologico grafico. Pertanto, in *Eddy* le ontologie sono facilmente disegnate come diagrammi, piuttosto che scritte come insiemi di assiomi espressi con i tipici formalismi ontologici poco *user-friendly*.

#### 4.2.2 La predisposizione per la pubblicazione dei prodotti semantici

Le ontologie a oggi pubblicate su NDC si suddividono in Ontologie *Core*, ossia:

- CLV (*Core Location Vocabulary*), l'ontologia del profilo applicativo italiano sugli indirizzi e i luoghi dedicata alla modellazione delle entità di tipo territoriale.
- COV (*Core Organization Vocabulary*), l'ontologia del profilo applicativo italiano sulle organizzazioni (pubbliche e private).
- CPV (*Core Person Vocabulary*), l'ontologia del profilo applicativo italiano sulle persone.

A queste si affiancano:

- *Ontologie di dominio*, che modellano particolari settori informativi di una amministrazione, come ad esempio i singoli processi dei Certificati anagrafici gestiti all'ANPR (Anagrafe Nazionale della Popolazione Residente);
- *Vocabolari controllati*, ossia classificazioni che esprimono le modalità di variabili categoriali;
- *Schemi dati*, ossia la rappresentazione delle strutture dati e le strutture operative di intercambio di dati, come ad esempio gli *e-service*.

#### 4. La semantica e le ontologie per il Catalogo Nazionale dei Dati pubblici: aspetti metodologici

L'operazione di pubblicazione dell'ontologia e quindi dei file delle ontologie, dei vocabolari e, più in generale, dei singoli concetti che la compongono, definiti anche "risorse", 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 NDC (descritto nel Capitolo 2);
- la predisposizione degli archivi degli *asset* semantici (tipicamente su *GitHub*) in una schematizzazione standard, secondo quanto indicato nelle linee guida del Catalogo nazionale.

La registrazione degli identificativi unici è necessaria per definire in modo univoco le risorse semantiche tra tutte le risorse del web. In particolare, gli identificativi univoci sono rappresentati da URI (*Uniform Resource Identifier*), che successivamente sono stati generalizzati in IRI (*International Resource Identifier*) per includere anche caratteri non ASCII.

In conformità a quanto già sviluppato in OntoPiA, la radice delle URI utilizzate per NDC è la seguente: <https://w3id.org/italia/>, dove, nello specifico:

- secondo lo schema di OntoPiA possono essere registrate: le ontologie sotto la radice "onto", i vocabolari controllati nella radice "controlled-vocabulary" e lo schema dati nella radice "data";
- secondo lo schema proprietario di specifiche PP.AA., i diversi prodotti semantici debbono essere registrati con specifiche sotto radici. Ad esempio: Inail ha registrato delle URI proprietarie dedicate al processo amministrativo delle indennità assicurative per incidenti sul lavoro sotto il *name-space* "work-accident" (<https://w3id.org/italia/work-accident>).

La pubblicazione nel Catalogo richiede, come introdotto precedentemente, la predisposizione per ogni prodotto semantico di metadati che ne descrivano:

- l'ambito tematico;
- il processo di definizione in termini di responsabilità, *editing* e gestione;
- il livello di aggiornamento e le parole chiavi di ricerca.

Questi metadati sono modellati dall'ontologia DCAT-AP\_IT, profilo italiano dell'ontologia internazionale DCAT, che consente di garantire interoperabilità semantica anche della definizione di metadati descrittivi dei prodotti semantici del Catalogo.

Quindi questi metadati favoriscono le operazioni di *harvesting* (indicizzazione sul portale) dei prodotti semantici, anche verso portali diversi da quelli di pubblicazione del Catalogo (ad esempio gli *asset* semantici del progetto LEO – *Livestock Environment Opendata*, pubblicati sul portale *Open data* di riferimento, avendo rispettato tutti i vincoli formali definiti dalle linee guida di NDC e dall'*harvester* possono essere pubblicati anche sul portale NDC), aumentandone l'interoperabilità semantica.

Questo grazie alla funzione di indicizzazione dei contenuti semantici, resa disponibile dall'*harvesting*, che favorisce il reperimento e la navigazione delle informazioni per gli utenti finali del Catalogo come descritto nel Capitolo 5.

In sintesi, lo sviluppo di prodotti semantici – ontologie, vocabolari controllati e schemi dati – richiede una sinergia di esperti: tematici, semantici, tecnico-giuridici.

Gli esperti cooperano e interagiscono nei vari ambiti descritti nei Capitoli precedenti per produrre file che, pubblicati sul portale NDC, rendono i contenuti semantici disponibili a una ampia platea, che li riutilizza per la creazione di prodotti che rendono i dati interoperabili.

L'interoperabilità semantica consente i servizi di scambio dati in un linguaggio unico e formale che abilita funzioni ampie, fino a quella di interrogazione dei dati definita dal paradigma di OBDA (Kontchakov, R., *et al.*, 2013). Lo sviluppo di Ontologie per modellare il dominio dei dati amministrativi consente attraverso OBDA non solo di eseguire *query* sui dati in modo trasparente rispetto alla struttura dei dati di ogni organizzazione (Geisler, S., *et al.*, 2021), ma anche di fornire prodotti avanzati di Intelligenza Artificiale (AI) come il *reasoner* automatico (Lenzerini, M., 2023). Questo servizio è in grado di evidenziare incoerenze sui dati o estrarre *pattern* conoscitivi non noti.

La frontiera, non lontana, è rappresentata dall'uso dell'AI per generare in modo automatico l'ontologia o per creare *query* sulle ontologie partendo da un testo scritto in modalità verbosa e non tecnica.

Anche in questo caso, aver definito un modello operativo e uno standard nella rappresentazione della semantica sarà un punto di riferimento per la redazione automatica di prodotti semantici.

---

## 5. LA PIATTAFORMA DEL CATALOGO<sup>1</sup>

### 5.1 Obiettivi dell'infrastruttura

Il perseguimento dell'interoperabilità dei sistemi informativi della P.A. presuppone, come visto, la predisposizione di diversi *layer* abilitativi in grado di soddisfare le esigenze derivanti dall'integrazione e dall'interazione di diversi ambienti digitali. Per ciò che concerne il livello semantico, oltre all'analisi concettuale dei domini e alla relativa rappresentazione con modelli di dati (cfr. Capitoli 3 e 4) – che portano alla disponibilità di risorse semantiche per ciascun dominio di interesse coinvolto nell'interscambio di informazioni – risulta necessario assicurare un sistema efficiente di utilizzo delle risorse stesse, ricorrendo a soluzioni in grado di garantire la possibilità, per erogatori e fruitori di *e-service*, di accedervi agevolmente e sfruttarle per lo sviluppo di API.

Nell'ambito del metodo sviluppato per la progettazione del Catalogo Nazionale della Semantica dei Dati (*National Data Catalog* - NDC, cfr. Capitolo 1), l'obiettivo dell'infrastruttura informatica del Catalogo è quello di rendere disponibili, tramite il portale [schema.gov.it](https://schema.gov.it), le descrizioni semantiche dei dati detenuti dalle amministrazioni pubbliche, in modo da abilitare il successivo sviluppo di API funzionali agli interscambi informativi tra soggetti erogatori e soggetti fruitori della PDND. L'implementazione di [schema.gov.it](https://schema.gov.it) è, dunque, una misura propedeutica alla realizzazione del GaaP (*Government as a Platform*, cfr. Capitolo 2), il modello innovativo per costruire servizi pubblici digitali.

Infatti, prima ancora dell'erogazione di *e-service*, da svilupparsi in forma collaborativa tra utente e amministrazione, occorre perseguire l'obiettivo di rendere effettivamente interoperabili le banche dati delle pubbliche amministrazioni, in modo da assicurare, innanzitutto, il pieno sfruttamento del patrimonio informativo pubblico e, quindi, il principio *once only*. Tale principio rappresenta, paradigmaticamente, il primo passo di vera emancipazione dai tradizionali schemi di espressione dell'attività amministrativa.

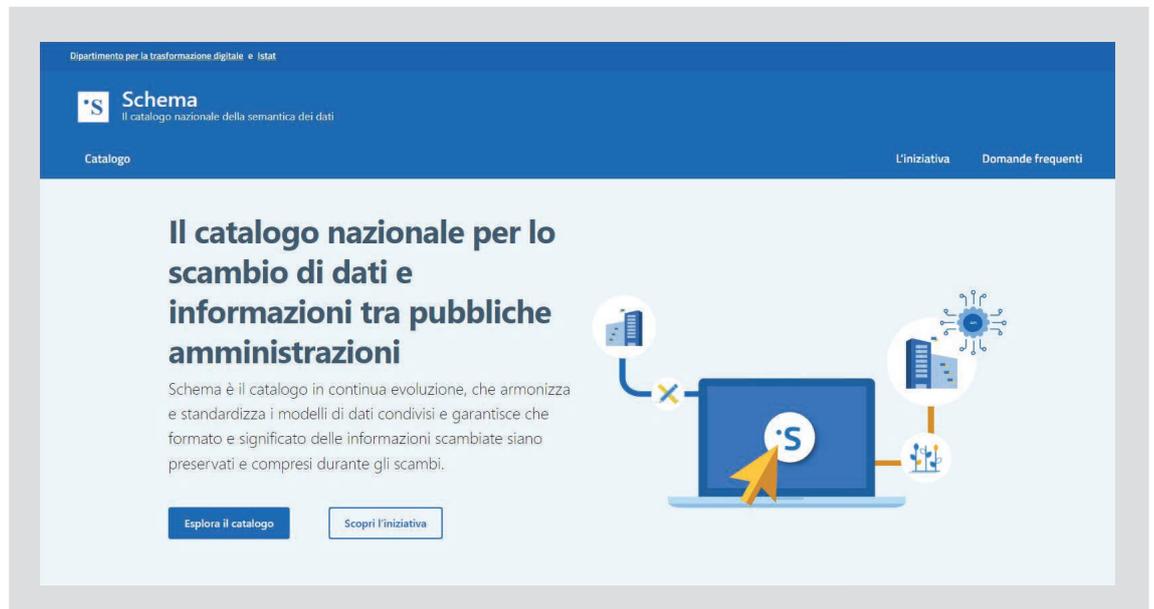
### 5.2 Schema.gov.it

Entrando più nel dettaglio tecnico, le attività di sviluppo digitale del Catalogo Nazionale Dati finora implementate si sono concretizzate, già a partire dal 30 giugno 2022, con la realizzazione di un portale, accessibile al [link schema.gov.it](https://schema.gov.it).

Il portale mette a disposizione numerose sezioni e funzionalità, per la cui esplorazione è possibile avere informazioni di dettaglio già dalla *Home page* (Figura 5.1).

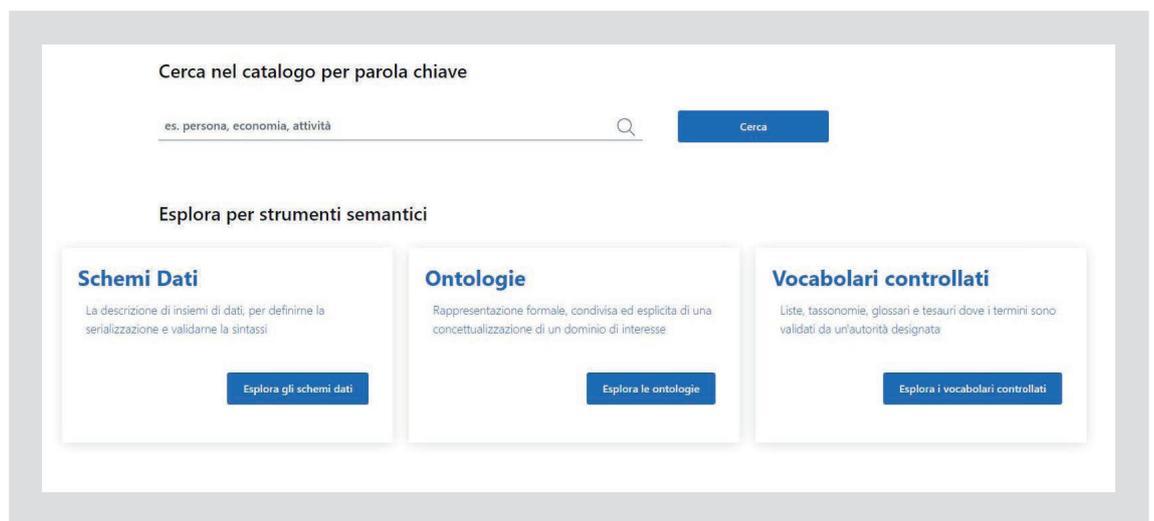
---

<sup>1</sup> Di Sara Letardi e Marco Polizzi.

Figura 5.1 - Catalogo Nazionale della Semantica dei Dati: la *Home page*

L'utente può ricercare gli strumenti del Catalogo per tipologia: schemi dati, ontologie e vocabolari controllati. È possibile effettuare una ricerca all'interno del Catalogo per parole chiave o per strumenti semantici. Cliccando, ad esempio, su "Vocabolari controllati", si accede a una pagina di dettaglio con la lista completa dei vocabolari (Figura 5.2).

Figura 5.2 - Catalogo Nazionale della Semantica dei Dati: le sezioni

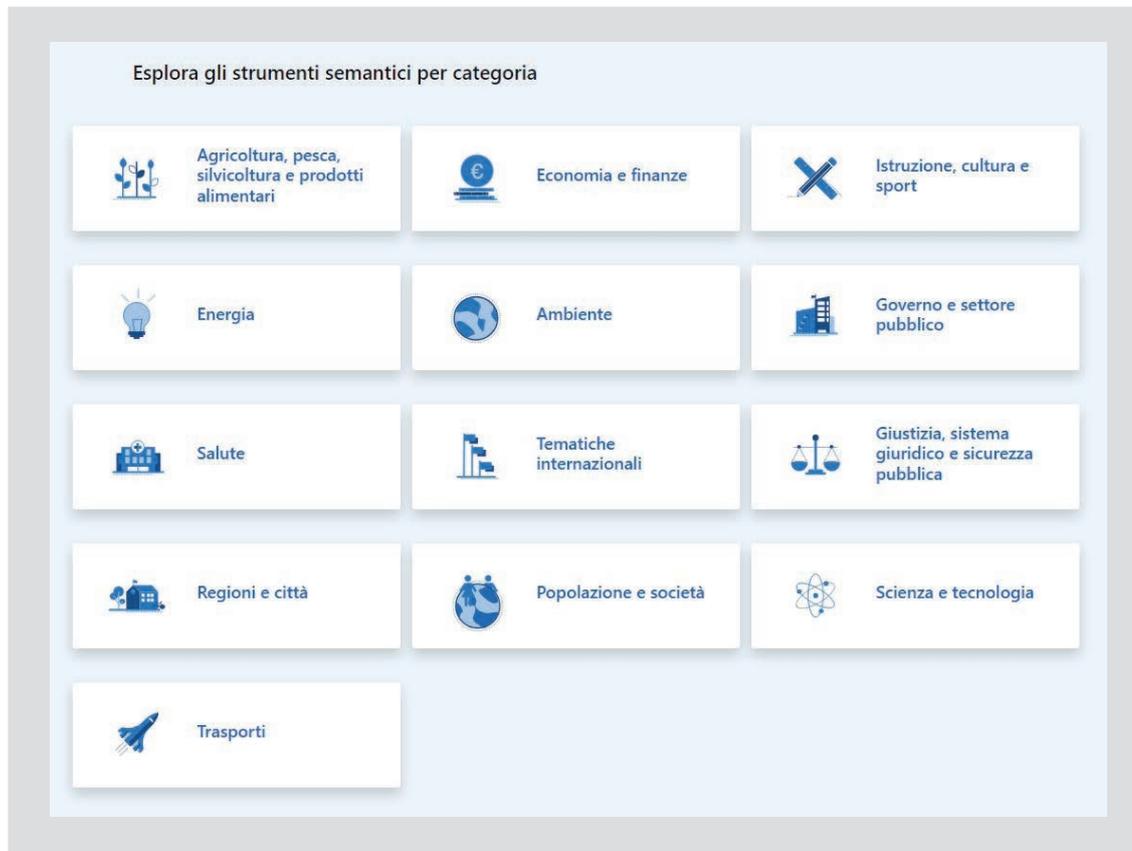


Si può inoltre scegliere di esplorare il Catalogo per categorie tematiche (Figura 5.3). Attualmente, l'utente può esplorare tredici categorie. All'interno di ciascuna categoria troverà i relativi vocabolari controllati, le ontologie e gli schemi dati.

In fondo alla *Home page* è presente una sezione che descrive come contribuire, ad esempio come utilizzare e pubblicare contenuti aggiuntivi per il Catalogo.

## 5. La piattaforma del Catalogo

Figura 5.3 - Catalogo Nazionale della Semantica dei Dati: le categorie tematiche



Il portale [schema.gov.it](https://schema.gov.it) utilizza il dominio `w3id` per ottenere URI persistenti utili alla rete italiana di ontologie e vocabolari controllati per le pubbliche amministrazioni.

In questo contesto è opportuno sottolineare che <https://w3id.org/> è un sito web di registrazione di identificatori web semantici (URI) gestito dal *World Wide Web Consortium* (W3C). L'obiettivo principale di questo sito è quello di fornire uno spazio per la registrazione degli URI che vengono utilizzati in contesti semantici, come ad esempio nell'ambito dei dati aperti, delle ontologie e delle risorse semantiche.

In pratica, questo sito permette a chiunque segua le regole di registrare un URI semantico univoco che possa essere utilizzato per identificare in modo univoco un concetto, una risorsa o un dato all'interno di un contesto semantico specifico. Gli URI registrati su <https://w3id.org/> sono conformi alle specifiche del W3C e vengono utilizzati in diverse applicazioni web semantiche. Ad esempio, se si sta creando un'ontologia per descrivere i dati di un certo dominio, è possibile registrare un URI semantico univoco per ogni concetto dell'ontologia su <https://w3id.org/>. In questo modo, gli URI registrati possono essere utilizzati da altre ontologie e risorse semantiche per creare connessioni tra i dati e migliorare l'interoperabilità tra diverse fonti di dati. In sintesi, <https://w3id.org/> fornisce uno spazio di registrazione per identificatori semantici univoci e standardizzati, che sono utilizzati per identificare in modo preciso le risorse all'interno di contesti semantici specifici.

Il flusso di informazioni relative agli *asset* semantici che sono indicizzati nel portale [schema.gov.it](https://schema.gov.it) si svolge secondo le seguenti modalità: l'amministrazione pubblica che desidera rendere disponibile nel Catalogo le proprie risorse semantiche deve effettuare una richiesta diretta agli

amministratori di NDC e inserire le proprie risorse semantiche nel *repository* messo a disposizione dagli amministratori del Catalogo, assicurandosi di aver seguito le relative linee guida. In alternativa, l'amministrazione pubblica ha la possibilità di rendere disponibile un proprio *repository* tramite la condivisione del relativo URL, che verrà gestito dal *team* tecnico del portale.

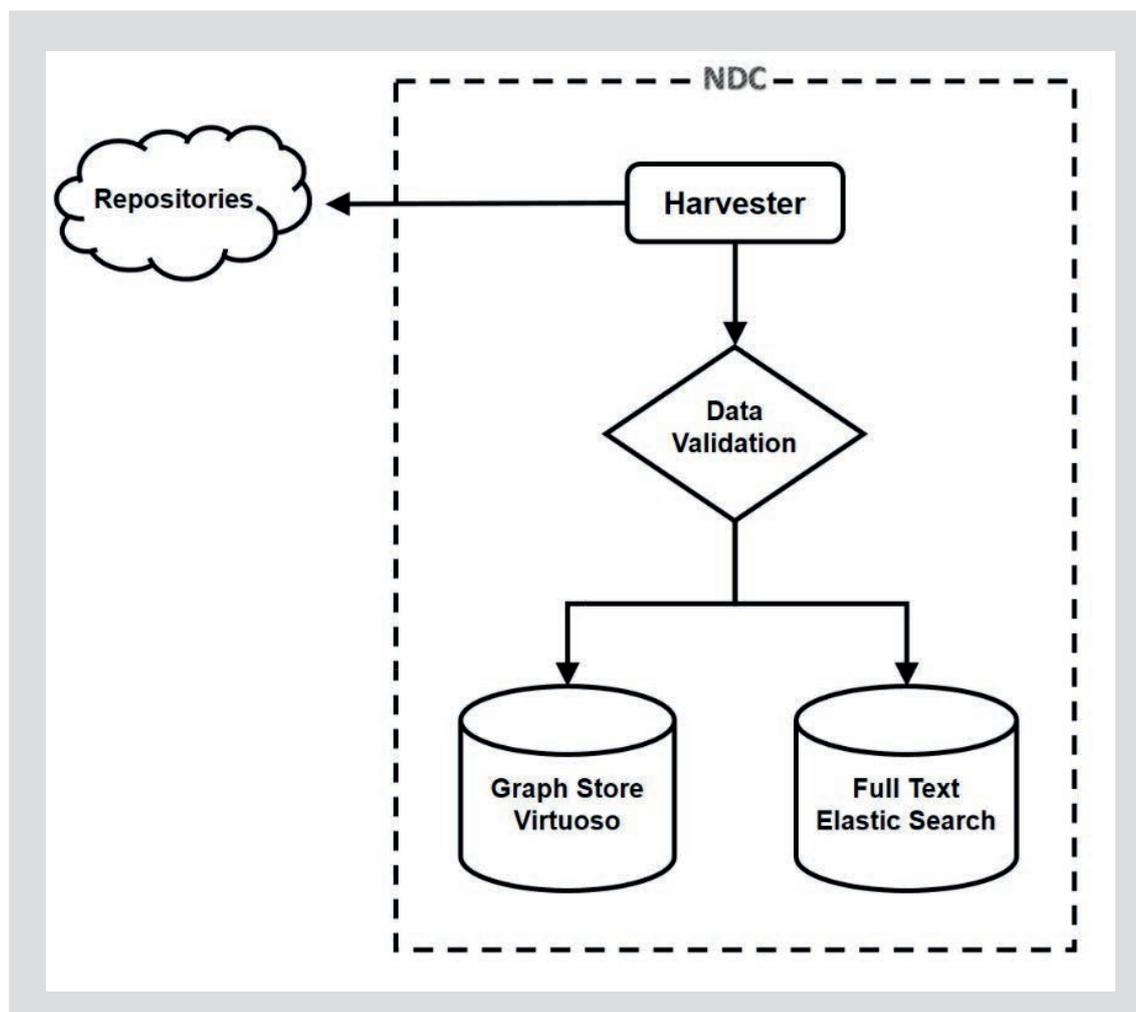
### 5.3 Il processo di *harvesting*

Un aspetto molto importante della fase di evoluzione e popolamento dei contenuti del Catalogo è rappresentato dal processo di *harvesting*, ossia di quell'insieme di attività di acquisizione dei contenuti semantici provenienti dai *repository* resi disponibili da istituzioni esterne e, in generale, dalle PP.AA.

Il processo di *harvesting* non altera in alcun modo il contenuto dei file dei *repository*. Pertanto, il contenuto dei file, la congruenza degli URI e altro ancora sono in carico a coloro (redattori) che redigono i file da processare.

Nella Figura 5.4 è riportato uno schema descrittivo del processo di *harvesting*, che rappresenta il processo fondante del Catalogo, poiché consente di aggiornare continuamente le risorse semantiche indicizzate al suo interno.

Figura 5.4 - Catalogo Nazionale della Semantica dei Dati: il processo di *harvesting*



## 5. La piattaforma del Catalogo

---

Il processo di *harvesting* è effettuato dal componente informatico denominato *harvester*, che visita *repository* esterni di tipo *Git* al fine di indicizzarne i metadati presenti e copiarne i relativi Modelli RDF (*Resource Description Framework*). L'attuale implementazione del componente *harvester* prevede un aggiornamento giornaliero tramite la schedulazione di un *job* sviluppato *ad hoc*.

Il processo di *harvesting* coinvolge i *Git* specificatamente configurati in [schema.gov.it](https://schema.gov.it). Al fine di mantenere i contenuti aggiornati, il processo di *harvesting* accede con frequenza giornaliera ai *repository Git*.

L'*harvester* accede direttamente ai *repository*, ne copia i metadati e i modelli RDF.

L'amministrazione pubblica che intende esporre i suoi *asset* semantici su [schema.gov.it](https://schema.gov.it) ha tre possibilità:

- fornire agli amministratori del Catalogo gli URL del proprio *Git* che sarà processato dall'*harvester*;
- fornire agli amministratori del *GitHub*/DAF i propri *asset* semantici;
- richiedere la gestione dei dati e della semantica all'Istat per la creazione dei propri *asset* semantici e la pubblicazione su *GISE*, il sistema automatico che si occupa delle operazioni di *data entry*.

Il portale [schema.gov.it](https://schema.gov.it), quindi, cataloga i *repository* di tipo *Git* dopo un esplicito accordo con i loro *owner*, verifica che i *Git* da sottoporre a *harvesting* presentino ontologie, schemi e vocabolari controllati redatti secondo le linee guida della semantica per la P.A. e che seguano le regole informatiche con cui l'*harvester* processa i contenuti semantici. Il processo di *harvesting* opera quindi in modo automatico procedendo al controllo sintattico delle ontologie, schemi e vocabolari controllati a ogni aggiornamento, ma non ne controlla il contenuto semantico.

### 5.4 La qualità dei dati

Poiché esistono rischi di inconsistenza di [schema.gov.it](https://schema.gov.it) dovuti all'accumulo successivo di materiali semantici, si pone un problema di qualità dei dati. Il Progetto NDC affronta in modo strutturato questo aspetto, proponendo tra le possibili soluzioni i seguenti strumenti di controllo:

- *auditing* manuale e periodico dei materiali raccolti dall'*harvester*. Assegnazione di *status* "in review" ai materiali semantici raccolti dall'*harvester* e non consistenti con i precedenti;
- validazione automatica del materiale semantico effettuata in fase di *harvesting*. Questo strumento sarà altamente innovativo e richiederà avanzate capacità di ricerca, pertanto sarà introdotto sul medio termine.

Analizzando temporalmente il processo di *harvesting*, è possibile suddividerlo nei seguenti *step*:

1. clone del contenuto del *repository* in una cartella temporanea;
2. eliminazione dei dati precedentemente salvati per il *repository*;
3. ricerca degli *asset* semantici nelle tre cartelle pertinenti, associate ai tipi di *asset* semantici (ontologie, vocabolari controllati e schemi), le quali devono contenere il file *turtle* da processare;
4. estrapolazione del modello RDF per ogni file *turtle* (in questa fase, avviene la validazione

- sintattica del file; pertanto, se sono presenti errori di sintassi, verranno registrati sul *log*, il file non verrà processato e si passerà al file successivo);
5. estrapolazione dei metadati dal modello RDF precedentemente estratto in fase di analisi, in modo che possano essere ricercati e resi fruibili tramite REST API e tramite interfaccia utente. In questa fase avviene anche la validazione della descrizione semantica del file; pertanto, se all'interno del file sono state omesse informazioni obbligatorie, anche in questo caso l'evento verrà registrato sul *log*, il file non verrà processato e si passerà al file successivo. Per ogni tipologia di *asset* semantico, vengono applicate regole di "validazione" diverse;
  6. salvataggio dei metadati estrapolati al punto precedente sul *server Elasticsearch*;
  7. salvataggio delle informazioni contenute nel modello RDF estrapolato al punto 4 su un *database server*;
  8. importazione dei vocabolari controllati, qualora vengano rilevati file *csv*, in un indice creato *ad hoc* in *Elasticsearch*, per il recupero successivo;
  9. eliminazione del contenuto del *repository* clonato al punto 1, una volta processati tutti i file contenuti nelle tre cartelle pertinenti, e termine dell'*harvesting*.

L'*harvester* può gestire e processare più di un *repository* che sia registrato e sia coerente con le regole del *w3id* processate; pertanto, il processo di *harvesting* così come descritto nei precedenti punti viene eseguito ciclicamente, ossia ripetuto per ogni *repository* acquisito dall'*harvester*.

## 5.5 Profili di sicurezza

Un ulteriore aspetto di fondamentale importanza per l'architettura del NDC è quello legato alla sicurezza, che garantisce l'integrità, la riservatezza e la disponibilità dei dati secondo lo standard ISO 27002. L'adozione di un robusto standard di sicurezza è essenziale per proteggere i dati sensibili e garantire la continuità delle operazioni.

Nelle applicazioni NDC la sicurezza è basata sullo standard OWASP (*Open Web Application Security Project*). OWASP rappresenta una pietra miliare nella gestione della sicurezza delle applicazioni web ed è ampiamente riconosciuto per le sue pratiche di sicurezza avanzate. Le tecniche OWASP, per la qualità del codice, sono considerate tra le *best practice* nell'ambito della prevenzione delle vulnerabilità di sicurezza.

Le vulnerabilità software potrebbero potenzialmente causare gravi conseguenze, tra cui perdite di dati, interruzioni dei servizi, accessi non autorizzati a informazioni sensibili e altri problemi critici. Per evitare tali problematiche, il software sottoposto a revisione è costantemente aggiornato per utilizzare *stack* tecnologici moderni e per ridurre, mitigare o eliminare, ove possibile, le vulnerabilità che potrebbero emergere durante le analisi di sicurezza.

Infine, il software NDC è soggetto periodicamente a ulteriori verifiche di sicurezza attraverso strumenti specializzati (Polizzi, M., e S. Letardi, 2023).

## RIFERIMENTI BIBLIOGRAFICI

- 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, *National Information Standards Series*.
- 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”. *Lecture Statistiche – Metodi*. Roma, Italia: Istat. <https://www.istat.it/it/archivio/259025>.
- 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-istat-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”. *Lecture Statistiche – Metodi*. Roma, Italia: Istat. <https://www.istat.it/it/archivio/278388>.
- 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. 9<sup>th</sup> 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 – 12<sup>th</sup> 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](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/it/archivio/278388>.
- Rosamond, B. 2003. "(European) Integration Theory, EU Studies and the Sociology of Knowledge". Paper presented to the 8<sup>th</sup> 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. 10<sup>th</sup> 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 Santaniello), 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 – UNECE. 2000. "Terminology on Statistical Metadata". Conference of European Statisticians, *Statistical Standards and Studies*, N. 53. Geneva, Switzerland: United Nations Publication.
- 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>.