

## A look at metadata processing beyond Libraries

Valdo Pasqui<sup>(a)</sup>

a) University of Florence, <https://orcid.org/0000-0002-4105-2208>

**Contact:** Valdo Pasqui, [valdo.pasqui@unifi.it](mailto:valdo.pasqui@unifi.it)

**Received:** 18 June 2022; **Accepted:** 30 June 2022; **First Published:** 15 September 2022

### ABSTRACT

Since many years libraries, archives and museums, the institutions entrusted with the dissemination and conservation of cultural heritage, contributed to metadata standards definition, meta-dating methodologies and metadata representation in different syntaxes, by participating to national projects and international initiatives.

Management systems and catalogs used in these contexts borrow from information and telecommunication technology tools, methodologies and techniques to generate, organize, share and use various types of metadata. But the Information and Communication Technology (ICT) area is not only a supplier of technological tools and solutions as it also constitutes a rich basin in which metadata plays a fundamental role in designing architectures, modeling information systems and implementing services.

Through an overview that includes national and European initiatives, especially in the digital transformation process of the public sector, this paper aims to offer a look to metadata beyond the traditional boundary of libraries and other cultural institutions and to underline some relevant aspects such as standardization, sharing, reuse and metadata quality assessment.

This framework highlights the need to carry on with the path of cooperation between different functional domains and organizational contexts in order to consolidate and extend the (re)use of metadata schemes, ontologies and controlled vocabularies both in the redesign of digitized processes and in the implementation of services supporting them. By undertaking since design early stages a multidisciplinary approach based on metadata standards can ensure greater flexibility and higher interoperability. This vision requires the enhancement of intersectorial skills that meld metadata methodologies and syntaxes representation basic knowledge with the ability to model functional domains using metadata schemes and ontologies.

### KEYWORDS

Metadata; Ontologies; Controlled vocabularies; Open Data; Web Services; API; Interoperability.

## Uno sguardo alla metadattazione oltre le Biblioteche

### ABSTRACT

I metadattati, le metodologie di metadattazione e le sintassi per rappresentare i metadattati sono ambiti nei quali le biblioteche, gli archivi e i musei, istituzioni alle quali è affidata la diffusione e la conservazione del patrimonio culturale, da molti anni hanno contribuito alla definizione di standard e allo sviluppo di metodologie grazie anche a progetti nazionali e iniziative internazionali. I sistemi gestionali e i cataloghi di cui si avvalgono questi contesti mutuano dalle tecnologie informatiche e delle telecomunicazioni tecniche e strumenti che consentono di generare, organizzare, condividere e usare le varie tipologie di metadattati. Ma l'area Information and Communication Technology (ICT) non è solo un fornitore di strumenti e soluzioni tecnologiche poiché costituisce anche un ricco bacino nel quale i metadattati svolgono un ruolo fondamentale per il disegno delle architetture, per la modellazione dei sistemi informativi e per l'implementazione dei servizi.

Attraverso una panoramica che comprende iniziative nazionali e europee, soprattutto nell'ambito della trasformazione digitale del settore pubblico, questo contributo intende offrire uno sguardo oltre al confine dei metadattati del mondo bibliotecario e delle altre istituzioni culturali e mettere in evidenza alcuni aspetti rilevanti come la standardizzazione, la condivisione, il riuso e la qualità dei metadattati.

Dal quadro esaminato emerge la necessità di proseguire nel cammino intrapreso per promuovere la cooperazione tra i diversi domini funzionali e i vari contesti organizzativi al fine di consolidare e ampliare l'uso degli schemi di metadattati, delle ontologie e dei vocabolari controllati nella progettazione dei processi digitalizzati e dei servizi che li supportano. Un approccio multidisciplinare che si fonda sull'uso di schemi di metadattati può assicurare fin dalle prime fasi della progettazione maggiore flessibilità e una più elevata interoperabilità. Questa visione richiede lo sviluppo di competenze intersectoriali che uniscano alla conoscenza delle metodologie di metadattazione e delle sintassi di rappresentazione dei metadattati anche la capacità di modellare i domini funzionali avvalendosi degli schemi di metadattati e delle ontologie.

### PAROLE CHIAVE

Metadata; Ontologie; Vocabolari controllati; Open Data; Web Services; API; Interoperabilità.

## 1. Spigolature in retrospettiva sulla metadattazione

Nell'affascinante libro “Payprus. L’infinito in un giunco” (Vallejo 2021) la studiosa spagnola Irene Vallejo ricorda che le tavolette di argilla usate dai Sumeri cinquemila anni fa sono state la prima forma di libro. A Nippur, nel sud della Mesopotamia, gli scavi archeologici hanno portato alla luce tavolette usate come cataloghi nelle quali ogni opera veniva identificata attraverso la prima riga del testo o un breve riassunto del contenuto, talvolta erano riportati anche il nome dell’autore, altri dati e per evitare la dispersione dei testi lunghi veniva indicato il numero delle tavolette che componevano l’opera. Nella biblioteca di Khattusa, capitale del regno ittita situata nell’odierna Turchia, sono state rinvenute tavolette contenenti i cataloghi di alcune collezioni. Si tratta di alcuni degli esempi più antichi di metadattazione a testimonianza di come, già con le prime tipologie di supporti fisici assimilabili ai libri, gli esseri umani abbiano avvertito la necessità di descrivere i libri e organizzare tali descrizioni in cataloghi. L’autrice afferma che “Un catalogo non è una semplice appendice della biblioteca: è il concetto su cui si basa, ne è il nesso e il culmine”. Questa essenza e centralità del catalogo si afferma e consolida proprio grazie ai metadati e alla metadattazione per mezzo dei quali prende forma, evolve nel tempo e svolge la propria funzione.

L’evoluzione degli strumenti e delle soluzioni offerte dall’ICT ha consentito non solo di strutturare e gestire in modo sempre più articolato e ricco i metadati delle biblioteche, per esempio passando dall’approccio basato sui record dei formati MARC a quello orientato alle relazioni basato su linked data, ma ha anche favorito l’uso dei metadati e lo sviluppo dei modelli di metadattazione in molti altri bacini di applicazione. La diffusione delle reti di telecomunicazione e del Web, la disponibilità dei dispositivi di memorizzazione (storage) che consentono l’acquisizione e l’archiviazione di grandi volumi di dati (database, testi, video, immagini, big data), l’elevata potenza di calcolo e le tecnologie del Cloud hanno consentito la digitalizzazione di processi sempre più complessi, distribuiti in rete e tra loro interoperanti, una trasformazione nella quale i metadati svolgono un ruolo fondamentale e che è avvenuta in tempi molto rapidi e in un arco temporale molto limitato se paragonati alla storia dei libri e delle biblioteche. In questo recente percorso si iscrive anche una curiosa storia che si colloca a cavallo tra tecnologia e linguistica. Nel 1955, quando negli Stati Uniti i primi calcolatori venivano chiamati Electronic Data Processing System (EDPS), denominazione poi sostituita da “computer”, ormai entrato a far parte del lessico comune anche nella lingua italiana, François Girard, responsabile del settore pubblicitario della IBM France, scrisse una lettera al suo ex insegnante Jacques Perret, professore di filologia latina presso la Sorbonne, per chiedergli suggerimenti sul nome francese da attribuire al nuovo modello di calcolatore che era in fase di realizzazione presso la fabbrica Corbeil-Essonnes (IBM 2014). La risposta dell’accademico esamina varie ipotesi e propone la parola “ordinateur”, anzi la declinazione al femminile “ordinatrice” aggettivata con “électronique” sia perché allora IBM aveva adottato alcuni nomi femminili come selezionatrice e tabulatrice, sia perché così sarebbe stato più chiara la differenziazione dal significato ecclesiastico dei termini “ordinare” e “ordinazione”. Perret motiva la proposta citando anche il Littré, Dizionario della lingua francese, in cui il termine “ordinateur” viene riferito a Dio che mette ordine nel mondo. La scelta dell’IBM sarà di adottare il termine “ordinateur” che da allora è il vocabolo correntemente usato in francese per denominare i computer.

Di fronte alla complessità delle informazioni e alla quantità dei dati questa aspirazione umana di “mettere ordine” si attua proprio attraverso gli schemi di metadati, la metadattazione e i cataloghi. I metadati sono stati usati fin dall’inizio anche nell’informatica sebbene assumendo nomi diversi come attributi, descrittori, proprietà. Il Multics, sviluppato nella seconda metà degli anni Sessanta in collaborazione tra i laboratori Bell, la General Electric e il MIT, è stato uno dei primi sistemi operativi che ha ispirato molti sistemi successivi tra cui UNIX. Una delle innovazioni di quel progetto fu l’ideazione del file system come il componente del sistema operativo cui è affidata la gestione dei dati e dei dispositivi di registrazione. Secondo una definizione classica (Wiederhold 1977) un file è un insieme di record simili composti di campi che contengono dati elementari. Nell’ambito bibliotecario ancora oggi i formati MARC si rifanno a questo approccio poiché sono stati concepiti proprio agli inizi dello sviluppo dell’automazione. Nei sistemi operativi dei nostri computer e tablet e nei servizi di condivisione come Google Drive, PCloud o Dropbox siamo abituati a chiamare file anche un documento PDF o Word e un’immagine JPEG, ovvero una qualunque sequenza di bit strutturata in base a un determinato formato. Le caratteristiche di questi file sono descritte da proprietà come il nome, la dimensione, il formato e i diritti di accesso/modifica, a loro volta questi metadati sono organizzati in cataloghi, chiamati “files directories” o “volume table of contents” (VTOC), utilizzati dal file system del sistema operativo per l’identificazione, la localizzazione e l’accesso ai dati. L’utente del sistema ne ha visibilità anche in forma leggibile come in Windows attraverso Esplora file o, a più basso livello, il classico comando “dir” del DOS (vedi Figura 1).

```
Directory di D:\DOCS\Articoli\Metadattazione_maggio_2022
14/04/2022 10:18 <DIR>          S301L\Valdo      .
14/04/2022 10:18 <DIR>          S301L\Valdo      ..
31/03/2022 10:51          301.710 S301L\Valdo     16 avril 1955 _ _Que diriez-vous d'ordinateur __.pdf
10/04/2022 19:00           66.433 S301L\Valdo     Appunti.docx
07/04/2022 18:23        12.745.981 S301L\Valdo     Bagley.pdf
31/03/2022 12:25         1.281.186 S301L\Valdo     Core_Vocabularies-Business_Location_Person_v1.00_Specif
11/04/2022 18:37           18.731 S301L\Valdo     Fwd I Seminario di JLIS.it del 19 maggio 2022 - Valdo
21/03/2022 22:26         113.388 S301L\Valdo     Guerrini_BibliotecheOggi_Metadattazione_2022.docx
12/03/2022 20:09         123.972 S301L\Valdo     Progetto DAFNE (progetto MIUR) Progetti e realizzazioni
11/04/2022 18:35         536.611 S301L\Valdo     Submissions JLIS.it.pdf
10/04/2022 17:33         552.706 S301L\Valdo     T-REC-X.500-200811-S!!PDF-E.pdf
08/04/2022 18:30           2.861 S301L\Valdo     testo_email.txt
14/04/2022 10:18          43.954 S301L\Valdo     Testo.docx
24/03/2022 19:26         1.647.254 S301L\Valdo     Understanding metadata NISO.pdf
      12 File      17.434.787 byte
       2 Directory 184.197.332.992 byte disponibili
```

Figura 1 Metadati che descrivono una directory DOS mostrati dal comando “dir”.

I file testo, immagini, audio e video (es. TIFF, JPEG, MP3, WAV, DOCX, PDF) riservano una porzione ai metadati (descrittivi, tecnici e di struttura) grazie ai quali i sistemi operativi li possono riconoscere e gestire. La specifica del formato PDF/A (PDF/A 2005), il cui uso è contemplato dalle norme nazionali in materia di conservazione degli archivi e dei documenti, fa sì che i documenti generati in conformità a tale formato contengano tutte le informazioni relative a testo, immagini raster e vettoriali, font di caratteri, colori, etc. necessarie per essere interpretati e visualizzati senza dipendere da dati residenti in sorgenti esterne.

Il termine metadata, “dati sui dati” (data about data), è originato in ambito ICT e la sua paternità è attribuita a Philip Bagley che nel 1968 propose alcuni concetti innovativi per la progettazione dei linguaggi di programmazione (Bagley 1968) quando i linguaggi più evoluti (es. Algol, Cobol, Fortran, PL/I) avevano una limitata capacità di gestire strutture di dati complesse. Nel seguente passaggio:

As important as being able to combine data elements to make composite data elements is the ability to associate explicitly with a data element a second data element which represents data “about” the first data element. This second data element we might term a “metadata element”. Examples of such metadata elements are: an identifier, a domain “prescriptor” which specifies from what domain the values of the first element must be taken, an access code which limits the which the first data element can be accessed

sono espressi il concetto di “metadata element”, ormai entrato nel linguaggio corrente dei principali schemi di metadati, la possibilità di restringere il dominio di valori che il dato descritto può assumere, l’identificatore, gli elementi compositi per rappresentare dati strutturati, le relazioni tra elementi, i domini di valori e i modelli utilizzati per raggruppare i metadati comuni ad una classe di elementi compositi. Si tratta di principi poi ripresi e sviluppati nella modellazione a oggetti e nei formalismi di definizione dei metadati e degli schemi di metadati come XML Schema.

Infine in questo breve excursus non si può rinunciare a citare la famiglia degli standard X.500 (ISO 9554), sviluppati<sup>1</sup> negli anni Novanta per assicurare l’interoperabilità dei directory services, servizi concepiti per gestire dati relativi a persone, organizzazioni (es. elenchi telefonici “pagine bianche”, “pagine gialle”, contatti aziendali, indirizzari etc.), risorse informatiche (computer, stampanti, dispositivi) e servizi applicativi. Attualmente l’autenticazione federata, il controllo degli accessi a servizi e dispositivi, l’infrastruttura a chiave pubblica (PKI) per i certificati digitali usati ai fini della trasmissione sicura (es. HTTPS) e la firma digitale si avvalgono di sistemi che implementano il Lightweight Directory Access Protocol (LDAP) basato sugli standard X.500. Il modello definito da X.500 prevede un insieme distribuito di Directory Information Tree (DIT) organizzati secondo una struttura gerarchica ad albero costituita da oggetti, chiamati entry, ciascuno dei quali è un insieme di attributi che hanno un tipo e assumono uno o più valori. Il directory schema definisce le classi di oggetti e per ciascuna gli attributi e se questi sono obbligatori o opzionali. Ogni oggetto (entry) può essere associato ad una o più classi di oggetti e ha un nome univoco detto distinguished name (DN), costituito dalla concatenazione di alcuni attributi dalla radice dell’albero fino alla entry. L’insieme delle definizioni dei tipi degli attributi, delle classi di oggetti e di altre informazioni che permettono di operare nel database che implementa la struttura gerarchica formano lo schema. Il documento (RFC2252 1997) del 1997 e i suoi successivi aggiornamenti formalizzano i tipi degli attributi, le sintassi e le classi di oggetti, a ciascuno di questi elementi è associato un identificatore, object identifier (OID) univoco e un nome descrittivo. Gli standard X.520 e X.521 definiscono insiemi di attributi e

---

<sup>1</sup> Lo standard è stato elaborato dal Consultative Committee for International Telephony and Telegraphy (CCITT), poi diventato International Telecommunication Union (ITU-T), in collaborazione con The International Organization for Standardization (ISO).

classi di oggetti che permettono di rappresentare le persone e le organizzazioni. I documenti (RFC 2256 1997), (RFC 2798 2000) e (RFC 4519 2006) definiscono una serie di attributi, alcuni dei quali sono riportati nella Tabella 1, e un insieme di classi di oggetti come “organization”, “organizationalPerson”, “person”, “device” che aggregano gli attributi base e permettono di descrivere oggetti appartenenti a determinate categorie.

Attributo	Descrizione
c	Country code / Codice della nazione
cn	Person's full name /Nome e Cognome
sn	Surname (family name) /Gognome
givenName	Part of a person's name which is not their surname nor middle name /Nome
l	LocalityName (e.g. city) /Città
st	stateOrProvinceName /Stato o Provincia
street	Street / Indirizzo
o	organizationName /Denominazione dell' organizzazione
title	Title of a person in their organizational context / Ruolo della persona nel contesto organizzativo
businessCategory	the kind of business performed by an organization / Attività dell'organizzazione
postalAddress	Indirizzo postale
postalCode	Codice postale
telephoneNumber	Numero telefonico
mail	Indirizzo di posta elettronica
mobile	Numero telefonico mobile
userPassword	Password
userCertificate	Certificato digitale dell'utente
cACertificate	Certificato digitale della Certification Authority

Tabella 1. Alcuni attributi usati negli schemi LDAP.

Nell'ambito delle istituzioni accademiche e di ricerca sono stati consolidati gli schemi eduPerson (eduPerson 2016), eduOrg e Schema for Academia (SCHAC) (SCHAC 2022) che consentono di descrivere persone ed enti che operano nei contesti della formazione superiore<sup>2</sup>.

I concetti di identificatore univoco, per identificare gli attributi e gli oggetti, di classi astratte di oggetti, per descrivere oggetti più complessi a partire da elementi descrittivi elementari (attributi), di schema, per organizzare gli attributi/metadati attraverso meccanismi di inclusione e riuso di insiemi già definiti, sono analoghi ai principi fondanti delle correnti modalità di metadattazione basate su XML e su XML Schema Definition.

<sup>2</sup> Attualmente la manutenzione e lo sviluppo di questi schemi e dei servizi di cooperazione interistituzionale correlati è svolto dal Research and Education FEDerations group (REFEDS) (REFEDS n.d.) che a sua volta opera nell'ambito di GÉANT (GÉANTn.d.) costituito dalla European National Research and Education Networks (NRENs).



## 2. Metadati e Metadazione nell'ecosistema digitale

Nel paragrafo precedente abbiamo visto come la metadazione abbia accompagnato fin dall'inizio lo sviluppo dell'ICT risultando nella definizione di insiemi (schemi) di metadati per descrivere oggetti del mondo reale e oggetti nativi del mondo digitale come i file, i data set, i dispositivi, i servizi e le applicazioni. I metadati sono ormai pervasivi nello sviluppo dei sistemi informativi e nella digitalizzazione dei processi sia nell'ambito detto business to business (B2B) sia nel settore pubblico attraverso i piani per la trasformazione digitale della Pubblica Amministrazione.

Le due organizzazioni W3C e OASIS (OASIS n.d.) hanno definito standard, metodologie e strumenti per la produzione, la gestione e la condivisione dei metadati. Risale ai primi anni Duemila la diffusione di XML, la definizione di XML schema per descrivere i documenti XML e la successiva specifica XML Schema Definition (XSD) (XSD 2012) che hanno consentito la formalizzazione, l'importazione e il riuso di molteplici schemi di metadati. Il termine "namespace", che in XML esprime il principio della modularità mediante la condivisione di elementi e attributi ("markup vocabulary") definiti in moduli software distinti, è ormai diventato di uso corrente per indicare l'importazione e il riutilizzo di insiemi di metadati. Parallelamente, l'attività di sviluppo e di standardizzazione del Web Semantico, ispirata dal modello Linked Data ideato da Tim Berners-Lee (Berners-Lee 2006), ha portato alla definizione nel 2014 del Resource Description Framework (RDF) che rappresenta l'informazione del Web come un grafo connesso (RDF 2014) costituito da triple soggetto, oggetto, predicato. Gli RDF Schema (RDFS) permettono di formalizzare classi, tipi di dati, domini delle proprietà, relazioni gerarchiche tra le classi e le proprietà, detti anche RDF Vocabulary, e sono diventati una delle modalità più diffuse per definire schemi di metadati. Recentemente si è affermata la notazione JSON-LD (JavaScript Object Notation for Linked Data), una "raccomandazione" nell'ambito del W3C che specifica come codificare (serializzare) i linked data secondo la sintassi JSON (JavaScript Object Notation) molto diffusa nell'ambito degli sviluppatori delle interfacce programmatiche (API). Questi modelli e le relative specifiche hanno agevolato l'utilizzo dei metadati e la definizione di schemi metadazione in molti settori dell'ICT che per comodità di esposizione e in base alle finalità di applicazione possono essere schematizzati in tre ambiti:

- a) assicurare l'interoperabilità dei dati tra domini funzionali diversi
- b) facilitare la produzione, diffusione e il riuso dei data aperti (Open Data)
- c) agevolare l'interoperabilità applicativa tra i servizi di rete.

Per ciascuna delle tre aree individuate esamineremo alcuni aspetti relativi alla definizione e all'uso dei metadati con un'attenzione prioritaria per le iniziative rivolte al settore pubblico in ambito europeo e italiano. Giova ricordare l'esistenza di altri ambiti come quello dei siti Web, dei social network, dei servizi multimediali per la diffusione di video e musica e della gestione dei diritti d'autore nei quali sono stati definiti schemi di metadati (es. FOAF, Schema.org, ODRL, PRISM) che per ragioni di spazio non è possibile trattare in questo contesto.

## 3. Metadati e interoperabilità dei dati

La necessità di assicurare l'interoperabilità dei dati appartenenti ad ambiti organizzativi e a domini funzionali eterogenei è un'esigenza ormai prioritaria per le pubbliche amministrazioni (e-government), nel settore commerciale e in quello della ricerca, oltre che per la conservazione e valorizzazione

zione del patrimonio culturale. Le ontologie, i vocabolari controllati, il riuso di schemi di metadati sono punti di riferimento imprescindibili per ogni progetto che intenda soddisfare tali requisiti.

### 3.1 Il contesto europeo

La “Strategia europea per i dati” sottolinea la centralità dei dati per il futuro dell’economia europea e esprime la volontà di creare uno “spazio unico dei dati”, cioè un mercato unico che favorisca lo sviluppo di nuovi prodotti e servizi basati sui dati (personali, non personali e dati rilevanti in ambito commerciale e industriale) accessibili in modo sicuro (EUDati 2020). Una delle prime iniziative europee con queste finalità è stata la Direttiva europea “Infrastructure for spatial information in Europe (INSPIRE)” volta a creare un’infrastruttura europea di supporto all’attuazione delle politiche ambientali e di salvaguardia del territorio (INSPIRE, 2007), attraverso la condivisione e l’accesso pubblico ai dati territoriali e ai dati inerenti le attività che impattano sull’ambiente. Tra le regole di implementazione comuni la Direttiva individua i metadati poiché essi assicurano l’interoperabilità tra i dati spaziali/territoriali e i servizi che li rendono disponibili (INSPIRE\_TG 2022).

Un altro obiettivo delle strategie messe in atto a livello europeo e rafforzate dalla Dichiarazione di Tallin del 2017 (TALLIN 2017) è realizzare il principio once-only mediante lo sviluppo di servizi pubblici digitali transnazionali (cross-border) grazie a i quali i cittadini non siano più costretti a fornire alle amministrazioni documenti e dati già in possesso di altre amministrazioni o autorità pubbliche. A tale finalità sono riconducibili due iniziative:

- la piattaforma collaborativa Joinup, realizzata per favorire la condivisione e lo sviluppo di servizi interoperabili (Joinup\_CV 2022);
- il Single Digital Gateway (SDG), attivo dal Dicembre 2020, un portale concepito per aiutare i cittadini e le imprese a trovare informazioni sulle norme, i diritti e i procedimenti sia europei sia nazionali con l’obiettivo di attivare entro il 2023 un insieme di 21 procedure amministrative e di rendere pienamente accessibili a livello transnazionale i servizi online nazionali (SDG n.d.).

Questi progetti sono volti a promuovere la cultura del riuso dei dati nelle amministrazioni, prevedono la creazione di registri autoritativi (basic registries) che assicurino l’autenticità dei dati, la loro reperibilità, qualità e accessibilità, e la condivisione di schemi di metadati, vocabolari controllati e ontologie che permettono il funzionamento dell’intera infrastruttura di servizi. Nell’ambito Joinup si segnala il lavoro che ha condotto alla definizione degli e-Government Core Vocabularies (e-GOV n.d.) che comprendono:

- Core Business Vocabulary - Rappresenta le caratteristiche delle entità legali (“trading bodies” formalmente registrate da qualche autorità e presenti in qualche registro) come il nome legale, l’attività, l’indirizzo, la tipologia etc.
- The Core Location Vocabulary - Coglie le caratteristiche dei luoghi come l’indirizzo, l’identificatore geografico (un URI definito da GeoNames.org oppure un DBpedia URI), il nome geografico (secondo le specifiche INSPIRE) etc.
- The Core Person Vocabulary - Esprime le caratteristiche fondamentali di una persona intesa come individuo reale (nome, sesso, data di nascita, etc.)

- The Core Public Organization Vocabulary (CPOV) - Descrive le organizzazioni pubbliche che operano nell'Unione Europea (organizzazione, le unità organizzative, le unità amministrative territoriali, i contatti, le ore di apertura etc.)
- Core Public Service Vocabulary Application Profile (CPSV-AP) - Descrive le caratteristiche di servizi offerti dalle pubbliche amministrazioni (titolo, descrizione, input, output, fornitori, luoghi, etc.)
- The Core Evidence and Criterion Vocabulary - Descrive i principi e le modalità che un'entità privata deve soddisfare per essere qualificata a realizzare servizi pubblici o a partecipare alle forniture pubbliche. Per "criterio" si intende una regola usata per valutare, giudicare o testare e per "evidenza" il modo attraverso il quale un criterio può essere comprovato.

Questi vocabolari, espressi in sintassi RDF, sono modelli riusabili e estensibili che catturano e descrivono le caratteristiche di entità tipiche del settore pubblico e che possono essere usati per:

- avviare la progettazione dei modelli logici e concettuali di nuovi sistemi informativi
- modellare dati specifici usati per l'interscambio tra sistemi informativi esistenti
- integrare dati provenienti da differenti sorgenti
- pubblicare dati aperti attraverso cataloghi, registri e portali pubblici (vedi il paragrafo seguente)

al fine di assicurare il riuso e la condivisione dei concetti base, l'interoperabilità semantica dei modelli concettuali e l'utilizzo di sintassi standard per esprimerle.

### 3.2 Il contesto italiano

La centralità dei dati è uno degli elementi cardine del Piano Triennale per l'informatica nella Pubblica Amministrazione fin dalla sua prima versione (2017 – 2019) che ha impostato una strategia basata sull'individuazione delle basi dati di interesse nazionale e sulla condivisione di modelli dati (ontologie) e vocabolari controllati. La versione 2022-2023 del Piano, nell'ambito dell'Obiettivo 2.2 volto a migliorare la qualità dei dati e dei metadati, ha introdotto una nuova linea d'azione che fissa per Giugno 2022 il rilascio del "National Data Catalog per l'interoperabilità semantica" a cura del Dipartimento per la Trasformazione Digitale (PTIPA 2021).

Il Portale dei dati aperti della pubblica amministrazione (cfr. il prossimo paragrafo), il Sistema dei Registri INSPIRE Italia (INSPIRE\_REG\_IT, n.d.), l'insieme delle ontologie OntoPiA e i vocabolari relativi ad alcuni settori pubblici costituiscono un insieme di strumenti concepiti allo scopo di normalizzare i dati in possesso della PA, offrire alle imprese e ai privati punti di riferimento ufficiali per il popolamento delle loro basi di dati e agevolare lo scambio di dati tra le PA.

La Figura 2 è una rappresentazione funzionale della rete di ontologie OntoPiA nella quale si rintracciano quelle relative alle Persone, Organizzazioni e Luoghi, la Cultural Heritage Ontology che sistematizza le ontologie sviluppate nel biennio 2017-2018 dal progetto ArCo del MiC (ArCo 2017), oltre alle classificazioni per il settore cultura, le tipologie di luoghi di interesse culturale e alcune classificazioni per l'ambito universitario (ruoli accademici, settori scientifico-disciplinari).



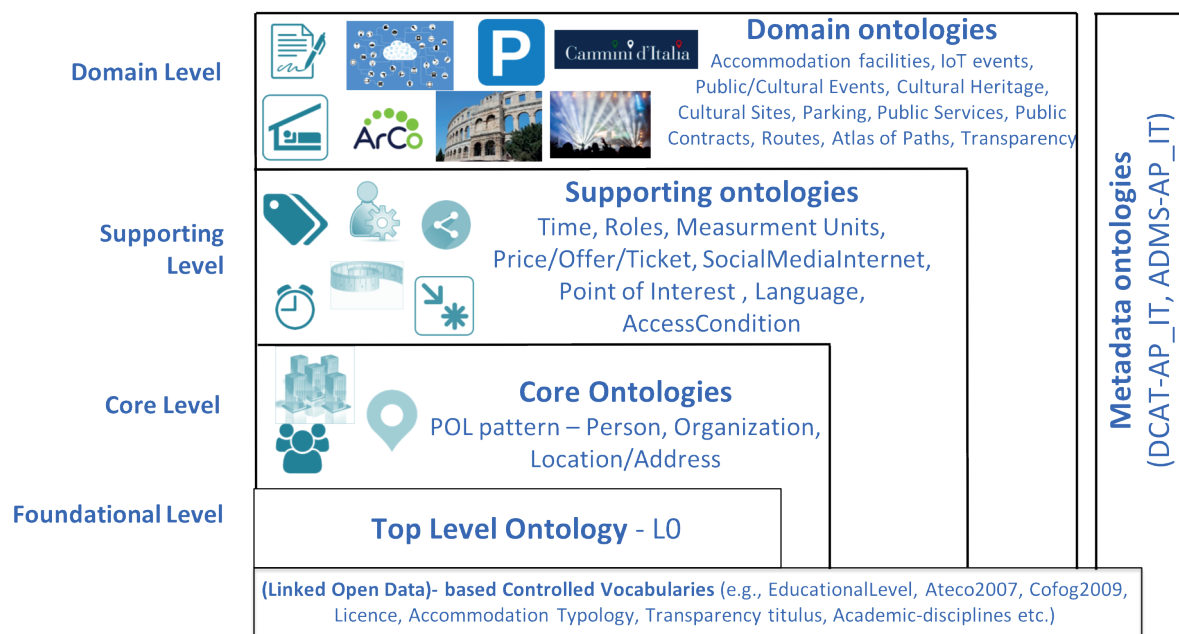


Figura 2 Rete OntoPiA tratta da <https://github.com/italia/daf-ontologie-vocabolari-controllati>

il PNRR comprende anche una specifica linea d'azione ("misura") denominata "1.3.2 Sportello Digitale Unico" volta alla implementazione della componente italiana del Single Digital Gateway, alla quale sono interessati 230 enti (PA Centrali e Locali, Università e Istituti di Ricerca) di cui 55 coinvolti nel progetto che prevede di realizzare 10 procedure online entro dicembre 2022 (target nazionale) e giungere a 21 entro dicembre 2023 (target finale europeo). L'AgID è il soggetto attuatore (AgID, 2022) di questo progetto, declinato come processo di reingegnerizzazione e introduzione di nuove tecnologie digitali. Le Università hanno attivato un tavolo tecnico al fine di impostare le soluzioni architettoniche e funzionali più idonee per l'attivazione dei servizi relativi all'orientamento, all'iscrizione ai test ai corsi a numero programmato e alle immatricolazioni che tengano conto delle complesse e differenti articolazioni degli Atenei. I metadati e i vocabolari controllati svolgono un ruolo abilitante per la realizzazione dei servizi SDG ai fini della normalizzazione e riconciliazione di dati e procedure eterogenei.

### 3.3 Esempi di interoperabilità tramite metadati

Avvalersi di schemi di metadati consolidati agevola l'integrazione dei dati tra domini diversi anche attraverso la mappatura di dati eterogenei verso schemi comuni. Un esempio assai diffuso di questa pratica sono i metadati Dublin Core spesso usati per trasformare dati bibliografici provenienti da sorgenti eterogenee per esempio nei processi di raccolta tramite protocolli di harvesting come il ben noto OAI-PMH. In ambito informatico è molto diffusa la pratica della mappatura per riconciliare tra loro i dati provenienti da basi di dati differenti per poter effettuare elaborazioni complesse, trasversali e che richiedono l'integrazione e la correlazione di dati appartenenti a diversi domini applicativi.

Nei sistemi di Business Intelligence (BI) e Data Warehouse (DW) e nell'analisi di grandi volumi di dati (big data) esistono processi, metodologie e linguaggi, denominati Extract-Transform-Load (ETL) che permettono di interpretare, elaborare, trasformare, riconciliare e normalizzare dati provenienti da sorgenti diverse. I metadati che descrivono i dati e le procedure (raccolta, estrazione, pulizia, trasformazione, mappatura dei tipi e caricamento dei dati, struttura delle aree intermedie usate per la trasformazione, modellazione logica fisica del database del DW, uso degli strumenti di analisi e reportistica) guidano questi articolati e complessi processi al punto che si usa anche l'espressione metadata driven ETL.

Un altro esempio di metadati concepiti ai fini dell'interoperabilità semantica è la recente diffusione nell'ambito dell'istruzione superiore di pacchetti informativi digitali contenenti metadati che certificano le abilità, le competenze e i titoli conseguiti da una persona (es. certificazioni linguistiche e informatiche/digitali, crediti e titoli accademici), il soggetto che li ha conseguiti e quello che li ha emessi. Lo IMS Global Learning Consortium (IMS Global/IMS) dal 2018 ha definito gli Open Badge la cui specifica più recente (OpenBadges 2021) include lo Open Badges Vocabulary che formalizza classi, proprietà e valori per esprimere gli Open Badge come linked data rappresentati in sintassi JSON-LD. Grazie ai linked data la specifica garantisce l'estendibilità, il multilinguismo e l'uso di identificatori unici per identificare gli oggetti descritti e i valori delle proprietà. La conformità a questo schema di metadattazione consente agli istituti di formazione pubblici e privati di emettere certificazioni e titoli che possono essere interpretati e elaborati automaticamente ai fini della valutazione e del riconoscimento in contesti lavorativi e formativi diversi, a vantaggio dei soggetti che detengono i badge per o quali così si semplifica e si velocizza il conferimento dei dati.

Il Sistema Pubblico di Identità Digitale (SPID) permette di accedere a tutti i servizi online della Pubblica Amministrazione italiana con un'unica Identità Digitale. Oltre agli utenti (cittadini e imprese) il sistema comprende due tipologie di attori: i) i gestori dell'identità digitale (identity provider o IdP), accreditati da AgID per la gestione, la creazione e il rilascio agli utenti delle identità digitali; ii) i fornitori di servizi (service provider o SP), organizzazioni pubbliche o private che consentono l'accesso protetto ai propri servizi tramite l'identità digitale. Ciascuna di queste entità è descritta da un insieme di metadati (SPID\_RT n.d.) conformi allo standard SAML v2.0 (SAML-Metadata). I metadati sui gestori e i fornitori dei servizi sono consultabili e scaricabili in formato aperto da un registro pubblico (SPID\_REG n.d.). L'interazione tra gli Identity Provider e i Service Provider avviene tramite asserzioni espresse in SAML (SAML n.d) che contengono metadati che descrivono i soggetti, persone fisiche, enti o aziende (SPID\_AT n.d.). Quando un utente accede ai servizi di una pubblica amministrazione (es. Agenzia delle Entrate, INPS, etc.) viene rinvio al proprio provider SPID e dopo aver inserito le proprie credenziali è reindirizzato al servizio iniziale insieme ad un pacchetto di metadati<sup>3</sup> che lo identificano. Grazie a questi metadati i fornitori di servizio possono riconoscere l'identità dei soggetti, attuare politiche di controllo degli accessi e realizzare controlli di tipo autorizzativo.

---

<sup>3</sup> L'accesso al portale INPS con SPID comporta, previa autorizzazione esplicita dell'utente, il trasferimento del seguente pacchetto di metadati: Luogo di Nascita, Sesso, Numero di telefono mobile, Cognome, Domicilio digitale, Nome, Codice identificativo SPID, Data di nascita, Provincia di nascita, Codice Fiscale, Indirizzo di posta elettronica.

## 4. Metadati come elemento abilitante per il riuso dei dati aperti

I dati ad accesso aperto comprendono una molteplicità di tipologie e di produttori, da quelli generati nell'ambito amministrativo ai dati prodotti nell'ambito della ricerca finanziata con fondi pubblici secondo i principi FAIR (FAIR n.d.). Per condividere dati aperti non è sufficiente pubblicare file in formato aperto (es. csv, xml, json etc.), ma occorre corredarli di metadati che ne consentano la scoperta e il riuso attraverso cataloghi e motori di ricerca. Il Codice per l'Amministrazione Digitale (CAD 2021) all'Articolo 1 lettera l-ter precisa che i dati aperti "sono adatti all'utilizzo automatico da parte di programmi per elaboratori e sono provvisti dei relativi metadati".

La Direttiva Europea sui Dati Aperti (Directive (EU) 2019/1024 2019), entrata in vigore il 19 Luglio 2019, si rivolge alle pubbliche amministrazioni, agli enti e alle imprese che operano in questo ambito con l'intento di promuovere l'utilizzo dei dati aperti e il riuso dei dati fornendo un quadro di riferimento omogeneo e condiviso che include la disponibilità di interfacce programmatiche (API) aperte e l'utilizzo di formati di dati aperti riconosciuti a livello internazionale. Il "portale ufficiale dei dati europei" data.europa.eu mette a disposizione un rilevante patrimonio di dati aperti, 174 Cataloghi, 36 Paesi e 1.444.438 Set di dati (al 21 aprile 2022), e include due cataloghi italiani:

- il Catalogo Italiano dei Dati Territoriali (Spatial Data Portal Italy) creato per dare seguito alla direttiva europea INSPIRE (<https://geodati.gov.it/geoportale>)
- il Portale dei dati aperti della Pubblica Amministrazione (Open Data Portal) dati.gov.it (<https://www.dati.gov.it/>)

Questa iniziativa europea consente di estrapolare tre punti che costituiscono una linea guida per qualunque progetto relativo alla produzione di metadati:

- a) l'uso di uno schema di metadattazione di riferimento comune;
- b) il consistente riuso di termini e metadati mutuati da altre specifiche per la definizione di tale schema;
- c) l'attenzione per la qualità dei metadati attraverso una metodologia di analisi e valutazione basata su specifiche metriche.

Al fine di garantire il corretto harvesting dei metadati e dei dati e per accrescere la capacità di scoperta (discoverability) dei datasets e dei servizi i vari cataloghi si devono conformare al "DCAT Application Profile for data portals in Europe" (DCAT-AP), definito da un'iniziativa congiunta di tre organismi europei (DG CONNECT, EU Publications Office, Interoperable Europe Programme). Questo profilo è basato sul Data Catalog Vocabulary (DCAT), un modello e un vocabolario espresso in RDF per descrivere i dataset e i servizi offerti da cataloghi pubblicati sul web (DCAT 2020), e consente di descrivere i dataset delle pubbliche amministrazioni europee e i relativi cataloghi in modo da supportare lo scambio dei metadati e le ricerche tra i portali dei dati aperti del settore pubblico (DCAT-AP 2022). Il profilo prevede una serie di classi (Agent, Catalogue, Dataset, Data Service, Document, Media type, Relationship, Provenance statement, Rights statement etc.) ciascuna delle quali ha delle proprietà (obbligatorie, raccomandate, opzionali), identificate tramite un URI, in gran parte mutate da namespace standard: Asset Description Metadata Schema (ADMS), Location Core Vocabulary (LOCN), DCMI Metadata Terms (DCTT), FOAF, Open Digital Rights Language (ODRL), W3C PROV Namespace, Schema.org, Simple Knowledge Organization System (SKOS), Software Package Data Exchange (SPDX), Vocabulary of a Friend (VOAF), vCARD.

Il portale europeo attua anche una metodologia di misurazione della qualità dei metadati basata sui criteri FAIR e mediante alcuni strumenti di analisi e valutazione viene periodicamente generato un rapporto che classifica la qualità dei metadati prodotti dai servizi e cataloghi attivati dai vari Paesi. La Figura 3 mostra gli indicatori di qualità, raggruppati nei cinque ambiti FAIR, che vengono misurati dallo strumento di valutazione della qualità dei metadati denominato Metadata Quality Assessment (MQA) e il punteggio riportato dal portale italiano dei dati aperti.



Figura 3 Dashboard del portale europeo relativa alla valutazione della qualità del portale italiano dei dati aperti dati.gov.it (<https://data.europa.eu/mqa/catalogues/dati-gov-it/?locale=en>)

Tra gli indicatori dell'interoperabilità Format misura se nella descrizione dei dataset è usata il meta-dato "dct:format", il valore assunto da Machine readable deriva dalla verifica se il formato del dataset è uno di quelli previsti (JSON, CSV, XLS, XLSX, vari formati RDF etc.), Non proprietary misura se il formato del file appartiene a quelli classificati come non proprietari (es. TXT, RTF, PNG, ZIP, GZIP, JPEG2002, TAR, TIFF etc.), l'indicatore Format/Media type from vocabulary valuta se i valori dei due metadati dct:format e dcat:mediaType appartengono rispettivamente ai vocabolari controllati previsti da data.europa e ai Media Types definiti da IANA (Media Types 2002).

La valutazione dei due portali italiani nel rapporto del 3 Aprile 2022 è “appena sufficiente” per dati.gov.it e “scarso” per il catalogo dei dati territoriali a dimostrazione del fatto che occorrono ulteriori investimenti (organizzativi, infrastrutture, risorse umane) per migliorare la metadattazione dei dati aperti e per renderli effettivamente utilizzabili. Il CAD, con l’Art.1 e il Titolo 5, le Linee Guida Nazionali per la Valorizzazione del Patrimonio Informativo Pubblico (LGNVPIP 2020) e le varie versioni del Piano Triennale per l’informatica nella Pubblica Amministrazione hanno promosso la produzione e la diffusione dei dati aperti, attraverso l’attivazione del portale italiano e la definizione del profilo italiano dei metadati DCAT-AP\_IT a cui dovrebbero conformarsi tutte le amministrazioni per alimentare il catalogo nazionale (DCAT-AP\_IT 2022). Lo scorso 15 dicembre 2021 è entrato in vigore il decreto legislativo n. 200/2021 che recepisce la direttiva (UE) 2019/1024 sui Dati Aperti e l’AgID ha attivato il coordinamento di un tavolo tecnico con i Responsabili della Transizione per definire delle Linee guida contenenti regole tecniche relative al tema dei dati aperti, la cui adozione secondo l’attuale versione del Piano Triennale (CAP2.LA08) è prevista per dicembre 2022. Infine, un’ulteriore interessante iniziativa da citare è il portale Ocds (OCDS\_Port n.d.) realizzato dall’Agenzia Nazionale Anticorruzione (ANAC) che consente di consultare e confrontare le informazioni degli appalti pubblici contenute nella banca dati ANAC<sup>4</sup>. Il portale segue le specifiche dello standard OCDS, elaborate dall’organizzazione non-profit Open Contracting Partnership, che comprende alcuni schemi di metadati concepiti per modellare i soggetti e rappresentare tutte le entità dei contratti e che sono espressi in JSON (OCDS, n.d.).

## 5. Metadati e interoperabilità applicativa

La disponibilità di infrastrutture di rete sempre più performanti per velocità, volumi di traffico sostenibile e ridondanza ha consentito di digitalizzare molti processi sia delle pubbliche amministrazioni sia nelle aziende private, utilizzando modalità di cooperazione basate sull’integrazione dei servizi distribuiti in rete e operanti nel Cloud. Le applicazioni interagiscono e cooperano tra di loro grazie alle interfacce programmatiche (API) e a web services che consentono di esporre funzionalità per la creazione, la modifica e la cancellazione dei dati.

Le architetture orientate ai servizi si sono particolarmente affermate a partire dagli anni Duemila e la W3C “Web services Architecture” definisce un Web service come un “software system designed to support interoperable machine-to-machine interaction over a network” (W3CWS 2004). Inizialmente i WS erano basati sul protocollo SOAP, concepito per scambiare messaggi in formato XML, tra un mittente e un ricevente (sender e receiver), in modo indipendente dal sottostante protocollo di rete (di solito HTTP/HTTPS). Più di recente si sono diffusi i web service realizzati in conformità allo stile architetturale Representational State Transfer (REST) (Fielding 2000) che ormai è il più comunemente utilizzato per sviluppare interfacce programmatiche dette Web API o REST API. REST si basa su alcuni principi tra i quali il più pertinente per la tematica che stiamo trattando è il concetto astratto di risorsa, “Any information that can be named can be a resource”. Ogni risorsa è identificata da un identificatore (URI), in un determinato istante di tempo assume

---

<sup>4</sup> ANAC ha lanciato una competizione (datathon) nella quale i concorrenti sono stati invitati a proporre, entro il 22 maggio 2022, esempi di utilizzo dei dati Ocds (Open Contracting Data Standard) dell’Autorità.



uno stato chiamato “resource representation” costituito dai dati (es. un’immagine, un documento), i metadati che descrivono la risorsa e i link ipermediali che permettono di modificarne lo stato. In entrambe gli ambiti, SOAP e REST, sono state definite modalità di descrizione dei servizi. Il Web Services Description Language (WSDL) è solitamente utilizzato per pubblicare le specifiche dei web service SOAP in base a XML Schema avvalendosi di elementi come types (descrive i tipi di messaggi che il servizio invia e riceve), interface (attraverso gli elementi operation, input e output descrive le funzioni astratte che il Web service fornisce), binding (indica con quale protocollo e quali operazioni si accede al servizio), service (mediante l’elemento endpoint il cui valore è un URI descrive dove accedere al servizio). La specifica WSDL 2.0 (WSDL 2.0 2007) consente anche la descrizione dei web services di tipo RESTful. La OpenAPI Specification (OAS) è uno schema di metadate appositamente concepito per descrivere, attraverso un OpenAPI document in sintassi JSON, questa tipologia di web services avvalendosi di una serie di oggetti (OpenAPI Object, Info Object, Contact Object, License Object, Server Object, Paths Object, Operation Object, etc.) e delle relative proprietà/metadati (OAS n.d.). I metadati strutturati di WSDL e OAS permettono di testare e integrare rapidamente le API esposte da altri servizi e di generare, pubblicare e condividere nuove interfacce programmatiche consentendo agli sviluppatori di generare automaticamente il codice applicativo (es. classi e metodi Java, librerie PHP) senza doversi preoccupare dei dettagli e di svolgere attività soggette ad errore.

Nel contesto nazionale nel corso degli ultimi anni si è assistito ad un notevole sviluppo di sistemi orientati ai servizi, basti pensare alle modalità di pagamento elettronico (dall’internet banking all’uso delle carte di credito tramite POS virtuali fino al sistema di pagamento per le Pubbliche amministrazioni PagoPA), alle transazioni commerciali business-to-business (B2B), al recupero delle attestazioni ISEE, al Sistema informativo sulle operazioni degli enti pubblici (SIOPE+) e l’Anagrafe Nazionale della Popolazione Residente (ANPR). Il Sistema d’Interscambio (SdI) dell’Agenzia delle Entrate per l’invio e la ricezione delle fatture elettronica con le PA (SDI n.d.) e il Nodo di smistamento degli ordini di acquisto delle amministrazioni pubbliche (NSO) gestito dalla Ragioneria dello Stato (NSO n.d.) espongono API che permettono ai gestionali locali (es. i sistemi di gestione ordini e contabilità) di integrarsi per gestire lo scambio delle fatture e degli ordini in formato elettronico. È utile notare come da un lato gli oggetti digitali così scambiati (certificati, dati relativi al reddito, incassi e pagamenti delle pubbliche amministrazioni ai propri tesoreri/cassieri, fatture, ordini di acquisto etc.) sono accompagnati da metadati strutturati che ne permettono la corretta costruzione e interpretazione e dall’altro i servizi invocati per cooperare sono a loro volta descritti tramite appropriati metadati che permettono lo sviluppo delle applicazioni (programmi) che li utilizzano. Il Dipartimento per la trasformazione digitale del Ministero per l’innovazione tecnologica e la transizione digitale (MITD) in collaborazione con l’AgID ha definito le linee guida e il Modello di Interoperabilità allo scopo di rendere possibile la collaborazione tra le Pubbliche Amministrazioni e tra queste e soggetti terzi attraverso interfacce programmatiche (API) conformi agli standard consolidati<sup>5</sup>. Inoltre Il Dipartimento supervisiona la realizzazione della Piattaforma Digitale Nazionale

---

<sup>5</sup> AgID con la Determinazione n. 547 del 1 ottobre 2021, ha adottato le Linee guida sull’interoperabilità tecnica delle Pubbliche Amministrazioni e le Linee guida Tecnologie e standard per la sicurezza dell’interoperabilità tramite API dei sistemi informatici [https://www.agid.gov.it/sites/default/files/repository\\_files/547\\_dt\\_dg\\_n\\_547\\_1\\_ott\\_2021\\_adozione\\_lg\\_interoperabilit\\_tecnica\\_e\\_sicurezza.pdf](https://www.agid.gov.it/sites/default/files/repository_files/547_dt_dg_n_547_1_ott_2021_adozione_lg_interoperabilit_tecnica_e_sicurezza.pdf).

Dati (PDND), affidata alla società PagoPA Spa, che renderà possibile l'interoperabilità dei sistemi informativi e delle basi di dati delle pubbliche amministrazioni e dei gestori di servizi pubblici<sup>6</sup>.

## 6. Considerazioni e prospettive

Le iniziative citate e gli esempi esaminati nei paragrafi precedenti mostrano come i metadati abbiano ormai raggiunto un elevato livello di pervasività che si estende dai ben noti strumenti di ricerca e cataloghi in rete fino all'interoperabilità semantica tra diversi contesti applicativi e domini funzionali eterogenei. Questo scenario comprende sistemi e servizi che interagiscono e cooperano secondo modalità guidate dai metadati per integrazione diretta, attraverso web service e interfacce programmatiche, oppure mediante la mappatura di diverse sorgenti dati in uno schema di metadati condiviso o, infine, attraverso processi di trasformazione, riconciliazione e integrazione di dati in origine diversi avvalendosi di specifici insiemi di metadati.

Gli schemi di metadati e le ontologie sono gli strumenti attraverso i quali è possibile realizzare soluzioni come il Single Digital Gateway, gli Open Badge e i sistemi di identità federata che permettono di ridisegnare molti processi adottando i principi dell'only-once e della cooperazione interistituzionale. I metadati sono un fattore abilitante per l'evoluzione dell'ecosistema digitale formato dalla rete di servizi e applicazioni e per implementare processi complessi nei quali operano in ambito transnazionale enti e soggetti del settore pubblico e di quello privato. Svariate iniziative e progetti, a livello europeo e italiano, hanno dato luogo alla definizione di schemi di metadati e ontologie che coprono porzioni significative dell'universo informativo di questo bacino di applicazione e hanno promosso la diffusione dei metadati e la conoscenza delle tecniche di metadattazione. Tuttavia i sistemi informativi e le applicazioni delle Pubbliche Amministrazioni usano ancora in modo insufficiente e parziale schemi di metadati standardizzati. Tra le cause di questa carenza si possono annoverare l'esistenza di molte applicazioni datate, la tendenza delle aziende a privilegiare soluzioni verticali e poco standardizzate per mantenere vincolati i clienti ai propri prodotti (fenomeno noto come vendor lock-in) e la maggiore priorità degli adeguamenti dei software alle linee guida sulla usabilità, l'accessibilità e la sicurezza. Ma nel settore pubblico vanno posti in evidenza anche altri due fattori inibitori: la nota inerzia al cambiamento e i limiti delle risorse economiche disponibili per lo sviluppo di soluzioni innovative, per la formazione delle competenze e per l'acquisizione di nuove risorse umane. Alcune linee d'azione del PNRR sono destinate a dare un impulso al processo di cambiamento, ma resta da chiarire quali risorse saranno effettivamente disponibili dopo il 2026 per garantire la gestione, la manutenzione e l'evoluzione di quanto realizzato, dato che fino ad oggi tutti i piani di trasformazione digitale e innovazione nelle PA sono stati declinati escludendo costi aggiuntivi, anzi prevedendo riduzione di spesa per gli investimenti nel settore ICT.

Occorre ancora molto lavoro e impegno affinché la cultura dei metadati si consolidi, ovvero af-

---

<sup>6</sup> In base al cronoprogramma del Piano Triennale è previsto che entro Maggio 2022 AgID emetta le Linee Guida sulla Piattaforma Digitale Nazionale Dati e che entro Dicembre 2022 siano completati gli sviluppi tecnologici delle componenti della Piattaforma Digitale Nazionale Dati (PDND) e del Catalogo Nazionale dei Dati necessarie per integrare il Single Digital Gateway. Il 16 giugno 2022 AgID ha pubblicato la consultazione pubblica fino al 17 luglio le "Linee Guida relative alle regole tecniche sull'apertura dei dati e il riutilizzo dell'informazione del settore pubblico" <https://www.agid.gov.it/agenzia/stampa-e-comunicazione/notizie/2022/06/16/open-data-consultazione-linee-guida>.

finché si diffonda e permei un atteggiamento mentale che non consideri la metadattazione soltanto come un insieme di strumenti e di tecniche ma come un approccio sistematico che deve accompagnare l'intero ciclo della trasformazione digitale iniziando dalla fase di riorganizzazione e di progettazione dei processi, proseguendo con la reingegnerizzazione e lo sviluppo dei sistemi informativi fino a giungere alla realizzazione e all'attivazione di nuove applicazioni.

Il primo passo consiste nella modellazione dei processi e dei sistemi informativi attraverso il riuso di schemi di metadati, ontologie e vocabolari controllati già consolidati. Esiste già un consistente insieme di schemi ma occorre promuovere la conoscenza dei contesti e le finalità delle iniziative che li hanno realizzati, costruendo e alimentando repertori, come quello europeo di Joinup e il portale delle ontologie italiano, che consentano ai soggetti che si occupano dell'analisi e della modellazione di avvalersi dei modelli già disponibili eventualmente estendendoli con nuove definizioni per coprire ambiti ancora non trattati.

Il secondo passo comprende lo sviluppo di strategie e l'attuazione di metodologie per la misurazione e il controllo costanti della qualità dei metadati prodotti e utilizzati. Se infatti i metadati non rispettano gli schemi e i vocabolari controllati adottati, sono creati in modo parziale o vengono aggiornati in modo irregolare l'intero ecosistema che su di essi si basa non può funzionare in modo soddisfacente.

Il terzo passo include la collaborazione intersettoriale e lo sviluppo di specifiche competenze professionali. Molto spesso nelle aziende che producono software non si dispone delle adeguate conoscenze dei domini funzionali per cogliere le specifiche e delle competenze per razionalizzare i processi riprogettandoli e semplificandoli avvalendosi delle potenzialità offerte dai metadati e dalle ontologie. Il settore informatico è caratterizzato da un elevato livello di tecnicismo e da una forte specializzazione, una logica che rischia di essere poco fruttuosa ai fini della revisione dei processi e della modellazione dei sistemi informativi. Al contrario è necessario un approccio intersettoriale e orientato al riuso dei metadati che coinvolga figure adeguatamente preparate in grado di pensare in termini di schemi e di ontologie. Per poter attuare questa trasformazione occorre investire maggiormente, anche a livello accademico, per creare figure professionali dotate di competenze sull'uso e la definizione di schemi di metadati, di vocabolari controllati e la costruzione di ontologie. Alcuni argomenti (es. linked data, RDF, OWL) vengono già trattati in alcuni corsi di informatica o ingegneria informatica (per lo più di secondo livello), tuttavia le figure professionali da formare non dovrebbero necessariamente appartenere all'ambito scientifico e tecnologico, ma potrebbero provenire anche dall'ambito umanistico attraverso un'adeguata formazione sui linked data, sulle sintassi di rappresentazione e sugli standard esistenti. Un'impostazione multidisciplinare e un nuovo umanesimo declinato sulla base delle potenzialità che il contesto ICT mette a disposizione possono assicurare lo sviluppo di competenze trasversali e intersettoriali che uniscano all'apprendimento delle sintassi e delle tecniche di rappresentazione dei metadati le conoscenze per progettare secondo astrazioni e modelli che si avvalgono degli schemi e delle ontologie. Nuove figure professionali con questo tipo di preparazione possono sorgere da una forte collaborazione tra l'accademia, le aziende che operano nel mercato ITC e il settore pubblico e privato che chiedono soluzioni volte a semplificare e razionalizzare i propri processi, assicurando così le necessarie sinergie col mondo del lavoro e l'aggiornamento costante del personale.

## Riferimenti bibliografici

AgID (Agenzia per l'Italia Digitale).2022. *Attuazione Misure PNRR*. <https://www.agid.gov.it/it/agenzia/attuazione-misure-pnrr>

ArCo. 2017. MiC (Ministero della Cultura). *Progetto Arco – Architettura della conoscenza*. Ultimo accesso 5 Maggio 2022. <https://dati.beniculturali.it/progetto-arco-architettura-della-conoscenza/>

Bagley, Philip.1968. *Extension of programming language concepts*. Philadelphia: University City Science Center. Archived (PDF) from the original on November 30, 2012. <https://web.archive.org/web/20121130091053/http://www.dtic.mil/dtic/tr/fulltext/u2/680815.pdf>

CAD. 2021 *Decreto Legislativo 7 marzo 2005, n. 82 - Codice dell'amministrazione digitale*. Successive modifiche (*Ultimo aggiornamento all'atto pubblicato il 31/12/2021*). <https://www.normattiva.it/uri-res/N2Ls?urn:nir:stato:decreto.legislativo:2005-03-07;82!vig=>

DCAT. 2020. *Data Catalog Vocabulary (DCAT) - Version 2 - W3C Recommendation 04 February 2020*. <https://www.w3.org/TR/vocab-dcat-2/>

DCAT-AP. 2022. *Data Catalog Vocabulary (DCAT) Application Profile for data portals*. <https://joinup.ec.europa.eu/collection/semantic-interoperability-community-semic/solution/dcat-application-profile-data-portals-europe/release/210>

DCAT-AP\_IT. 2022. *DCAT-AP\_IT v1.1 – Profilo italiano di DCAT-AP*. AgID (Agenzia per l'Italia Digitale). <https://dati.gov.it/content/dcat-ap-it-v10-profilo-italiano-dcat-ap-0>

Directive (EU) 2019/1024. 2019 *Directive (EU) 2019/1024 of the European Parliament and of the Council of 20 June 2019 on opendata and the re-use of public sector information*. Ultimo accesso 5 Maggio 2022. <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A32019L1024>

eduPerson. 2016. *eduPerson Object Class Specification (201602)*. Internet2. February 2016. <http://software.internet2.edu/eduperson/internet2-mace-dir-eduperson-201602.html>

e-GOV. n.d. *Releases for e-Government Core Vocabularies solution*. Ultimo accesso 5 Maggio 2022. <https://joinup.ec.europa.eu/collection/semantic-interoperability-community-semic/solution/e-government-core-vocabularies>

EUDati. 2020. *Comunicazione della Commissione al Parlamento Europeo, al Consiglio, al Comitato Economico e Sociale Europeo e al Comitato delle Regioni. Una strategia europea per i dati*. Document 52020DC0066.2020.<https://eurlex.europa.eu/legal-content/IT/TXT/?qid=1593073685620&uri=CELEX%3A52020DC0066>

FAIR. n.d. *FAIR (Findability, Accessibility, Interoperability, Interoperabilità and Reusability) Principles*. Ultimo accesso 5 Maggio 2022. <https://www.go-fair.org/fair-principles/>

Fielding, Roy T. 2000. *Architectural Styles and the Design of Network-based Software Architectures* University of California, Irvine, 2000. <https://www.ics.uci.edu/~fielding/pubs/dissertation/top.htm>

GÉANT.n.d. Ultimo accesso 5 Maggio 2022. <http://www.geant.org/>

IBM.2014. *100 ans d'IBM en France*. <http://centenaireibmfrance.blogspot.com/2014/04/1955-terme-ordinateur-invente-par-jacques-perret.html>



INSPIRE. 2007. *Directive 2007/2/EC of March 2007 establishing an Infrastructure for Spatial Information in the European Community (INSPIRE)*. <https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=CELEX%3A32007L0002>

INSPIRE\_REG\_IT. n.d. *INSPIRE Italia Registry*. Ultimo accesso 5 Maggio 2022. <https://registry.geodati.gov.it/registry>

INSPIRE\_TG. 2022 *Technical Guidance for the implementation of INSPIRE dataset and service metadata based on ISO/TS 19139:2007*. <https://github.com/INSPIRE-MIF/technical-guidelines/blob/2022.1/metadata/metadata-iso19139/metadata-iso19139.adoc#conformance-classes-for-spatial-data-services>

Joinup-CV. 2022. *e-Government Core-Vocabularies*. <https://joinup.ec.europa.eu/collection/semantic-interoperability-community-semic/solution/e-government-core-vocabularies/release/200-0>

LGNVPIP. 2020. *Linee Guida Nazionali per la Valorizzazione del Patrimonio Informativo Pubblico* 13 Febbraio 2020. <https://docs.italia.it/media/pdf/lg-patrimonio-pubblico/bozza/lg-patrimonio-pubblico.pdf>

Media Types. 2022. *Media Types*. Internet Assigned Numbers Authority (IANA) 2022. <https://www.iana.org/assignments/media-types/media-types.xhtml>

NSO n.d. *Nodo di smistamento degli ordini di acquisto delle amministrazioni pubbliche (NSO)*. MEF (Ministero dell'Economia e Finanze) Ragioneria Generale dello Stato. [https://www.rgs.mef.gov.it/VERSIONE-I/e\\_government/amministrazioni\\_pubbliche/acquisti\\_publici\\_in\\_rete\\_apir/nodo\\_di\\_smistamento\\_degli\\_ordini\\_di\\_acquisto\\_delle\\_amministrazioni\\_pubbliche\\_nso/](https://www.rgs.mef.gov.it/VERSIONE-I/e_government/amministrazioni_pubbliche/acquisti_publici_in_rete_apir/nodo_di_smistamento_degli_ordini_di_acquisto_delle_amministrazioni_pubbliche_nso/)

OAIS. 2012. *Reference Model For An Open Archival Information System (OAIS)*. Consultative Committee for Space Data Systems. June 2012. Standard ISO 14721:2012. <https://public.ccsds.org/pubs/650x0m2.pdf>

OASIS. n.d. *Organization for the Advancement of Structured Information Standards*. Ultimo accesso 5 Maggio 2022. <https://www.oasis-open.org/>

OAS. n.d. *OpenAPI Specification Version 3.0.3* Ultimo accesso 5 Maggio 2022. <https://swagger.io/specification/>

OpenBadges 2.0. 2021. *Open Badges v2.0 IMS Final Release*. <https://www.imsglobal.org/sites/default/files/Badges/OBv2p0Final/index.html>

OCDS. n.d. *The Open Contracting Data Standard*. Ultimo accesso 5 Maggio 2022. <https://www.open-contracting.org/data-standard/>

OCDS\_Port. n.d. *Portale dei dati aperti dell'Autorità Nazionale Anticorruzione*. Ultimo accesso 5 Maggio 2022. <https://dati.anticorruzione.it/opendata/ocds>

PDF/A. 2005. *ISO 19005-1:2005 Document Management - Electronic document file format for long term preservation - Part 1: Use of PDF 1.4 (PDF/A-1)*. <https://www.iso.org/standard/38920.html>

PTIPA. 2021. *Piano triennale per l'informatica nella PA – Aggiornamento 2021 – 2023 (versione notificata alla Commissione Europea)* - Agenzia per l'Italia Digitale e Dipartimento per la Trasfor-



mazione Digitale. Roma, ottobre 2021. [https://www.agid.gov.it/sites/default/files/repository\\_files/piano\\_triennale\\_per\\_linformatica\\_nella\\_pubblica\\_amministrazione\\_2021-2023.pdf](https://www.agid.gov.it/sites/default/files/repository_files/piano_triennale_per_linformatica_nella_pubblica_amministrazione_2021-2023.pdf)

RDF. 2014. *RDF 1.1 Concepts and Abstract Syntax*. W3C Recommendation 25 February 2014. <https://www.w3.org/TR/rdf11-concepts/>

REFEDS. n.d. REFEDS (the Research and Education FEDerations group). Ultimo accesso 5 Maggio 2022. <https://refeds.org/>

RFC 2252. 1997. *Lightweight Directory Access Protocol (v3): Attribute Syntax Definitions*. RFC 2252 December 1997. <https://datatracker.ietf.org/doc/html/rfc2252>

RFC 2256. 1997. *A Summary of the X.500(96) User Schema for use with LDAPv3*. RFC 2256 December 1997. <https://www.ietf.org/rfc/rfc2256.txt>

RFC 2798. 2000. *Definition of the inetOrgPerson LDAP Object Class*. RFC 2798 April 2000. <https://www.ietf.org/rfc/rfc2798.txt>

RFC 4519. 2006. *Lightweight Directory Access Protocol (LDAP): Schema for User Applications*. RFC 4519 June 2006. <https://www.ietf.org/rfc/rfc4519.txt>

SAML. n.d. *Security Assertion Markup Language (SAML) v2.0*. Ultimo accesso 5 Maggio 2022. <https://www.oasis-open.org/standards/>

SHACH. 2022. *SCHAC - SCHEMA for Academia*. <https://wiki.refeds.org/display/STAN/SCHAC>

SDG. n.d. *The single digital gateway and Your Europe*. Ultimo accesso 5 Maggio 2022. [https://ec.europa.eu/growth/single-market/single-digital-gateway\\_it](https://ec.europa.eu/growth/single-market/single-digital-gateway_it)

SDI n.d. *Fatturazione elettronica verso la Pubblica amministrazione. Sistema di Interscambio*. Ultimo accesso 5 Maggio 2022. <https://www.fatturapa.gov.it/it/sistemainterscambio/cose-il-sdi/>

SPID\_AT. n.d. *SPID Tabella attributi*. Ultimo accesso 5 Maggio 2022. [https://www.agid.gov.it/sites/default/files/repository\\_files/regole\\_tecniche/tabella\\_attributi\\_idp\\_v1\\_0.pdf](https://www.agid.gov.it/sites/default/files/repository_files/regole_tecniche/tabella_attributi_idp_v1_0.pdf)

SPID\_REG. n.d. *SPID Registry*. Ultimo accesso 5 Maggio 2022. <https://registry.spid.gov.it/>

SPID\_RT.n.d.*SPIDMetadata*.<https://github.com/italia/spid-regole-tecniche/blob/96287a228c8db-f817a4b0b1c41e234d4e231a830/metadata.rst>

TALLIN. 2017. *Tallinn Declaration on eGovernment*. [https://ec.europa.eu/newsroom/document.cfm?doc\\_id=47559](https://ec.europa.eu/newsroom/document.cfm?doc_id=47559)

Berners-Lee, Tim. 2006. *Linked Data—Design Issues*. <http://www.w3.org/DesignIssues/Linked-Data.html>

Vallejo, Irene. 2021. *Payprus. L'infinito in un giunco. La grande avventura del libro nel mondo antico*. Firenze: Giunti Editore/Bompiani.

Wiederhold, Gio. 1997. *Database Design*, Tokio: McGraw-Hill computer science series.

W3CWS. 2004. *Web Services Glossary W3C Working Group Note*. 11 February 2004. <https://www.w3.org/TR/2004/NOTE-ws-gloss-20040211/#webservice>

WSDL 2.0. 2007 *Web Services Description Language (WSDL) Version 2.0 Part 1: Core Language*.  
<http://www.w3.org/TR/wsdl20>

XSD. 2012. *W3C XML Schema Definition Language (XSD) 1.1 Part 1: Structures W3C Recommendation*. 5 April 2012 <https://www.w3.org/TR/xmlschema11-1/> e *W3C XML Schema Definition Language (XSD) 1.1 Part 2: Datatypes W3C Recommendation* 5 April 2012 <https://www.w3.org/TR/xmlschema11-2/>