



UNIVERSITÀ DI PISA

Corso di Laurea in Informatica Umanistica

Relazione

**Linked Open Data per il Cultural Heritage
italiano: conversione e pubblicazione nel
Semantic Web del fondo antico del Comune di
Novi Ligure**

Candidato: *Diletta Goglia*

Relatore: *Andrea Marchetti*

Correlatore: *Enrica Salvatori*

Anno Accademico 2017-2018

Sommario

Elenco delle figure	3
Elenco delle Tabelle	5
Introduzione	6
1. Linked Open Data	9
1.1 Rappresentazione della conoscenza	9
1.2 RDF (Resource Description Framework)	10
1.3 Ontologie e vocabolari	13
1.4 Il Web semantico	16
1.4.1 Evoluzione e sviluppo del web semantico dal 1999 ad oggi	18
1.5 Cosa sono i Linked Data	22
1.6 Is your LOD 5 star?	24
1.7 I vantaggi dell'utilizzo di Linked Data.	25
2. Linked Data e beni culturali	28
2.1 I beni culturali	28
2.2 I Linked Data nell'universo dei beni culturali	29
2.2.1 La comunicazione del patrimonio culturale	30
2.2.2 I vantaggi dei Linked Data per le istituzioni culturali	31
2.2.3 I Linked Data tra le nuove tecnologie emergenti	34
2.2.4 Ripensare la conoscenza	38
2.3 Panoramica Internazionale	40
2.3.1 L'Authority Control	41
2.3.2 Standard di metadati internazionali	42
2.4 Panoramica nazionale	46
2.5 Progetti e istituzioni per la gestione in ambito ICT dei beni culturali in Italia	48
2.5.1 ArCo	48
2.5.2 CulturalItalia	49
2.5.3 Cultural-ON	51

2.5.4	Catalogo generale dei beni culturali	52
2.5.5	DCAT-AP-IT	53
2.6	Standard di metadati nazionali	54
2.6.1	Standard del SAN	54
2.6.2	Standard dell'ICCU	56
2.6.3	Standard dell'ICCD	57
2.6.4	Altri standard	59
2.7	Standard per la descrizione di beni culturali	63
2.8	Standard presenti nel portale Mèmorea	63
2.8.1	“Esportatori di dati” su Mèmorea	64
2.9	Riepilogo degli standard nazionali e internazionali	65
3.	Analisi dello stato dell'arte sui Linked Open Data	67
3.1	Lo stato degli Open Data in Italia	67
3.1.1	La posizione del nostro paese nella classifica europea	67
3.1.2	Punti di forza e punti deboli dell'open data italiano	70
3.1.3	Il portale nazionale italiano dati.gov.it	72
3.1.4	Confronto sulla qualità dei metadati tra il portale dati.gov.it e gli altri cataloghi open europei.	73
3.1.5	Riepilogo sugli open data in Italia	76
3.1.6	Quanti open data sono anche Linked Data?	77
3.2	Lavori simili	77
3.2.1	Amsterdam Museum Linked Open Data	78
3.2.2	SHARE Catalogue	82
3.2.3	Progetto ITACH@	83
3.2.4	OpenERCH	87
3.2.5	Archivio Fotografico della Fondazione Federico Zeri	88
3.2.6	L'esperienza LOD del Museo Galileo: il progetto MINERV@	92
3.2.7	LinkedUp Data Catalogue	94
3.2.8	Progetto LOD del COBIS	96
3.3	Riepilogo dello stato dell'arte	100

4. Pubblicare dati secondo il paradigma LOD	101
4.1 Linking Open Data Project	103
4.1.1 Evoluzione del progetto	103
4.2 Contribuire al Linking Data	106
4.3 Modelli di pubblicazione di Linked Data	108
4.4 Fasi del processo di conversione e pubblicazione dei dati	111
4.5 Mèmora	113
4.5.1 Descrizione del portale	113
5. Il Fondo antico del Comune di Novi Ligure	115
5.1 Descrizione del Fondo	115
5.2 Esportazione e pulizia dei dati da Mèmora	119
5.2.1 Individuazione delle entità	119
5.2.2 Recupero dei dati: esportazione dal portale	120
5.2.3 Conversione e pulizia dei dati	123
5.3 Arricchimento con entità estratte dal titolo delle opere attraverso un tool di Named Entity Recognition	124
5.4 Costruzione delle tabelle e import in un database relazionale	127
5.4.1 L'ambiente di sviluppo	127
5.4.2 Scelta dei dataset esterni	127
5.4.3 Collegamenti ai dataset esterni	131
5.4.4 Disambiguazione	133
5.4.5 Arricchimento del dataset: metadattazione e inferenza	137
5.4.6 Creazione del database relazionale	138
5.5 Scelta dell'ontologia per la rappresentazione dei dati	142
5.5.1 D2RQ	144
5.5.2 Creazione degli URI	147
5.5.3 Dereferenziazione degli URI	148
5.5.4 Grafi esplicativi	150
5.6 Esportazione dell'archivio in Linked Open Data attraverso D2RQ	153
5.7 Costruzione del sito web per la navigazione delle informazioni sul fondo	154

6. Conclusioni e sviluppi futuri	158
6.1 Aderenza ai principi del Linked Data e scoperta informativa	158
6.2 Proposte di sviluppo futuro	160
6.3 Considerazioni finali	161
Bibliografia	163
Libri	163
Articoli	164
Slide / altro	170
Sitografia	172

Elenco delle figure

Fig. 1.1 Esempio di una tripla RDF rappresentata con un grafo.

Fig. 1.2 Esempi di ink RDF interni ed esterni.

Fig. 1.3 *Ontology Spectrum*: diagramma dei vari tipi di ontologia in base alla forza semantica.

Fig. 1.4 Continuum di espressività delle ontologie RDFS, SKOS e OWL: dalla più semplice alla più complessa).

Fig. 1.5 *Semantic Web Wedding Cake*.

Fig. 1.6 Le quattro principali ondate di evoluzione del web.

Fig. 1.7 Confronto tra il web tradizionale e il web semantico.

Fig. 1.8 Five Star Classification.

Fig. 2.1 Gartner's *Hype Cycle for Advanced Analytics and Data Science* 2015.

Fig. 2.2 Gartner's *Hype Cycle for Emerging Technologies*, 2018.

Fig. 2.3 Fornitori di dati verso Europeana.

Fig. 2.4 *Il portale CulturalItalia nel quadro delle collaborazioni europee*.

Fig. 2.5 CulturalItalia, l'aggregatore nazionale.

Fig. 2.6 La figura illustra alcuni elementi principali dell'ontologia Cultural-ON. La figura non è esaustiva ma è inserita per aiutare a comprendere alcune relazioni tra concetti dell'ontologia.

Fig. 2.7 Tracciati catalografici dell'ICCD.

Fig. 2.8 CIDOC-CRM: architettura di integrazione delle informazioni

Fig. 3.1 Portale Europeo dei Dati: nel 2018 l'Italia è tra i "trend-setters".

Fig. 3.2 *Open data portals: top 10 countries*

Fig. 3.3 La collocazione geografica dei portali italiani.

Fig. 3.4 e 3.5 La posizione dell'Italia tra i "trend-setters" nell'*Open Data Maturity clustering*.

Fig. 3.6: controllo della qualità dei metadati del portale nazionale italiano dati.gov.it.

Fig. 3.7 Accessibilità: portale dati.gov.it.

Fig. 3.8 Accessibilità: panoramica generale di tutti i metadati disponibili per il portale europeo di dati (tutti i cataloghi).

Fig. 3.9 Leggibilità automatica: portale dati.gov.it.

Fig. 3.10 Leggibilità automatica: panoramica generale di tutti i metadati disponibili per il portale europeo di dati (tutti i cataloghi).

Fig. 3.11 Lo stato dell'*Open Data* italiano nel 2018.

Fig. 3.12 Parti costitutive del dataset LOD dell'Amsterdam Museum.

Fig. 3.13 Workflow di pubblicazione di dati eterogenei in Linked Data nella piattaforma OpLiDaF.

Fig. 3.14 Architettura del progetto LOD del CoBiS

Fig. 3.15 Sito web del progetto LOD del CoBiS.

Fig. 4.1 Un esempio di scheda di catalogazione nel portale Mèmora.

Fig. 4.2 Crescita del numero di serie di dati pubblicati sul Web come *Linked Data*.

Fig. 4.3 LOD Cloud: agosto 2014.

Fig. 4.4 LOD Cloud: settembre 2017.

Fig. 4.5 LOD Cloud: marzo 2019.

Fig. 4.6 Pattern di pubblicazione di Linked Data più comuni sotto forma di flussi di lavoro, da dati strutturati o contenuti testuali fino a dati collegati pubblicati sul Web.

Fig. 4.7 Riassunto dell'intero processo di pubblicazione, partendo dai dati strutturati fino ai *linkable*.

Fig. 4.8 Un altro modello per il workflow di pubblicazione dei *Linked Data*. A sinistra lo schema proposto dall'Agenzia per l'Italia Digitale (AGID) nel 2012 e a destra quello di A. lacono nel 2016.

Fig. 5.1 Il Fondo antico del Comune di Novi Ligure nella gerarchia di Mèmora.

Fig. 5.2 Esempio di una scheda di catalogo (unità documentaria) del portale Mèmora.

Fig. 5.3 Esempio di un'unità documentaria esportata nel formato MINI-icar-san (file XML): il tracciato raggruppa e rende interoperabili gli standard EAD, EAC.CPF e SCONS.

Fig. 5.4 Esempi di immagini digitali degli oggetti materiali dell'archivio.

Fig. 5.5 File excel contenente tutte le informazioni rilevanti sulle unità documentarie e le entità d'interesse.

Figg. 5.6 e 5.7 Due esempi di Named Entity Recognition con Dandelion API.

Fig. 5.8 Esempio della voce "Ernst Cassirer" (entità-persona) in VIAF.

Fig. 5.9 Un esempio di disambiguazione in Google Refine.

Fig. 5.10 Esempio di punteggio di riconciliazione in Google Refine.

Fig. 5.11 Esempio di colonne contenenti link.

Fig. 5.12 Processo completo di creazione di *Linked Data* con Google Refine.

Fig. 5.13 Vista delle tabelle di mapping e delle relazioni all'interno del database.

Fig. 5.14 Rappresentazione delle relazioni M-M nel file di mapping del database secondo modelli ontologici.

Fig. 5.15 Diagramma dell'architettura del server D2R.

Figg. 5.16 e 5.17 Grafi RDF dell'archivio LOD.

Figg. 5.18 - 5.22 Il sito web per il Fondo Antico del Comune di Novi Ligure.

Fig. 5.23 Endpoint SPARQL .

Elenco delle Tabele

Tab. 1.1 Confronto tra il web classico (“di documenti”) e il web semantico (“di dati”).

Tab. 1.2 Vantaggi di pubblicazione dei dati secondo il paradigma LOD, rispetto al paradigma di pubblicazione tradizionale.

Tab. 2.1 Standard di metadati internazionali.

Tab. 2.2 Esportatori di dati nella piattaforma Mèmora.

Tab. 2.3 Riepilogo degli standard di metadati.

Tab. 3.1 Riepilogo dello stato dell’arte.

Tab. 5.1 Entità, formati e tag corrispondenti.

Tab. 5.2 Dataset esterni da cui sono state recuperate le entità

Tab. 5.3 Tabella riassuntiva dei vocabolari e delle ontologie utilizzati.

Tab. 5.4 Esempio di *tripla letterale*.

Tab 5.5 Esempio di *link RDF interno*.

Tab. 5.6 Esempio di *link RDF esterno*.

Introduzione

Negli ultimi anni abbiamo assistito ad un enorme e rapido aumento non solo della quantità ma anche della varietà di risorse disponibili online (testi, suoni, immagini, video, 3D, ecc...). È nata perciò la necessità di descrivere questi dati in maniera appropriata per permettere agli utenti di trovarli sul web e accedervi (*search and retrieval*), renderli disponibili a istituzioni e organizzazioni per creare servizi e permettere anche a fruitori non umani di “consumarli” (es. motori di ricerca, applicazioni, ecc...)

Specialmente in virtù di questo apporto sempre più decisivo delle tecnologie digitali, le istituzioni culturali hanno avvertito fortemente la necessità di estendere l’accesso al proprio patrimonio, aprendo l’orizzonte conoscitivo attraverso il confronto e la collaborazione con altre realtà.

Gli scenari per lo sviluppo e la trasmissione della conoscenza, nonché per un nuovo ripensamento del concetto stesso di conoscenza, trovano condizioni ideali in “ambienti” come il web semantico.

Per queste ragioni, archivi e musei oggi si preoccupano di convertire l’informazione in dati *aperti* e soprattutto *collegati*, superando l’approccio limitante e ormai obsoleto dei “silos di informazione”. I dati infatti acquistano valore quando sono interconnessi con altri dati, e questo arricchimento è possibile grazie all’applicazione dei principi e delle tecnologie *Linked Open Data* (LOD).

Far uscire fuori i dati locali e renderli aperti e universalmente accessibili significa offrire opportunità di visibilità anche a dataset di piccola dimensione o che fanno riferimento ad una località geografica circoscritta.

Inoltre l’integrazione dei dati con quelli di altre istituzioni non solo aumenta il loro potenziale informativo ma li rende più completi e riusabili in contesti anche molto diversi da quelli d’origine.

La *mission* delle istituzioni culturali oggi è quindi quella di trattare e comunicare le informazioni sui beni culturali partecipando alla costruzione del web semantico e alla divulgazione della conoscenza del *Cultural Heritage*.

Tutti i principi che costituiscono il paradigma dei *Linked Data*, applicati nel mondo del *Cultural Heritage*, ne consentono una comunicazione efficace, arricchita e di ampio raggio, raggiungibile anche dai nuovi fruitori di informazione: le macchine.

In quest'ottica l'obiettivo del progetto consiste nella creazione di un dataset *Linked Open Data* relativo alla descrizione di beni artistici e culturali.

Lo scopo finale è quello di pubblicare tali dati all'interno della rete *LOD Cloud*, rendendoli disponibili all'utilizzo da parte anche delle macchine: l'opportunità offerta da questo nuovo metodo di pubblicare dati sul web, è l'esempio una radicale trasformazione del rapporto tra utente (umano e non) e universo dei beni culturali.

Il progetto vuole perciò presentarsi come esempio, in scala, della *rivoluzione copernicana dei Linked Open Data* nel mondo dei beni culturali.

- Il Capitolo 1 passa in rassegna il vasto mondo dei *Linked Open Data* e del web semantico, illustrandone i principi, l'evoluzione nel tempo e approfondendo i concetti di RDF, ontologie, vocabolari.
- Il Capitolo 2 scende nel dettaglio del mondo dei beni culturali, e in particolar modo dell'approccio *Linked Open Data* applicato all'interno di questo mondo, evidenziando i vantaggi che vi può apportare. Viene fatta una panoramica generale degli standard di metadati internazionali e nazionali, con particolare attenzione ad alcuni standard d'interesse utilizzati in Italia per i beni culturali.
- Il Capitolo 3 descrive lo stato dell'arte degli *Open Data* e dei *Linked Data* in Italia per trarne qualche informazione quantitativa in modo da valutare l'adozione di tali approcci. Inoltre passa in rassegna alcuni lavori simili al progetto svolto.
- Il Capitolo 4 presenta una panoramica generale del *Linked Open Data Project* e della sua evoluzione nel tempo. Illustra inoltre alcuni modelli e approcci alla pubblicazione di dati LOD presenti in letteratura, e si sofferma sulla descrizione del portale Mèmora, la fonte dei dati del progetto.

- Il Capitolo 5 descrive nel dettaglio tutte le fasi del lavoro di conversione e pubblicazione dei dati del Fondo antico di Novi Ligure in *Linked Open Data*.
- Il Capitolo 6, infine, presenta una breve valutazione del lavoro svolto ed espone i possibili sviluppi futuri e le considerazioni conclusive.

1. Linked Open Data

1.1 Rappresentazione della conoscenza

Per descrivere i dati, si usano i *metadati*, cioè informazioni (marcatori) che li descrivono.

Ne esistono diversi tipi, a seconda della tipologia di risorsa da descrivere (amministrativi e gestionali, descrittivi, di conservazione, tecnici).

Gli standard di metadati nazionali e internazionali più usati nel mondo delle biblioteche e dei beni culturali sono i seguenti: MARC, DC, MODS, METS, MIX, EAD, ISBD, RDA, VRA, ISAAR, ISAD, ISNI, NIERA.

Ciascuno di essi è descritto nel dettaglio al [capitolo 2](#).

1.2 RDF (*Resource Description Framework*)¹

RDF è un modello proposto dal *World Wide Web Consortium* (W3C)² per la codifica, lo scambio e il riutilizzo di *dati strutturati* e consente l'*interoperabilità semantica* tra applicazioni che condividono le informazioni sul Web.

*La crescente interoperabilità fra i dati è certo un avanzamento molto rilevante, dal punto di vista tecnico [...] Tuttavia, come sempre, la tecnica non basta. Perché sia giustificato l'aggettivo "semantico", occorre che i dati siano interoperabili non solo in quanto stringhe di bit ma anche in quanto significati. Occorre cioè riconoscere e trattare problemi di omografia e sinonimia fra termini e di relazioni fra concetti*³

¹ Standard RDF, ultima versione

<https://www.w3.org/TR/2014/NOTE-rdf11-primer-20140624/>

² W3C <https://www.w3.org/>

³ C. Gnoli, *Connettere la conoscenza: quattro livelli di interoperabilità*, in *Biblioteche oggi*, vol.32, n°5, 2014, pp. 9-10

Il modello di dati RDF⁴ è progettato per la rappresentazione integrata di informazioni che provengono da più fonti e rappresenta le informazioni come *grafi* diretti etichettati (vedi fig. sotto).

La descrizione di una risorsa è rappresentata da un numero di triple⁵; le tre parti di ogni tripla sono *soggetto*, *oggetto* (nodi del grafo) e *predicato* (arco del grafo), identificati rispettivamente da tre *URI*.

Gli URI del soggetto e dell'oggetto identificano le risorse correlate, mentre quello del predicato definisce il tipo di relazione tra le risorse.

In particolare, gli URI predicati sono raccolti in vocabolari utilizzati per rappresentare informazioni su un determinato dominio.



Fig. 1.1 Esempio di una tripla RDF rappresentata con un grafo.

RDF fornisce un modo per descrivere entità del mondo (persone, luoghi o concetti astratti) e il modo in cui esse si relazionano (si “collegano”) ad altre entità, così da offrire ai calcolatori una conoscenza del nostro universo.

RDF permette di rappresentare ma anche di pubblicare queste informazioni sul Web in una forma che gli altri (umani o macchine) possono scoprire e riutilizzare.⁶

⁴ Modello di dati RDF: <https://www.w3.org/TR/rdf-concepts/>

⁵ Non ci occupiamo qui di *triple letterali* ma di *collegamenti RDF*: per approfondire <http://linkeddatabook.com/editions/1.0/#htoc16>

⁶ Per approfondire: T. Baker, *La traduzione dei dati nel linguaggio del web semantico*, in *Rivista italiana di biblioteconomia, archivistica e scienza dell'informazione*, vol.4, n°1, 2013, <https://www.jlis.it/article/view/6308/7866>

I collegamenti RDF possono essere interni ed esterni:

- I link RDF interni collegano le risorse all'interno di una singola fonte (*repository*) di dati collegati. Pertanto, gli URI di soggetto e oggetto hanno lo stesso *namespace*.
- I link RDF esterni connettono risorse provenienti da diverse origini di dati collegati. Gli URI di soggetto e oggetto hanno *namespace* diversi. I link esterni sono fondamentali per lo sviluppo e il mantenimento del *web of data* (cap. 1.4).

Le caratteristiche principali dei link RDF che vale la pena notare in questo contesto sono quelle che li differenziano dai link HTML (<a> . . .)⁷:

- I link RDF collegano entità, non solo documenti.
- I link RDF sono tipizzati. Ciò significa che non lasciano all'utente l'inferenza della natura della loro relazione, ma consentono all'editore di dichiarare esplicitamente in tipo di connessione. Ad esempio, il collegamento "amico di" è un tipo di relazione tra due persone. Rendendo espliciti questi collegamenti, RDF permette alle applicazioni di percorrerli e scoprire una quantità sempre maggiore di dati.

Rendere espliciti i link tra diversi dataset, specialmente a livello semantico, richiede la definizione rigorosa di tutte le caratteristiche "astratte" del sistema di dati e metadati. Questo processo richiede l'individuazione di vocabolari e ontologie di riferimento.

⁷ C. Bizer, T. Heath, *Linked Data: Evolving the Web into a Global Data Space (1st edition)*, 2011, cap. 2.1.

A. Iacono, *Introduzione a Linked Open data e Web semantico*, 2016, Università La Sapienza, Roma, pp.36-37

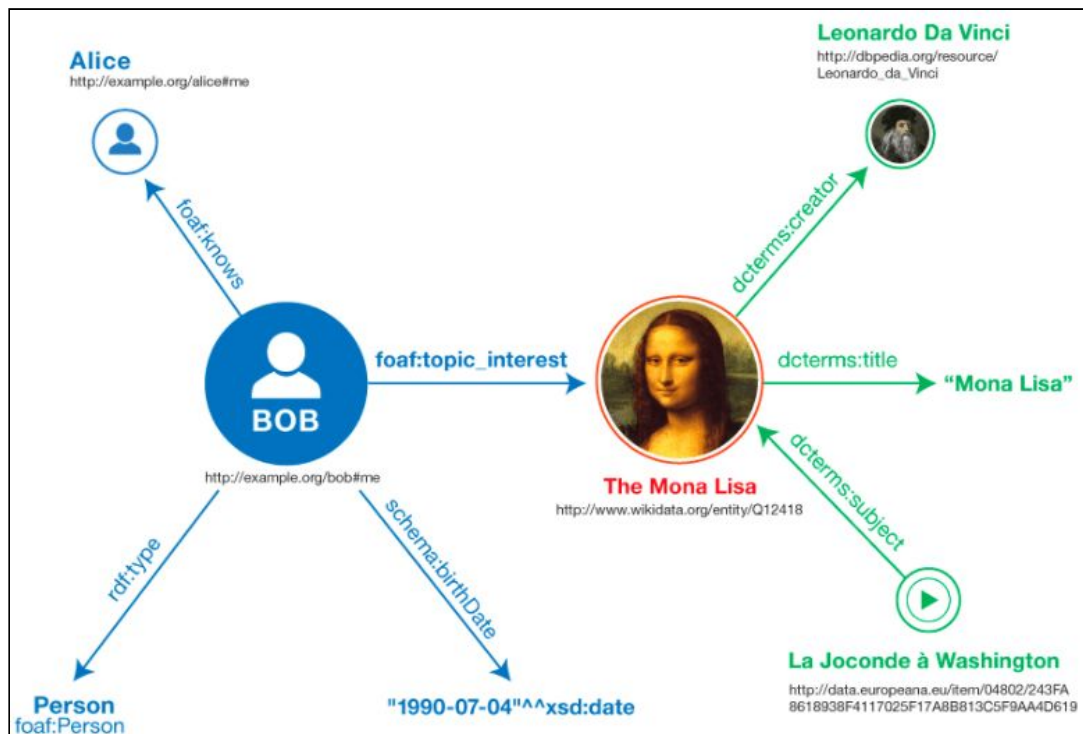


Fig. 1.2 Esempi di ink RDF interni ed esterni

1.3 Ontologie e vocabolari

La distinzione tra ontologie e vocabolari , nell’ambito della rappresentazione formale della conoscenza, non è netta. Generalmente:

- Un *vocabolario* è una lista chiusa di termini da usare come valori nella tripla RDF.
- Una *ontologia* è una rappresentazione formale per concettualizzare un dominio di interesse, cioè per descrivere relazioni / modelli concettuali tramite classi e proprietà. Serve per descrivere il modo in cui diversi schemi vengono combinati in una struttura dati (vocabolario) contenente tutte le entità rilevanti e le loro relazioni all’interno del dominio.⁸

⁸ Mentre i meccanismi di classificazione si occupano dell’accesso all’informazione, basandosi su criteri predefiniti codificati con elementi “sintattici”, le ontologie si concentrano sul significato (“semantica”) dei termini e sulla natura e struttura di un dominio. (O. Signore, O. Missikoff, P. Moscati, *La gestione della conoscenza in archeologia: modelli, linguaggi e strumenti di modellazione concettuale dall’XML al Semantic Web*, in *Archeologia e Calcolatori*, n°16, 2005, pag. 294)

I *vocabolari controllati*⁹ sono insiemi di termini prescelti in quanto d'uso attestato da una o più fonti (*authorities*), raccolti e ordinati logicamente tra loro allo scopo di riferire le varianti (ortografiche, sincroniche, diacroniche, ...). Nel caso in cui i termini siano connessi tra loro anche in relazione gerarchica o associativa, i vocabolari controllati prendono il nome di *thesauri*.¹⁰¹¹

*[...] thesauri, mappe concettuali e ontologie non possono essere usati se non in stretta correlazione l'uno con gli altri. Attraverso uno stretto legame tra le diverse autorità terminologiche si può realizzare un effettivo controllo della conoscenza senza il quale Internet rischia di diventare un'immensa raccolta di dati e di informazioni difficilmente utilizzabili.*¹²

Le varie istituzioni hanno preferenze diverse sui vocabolari da utilizzare per rappresentare i propri contenuti: standard differenti vengono quindi utilizzati parallelamente.

Nonostante questa generale apertura, è considerata una buona pratica riutilizzare, ove possibile, termini da vocabolari RDF noti al fine di rendere più semplice l'elaborazione delle applicazioni *client*. Solo se questi vocabolari non forniscono i termini richiesti, gli editori di dati possono definire una nuova terminologia specifica per i dati.

Tra quelli più utilizzati ci sono:

- SKOS¹³ (*Simple Knowledge Organization System*), per esprimere thesauri e gerarchie concettuali (tassonomie);

⁹ Vocabolari controllati dell'ICCD:

<https://github.com/ICCD-MiBACT/Standard-catalografici/tree/master/strumenti-terminologici>

¹⁰ L. Corti, *Beni culturali, standards di rappresentazione, descrizione e vocabolario*, in: *Informatica e beni culturali*, Franco Cosimo Panini Editore, Modena, 1992, pag. 193.

"Thesaurus: costruzione ramificata di un vocabolario, specifica per disciplina" (Sapienza Università di Roma, Banca dati LOD della Biblioteca di Economia "Enrico Barone", *Il web semantico oggi*, pag.2)

¹¹ Nei sistemi di *Information Retrieval*, la valutazione dei valori di "precisione" e "recupero" avviene (anche) in rapporto al vocabolario controllato (vedi:

https://en.wikipedia.org/wiki/Precision_and_recall).

¹² P. Capitani, *Il Knowledge Management*, in: *Associazione Italiana Formatori*, Franco Angeli Editore, Milano, 2006, pag.47

¹³ <https://www.w3.org/TR/skos-reference/> , <https://www.w3.org/2004/02/skos/>

- DCMI (*The Dublin Core Metadata Initiative*) *Metadata Terms*¹⁴: definisce attributi generici come “titolo”, “autore”, “data”, ecc ...
- FOAF¹⁵ (*Friend-of-a-Friend*): definisce termini per descrivere persone, le loro attività e le loro relazioni con altre persone e oggetti.

OWL¹⁶ (*Web Ontology Language*) e RDF Schema (RDFS)¹⁷ sono invece delle ontologie.

RDFS è un insieme di classi e proprietà RDF che estende il vocabolario base di RDF e viene utilizzato per strutturare altri vocabolari RDF; diverse delle sue componenti sono incluse in OWL che è più espressivo.

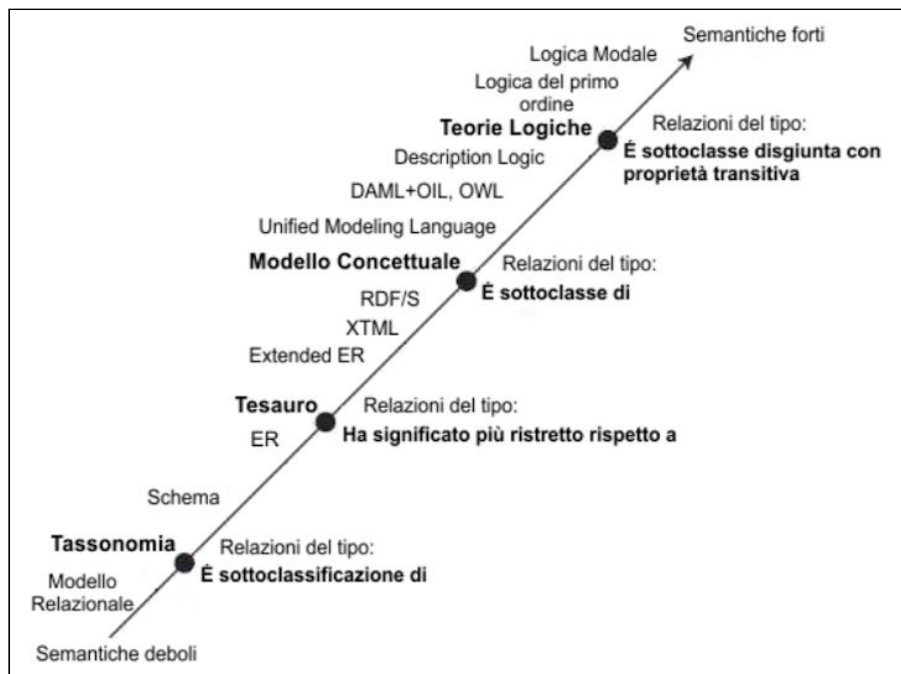


Fig. 1.3 *Ontology Spectrum*: diagramma dei vari tipi di ontologia in base alla forza semantica.¹⁸

Le tecnologie messe a disposizione dal *Semantic Web* si collocano tra i punti più alti di questo diagramma, permettendo di esprimere in modo completo il significato ontologico dei dati spiegando come questi devono essere interpretati.

¹⁴ <http://www.dublincore.org/specifications/dublin-core/dcmi-terms/>

¹⁵ <http://xmlns.com/foaf/spec/>

¹⁶ <https://www.w3.org/TR/owl-features/>

¹⁷ <https://www.w3.org/TR/rdf-schema/>

¹⁸ M. C. Daconta, L. J. Obrst, K. T Smith, *The Semantic Web: A Guide to the Future of XML, Web Services, and Knowledge Management*, Wiley, 2003.

Collettivamente, SKOS, RDFS e OWL forniscono un “continuum di espressività”. Nel contesto dei *Linked Data* è sufficiente esprimere i vocabolari in RDFS¹⁹.



Fig. 1.4 Continuum di espressività delle ontologie RDFS, SKOS e OWL: dalla più semplice alla più complessa)

Infine, vale la pena citare *Schema.org*²⁰, un repository di ontologie creato da Google.

1.4 Il Web semantico²¹

Usare lo standard RDF, vuol dire passare da un web di documenti ad un *web di dati*.

Come è stato già accennato, i link RDF esterni permettono di collegare dati provenienti da diverse fonti (*repositories*): di conseguenza, collegano tra loro anche le fonti stesse, all'interno di un unico spazio globale interconnesso,²² è possibile quindi immaginare tutti i *Linked Data* come un “gigantesco grafo globale”²³.

¹⁹ Heath & Bizer, 2011, cap. 4.4.1; Iacono 2016, pag. 88

²⁰ <https://schema.org/>

²¹ *Semantic Web*: <https://www.w3.org/standards/semanticweb/>

²² Nel web, molti server diversi sono responsabili di rispondere alle richieste di dereferenziazione degli URI HTTP in molti *namespace* diversi e, nel contesto dei *Linked Data*, restituiscono descrizioni RDF delle risorse identificate da questi URI. Pertanto, se un collegamento RDF connette gli URI in diversi *namespace*, alla fine collega le risorse in diversi set di dati (Heath & Bizer, 2011, cap. 2.1).

²³ *Giant Global Graph* (GGG) è un termine coniato da T. B. Lee, apparso per la prima volta nel 2007 sul suo blog.

Il termine *Semantic Web* (in italiano, “web semantico”) è stato proposto per la prima volta nel 1999 da Tim Berners-Lee²⁴ nel suo libro *Weaving the web*²⁵, e fa riferimento proprio a questo “spazio globale di dati” in cui la conoscenza codificata è rappresentata in modo comprensibile dal computer, fornendo le descrizioni delle risorse presenti sul web in “modo formale”, secondo standard e modelli comuni.

Questa conoscenza può così essere utilizzata dagli “agenti intelligenti”, applicazioni che, grazie all’identificazione tramite URI (univoca e indipendente dalla lingua) e ad una rappresentazione formalizzata della conoscenza (in RDF), sono in grado di realizzare associazioni e di identificare equivalenze terminologiche. Sono quindi capaci di comprendere il significato dei testi presenti sulla rete, nonché di eseguire ragionamenti deduttivi sulla base delle relazioni logiche che collegano le informazioni, operando sull’intero spazio di dati²⁶.

Questo rappresenta l'essenza dei *Linked Data*.

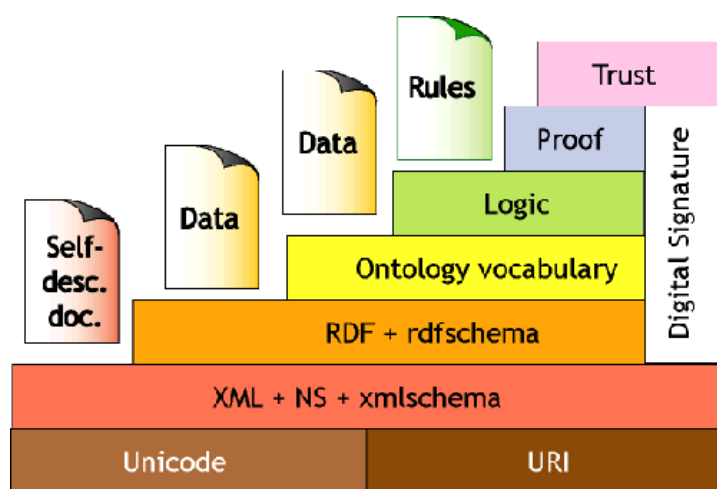


Fig. 1.5 *Semantic Web Wedding Cake*

Questa rappresentazione del web semantico, partendo dai caratteri ammessi (Unicode) e attraverso l’identificazione delle risorse sul web (URI), arriva ad analizzare l’ambito sintattico prima (XML) e semantico poi (RDF), per terminare con le ontologie, la possibilità di operare ragionamenti (livello logico) e il “trust”, cioè l’affidabilità delle risorse sul web.²⁷

²⁴ T. B. Lee: <https://www.w3.org/People/Berners-Lee/>

²⁵ Tutte le pubblicazioni di T. B. Lee: <https://www.w3.org/People/Berners-Lee/Publications.html>

²⁶ Ad esempio, un'applicazione che ha cercato un URI e ha recuperato dati RDF che descrivono una persona può seguire i collegamenti da tali dati a dati collocati su diversi server Web, descrivendo, ad esempio, il luogo in cui la persona vive o la società per la quale la persona lavora. (Heath & Bizer, 2011, cap. 2.1)

²⁷ O. Signore et al., 2005, pag. 292

RDF, nell'ottica del web semantico, fornisce al computer un modello concettuale della "porzione di universo" che vogliamo rendergli comprensibile.²⁸

L'idea alla base del web semantico, che lo distingue dal web classico, è proprio questa evoluzione dal concetto di *machine readable* a quello di *machine understandable*.

1.4.1 Evoluzione e sviluppo del web semantico dal 1999 ad oggi

*I have a dream for the Web [in which computers] become capable of analyzing all the data on the Web – the content, links, and transactions between people and computers. A 'Semantic Web', which should make this possible, has yet to emerge, but when it does, the day-to-day mechanisms of trade, bureaucracy and our daily lives will be handled by machines talking to machines. The 'intelligent agents' people have touted for ages will finally materialize.*²⁹

Gli "agenti intelligenti" di cui parla Lee possono:

- comprendere i dati (*machine understandable*);
- integrarli e metterli in relazione, attraverso operazioni logiche;
- elaborarli automaticamente per ricavare inferenze e deduzioni, cioè attivare processi, attraverso i quali, partendo da conoscenze note, si giunge ad ottenere nuove conoscenze (*machine processable*);
- guidare l'utente direttamente verso l'informazione ricercata, sulla base di questi percorsi creati a partire dall'informazione stessa;
- sostituirsi all'utente spostandosi da un nodo all'altro nel "grafo della conoscenza", collegando logicamente elementi diversi dell'informazione richiesta.

²⁸ V. Casarosa, *RDF e RDF Schema - Resource Description Framework*, in *Biblioteche Digitali*, 2018

²⁹ T. B. Lee, *Weaving the Web: The original design and ultimate destiny of the World Wide Web (1st edition)*, Thomson Learning Emea, 2000 [1999]

Nel 2006, sempre Lee, pubblica un articolo³⁰ in cui esprime alcune riflessioni sul web semantico:

The Semantic Web isn't just about putting data on the web. It is about making links, so that a person or machine can explore the web of data. With linked data, when you have some of it, you can find other, related, data.

Secondo Tim Berners-Lee, uno degli elementi fondamentali del web semantico nel futuro sarà la compresenza di più ontologie. Se si vuole un sistema dinamico in grado di raffinarsi e funzionare su scala universale, bisognerà “pagare il prezzo di una certa dose di incoerenza”.

Nell'articolo *The Semantic Web Revisited* (2006), Berners-Lee, Shadbolt e Hall hanno ridefinito alcuni aspetti del web semantico in funzione della dinamicità del *World Wide Web*. L'idea era quella di sviluppare e migliorare le ontologie in maniera collaborativa, grazie all'intervento di comunità.

In seguito, questo nuovo approccio si è orientato sempre più a far sì che i dati vengano strutturati sotto forma di *Linked Data* anziché sotto forma di ontologie.

Quindi se in una prima fase (1999-2006) il web semantico era dominato da un approccio *top down*, basato sul concetto di ontologia come sistema di rappresentazione della conoscenza *a priori*, dal 2006 si passa ad una seconda fase in cui i *Linked Data* delineano un approccio differente, di tipo *bottom up*: i collegamenti tra le cose e i concetti si formano dal basso, rendendo i dati interoperabili tra loro³¹.

Le ontologie oggi servono ad inquadrare i dati proprio attraverso una metodologia *bottom up*, si parte da ciò che si vuole rappresentare e si decide come rappresentarlo.

³⁰ T. B. Lee, *Linked Data*, in *Design Issues* (W3C), 2006

³¹ Iacono 2016, pag.16

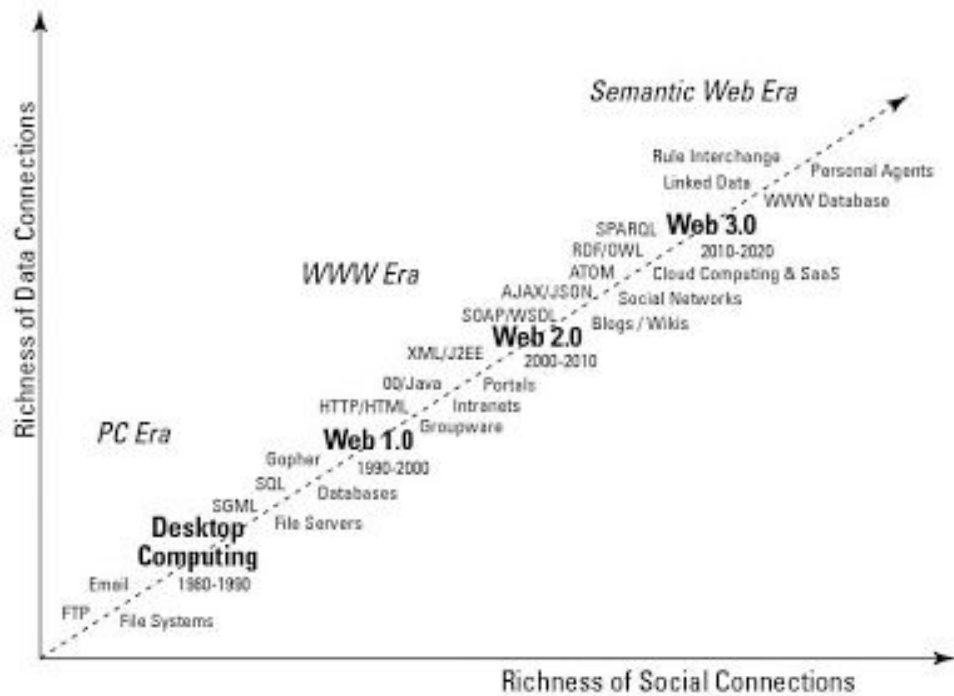


Fig. 1.6 Le quattro principali ondate di evoluzione del web

Nell'articolo del 2009³², sempre di T. B. Lee, emerge chiaramente che i *Linked Data* vengono considerati come le nuove colonne portanti su cui costruire il web semantico.

Il web semantico oggi è un movimento collaborativo guidato dal W3C. Incoraggiando l'inclusione di contenuti semantici nelle pagine web, esso mira a convertire sempre di più il web corrente in una "rete di dati".

³² C. Bizer, T. Heath, T. B. Lee, *Linked Data - The Story So Far*, 2009

	Web of Document	Web of Data
Oggetti	documenti	cose (e loro descrizioni)
Grado di strutturazione degli oggetti	basso	alto
Collegamenti tra	documenti	cose (inclusi documenti)
Tipo di collegamenti	non tipizzati (<i>untyped links</i>)	tipizzati (<i>typed links</i>)
Contenuto semantico	implicito	esplicito
Progettato per	fruizione umana	<i>machines first, humans later</i>

Tab. 1.1 Confronto tra il web classico (“di documenti”) e il web semantico (“di dati”)³³.

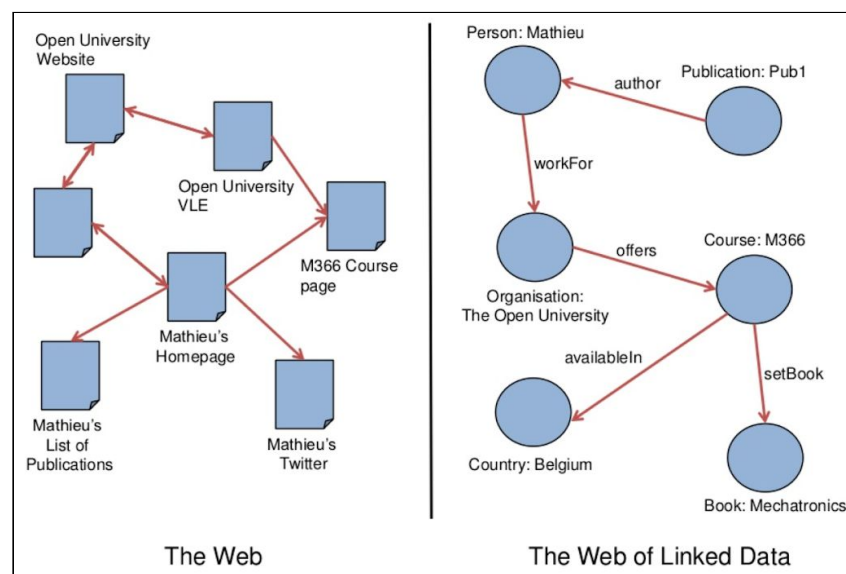


Fig. 1.7 Confronto tra il web tradizionale e il web semantico³⁴

³³ Heath, Bizer, Lee 2008; Guerrini et. al 2015

³⁴ M. d’Aquin, *Semantic Web, Linked Data and Education: A Perfect Fit?*, alla conferenza *Metadata and Semantics Research Conference (MTSR)*, Cadiz, Spagna, 2012, pag. 8

1.5 Cosa sono i *Linked Data*

I *Linked Data* sono profondamente legati al concetto di “dati aperti”, che ne costituisce la base intellettuale, tanto che si parla più propriamente di *Linked Open Data* (LOD), a identificare dati non solo collegati ma, appunto, aperti.

Mentre il paradigma dei Linked Data mira ad abbattere le barriere tecnologiche che impediscono la libera condivisione dei dati per creare uno spazio unico e libero di condivisione e di riutilizzo dei dati nel Web, l'Open Data promuove iniziative volte a eliminare le barriere economiche, sociali e culturali alla libera condivisione dei dati. I due concetti sono profondamente legati: gli Open Data costituiscono il sostrato intellettuale su cui fa perno il paradigma Linked Data. ³⁵

Gli *Open Data* o “dati aperti” sono dati pubblicati sul web in modo da essere liberamente accessibili a tutti, le cui eventuali restrizioni sono l'obbligo di citare la fonte o di mantenere la banca dati sempre aperta. Possono essere utilizzati e ridistribuiti da chiunque.

Con *Linked Data*³⁶ o “dati collegati” ci si riferisce ad un insieme di *best practice* per la pubblicazione e il collegamento dati strutturati sul web, permettendo così ad un essere umano o ad una macchina, di trovare dati correlati provenienti da fonti diverse.

Queste *best practice* sono state introdotte da Tim Berners-Lee, l'inventore del World Wide Web, e sono diventate note come i “principi dei *Linked Data*”:

1. *Usare gli URI per dare nomi alle cose* (non contenuti digitali, ma anche oggetti del mondo reale e concetti astratti).

³⁵ V. Lo Castro, *Web semantico e Linked Open Data: best practices, prospettive e criticità*, in *Nuovi Annali Della Scuola Speciale per Archivistici e Bibliotecari*, Firenze, Vol. 28, 2014, pag. 2013

³⁶ <http://linkeddata.org/>

2. *Usare gli URI HTTP, in modo che le persone possano cercare quei nomi.* Gli URI HTTP infatti :

- a. Identificano univocamente una risorsa/entità nel web in modo decentralizzato: ogni proprietario di un nome di dominio può creare nuovi riferimenti URI.
- b. Combinano questa identificazione univoca con un meccanismo di recupero semplice e veloce: servono quindi non solo come nome ma anche come mezzo di accesso alle informazioni che descrivono l'entità identificata.

Questo principio richiede che gli URI siano HTTP per poter essere *dereferenziati*, cioè che le macchine possano accedere (attraverso il protocollo HTTP) a una descrizione dell'oggetto/concetto identificato dall'URI.³⁷ Per costruire un URI dereferenziabile è necessario fornire una rappresentazione della risorsa che si vuole pubblicare non solo in formato HTML ma in formato RDF (*machine readable*).

3. *Fornire informazioni utili usando gli standard (RDF, SPARQL).* L'uso di un singolo modello per la pubblicazione di dati strutturati sul Web consente alle applicazioni di elaborarne il contenuto.
4. *Includere link (RDF) ad altri URI, in modo che possano venire scoperte più cose.*

L'idea centrale dei *Linked Data* è quindi quella di utilizzare gli URI HTTP non solo per identificare documenti web, ma anche per identificare entità del mondo reale.

Linked Open Data (LOD) sono quindi dati collegati che vengono rilasciati con una licenza aperta³⁸, che ne consente il riutilizzo gratuito.

³⁷ Qualsiasi URI HTTP dovrebbe essere dereferenziabile: i *client* HTTP possono cercare l'URI usando il protocollo HTTP e recuperare una descrizione della risorsa identificata dall'URI. Per approfondimenti sulla dereferenziazione:

<http://linkeddatabook.com/editions/1.0/#htoc11>

³⁸ Solitamente i dati aperti sono disponibili con una licenza CC-BY-SA. Questo significa che gli utenti possono includerli in qualsiasi altro lavoro (Creative Commons) a condizione che diano loro la giusta attribuzione ("created BY"). Se si creano opere derivate (come versioni modificate o estese di dati aperti), dovranno essere a loro volta concesse in licenza come CC-BY-SA (*Share Alike*).

1.6 Is your LOD 5 star? ³⁹

Esiste un sistema di classificazione “a stelle” per i dati pubblicati sul web, ideato e sviluppato da Tim Berners-Lee.

Più i dati sono “potenti” ⁴⁰ e facili da usare, più stelle si ottengono.

- ★ Disponibili sul web (in qualunque formato) ma con una licenza aperta
- ★★ Disponibili come dati strutturati *machine readable*
(ad es. Excel invece della scansione dell'immagine di una tabella)
- ★★★ come “2 stelle” ma con un formato non proprietario (*Open Data*)
(es. CSV invece di Excel)
- ★★★★ In base a quanto sopra, con l'uso degli standard aperti di W3C (RDF e SPARQL) per identificare le entità, in modo che gli altri possano collegarvi i propri dati
- ★★★★★ In base a quanto sopra, più collegamenti esterni a dati di altre
★ istituzioni/persone, in modo da fornire un contesto (*Linked Data*)

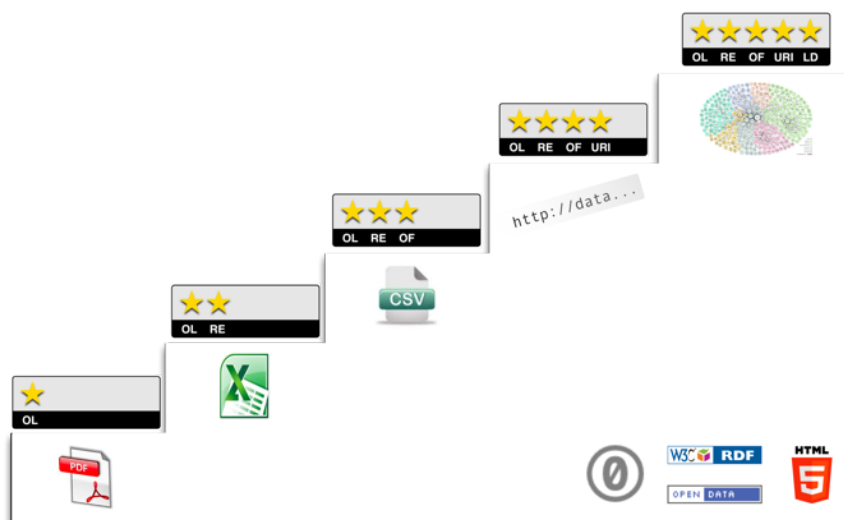


Fig. 1.8: Five Star Classification

³⁹ cit. T. B. Lee, *Linked Data*

⁴⁰ cit. T. B. Lee, *Linked Data*

Nel 2010 per i dati del governo è stato aggiunto un nuovo requisito, ovvero la presenza di metadati sui dati stessi e la disponibilità di accedere agli stessi da un catalogo principale. Ogni paese ha il proprio catalogo, indicato nel modo seguente: `data.gov.[sigla_paese]`⁴¹

1.7 I vantaggi dell'utilizzo di *Linked Data*.

Perché promuovere l'adozione di *Linked Data* al posto (o in aggiunta) delle consolidate tecniche di pubblicazione di dati sul web?

Tecniche standard di pubblicazione nel web	Vantaggi apportati dal LOD
Ampia varietà di modelli di dati diversi → l'eterogeneità risultante non facilita il processo di integrazione.	Modello di dati unificante: RDF garantisce l'identificazione univoca a livello globale delle entità e consente l'utilizzo di schemi diversi in parallelo per rappresentare i dati.
Le web API sono accessibili tramite diverse interfacce proprietarie: si basano su modelli di dati eterogenei e interfacce di accesso.	Meccanismo di accesso ai dati standardizzato: è previsto uno schema specifico per utilizzo del protocollo HTTP che consente l'accesso a fonti di dati utilizzando browser generici e la scansione dello spazio dati completo dai motori di ricerca.
Le web API e i dump di dati in formati proprietari rimangono "isole di dati" isolate nel web.	Rilevamento dei dati basato su collegamenti tra entità in diverse origini dati. Collegando tra di loro risorse provenienti da diversi dataset (collezioni di dati), collega anche i dataset stessi, così i repository vengono resi interoperabili. Questo approccio supera e rende ormai deprecati i "silos" di informazioni

⁴¹ Catalogo italiano: <https://www.dati.gov.it/>

	<p>chiuse, fornendo nuovi modi di accesso, convalida e utilizzo di dati provenienti da fonti diverse.</p>
<p>Espressività limitata.</p>	<p>Il LOD semplifica l'integrazione di dati provenienti da fonti diverse facendo affidamento su vocabolari condivisi, aventi definizioni recuperabili e termini (appartenenti a vocabolari diversi) collegati l'uno all'altro.</p> <p>E' quindi possibile rappresentare le informazioni in modo preciso grazie alla varietà di espressione che deriva dalla possibilità di utilizzare termini provenienti da vocabolari diversi.</p>
<p>Le applicazioni di mashup nel Web 2.0 funzionano e devono essere implementate necessariamente su un dataset fisso.</p>	<p>Le applicazioni possono scoprire nuove origini dati in fase di esecuzione seguendo i collegamenti RDF.</p> <p>Ogni tripla RDF, infatti, fa parte dello "spazio globale di dati" e quindi può essere utilizzata come punto di partenza per esplorare questo spazio. Ogni risorsa contiene link che puntano ad altri URI che, a loro volta, possono anche essere dereferenziati, e così via... possono a loro volta contenere collegamenti che puntano ad altri vocabolari, definendo così le mappature tra i vocabolari correlati.</p> <p>Questo è il modo in cui le descrizioni delle risorse sono intrecciate nel web of data ed è anche il modo in cui il web of data può essere navigato o scansionato da un crawler.</p> <p>I Linked Data facilitano così la scoperta delle risorse sul web da parte dei motori di ricerca e dagli utenti con i browser.</p>

Tab. 1.2 Vantaggi di pubblicazione dei dati secondo il paradigma LOD, rispetto al paradigma di pubblicazione tradizionale

Il LOD offre sicuramente un paradigma di pubblicazione più generico e flessibile, che rende più facile per i consumatori di dati (umani o non) scoprire ed integrare questi dati da un gran numero di fonti.

L'adozione al paradigma LOD consente alle istituzioni di:

- migliorare la propria presenza sul web (ovvero laddove oggi vengono ricercate la maggior parte delle informazioni)
- migliorare i servizi offerti ai propri utenti
- promuovere l'uso innovativo dei dati che detengono

Il *Semantic Web* oggi cerca di cambiare il panorama di Internet in vari modi:

- rimuovendo le barriere concettuali e intellettuali all'apertura dei dati e al loro utilizzo (*Open Data*);
- rimuovendo gli ostacoli tecnologici alla libera condivisione dei dati nel web (*Linked Data*);
- incoraggiando le aziende, le organizzazioni e gli individui a pubblicare liberamente i propri dati, in un formato standard aperto.

*We already see an increasing need and a rising obligation for people and organizations to make their data available.*⁴²

*[...] we believe the companies who choose to start exploiting Semantic Web technologies will be the first to reap the rewards.*⁴³

*Linked Open Data è un'espressione pratica del web semantico, utile e realizzabile per "liberare" e rendere interoperabile il sapere nel web. [...] In questo nuovo scenario di "know how" di modellazione della conoscenza, è importante che le istituzioni, in modo cooperante, si pongono il problema di come trasformare il loro "caos" di dati in una vero e proprio universo semantico di dataset interconnessi, cioè di come pubblicare e contestualizzare tali dati sul web, collegarli ad altri dati e renderli più utili e facilmente utilizzabili e, nel lungo periodo, molto più richiesti. La soluzione più incoraggiante (e largamente accettata dalla comunità internazionale) è, appunto, Linked Data.*⁴⁴

⁴² Shadbolt et al., 2006, pag. 100

⁴³ Hendler et al., 2002

⁴⁴ Solodovnik, 2011

2. Linked Data e beni culturali

2.1 I beni culturali

Si considerano “beni culturali” (più propriamente, *Cultural Heritage*) l’insieme di artefatti in senso lato (oggetti, pitture, sculture, edifici) ma anche tradizioni, costumi, che costituiscono il nostro patrimonio culturale.

Concretamente, per quanto riguarda l’Italia, possiamo dire che si tratta di musei, oggetti di valore storico-artistico, edifici, monumenti, chiese, quadri, mobili, arredi, oggetti di culto, tessuti, strutture urbanistiche, ecc...

Ogni soggetto che si trovi ad essere responsabile di beni culturali deve preoccuparsi della loro gestione che, in ambito ICT, significa:

- dal punto di vista amministrativo, catalogare i beni (o le loro rappresentazioni) in inventari (*repositories*) secondo schede organizzate in modo utile;
- da un punto di vista più utile al progresso, supportare all’attività di studio e ricerca (*search and retrieval*).

Mentre per “classificazione” si intende la raccolta di risorse correlate sotto un’etichetta comune (es. arte, scienza, medicina, ...), la catalogazione consiste nella “classificazione, registrazione e descrizione di un bene culturale”: crea cioè dei “surrogati delle risorse”, le loro descrizioni (*record* di catalogo), da utilizzare *al posto* delle risorse stesse⁴⁵.

Per descrivere e collocare in un archivio un “oggetto culturale” è quindi necessario ricorrere ad una sua *rappresentazione* (ovvero ad uno standard) nella quale sono menzionati tutti i suoi attributi.⁴⁶

⁴⁵ V. Casarosa, *Catalogazione - Schede bibliografiche, OPAC, MARC*, in *Biblioteche Digitali*, 2018, pag.5

⁴⁶ Corti, 1992, pag.20

Per rappresentare le risorse vengono utilizzati *standard di metadati*⁴⁷, specifici per ogni tipologia di bene: patrimonio archeologico, architettonico, storico artistico, librario, etnoantropologico.

Nel contesto delle applicazioni informatiche ai beni culturali, gli standard sono “l’insieme delle norme atte a strutturare l’informazione” e hanno lo scopo di uniformare i comportamenti di coloro che creano e che cercano rappresentazioni.⁴⁸

2.2 I *Linked Data* nell’universo dei beni culturali

Nel dinamico universo delle discipline inerenti ai beni culturali il termine GLAM è diventato consuetudinario, tanto che possiamo legittimamente considerare l’acronimo anglosassone per *Galleries Libraries Archives Museums* come espressione di un nuovo paradigma, un approccio innovativo per la tutela e la diffusione del patrimonio culturale.

Negli ultimi dieci anni, specialmente in virtù del sempre più decisivo apporto delle tecnologie digitali, le istituzioni culturali hanno avvertito fortemente la necessità di estendere su vasta scala e con modalità diverse l’accesso al proprio patrimonio, estendendo l’orizzonte conoscitivo attraverso il confronto e la collaborazione con altre realtà per proporre un nuovo concetto di collezione.

Questi promettenti scenari per lo sviluppo e la trasmissione della conoscenza trovano condizioni ideali in “ambienti” come il *Semantic Web*.

L’Agenda Digitale Europea⁴⁹ detta linee guida e indirizzi per rafforzare questa tendenza ed è ormai diffusa l’idea che il bene culturale debba oltrepassare i confini

⁴⁷ Alcune definizioni di “metadato”:

- *Machine-understandable information about Web resources or other things* (Berners-Lee 1997)
- *Data associated with objects which relieves their potential users of having to have full advance knowledge of their existence or characteristics; a user might be a program or a person* (L. Dempsey 1998)
- *Structured data about resources that can be used to help support a wide range of operations* (Michael Day 2001)
- *Structured data about data* (DCMI 2003)
- *Structured information that describes, explains, locates, or otherwise makes it easier to retrieve, use, or manage an information resource* (NISO 2004)

⁴⁸ Corti 1992, pag.8

⁴⁹ http://europa.eu/rapid/press-release_MEMO-10-200_it.htm

della sua fruizione fisica tradizionale assumendo una vasta dimensione “digitale e interconnessa”.

E' doveroso quindi spostare l'attenzione dal bene materiale in sé alle relazioni (immateriali) che esso intrattiene con altre risorse e concetti e lasciare al digitale (e in particolare al paradigma LOD) il compito di saldare queste due realtà, materiale e immateriale.

2.2.1 La comunicazione del patrimonio culturale

Per valorizzare al meglio il patrimonio culturale di una nazione è importante comunicarne il valore nel modo giusto.

La *comunicazione* applicata ai beni culturali è un tema molto importante: le prospettive di divulgazione, chiaramente, si ampliano se si sfrutta il potenziale dell'ICT⁵⁰.

Il contesto applicativo attuale è sempre più basato sul web: i servizi messi a disposizione degli utenti da parte delle istituzioni culturali (es. archivi navigabili, musei virtuali, Europeana⁵¹...) sono un modo molto valido per fare comunicazione sui beni culturali.

L'applicazione sempre più massiccia e pervasiva delle tecnologie ICT ai beni culturali ha condotto alla generazione di un'enorme quantità di dati artistici, archeologici, archivistici, librari, ecc... in formato digitale. Tali dati sono *eterogenei*, cioè si caratterizzano per la loro varietà nella tipologia dei media e dei formati, nella dimensione, nei livelli di accessibilità, nei requisiti software e hardware necessari per la loro fruizione.

Il diffondersi del fenomeno è accompagnato dalla compresenza di due tendenze opposte: se da un lato continuano a svilupparsi standard indipendenti, basati su metodologie e requisiti locali, che danno luogo ad una varietà di dataset non interoperabili, dall'altro le autorità nazionali avvertono il problema di gestione e

⁵⁰ “C'è una crescente attenzione per un uso efficace dell'ICT nel mondo dei beni culturali” (P. Paolini, N. Di Blas, F. Alonzo, *ICT per i beni culturali: esempi di applicazione in Mondo Digitale*, n°3, 2005, pag.60)

⁵¹ <https://www.europeana.eu/portal/it>

condivisione di una mole enorme di dati (appunto eterogenei) e iniziano a rispondere costituendo gruppi di lavoro per lo sviluppo di standard, liste di termini, thesauri, lessici e vocabolari specifici per i beni culturali.⁵²

2.2.2 I vantaggi dei *Linked Data* per le istituzioni culturali

Nel quadro del patrimonio culturale digitale, le cosiddette *Memory Institutions* (istituzioni della memoria) si occupano di gestire una grande massa di contenuti digitali che deve essere:

- immagazzinata e conservata a lungo termine in modo sicuro,
- accessibile in modo permanente,
- facilmente riutilizzabile nel tempo da diverse comunità di ricerca e gruppi di utenti.

L'obiettivo è di adottare una "infrastruttura digitale condivisa" per i beni culturali che porterà alla riduzione dei costi e al miglioramento dell'interoperabilità attraverso:

- l'adozione di procedure comuni,
- nuovi flussi di lavoro e metodologie di ricerca,
- sviluppo di nuove competenze e abilità.

Perciò è stato (ed è tuttora) necessario stabilire politiche e strategie comuni, che tengano conto di processi, servizi, strumenti e protocolli condivisi, per lo sviluppo di una *e-Infrastructure* per il patrimonio culturale.⁵³

Proprio in questo contesto:

*[...] la tecnologia offerta dai Linked Data è un'opportunità di straordinaria importanza.*⁵⁴

⁵² F. Grew, *From museum store to data warehouse: Archaeological archives for the twenty-first century*, in G. Lock, K. Brown (eds.), *On the Theory and Practice of Archaeological Computing*, Oxford University Committee for Archaeology, Oxford, 2000, pag. 56

⁵³ S. Di Giorgio, *Il portale Culturaltalia nel quadro delle collaborazioni europee*, Roma, 2016, pp. 7-8

⁵⁴ Crupi, 2013

La grande opportunità che i *Linked Data* offrono al mondo dei beni culturali è quella di *integrare* i dati strutturati dei cataloghi con informazioni provenienti da altri cataloghi⁵⁵ e da fonti terze, in modo da ottenere vantaggi in termini di⁵⁶:

- *Interoperabilità*: grazie all'utilizzo di formati standard, i LOD garantiscono l'interoperabilità tra diversi *repository* sia da un punto di vista formale (in quanto codificati in un formato comune), sia concettuale (in quanto definiti utilizzando classi omogenee).
- *Riduzione della duplicazione* delle informazioni: chi crea un *dataset* può collegarlo direttamente a *dataset* esistenti di cui non dispone direttamente; chi crea un *mashup*, un sito o un'applicazione di tipo ibrido, che includa dinamicamente informazioni o contenuti provenienti da più fonti, invece di importare i dati può collegarli. Meno lavoro, quindi, ma soprattutto dati sempre aggiornati.
- *Autorevolezza*: DBpedia⁵⁷ e Project Gutenberg⁵⁸ vengono spesso indicati come fonti di metadati autorevoli. Gli istituti culturali possono affermarsi anch'essi come fonti autorevoli di informazioni sul patrimonio culturale, realizzando una "spina dorsale" per lo sviluppo del web semantico.
- *Maggiore evidenza*: i LOD facilitano gli utenti nella scoperta dei contenuti, mettendo in evidenza i dati prodotti dagli enti e aumentando il traffico verso i siti web degli istituti produttori.
- *Nuovo pubblico*: quando gli utenti analizzano i dati e li utilizzano per creare applicazioni come API e *mashup*, propongono i contenuti ad un pubblico nuovo che difficilmente sarebbe raggiungibile senza queste soluzioni.
- *Migliore esperienza per gli utenti*: fornendo agli utenti informazioni di alta qualità e contestualmente utili, si migliorerà la loro esperienza di fruizione.

⁵⁵ Vedi, per esempio, il caso dei LOD di Europeana che alimentano dei database irlandesi:

<https://pro.europeana.eu/data/linked-logainm>

⁵⁶ http://www.culturaitalia.it/opencms/linked_open_data_it.jsp

⁵⁷ <https://wiki.dbpedia.org/>

⁵⁸ <http://www.gutenberg.org/>

- *Uso efficiente delle risorse*: la condivisione di dati fa sì che essi possano essere utilizzati in modo più efficiente permettendo agli utenti di contribuire ad arricchire i metadati. Questo comporta anche il riutilizzo diretto in settori come la formazione, la ricerca scientifica e il turismo culturale.

Parola chiave di tale processo è *interoperabilità*⁵⁹, non solo tecnologica ma soprattutto *semantica*, quella che nasce dall'incontro di istituzioni culturali diverse, con i loro linguaggi e il loro differente modo di classificare e rappresentare le "cose" del mondo.

Negli ultimi decenni, le istituzioni culturali hanno scoperto e messo a frutto sempre di più i vantaggi che derivano dall'utilizzo di metodologie e strumenti di gestione della conoscenza (*Knowledge Management*), sfruttando concetti chiave come rappresentazione della conoscenza (*Knowledge Representation*), metadati, modellazione concettuale, interoperabilità sintattica e semantica, ontologie, ecc ...⁶⁰

Il mondo degli archivi, dei musei e delle biblioteche⁶¹ entra nel mondo del *semantic web*, immettendo in esso una consolidata tradizione sul controllo di autorità dei dati, sulla sensibilità e capacità di gestire le informazioni catalogando il sapere, creando nuove connessioni semantiche tra i documenti, fornendo loro valore aggiunto attraverso la struttura collegata dei cataloghi, degli indici e degli *standard di classificazione e catalogazione*.

Una sensibilità, che oggi si traduce nella progettazione di nuovi contesti e strumenti per consentire un accesso intuitivo ai contenuti e un loro facile recupero, per fruitori umani e non. Ecco perché è fondamentale che i dati strutturati siano presenti in rete e accessibili con nuovi strumenti, compatibili con le tecnologie e gli standard del web.

⁵⁹ Interoperabilità significa, in questo contesto, "rendere accessibili e disponibili i dati, in modo tale da essere elaborabili da una macchina per consentire la loro integrazione e il loro riutilizzo in applicazioni differenti".

⁶⁰ Signore et. al, 2005, pag.291

⁶¹ Per approfondire sul tema dei LOD nelle biblioteche digitali: A. Iacono, *Dal record al dato. Linked data e ricerca dell'informazione nell'OPAC*, in *Rivista italiana di biblioteconomia, archivistica e scienza dell'informazione*, vol.5, n°1, 2014, <https://www.jlis.it/article/viewFile/9095/8662>

I *Linked Data* daranno vita a nuovi servizi basati sul bagaglio di conoscenze e di pratiche che sono parte integrante della tradizione del patrimonio culturale.

L'opportunità offerta da questo nuovo metodo di pubblicare dati sul web, comporterà una radicale trasformazione del rapporto tra utente (umano e non) e universo dei beni culturali:

- l'integrazione dei propri dati con quelli di altre istituzioni non solo aumenta il loro potenziale informativo ma li rende più completi e riusabili in contesti anche molto diversi da quelli d'origine;
- l'aggregazione e il collegamento con altre risorse web, anche se strutturate secondo standard differenti, consente di estendere le informazioni di contesto relative a ciascun dato;
- l'incontro con altri dati nel web accresce il numero degli strumenti di controllo terminologico a disposizione, rendendo più precise e pertinenti le fonti informative, la cui riconosciuta autorevolezza è criterio distintivo fondamentale per conferire legittimità e validità ai dati;
- far uscire fuori dal web dati locali e renderli aperti, universalmente accessibili, significa offrire opportunità di visibilità a culture minoritarie;
- l'integrazione dei dati catalografici all'interno del web semantico comporta, di riflesso, un arricchimento degli stessi cataloghi e la possibilità di offrire nuovi servizi basati sulla tecnologia e sul linguaggio del web.

La *rivoluzione copernicana dei Linked Data* consiste proprio nel fatto che i collegamenti tipici del web tradizionale, acquistano, nel contesto del *semantic web*, un ruolo semantico primario, una funzione predicativa che dà significato ai dati stessi, poiché comporta la frantumazione dell'informazione in singole componenti atomiche (entità) e rappresenta ed esprime i differenti tipi di relazione che esse possono intrattenere.

2.2.3 I Linked Data tra le nuove tecnologie emergenti

Questo paragrafo si occupa di analizzare la tecnologia dei *Linked Open Data* facendo riferimento ad un modello ciclico sviluppato da Gartner.

Gartner Inc.⁶² è una società per azioni multinazionale, leader mondiale nella consulenza strategica, ricerca e analisi nel campo della tecnologia dell'informazione (ICT)⁶³. L'attività principale consiste nel supportare le decisioni di investimento dei suoi clienti attraverso ricerca, consulenza, benchmarking, eventi e notizie.

Gartner ha sviluppato una metodologia chiamata modello *Hype Cycle* (lett. *ciclo dell'esagerazione*) per rappresentare graficamente la maturità, l'adozione e l'applicazione di specifiche tecnologie.⁶⁴ Questo modello è applicato a diversi settori (sicurezza, privacy, istruzione, *open software*, e-commerce, tecnologie blockchain, ecc...) è articolato in cinque fondamentali fasi del ciclo di vita di una tecnologia.

1. *Technology Trigger* (innesco della tecnologia): una nuova tecnologia potenzialmente dirompente viene avviata. I principi si dimostrano fondati e l'attenzione dei media scatena pubblicità. Spesso non esistono ancora prodotti utilizzabili e non c'è prova della validità commerciale della tecnologia.
2. *Peak of Inflated Expectations* (picco delle aspettative esagerate): la pubblicità dà luogo ad una serie di storie di successo, spesso accompagnate da molti di casi di fallimento. La maggior parte delle imprese non agiscono.
3. *Trough of Disillusionment* (fossa della disillusione): l'interesse nella tecnologia svanisce quando la sperimentazione e l'implementazione non producono i risultati sperati. I (pochi) produttori della tecnologia entrano in crisi o falliscono.

⁶² <https://www.gartner.com/en>

⁶³ https://it.wikipedia.org/wiki/Information_technology

⁶⁴ Come tutti i modelli di successo, anche questo ha ricevuto numerose critiche di cui viene qui elencata la letteratura e i riferimenti.

- Richard Veryard, *Technology Hype Curve*, in *demandingchange.blogspot.com*, 2005
<http://demandingchange.blogspot.com/2005/09/technology-hype-curve.html>
- Gerald Weinberg et al., *HypeCycle*, *AYE Conference*, 2003
<http://www.davewsmith.com/ayewiki/scribble.cgi?read=HypeCycle>
- Jorge Aranda, *Cheap shots at the Gartner Hype Curve*, in *catenary.wordpress.com*, 2006
<http://catenary.wordpress.com/2006/10/22/cheap-shots-at-the-gartner-hype-curve/>

Le critiche riguardano il fatto che il modello non rappresenta esattamente un ciclo, che non abbia fondamento scientifico, che non tiene in considerazione dei cambiamenti temporali relativi alla velocità alla quale la tecnologia si sviluppa e che il risultato finale non dipende dalla natura della tecnologia in sé. Inoltre il "ciclo" non produce benefici in termini di sviluppo e marketing di nuove tecnologie, ma fornisce solamente commenti su tendenze in atto. Con i termini (soggettivi e ambigui) "disillusione", "illuminazione" e "aspettative" non può essere descritto oggettivamente o chiaramente lo stato attuale di una tecnologia.

Gli investimenti continuano solo per quelle imprese sopravvissute grazie al miglioramento delle tecnologie di prodotto per soddisfare gli *early adopter*⁶⁵ (“utenti precoci”).

4. *Slope of Enlightenment* (salita dell'illuminazione): comincia a diffondersi ampiamente e la consapevolezza di come la tecnologia possa avvantaggiare le imprese in diversi modi. Gli sviluppatori creano prodotti di seconda e terza generazione. Un numero crescente di imprese finanzia progetti pilota, mentre quelle conservatrici restano prudenti.
5. *Plateau of Productivity* (altopiano della produttività): l'adozione della tecnologia prende sempre più piede e diventa tendenza. L'applicabilità e la rilevanza per il mercato di massa producono frutti.

L'*Hype Cycle* pubblicato a giugno del 2015 (fig. sotto) relativo al settore *Advanced Analytics and Data Science*, mostra che i *Linked Data* in quell'anno avevano già raggiunto il terzo step, ed erano quindi pronti per “risalire verso l'illuminazione” e giungere poi, tra i 5 e i 10 anni di tempo (cioè esattamente nel range dei nostri giorni) nel pieno della produttività.⁶⁶

Questo modello quindi conferma che i *Linked Data* sono (e saranno nei prossimi anni) una delle tecnologie più applicate, rilevanti e di tendenza nel settore informatico. Molte istituzioni infatti stanno assegnando fondi per lo “sforzo *Linked Data*” e per la produzioni di servizi relativi ad essi.

*Work continues in earnest to climb Gartner's Slope of Enlightenment.
And expectations are high.*⁶²

⁶⁵ https://it.wikipedia.org/wiki/Early_adopter

⁶⁶ Per quanto riguarda il mondo delle biblioteche digitali, i *Linked Data* rischiavano di rimanere nel Trough of Disillusionment fino a quando “MARC danced on its grave” (A. K. Pace, *Linked data in libraries: From disillusionment to productivity*, in OCLC.org, 2018) <http://www.oclc.org/blog/main/linked-data-in-libraries-from-disillusionment-to-productivity/>

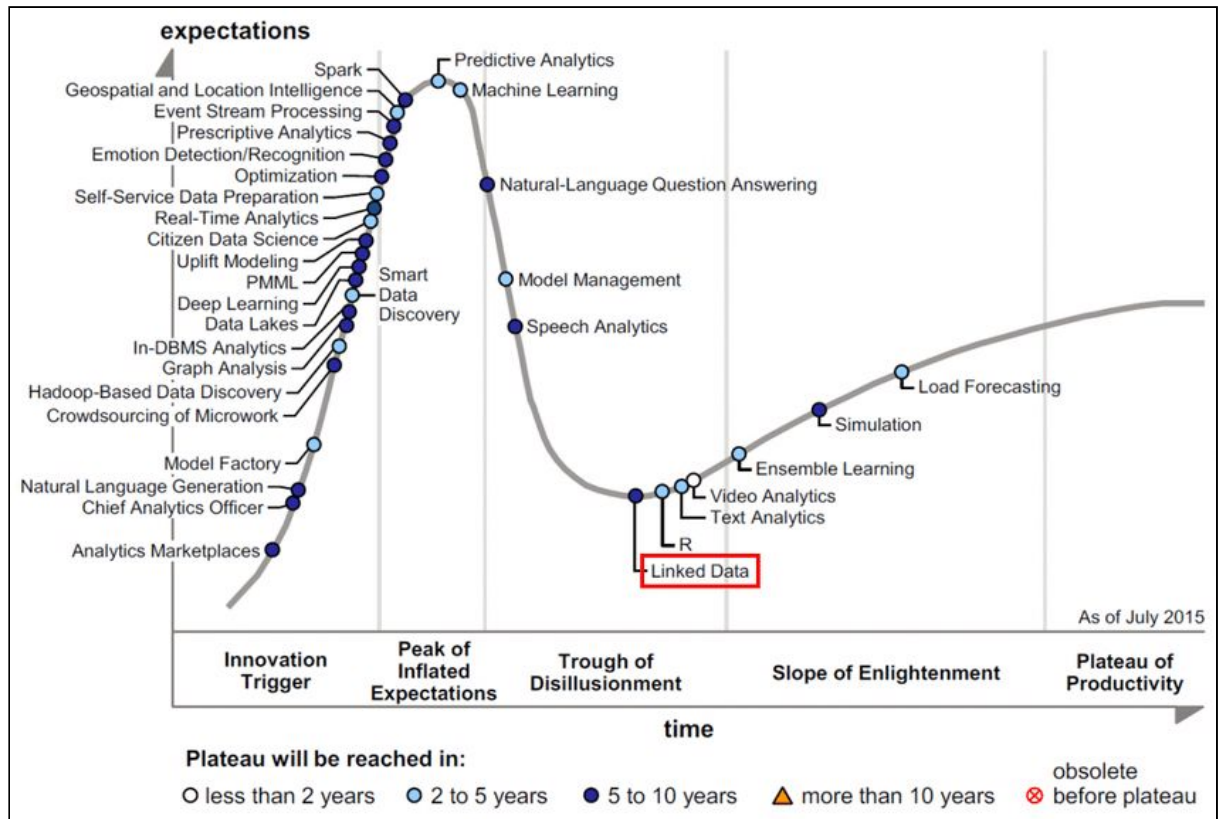


Fig. 2.1 Gartner's *Hype Cycle for Advanced Analytics and Data Science 2015*, from *Research Gate*⁶⁷

Nel ciclo del 2018 relativo alle tecnologie emergenti, sorprendentemente, i *Knowledge Graph* stanno ancora salendo la curva al picco delle aspettative gonfiate.

67

https://www.researchgate.net/figure/Gartners-Hype-Cycle-for-Advanced-Analytics-and-Data-Science-as-of-July-2015-3_fig1_326557001

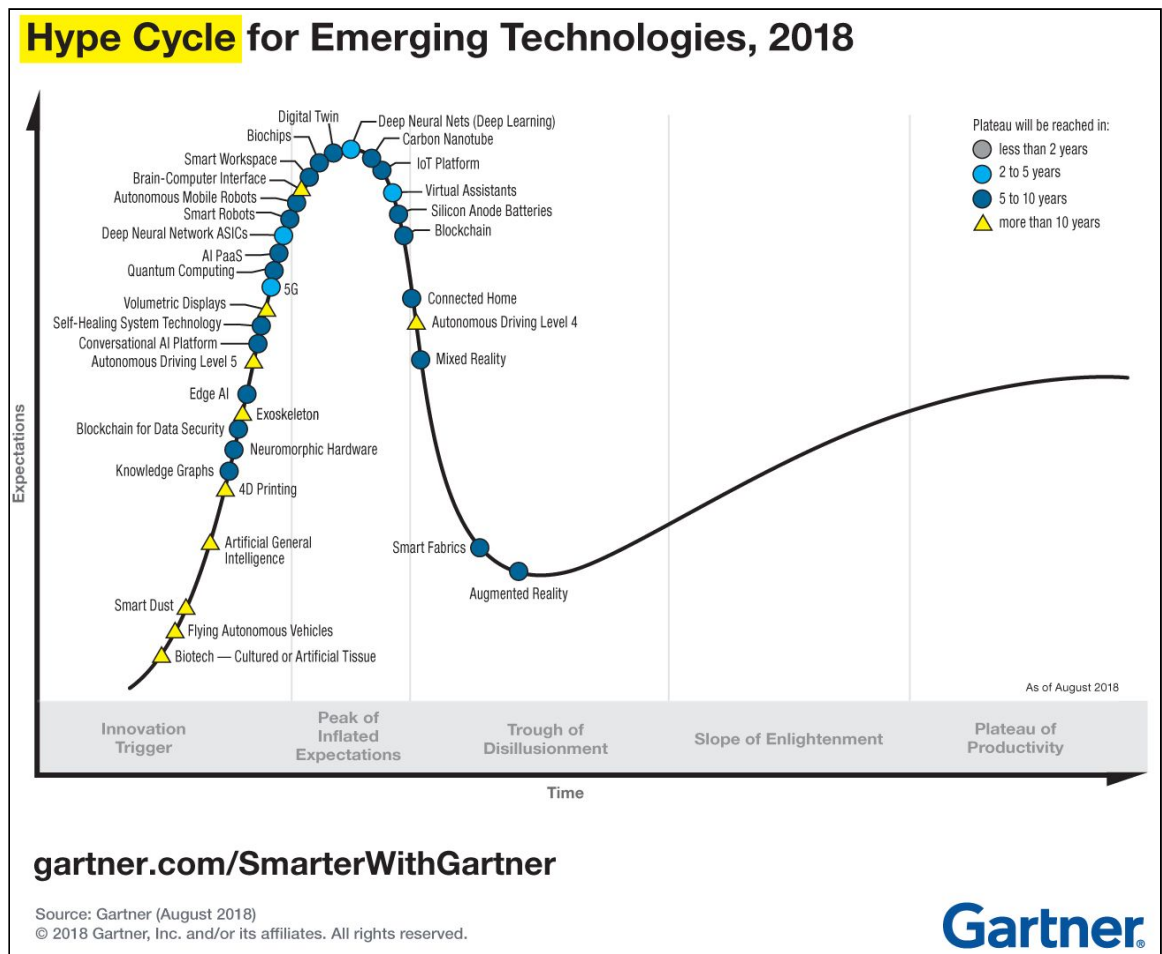


Fig. 2.2 Gartner's *Hype Cycle for Emerging Technologies, 2018*, from *Gartner.com*⁶⁸

2.2.4 Ripensare la conoscenza

Tutti i principi che costituiscono il paradigma dei *Linked Data*, applicati nel mondo del *Cultural Heritage*, *modificano i processi cognitivi* che hanno finora governato la nostra relazione con l'universo dei beni culturali e con gli strumenti che storicamente hanno mediato il rapporto tra utente e conoscenza (cataloghi, record, apparati indicali, ecc.).

*[...] the new semantic web technology may change the way knowledge is produced and shared.*⁶⁹

⁶⁸

<https://www.gartner.com/smarterwithgartner/5-trends-emerge-in-gartner-hype-cycle-for-emerging-technologies-2018/>

⁶⁹ J. Hendler, T .B. Lee, *Publishing on the Semantic Web*, in *Nature*, 2001 pp. 1023-1025

Per queste ragioni, archivi e musei oggi si preoccupano di convertire l'informazione in dati e metadati di qualità.

Non possiamo entrare nell'ambiente informativo, ricco e dinamico, del web del XXI secolo con dati basati sui principi del XIX secolo⁷⁰

La qualità di un archivio infatti non si misura tanto dalla quantità di dati che possiede, quanto dalla capacità di strutturare e di modellizzare questi dati, di renderli accessibili, conservando al contempo la stratificazione dei contesti, la relazione tra quello nuovo che si crea e il contesto di origine, oltre che con tutti gli altri documenti con cui essi stringono relazioni semantiche esplicite o implicite.

Stop hugging your data⁷¹: Tim Berners-Lee nel 2009 invitava così le istituzioni a rendere accessibili i propri dati e a farli uscire dai “silos di informazione”. I dati infatti acquistano valore quando sono interconnessi con altri dati.

Questo sforzo – già in atto – ha lo scopo di rendere identificabili in modo univoco i propri dati nel contesto del web e renderli disponibili ad essere leggibili, interpretabili e usabili dalle macchine.

La *mission* delle istituzioni culturali oggi è quella di trattare e comunicare le informazioni sui beni culturali partecipando alla costruzione del web semantico e alla divulgazione della conoscenza del *cultural heritage*.

Nuovi standard e metodi di rappresentazione sono stati definiti e applicati da queste istituzioni al fine di accrescere la visibilità e facilitare l'accesso delle risorse conservate negli archivi.

Linked Data becomes more powerful the more of it there is.⁷²

⁷⁰ Coyle, 2013

⁷¹ https://www.youtube.com/watch?v=_J8H-ZVJwk8

⁷² Byrne & Goddard 2010

2.3 Panoramica Internazionale

Presupposto imprescindibile alla base di qualsiasi condivisione delle informazioni è dunque l'esistenza di un "linguaggio comune": oltre all'interoperabilità tra sistemi, per cooperare, è necessario che le informazioni vengano raccolte in modo rigoroso e secondo poche ma precise regole, ovvero secondo standard condivisi.

La *International Organization for Standardization* (ISO) è la prima agenzia internazionale per la standardizzazione in (quasi) tutti i campi⁷³; a sua volta, ISO è suddiviso in una serie di comitati tecnici responsabili della definizione di standard in settori specifici.

Altre istituzioni culturali e organizzazioni d'importanza internazionale hanno lavorato per definire standard di metadati:

- *Library of Congress* (LoC)⁷⁴
- *World Wide Web Consortium* (W3C)⁷⁵
- *International Council of Archives* (ICA)⁷⁶
- *Society of American Archivists* (SAA)⁷⁷
- *American Library Association* (ALA)⁷⁸
- *International Federation of Library Associations and Institutions* (IFLA)⁷⁹
- Europeana⁸⁰, biblioteca digitale che riunisce contributi già digitalizzati da diverse istituzioni dei 28 paesi membri dell'Unione europea in 30 lingue e offre accesso a oltre 45 milioni di oggetti digitalizzati provenienti da musei, biblioteche, archivi e collezioni audiovisive di tutta Europa.

⁷³ Corti 1992, pag.184

⁷⁴ LoC: <https://www.loc.gov/>

⁷⁵ W3C: <https://www.w3.org/>

⁷⁶ ICA: <https://www.ica.org/en>

⁷⁷ SAA: <https://www2.archivists.org/>

⁷⁸ ALA: <http://www.ala.org/>

⁷⁹ IFLA: <https://www.ifla.org/>

⁸⁰ Europeana: <https://www.europeana.eu/portal/it>

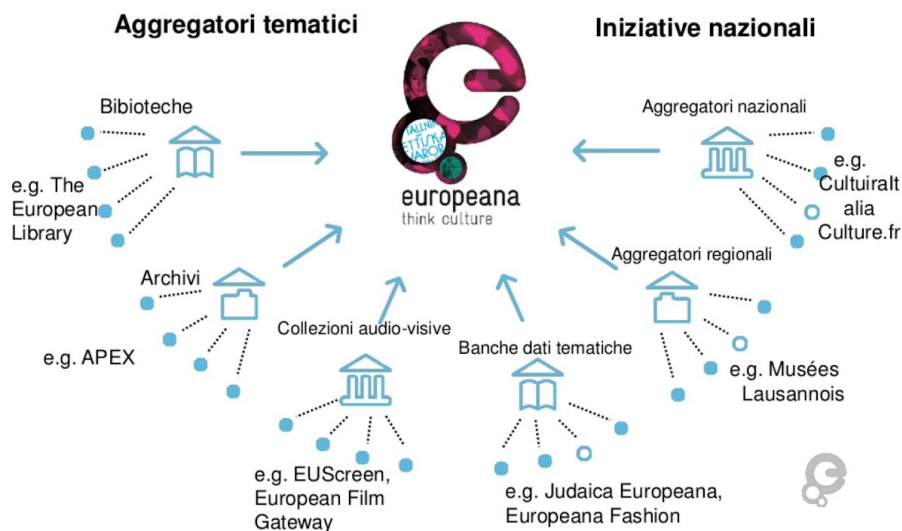


Fig.2.3 Fornitori di dati verso Europeana⁸¹

2.3.1 L'Authority Control

Prima di procedere con la descrizione e l'analisi degli standard, è necessaria una premessa circa l'attività denominata "controllo di autorità".

In ambito biblioteconomico e documentalistico, con "controllo di autorità" (dall'inglese *Authority Control*) si indica l'attività volta a organizzare un catalogo o un archivio utilizzando un singolo termine per ogni entità (autore, bibliografia, ecc.). Attraverso il rinvio da forme varianti del nome a questa unica forma del nome (chiamata "voce di autorità" o *authority record*) si consente l'accesso, per esempio, a tutte le opere di un autore a prescindere dal fatto che in una specifica opera egli sia nominato con forma diversa. Tale controllo viene effettuato anche per i titoli (con rinvii da forme di titolo varianti, traduzioni ecc.).

Il termine "autorità" sta ad indicare che le entità sono stabilite in una particolare forma che consente la loro applicazione coerente, facilitando il lavoro di catalogazione degli addetti ai lavori e la ricerca da parte dell'utente.

Il controllo di autorità è fondamentale per comprendere le funzioni di alcuni standard di metadati elencati di seguito e per descrivere il dataset VIAF ([cap. 5.4.2](#)) e l'attività di disambiguazione ([cap. 5.4.4](#))

⁸¹ S. Di Giorgio, 2016, pag. 44

2.3.2 Standard di metadati internazionali

Nella tabella seguente sono riportati, in ordine alfabetico, gli standard di metadati più conosciuti e utilizzati su scala internazionale.

Nome		Descrizione	Creatore
BIBO ⁸²	<i>Bibliographic Ontology</i>	E' un'ontologia per il web semantico per fornisce i concetti e le proprietà principali per la descrizione di citazioni e riferimenti bibliografici (ad esempio citazioni, libri, articoli, ecc.); è scritta in RDF e può essere per qualsiasi tipo di documento in RDF. ⁸³	LoC
DC ⁸⁴	<i>Dublin Core</i>	Standard per descrivere oggetti digitali e risorse sul web. ⁸⁵ Per esprimere il DC si può usare la sintassi XML (namespace "dc") oppure la sintassi RDF. <ul style="list-style-type: none"> • Simple DC: contiene 15 elementi di base. • Qualified DC: estensione del Simple DC. 	W3C
EAC-CPF ⁸⁶	<i>Encoded Archival Context for Corporate Body, Persons and Families</i>	Standard per la codifica in XML e l'interscambio di record di autorità basati sull'ISAAR (CPF). EAC-CPF può essere utilizzato in congiunzione con EAD per migliorarne le capacità di codifica di strumenti di ricerca,	Yale University Library

⁸² BIBO: <http://bibliontology.com/#>

⁸³ <https://github.com/structuredynamics/Bibliographic-Ontology-BIBO>

⁸⁴ Dublin Core: <http://dublincore.org/>

⁸⁵ A differenza delle schede MARC, che richiedono un bibliotecario di professione per essere sviluppate, le descrizioni in Dublin Core sono più semplici e anche più generiche, così da poter descrivere risorse di diversi tipi. MARC e Dublin Core infatti sono rappresentativi di due filosofie diametralmente opposte in termini di complessità: il primo è uno standard completo che può essere sviluppato solo da professionisti, mentre il secondo è volutamente minimalista per essere applicato a risorse eterogenee da persone non necessariamente esperte nella catalogazione (Casarosa, 2018)

⁸⁶ EAC-CPF: <https://eac.staatsbibliothek-berlin.de/>

		ma può anche essere utilizzato in combinazione con altri standard o per la codifica di file di autorità.	
EAD ⁸⁷	<i>Encoded Archival Description</i>	Standard XML per la codifica di supporti per la ricerca di archivi. Nato per creare uno standard per la descrizione degli archivi simile al MARC per la descrizione dei materiali bibliografici.	SAA in collaborazione con LoC
EDM ⁸⁸	<i>Europeana Data Model</i>	Per migliorare il modo in cui raccoglie, gestisce e pubblica metadati, Europeana ha sviluppato un nuovo standard, ispirato al Semantic Web e che riutilizza vocabolari e ontologie noti.	Europeana
ICP	Principi Internazionali di Catalogazione (ICP)		IFLA
	FRBR ⁸⁹	<i>Functional Requirements for Bibliographic Records</i>	
	FRAD ⁹⁰	<i>Functional Requirements for Authority Data</i>	
	FRSAD ⁹¹	<i>Functional Requirements for Subject Authority Data</i>	
	LRM ⁹²	<i>Library Reference Model: modello concettuale di riferimento ad alto livello che consolida i tre modelli precedenti.</i>	
ISAAR (CPF) ⁹³	<i>International Standard Archival</i>	Standard per l'elaborazione di record di autorità archivistici che offrono descrizioni di	ICA

⁸⁷ EAD: <https://www.loc.gov/ead/>

⁸⁸ EDM:

https://pro.europeana.eu/files/Europeana_Professional/Share_your_data/Technical_requirements/EDM_Documentation//EDM_Definition_v5.2.8_102017.pdf

⁸⁹ FRBR: https://www.ifla.org/files/assets/cataloguing/frbr/frbr_2008.pdf

⁹⁰ FRAD_IT: https://www.ifla.org/files/assets/cataloguing/frad/frad_2009-it.pdf

⁹¹ FRSAD:

<https://www.ifla.org/files/assets/classification-and-indexing/functional-requirements-for-subject-authority-data/frsad-final-report.pdf>

⁹² LRM:

https://www.ifla.org/files/assets/cataloguing/frbr-lrm/ifla-lrm-august-2017_rev201712.pdf

	<i>Authority Records for Corporate Bodies, Persons and Families</i>	entità (enti, persone e famiglie) coinvolte nella produzione e conservazione degli archivi.	(trad. italiana dell'ICAR)
ISAD (G) ⁹⁴	<i>General International Standard Archival Description</i>	Standard per la descrizione di archivi destinati alla registrazione di documenti prodotti da organizzazioni, persone e famiglie. ⁹⁵	
ISBD ⁹⁶	<i>International Standard for Bibliographic Description</i>	Standard per le descrizioni bibliografiche, avente lo scopo di rendere universalmente e facilmente disponibili i dati bibliografici relativi a tutti i tipi di risorse pubblicati in ogni paese. Lo scopo principale è fornire criteri uniformi per la condivisione delle informazioni tra istituzioni bibliografiche nazionali.	IFLA
ISNI ⁹⁷	<i>International Standard Name Identifier</i>	Standard per assegnare un numero identificativo univoco ai milioni di contributori ai lavori creativi e coloro che sono attivi nella loro distribuzione, inclusi ricercatori, inventori, scrittori, artisti, creatori visivi, artisti, produttori, editori, aggregatori e altro.	ISO
MARC ⁹⁸	<i>MAchine Readable Cataloguing</i>	Standard per un formato comune di scheda bibliografica digitale. Sviluppato allo scopo di facilitare la condivisione di risorse fra cataloghi. Raggruppa i campi di una scheda bibliografica in 10 categorie, e assegna un tag numerico a campi e sottocampi. Delle	LoC

⁹³ ISAAR(CPF)_IT: <http://www.icar.beniculturali.it/index.php?id=55>

⁹⁴ ISAD(G)_IT: <http://www.icar.beniculturali.it/index.php?id=54>

⁹⁵ Mappatura ISAD(G) - EAD3: [https://www.loc.gov/ead/EAD3taglib/EAD3.html#appendix-ISAD\(G\)toEAD3](https://www.loc.gov/ead/EAD3taglib/EAD3.html#appendix-ISAD(G)toEAD3)

⁹⁶ ISBD: <https://www.isbd.org/>

⁹⁷ ISNI: <http://www.isni.org/>

⁹⁸ MARC: <https://www.loc.gov/marc/>

		varie versioni rilasciate, quella attuale è il MARC21 (cioè il MARC del 21° secolo). ⁹⁹	
	<i>Universal MARC</i> (UNIMARC)	Esistono molte versioni nazionali del MARC: UKMARC, DANMARK, ecc... La versione standard è l'UNIMARC (<i>Universal MARC</i>). Esiste anche una versione <i>Human Readable</i> : HURC (<i>Human Readable MARC catalogue record</i>).	
	MARFXML ¹⁰⁰	Una DTD che descrive il MARC21 in XML. Rappresentazione dell'informazione bibliografica semanticamente equivalente a MARC21 ma che adotta XML come sintassi.	
METS ¹⁰¹	<i>Metadata Encoding and Transmission Standard</i>	Standard di metadati descrittivi, amministrativi e strutturali riguardanti oggetti digitali in una <i>digital library</i> ; basato su XML.	LoC
METS Rights ¹⁰²		Standard per la descrizione dei diritti d'uso utilizzati all'interno dei metadati METS	
MIX ¹⁰³	<i>NISO Metadata for Images in XML</i>	Standard di metadati per la descrizione di immagini; consiste in uno schema XML per la codifica di dati tecnici per gestire collezioni di immagini digitali	LoC
MODS ¹⁰⁴	<i>Metadata Object Description Schema</i>	Schema per descrizioni bibliografiche basato su XML. Consiste in un sottoinsieme di elementi MARC: è il DC del mondo bibliotecario (più semplice da usare anche da un meno esperto).	LoC

⁹⁹ Mappatura MARC21 - EAD3:

<https://www.loc.gov/ead/EAD3taglib/EAD3.html#appendix-MARC21toEAD3>

¹⁰⁰ MARFXML: <https://www.loc.gov/standards/marFXML/>

¹⁰¹ METS: <https://www.loc.gov/mets/>

¹⁰² METS Right: <https://www.loc.gov/standards/rights/METSRights.xsd>

¹⁰³ MIX: <https://www.loc.gov/mix/>

¹⁰⁴ MODS <https://www.loc.gov/standards/mods/>

RDA ¹⁰⁵	<i>Resource Description and Access</i>	<p>Standard per la descrizione e l'accesso alle risorse bibliografiche progettato per il mondo digitale. È un insieme di linee guida e istruzioni per la registrazione di dati bibliografici, indirizzato principalmente alle biblioteche, e ad altre istituzioni culturali come musei e archivi.</p> <p>Nel 2015 è stata pubblicata dall'ICCU, per conto del MiBAC, la traduzione italiana di RDA¹⁰⁶.</p> <p><i>Il nuovo codice per l'identificazione delle risorse [...] produce metadati strutturati e interconnessi [...] facendo sì che le biblioteche continuino a restare rilevanti, anche nell'epoca del web semantico.</i> (Buttò, 2016)</p>	ALA, LoC e altri
VRA Core ¹⁰⁷	<i>Visual Resources Association</i>	Standard per la descrizione di opere di cultura visiva e delle immagini che le documentano.	LoC

Tab. 2.1 Standard di metadati internazionali

2.4 Panoramica nazionale

*Il patrimonio italiano di beni culturali è tra i più rilevanti al mondo*¹⁰⁸

L'Italia ospita una delle più grandi collezioni di beni culturali al mondo, sparsi sul territorio nazionale e appartenenti ai più diversi soggetti (il Ministero per i Beni e le

¹⁰⁵ RDA: <http://www.rda-rsc.org/content/about-rda>

¹⁰⁶ RDA_IT:

http://www.iccu.sbn.it/opencms/export/sites/iccu/documenti/2015/RDA_Traduzione_ICCU_5_Novembre_REV.pdf

¹⁰⁷ VRA: <http://vraweb.org/> , <https://www.loc.gov/standards/vracore/>

¹⁰⁸ Paolini et al. 2005, pag.60

Attività Culturali, le Regioni, le Province, le città, la Chiesa Cattolica, i musei, privati cittadini...).

Le fonti d'informazioni sono altrettanto variegata: dai cataloghi e basi di dati di coloro che custodiscono / posseggono i beni, si va alle produzioni scientifiche delle università, all'editoria del settore artistico, alle produzioni legate al turismo, ecc..

Nelle nazioni dove i compiti di tutela e conservazione del patrimonio culturale sono assolti dallo stato, sono gli organi appositamente costituiti che hanno l'onere di compilare e diffondere le norme.¹⁰⁹

Il tentativo di standardizzare le informazioni del patrimonio culturale Italiano è giunto dal Ministero per i Beni e le Attività Culturali (MiBAC)¹¹⁰ e dai suoi organi, che hanno definito quali devono essere i dati dell'inventario, diversi a seconda del tipo di bene.

Gli organi¹¹¹ del MiBAC sono:

- l'Istituto per il Catalogo Unico delle biblioteche italiane e per le informazioni bibliografiche (ICCU)¹¹²
- l'Istituto Centrale per il Catalogo e la Documentazione (ICCD)¹¹³
- l'Istituto Centrale per gli Archivi (ICAR)¹¹⁴

L'obiettivo di avere un inventario "ragionevole" del patrimonio culturale italiano pubblico è ancora lontano [...] Inoltre, lo sforzo di catalogazione da solo NON E' SUFFICIENTE per l'attività di comunicazione e valorizzazione dei dati. Milioni di schede di catalogazione prodotte, e tuttora in produzione, sono difficilmente accessibili alla comunità di studiosi: "giacciono" in archivi e non vengono consultate¹¹⁵.

¹⁰⁹ Corti 1992, pag.185

¹¹⁰ MiBAC: <https://www.beniculturali.it/mibac/export/MiBAC/index.html>

¹¹¹ Organigramma del MiBAC:

<http://www.beniculturali.it/mibac/multimedia/MiBAC/images/Organigramma2019.jpg>

¹¹² ICCU: https://www.iccu.sbn.it/it/attivita-servizi/attivita-nazionali/pagina_0007.html

¹¹³ ICCD: <http://www.catalogo.beniculturali.it/opendata>

¹¹⁴ ICAR: <http://www.icar.beniculturali.it/index.php?id=67>

¹¹⁵ Paolini et al. 2005, pp.46-47

La questione, dunque, è: come modificare i cataloghi e i dati perché siano *del* web e non solo *sul* web?¹¹⁶

Proprio per questo è necessaria sempre di più l'applicazione di nuovi approcci tecnologici, come il paradigma dei *Linked Data*.

Il MiBAC e i suoi organi sono le principali istituzioni culturali italiane che aderiscono al progetto LOD¹¹⁷.

*Al MiBAC servono strumenti stabili per riconoscere la domanda, analizzarla, organizzarla, soddisfarla, così da creare valore (ossia servizi di qualità ai cittadini) dal riutilizzo dei dati aperti. Un gap che si vuole colmare con la pubblicazione del portale dei dati dei beni culturali.*¹¹⁸

2.5 Progetti e istituzioni per la gestione in ambito ICT dei beni culturali in Italia

2.5.1 ArCo¹¹⁹

Architettura della Conoscenza è un progetto che nasce nel 2017 dalla cooperazione tra l'ICCD e il CNR. Consiste in una *rete di ontologie*¹²⁰ per la strutturazione della conoscenza per i beni culturali. Serve a:

- sperimentare il nuovo paradigma LOD per la gestione dei dati;
- collegare entità ad *authority file* e dataset disponibili in LOD (VIAF, DBpedia, GeoNames, ...);
- sviluppare strumenti e metodi per la tutela e la valorizzazione del patrimonio artistico e culturale italiano;
- collegare e arricchire di informazioni i dati del MiBAC grazie ad archivi esterni di dati aperti;

¹¹⁶ M. Guerrini, T. Possemato, *Linked Data, un nuovo alfabeto del web semantico*, in *Rivista italiana di biblioteconomia, archivistica e scienza dell'informazione*, vol.4, n°1, 2013, pag. 11

¹¹⁷ Piattaforma di pubblicazione dei LOD del MiBAC: <http://dati.beniculturali.it/>

¹¹⁸ Moro, 2017

¹¹⁹ <http://dati.beniculturali.it/progetto-arco-architettura-della-conoscenza/>

¹²⁰ Ontologie di ArCo:

<https://github.com/ICCD-MiBACT/ArCo/tree/master/ArCo-release/ontologie>

- garantire interoperabilità con altre ontologie specifiche del dominio beni culturali, come CIDOC-CRM (vedi sotto);
- restituire i dati in modo semplice e facilmente riusabile da Google, Europea e Wikidata;
- integrare tutte le schede di catalogo definite dall'ICCD (cap. 2.6.3).

2.5.2 CulturalItalia¹²¹

L'aggregatore nazionale del patrimonio culturale italiano (promosso e gestito dal MiBAC).¹²²

E' un portale¹²³ che propone un accesso guidato al mondo della cultura italiana: raccoglie e organizza milioni di informazioni sulle risorse che compongono il patrimonio culturale del paese, mettendole a disposizione degli utenti.

Le informazioni sulle risorse culturali non sono prodotte dal portale, ma sono fornite direttamente dai soggetti che possiedono e gestiscono le risorse, tramite il protocollo OAI-PMH¹²⁴.

Una base di "metadati e risorse digitali", aggrega e organizza le informazioni provenienti dai fornitori di dati (*data providers*): una volta individuate le risorse di interesse, l'utente può consultarle direttamente presso la fonte dati.

CulturalItalia è un sistema "aperto", in quanto cresce e si sviluppa di pari passo con le nuove informazioni sulle risorse che arricchiscono il database.

¹²¹ http://www.culturalitalia.it/opencms/il_progetto_it.jsp?language=it&tematica=static

¹²² M. Piccinino, *Europeana e altri progetti europei dell'ICCU*, in *Digitalia*, anno VII, n°2, 2012, ISSN 1972-6201, pag.125

¹²³ <http://dati.culturalitalia.it/>

¹²⁴ OAI-PMH è un protocollo sviluppato da OAI (Open Archives Initiative) come infrastruttura di comunicazione per l'open access e definito per supportare lo scambio di metadati. È utilizzato per raccogliere (o collezionare) i metadati dei documenti in un archivio affinché i servizi possano essere costruiti utilizzando metadati da più archivi.

Il protocollo OAI è basato su HTTP per la raccolta di metadati da diversi repositories che supporta qualsiasi formato di metadati (ma il Dublin Core è obbligatorio per ogni data provider, "fornitore di dati"). Le risposte sono codificate in sintassi XML e formattate come risposte HTTP.

https://it.wikipedia.org/wiki/Open_Archives_Initiative_Protocol_for_Metadata_Harvesting
<http://www.openarchives.org/OAI/openarchivesprotocol.html>

CulturalItalia, in quanto aggregatore di risorse provenienti da tutti i settori del patrimonio, è il principale fornitore italiano di contenuti verso Europeana, e fa da tramite tra le istituzioni culturali italiane e il portale europeo.

I dati sono esposti in vari formati standard, come DC, EDM PICO e CIDOC-CRM.



Fig.2.4 Il portale CulturalItalia nel quadro delle collaborazioni europee, dalla conferenza tenuta presso la ex SSAB da Sara Di Giorgio (ICCU, responsabile gruppo tecnico CulturalItalia), aprile 2016¹²⁵

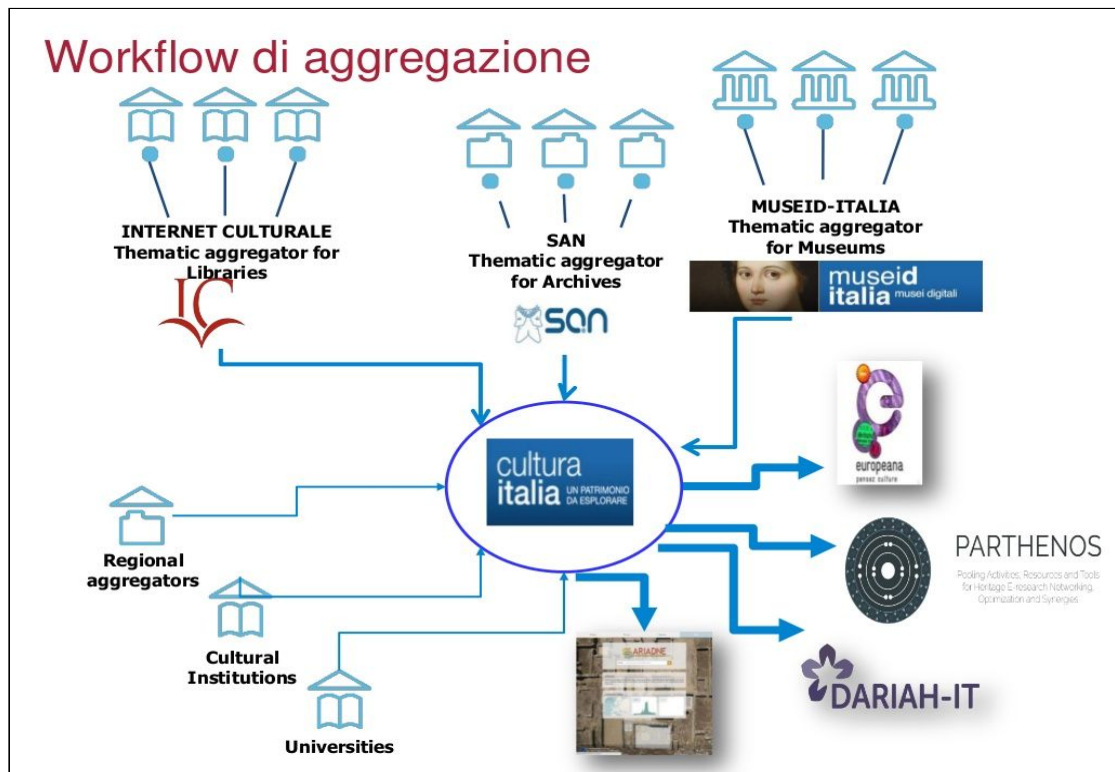


Fig. 2.5 CulturalItalia, l'aggregatore nazionale¹²⁶

¹²⁵ S. Di Giorgio, 2016, pp. 3-4

¹²⁶ S. Di Giorgio, 2016, pag. 18

2.5.3 Cultural-ON¹²⁷

E' un'ontologia che ha lo scopo di modellare i dati che caratterizzano gli istituti e i luoghi della cultura, come ad esempio i dati sugli enti o sulle persone, le sedi fisiche dei luoghi, eventuali materiali multimediali, ecc ...

Cultural-ON (*Cultural Ontology*)¹²⁸ è il risultato dell'impegno del MiBAC di "dotarsi di solidi strumenti per pubblicare dati aperti di qualità"¹²⁹, in risposta alla crescente esigenza di condivisione e circolazione dei dati conoscitivi del patrimonio culturale, nonché del riutilizzo di questi dati come incentivo per la crescita e lo sviluppo.

L'ICAR contribuisce ai LOD¹³⁰ per i beni culturali proprio secondo questa ontologia; a partire da essa sono stati modellati e pubblicati i *dataset open*¹³¹ del MiBAC relativi ai luoghi della cultura.

Cultural-ON è allineata con altre esistenti nel web dei dati come per esempio FOAF, Dublin Core, PRO, schema.org, ecc...

¹²⁷ http://dati.beniculturali.it/cultural_on/

¹²⁸ <http://dati.beniculturali.it/cultural-ON/ITA.html>

¹²⁹ L. Moro, *Mibact: Così gli Open Data della cultura creano crescita e sviluppo*, in *Agenda Digitale*, 2017

¹³⁰ <http://www.icar.beniculturali.it/index.php?id=67>

¹³¹ <http://dati.beniculturali.it/datasets/>

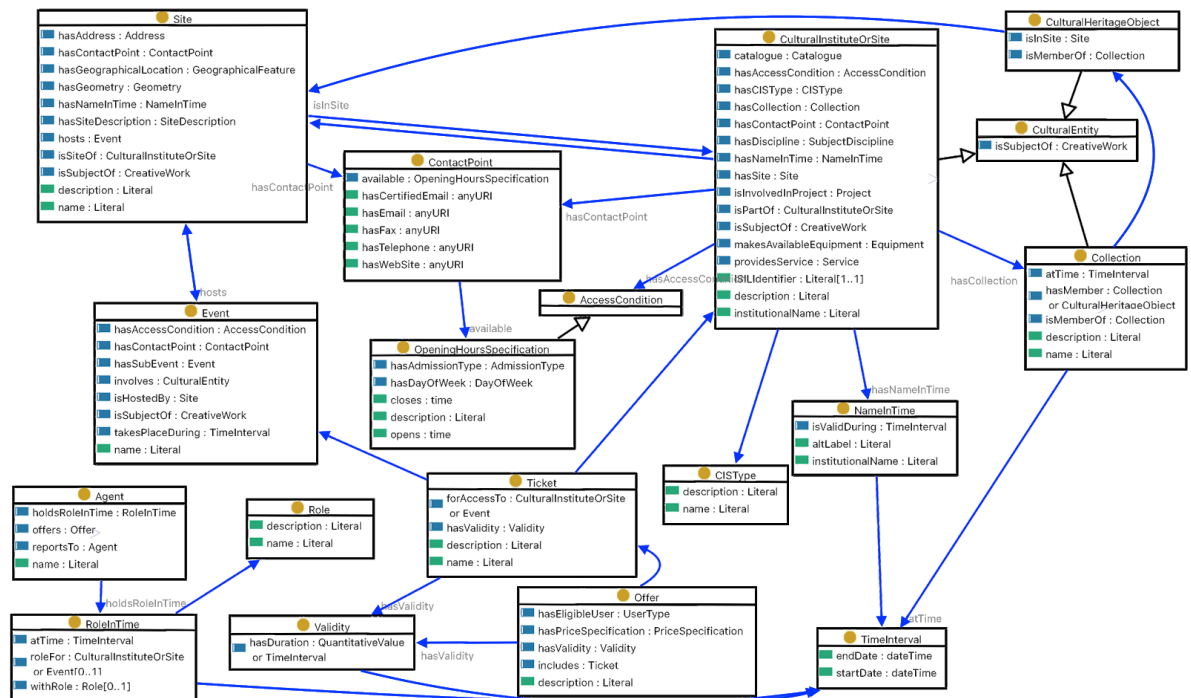


Fig. 2.6 La figura illustra alcuni elementi principali dell'ontologia Cultural-ON. La figura non è esaustiva ma è inserita per aiutare a comprendere alcune relazioni tra concetti dell'ontologia.

2.5.4 Catalogo generale dei beni culturali¹³²

È la banca dati che raccoglie e organizza le informazioni descrittive dei beni culturali catalogati in Italia, frutto delle attività di ricerca condotte da diverse istituzioni.

Le schede dei beni (quasi 3 milioni) sono conferite al Sistema Informativo Generale del Catalogo (SIGECweb)¹³³: monumenti, collezioni, raccolte, oggetti di interesse artistico e storico, reperti e siti archeologici, beni scientifici e naturalistici.

La definizione di un modello di interoperabilità per l'ICCD spetta proprio al Catalogo, che definisce regole, linee guida e standard a cui deve attenersi chiunque voglia interagire con esso.

¹³²

http://www.catalogo.beniculturali.it/sigecSSU_FE/Home.action?timestamp=1552668945708

¹³³ <http://www.iccd.beniculturali.it/it/sigec-web>

Tale modello si basa:

- sugli standard catalografici dell'ICCD¹³⁴
- sui vocabolari controllati in uso su SIGECweb pubblicati in SKOS¹³⁵
- sulle ontologie¹³⁶ e in particolare Cultural-ON e quelle del progetto ArCo.

Tra le ontologie utilizzate dal MiBAC per la pubblicazione dei LOD, citiamo:

- FOAF¹³⁷, per descrivere persone e le loro attività / relazioni.
- Dublin Core espresso in RDFS¹³⁸
- DCMI Metadata Terms¹³⁹
- Geonames¹⁴⁰
- Schema.org¹⁴¹

2.5.5 DCAT-AP-IT¹⁴²

Data Catalog Vocabulary - Application Profile è un profilo di metadattazione¹⁴³ che nasce dalla necessità di standardizzare le descrizioni condivise dei dati a livello nazionale, in conformità anche alle regole e agli standard di interoperabilità definiti a livello internazionale.

È un profilo applicativo del vocabolario *Data Catalog Vocabulary* (DCAT), vocabolario comune per armonizzare le descrizioni di oltre 258.000 set di dati raccolti da 67 portali di dati di 34 paesi, raccomandato del W3C per la descrizione dei *dataset* pubblicati in cataloghi.

Il profilo applicativo europeo ha come obiettivo quello di proporre un modello per la rappresentazione di metadati indipendenti dal particolare dominio applicativo.¹⁴⁴

¹³⁴ <http://github.com/ICCD-MiBACT/Standard-catalografici>

¹³⁵

<http://github.com/ICCD-MiBACT/Standard-catalografici/tree/master/strumenti-terminologici>

¹³⁶ Ontologie di riferimento: <http://dati.beniculturali.it/le-ontologie/>

¹³⁷ <http://xmlns.com/foaf/spec/>, <http://www.foaf-project.org/>

¹³⁸ <http://dublincore.org/schemas/rdfs/>

¹³⁹ <http://purl.org/dc/terms/>

¹⁴⁰ http://www.geonames.org/ontology/ontology_v3.1.rdf

¹⁴¹ <http://schema.org/>

¹⁴² <https://www.dati.gov.it/sites/default/files/linee-guida-cataloghi-dati-profilo-dcat-ap-it-2.pdf>

¹⁴³ https://docs.italia.it/italia/daf/linee-guida-cataloghi-dati-dcat-ap-it/it/stabile/dcat-ap_it.html

¹⁴⁴ Ontologia OWL DCAT-AP: <http://dati.gov.it/onto/dcatapit>

Le specifiche italiane sono state definite dall’Agenzia per l’Italia Digitale (AGID) la quale si occupa, tra le altre cose, di favorire l’interoperabilità semantica di dati e servizi.

DCAT-AP, secondo quanto riportato dalle specifiche, “intende essere uno strumento da utilizzare in un ambiente di tipo *Linked Data*”.¹⁴⁵

2.6 Standard di metadati nazionali

In Italia le regioni promuovono e coordinano la catalogazione secondo le metodologie nazionali definite in cooperazione con gli organi statali competenti, partecipando attivamente alla definizione degli standard.

2.6.1 Standard del SAN¹⁴⁶

Il Sistema Nazionale Archivistico (SAN) è l’aggregatore nazionale e il punto di accesso unificato alle risorse archivistiche e digitali statali e non statali, pubbliche e private; rende disponibile secondo il modello LOD, e attraverso un OAI-PMH *provider*, le proprie schede descrittive delle entità archivistiche (soggetti conservatori, soggetti produttori, profili istituzionali, complessi archivistici, ...).

Il Catalogo delle risorse archivistiche (CAT) del SAN è lo strumento di coordinamento e di integrazione delle descrizioni degli archivi italiani. Attraverso esso è possibile sapere quali risorse archivistiche esistono in Italia, chi le ha prodotte, dove sono conservate e come vi si accede.

L’ICAR si occupa di rendere disponibili i contenuti del SAN secondo il paradigma LOD tramite la messa a punto di strumenti e metodologie innovative finalizzati alla interoperabilità semantica, quali:

- L’ontologia SAN LOD (in formato OWL¹⁴⁷ o rappresentata in HTML¹⁴⁸);
- Il thesauro SAN (in formato SKOS¹⁴⁹ ed in formato grafico navigabile¹⁵⁰).

¹⁴⁵ I benefici del paradigma dei LOD, soprattutto in relazione al raggiungimento di obiettivi di interoperabilità semantica, sono già stati illustrati; per approfondire: https://www.agid.gov.it/sites/default/files/repository_files/documentazione_trasparenza/cdc-spc-gdl6-interoperabilitasemopendata_v2.0_0.pdf

¹⁴⁶ Standard SAN: <http://www.icar.beniculturali.it/index.php?id=107>

¹⁴⁷ Formato OWL dell’ontologia SAN LOD: <http://www.san.beniculturali.it/SAN/san-lod.owl>

¹⁴⁸ Rappresentazione HTML dell’ontologia SAN LOD: <http://dati.san.beniculturali.it/lode/aggiornato.htm>

¹⁴⁹ Formato SKOS del thesauro SAN: <http://www.san.beniculturali.it/SAN/tesaurosan.rdf>

Il SAN ha elaborato specifici standard e protocolli di comunicazione per realizzare l'interoperabilità fra esso e i sistemi archivistici nazionali e locali.

- CAT-SAN¹⁵¹, tracciato di esportazione dai sistemi aderenti al SAN delle schede descrittive dei complessi archivistici di primo livello (fondi, complessi di fondi) o di livello inferiore, dei soggetti conservatori, dei soggetti produttori e degli strumenti di ricerca.

Quelli che seguono sono basati su standard internazionali (adottandone le versioni più recenti) adattati alle finalità che si propone il CAT, al fine di ricomprendere informazioni non incluse nel tracciato CAT-SAN e che sono invece indispensabili per assicurare la più ampia interoperabilità possibile fra i sistemi di descrizione archivistica.

- EAD3¹⁵², per i *complessi archivistici* e per gli strumenti di ricerca.
- EAC-CPF, per la descrizione dei *soggetti produttori* e anche per la codifica delle entità/voci d'indice.¹⁵³
- METS-SAN¹⁵⁴, set di metadati descrittivi, tecnici e strutturali (*Metadata Encoding and Transmission Standard*) relativi agli *oggetti digitali* presenti nel SAN.
- SCONS2¹⁵⁵: versione 2 del tracciato SCONS, elaborato sulla base dello standard EAC, opportunamente modificato ed esteso per *soggetti conservatori e archivi*.

¹⁵⁰ Formato navigabile del thesauro SAN: <http://dati.san.beniculturali.it/skos/>

¹⁵¹ CAT-SAN: <http://www.icar.beniculturali.it/index.php?id=100>

¹⁵² EAD3: <https://github.com/SAA-SDT/EAD3/blob/v1.1.0/ead3.xsd> ,
<https://www.loc.gov/ead/EAD3taglib/EAD3.html>

¹⁵³ “Le relazioni fra le entità descritte in EAD3 e EAC-CPF sono state dotate di riferimenti, vocabolari e identificativi persistenti che le predispongono una loro autonoma elaborazione come LOD.” (ICAR, SAN, *Interoperabilità fra sistemi archivistici: tracciati EAD3, EAC-CPF, SCONS2, ICAR-IMPORT*, 2018)

¹⁵⁴ METS-SAN: <http://www.icar.beniculturali.it/index.php?id=101>

¹⁵⁵ SCONS2: <http://san.beniculturali.it/tracciato/scons2.xsd>

- ICAR-IMPORT¹⁵⁶: tracciato di *envelope* per le entità EAD3, EAC-CPF e SCONS2.

Raggruppa in un unico contenitore gli schemi sopra indicati e viene utilizzato per le procedure di *esportazione e importazione*.

- NIERA (EPF)¹⁵⁷.

Le *Norme italiane per l'elaborazione dei record di autorità archivistici di enti, persone, famiglie* accolgono l'indicazione proveniente dalla definizione dell'ISAAR (CPF) sulla necessità di descrivere separatamente le entità che sono in relazione con la documentazione archivistica e le informazioni di contesto che le riguardano. Questo standard nazionale è stato richiesto da ISAAR ed è associato ad esso; fornisce inoltre vocabolari controllati per la standardizzazione degli elementi strutturati dei record di autorità¹⁵⁸.

2.6.2 Standard dell'ICCU

- MAG¹⁵⁹

E' un *application profile* che ha l'obiettivo di fornire le specifiche formali per la fase di raccolta, di trasferimento e di disseminazione dei metadati e dei dati digitali nei rispettivi archivi. L'ultima versione disponibile ad oggi è la 2.0.1¹⁶⁰.

La finalità del MAG è quella di registrare un progetto di digitalizzazione, il quale fa riferimento a un oggetto reale e che ha prodotto una serie di copie digitali in vari formati (immagini, audio, video, ...).

Dunque le risorse principali schedate in un record MAG sono: il progetto di digitalizzazione, l'oggetto digitalizzato e le risorse digitali che ne derivano.

¹⁵⁶ ICAR-IMPORT: <http://san.beniculturali.it/tracciato/icar-import.xsd>

¹⁵⁷ NIERA (EPF): <http://www.icar.beniculturali.it/index.php?id=102>

¹⁵⁸ Per approfondire il controllo di autorità:
https://it.wikipedia.org/wiki/Controllo_di_autorit%C3%A0

¹⁵⁹ MAG: <https://www.iccu.sbn.it/export/sites/iccu/documenti/manuale.html>

¹⁶⁰ Versione 2.0.1:
<https://www.iccu.sbn.it/it/normative-standard/linee-guida-per-la-digitalizzazione-e-metadati/standard-mag-versione-2.0.1/index.html>

I MAG:

- comprendono informazioni generali sul progetto e sul metodo di digitalizzazione ed altre informazioni relative al singolo oggetto digitale;
- comprendono metadati descrittivi espressi Dublin Core;
- sono le descrizioni degli oggetti analogici (fonte) e risiedono nelle basi dati dei singoli sistemi;

Essendo un *application profile*, MAG interagisce e interopera con diversi standard internazionali metadati: in alcuni casi assume altri schemi di codifica (Dublin Core e MIX), in altri casi può essere trasformato in formati diversi (METS e MPEG-21).¹⁶¹

2.6.3 Standard dell'ICCD

Le schede di catalogo dell'ICCD sono organizzate in tre macrocategorie: beni mobili (ad esempio oggetti quali dipinti, sculture o reperti archeologici), beni immobili (ad esempio palazzi, chiese o siti e monumenti archeologici) e beni immateriali (ad esempio le tradizioni orali, le lingue, le arti performative, le pratiche sociali e rituali).

¹⁶²

All'interno di queste macrocategorie ci sono i settori disciplinari:

- beni archeologici
- beni architettonici e paesaggistici
- beni demoetnoantropologici
- beni fotografici
- beni musicali
- beni naturalistici
- beni numismatici
- beni scientifici e tecnologici
- beni storici e artistici

¹⁶¹ Mapping dal MAG al PICO:

https://www.iccu.sbn.it/export/sites/iccu/documenti/mapping_MAG-PICO-4.2-1.pdf

¹⁶² A. Negri, *Catalogazione dei beni culturali in Italia: metodi, strumenti e cooperazione tra sistemi informativi per la gestione della conoscenza*, in *Revista CPC, São Paulo*, n. 21, pp. 104-105

Ad oggi l'ICCD ha definito 30 tracciati catalografici, in relazione alle differenti tipologie di beni, che fanno riferimento ai suddetti nove settori disciplinari e che raccolgono in modo formalizzato le informazioni sui beni culturali (dati descrittivi, tecnico-scientifici, geografici, documentali) per le schede di catalogo.

Tipo scheda	Definizione	Settore disciplinare
A	Architettura	beni architettonici e paesaggistici
AT	Reperti antropologici	beni archeologici
BDI	Beni demoetnoantropologici immateriali	beni demoetnoantropologici
BDM	Beni demoetnoantropologici materiali	beni demoetnoantropologici
BNB	Beni naturalistici-Botanica	beni naturalistici
BNM	Beni naturalistici-Mineralogia	beni naturalistici
BNP	Beni naturalistici-Paleontologia	beni naturalistici
BNPE	Beni naturalistici-Petrologia	beni naturalistici
BNPL	Beni naturalistici-Planetologia	beni naturalistici
BNZ	Beni naturalistici-Zoologia	beni naturalistici
CA	Complessi archeologici	beni archeologici
CNS ^[1]	Centri/nuclei storici	beni architettonici e paesaggistici
D	Disegni	beni storici e artistici
F	Fotografia	beni fotografici
FF	Fondi fotografici	beni fotografici
MA	Monumenti archeologici	beni archeologici
MI	Matrici incise	beni storici e artistici

NU	Beni numismatici	beni numismatici
OA	Opere/oggetti d'arte	beni storici e artistici
OAC	Opere/oggetti d'arte contemporanea	beni storici e artistici
PG	Parchi/giardini	beni architettonici e paesaggistici
PST	Patrimonio scientifico e tecnologico	beni scientifici e tecnologici
RA	Reperti archeologici	beni archeologici
S	Stampe	beni storici e artistici
SAS	Saggi stratigrafici	beni archeologici
SI	Siti archeologici	beni archeologici
SM	Strumenti musicali	beni musicali
SMO	Strumenti musicali-Organo	beni musicali
TMA	Tabella materiali archeologici	beni archeologici
VeAC	Vestimenti antichi/contemporanei	beni storici e artistici

Fig. 2.7 Tracciati catalografici dell'ICCD

2.6.4 Altri standard

- PICO¹⁶³

Il *Portale Italiano della Cultura On-line* è un progetto tecnico-scientifico per CulturalItalia, avviato su incarico del MiBAC; consiste in un DC *Application Profile*¹⁶⁴ per recuperare le informazioni principali contenute in risorse digitali provenienti da vari *datasource* esterni, pertinenti ad enti pubblici e privati e

¹⁶³ PICO: <http://purl.org/pico/picoap1.0.xml>

¹⁶⁴ *We define application profiles as schemas which consist of data elements drawn from one or more namespaces, combined together by implementors, and optimised for a particular local application* (R. Heery and M. Patel. *Application profiles: mixing and matching metadata schemas*, <http://www.ariadne.ac.uk/issue25/app-profiles/>).

«Nell'ambito della DCMI, un profilo applicativo è la dichiarazione dei metadata terms che un'organizzazione una risorsa informativa, un'applicazione o una comunità di utenti usa nel suo schema di metadati». <http://dublincore.org/documents/usageguide/glossary.shtml>

relative a contenuti diversi per tipologia, formato ed informazioni, il cui minimo comune denominatore è quello di appartenere cultura italiana.

Un profilo di applicazione è composto da elementi presi da schemi di metadati preesistenti e combinati insieme per ottimizzare una determinata applicazione.¹⁶⁵ Questi schemi, che si basano su RDF e XML, servono a dichiarare gli elementi utilizzati per un progetto, indicando per ciascuna classe e proprietà i *namespace* di provenienza con riferimento agli URI dove risiedono le definizioni degli elementi.¹⁶⁶

PICO rende visibili tali risorse attraverso il portale Culturaitalia e consente l'interoperabilità e la ricerca incrociata tra contenuti provenienti da *repository* diversi.

Per tali finalità è stato definito questo schema unico, basato sul Qualified Dublin Core, in grado di recuperare i metadati relativi alle risorse d'interesse. Culturaitalia infatti, aggrega risorse provenienti dalle più autorevoli collezioni digitali del patrimonio culturale italiano, grazie a una rete di *providers* presenti in tutto il territorio nazionale in continua espansione.

Tutte le risorse presenti nel portale Culturaitalia sono classificate in base al PICO Thesaurus¹⁶⁷, un vocabolario controllato gerarchico, sviluppato in formato SKOS.

¹⁶⁵ R. Heery e M. Patel, *Application Profiles: Mixing and Matching Metadata Schemas*, *Ariadne*, n. 25 2000, <http://www.ariadne.ac.uk/issue25/app-profiles>

¹⁶⁶ S. Peruginelli, *I profili di applicazione*, ICCU, Roma, 2001
www.iccu.sbn.it/upload/documenti/Peruginelli.ppt

¹⁶⁷ PICO Thesaurus:
http://www.culturaitalia.it/opencms/export/sites/culturaitalia/attachments/thesaurus/4.3/thesaurus_4.3.0.skos.xml

- CIDOC-CRM¹⁶⁸

CIDOC *Conceptual Reference Model* (CRM) è un'ontologia¹⁶⁹ fondazionale¹⁷⁰ creata da International Council of Museums (ICOM) nel 2002 al fine di garantire l'interoperabilità semantica nel dominio del *Cultural Heritage*, è un modello evento-centrico, finalizzato a descrivere oggetti museali.

E' estensibile e fornisce definizioni e una struttura formale per descrivere i concetti impliciti ed espliciti e le relazioni utilizzate nella documentazione del patrimonio culturale e nella documentazione dei musei; è inteso a promuovere una comprensione condivisa delle informazioni sul patrimonio culturale, fornendo un quadro semantico comune ed estendibile su cui è possibile mappare qualsiasi informazione sul patrimonio culturale.

L'obiettivo generale del CIDOC-CRM è fornire un modello di riferimento e uno standard che i musei e altre istituzioni del patrimonio culturale possano utilizzare per descrivere le loro collezioni, al fine di migliorare la condivisione delle informazioni.

Grazie alla caratteristica dell'interoperabilità, CIDOC-CRM si trova in una posizione privilegiata per diventare un importante standard di informazioni e un modello di riferimento per le iniziative del *semantic web*.¹⁷¹

In seguito al successo della standardizzazione del CIDOC-CRM, nel 2006 è stata avviata l'iniziativa FRBRoo¹⁷² (*FRBR-object oriented*, "orientato agli

¹⁶⁸ CIDOC-CRM: <http://www.cidoc-crm.org/>

¹⁶⁹ <http://purl.org/NET/crm-owl#>

¹⁷⁰ Una *core ontology* si occupa di "fornire un modello globale ed estensibile in cui possono essere messi in corrispondenza ed integrati i dati provenienti da fonti eterogenee. Questa forma è in grado di fornire una singola base di conoscenza per strumenti e servizi *cross-domain* (*resource discovery, browsing, data mining*), riducendo la complessità combinatoria che deriverebbe dal tentativo di mettere in corrispondenza a due a due i singoli formati di metadati o le ontologie. [...] Un interessante esempio di core ontology è CIDOC CRM, costituita da circa 80 classi e 130 proprietà, che si pone come modello di riferimento per il patrimonio culturale" (Signore et al. pp.294-95)

¹⁷¹ Mapping dello schema di metadati PICO-AP / CIDOC-CRM: http://www.culturaitalia.it/opencms/export/sites/culturaitalia/attachments/MappingPico_CIDOC.pdf

¹⁷² FRBRoo: <https://it.wikipedia.org/wiki/FRBRoo>

oggetti"), per armonizzare l'ontologia con i requisiti funzionali per i record bibliografici (FRBR).

CIDOC-CRM è già certificato come standard serie ISO 21127:2006 ed impiegato come tale nella concettualizzazione delle risorse nel dominio dei beni culturali e discipline umanistiche.

"... lessici, tassonomie, vocabolari strutturati e thesauri non si prestano adeguatamente a collegare tra di loro alberi concettuali diversi, mentre approcci formalmente più solidi (tra i quali va senz'altro citato CIDOC CRM) promettono di rappresentare la conoscenza in modo da poter supportare ragionamento e deduzioni "

173

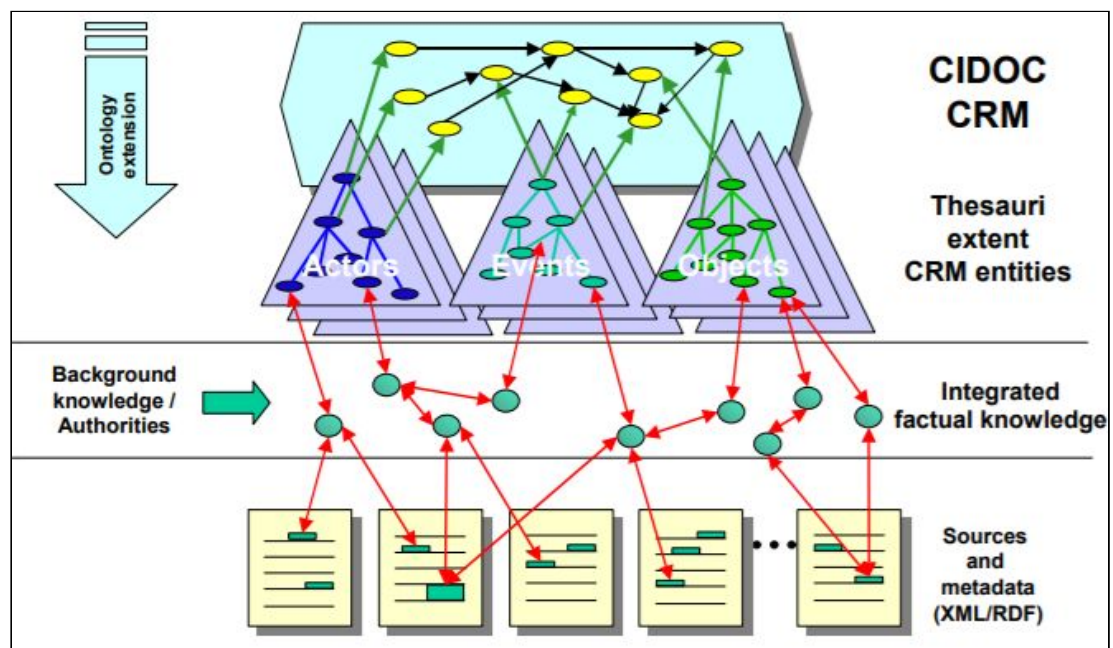


Fig. 2.8 CIDOC-CRM: architettura di integrazione delle informazioni¹⁷⁴

¹⁷³ Signore et al. pp.294-95

¹⁷⁴ M. Doerr, *The CIDOC CRM – an Ontological Approach to Semantic Interoperability of Metadata*, in *ResearchGate*, 2003

2.7 Standard per la descrizione di beni culturali

Il Dipartimento di Beni culturali dell'Università Cattolica di Roma, individua i seguenti "metadati descrittivi per i beni culturali ed in particolare per i beni artistici": Dublin Core, ICCD, PICO, EAD e MODS.

I musei digitali del portale Culturaitalia utilizzano come standard: schede di catalogo ICCD, VRA Core, METS¹⁷⁵.

2.8 Standard presenti nel portale Mèmora

La catalogazione del portale Mèmora ([cap. 4.5](#)) comprende: schede di catalogazione dei beni e schede *authority* (standard ICCD), schede di ambito archivistico (standard ISAAR/ISAD), schede di entità (norme NIERA), MAG (standard dell'ICCU), CIDOC-CRM e PICO (standard di Culturaitalia).

Gli standard presenti sul portale sono ripresi da:

- standard ICCD ([cap. 2.6.3](#)): schede A, OA, F, S;
- gran parte degli standard definiti dall'ICAR e dal SAN ([cap. 2.6.1](#)): CAT-SAN, METS-SAN, SCONS2, ICAR-IMPORT;
- alcuni standard internazionali ([cap. 2.3.2](#)): EAC-CPF, EAD 3, METS, METS Rights, MIX (usato, in questo contesto, per la descrizione dei metadati tecnici per oggetti digitali utilizzati all'interno dei metadati METS)

¹⁷⁵ Mapping tra le Schede di Catalogo ICCD e VRA Core 4.0:
http://www.culturaitalia.it/opencms/export/sites/culturaitalia/attachments/museiditalia/profiles/mets/Mapping_SchedeICCD_VRA.pdf

2.8.1 “Esportatori di dati” su Mèmora

Sono disponibili 12 mappature per esportare dati da Mèmora in formato XML, basati sugli standard presenti nel portale stesso.¹⁷⁶

Nome / codice	Tipologia	Standard definito da
Ambito Patrimonio Culturale - ICCD versione 3.0:		
scheda_A	oggetti/schede	ICCD
scheda_OA		
scheda_F		
scheda_S		
IMMFTAN (relazioni tra oggetti e oggetti digitali) ¹⁷⁷		
INFORMA (dati anagrafici dell'ente) ¹⁷⁸		
Ambito Archivi e Patrimonio Documentale:		
CAT-SAN ¹⁷⁹ (oggetti, livelli gerarchici, soggetti conservatori, soggetti produttori, relazioni)	oggetti/schede	ICAR - SAN
SCONS2 (soggetto conservatore e delle relazioni)	entità ¹⁸⁰	
EAC-CPF (soggetto produttore e delle relazioni)		
MINI-EAD3 (oggetti comprensivi dei livelli gerarchici e delle relazioni)	oggetti/schede	LoC
MINI-icar-san (oggetti, livelli gerarchici, soggetti conservatori, soggetti produttori, relazioni, oggetti digitali collegati) ¹⁸¹ , che raggruppa e rende interoperabili i seguenti tracciati: EAD3, EAC-CPF, SCONS.		
url_media	media	

Tab. 2.2 Esportatori di dati nella piattaforma Mèmora

¹⁷⁶ Mèmora, *Manuale d'uso*,

http://www.memora.piemonte.it/dwd/manuale_memora_V03_2018.pdf

¹⁷⁷ <http://www.iccd.beniculturali.it/getFile.php?id=293> pag. 19

¹⁷⁸ <http://www.iccd.beniculturali.it/getFile.php?id=293> pag. 17

¹⁷⁹ “[...] CAT-SAN costituisce ormai un punto di riferimento consolidato nell'assicurare l'interoperabilità fra i principali sistemi archivistici italiani e il Sistema Archivistico Nazionale”
<http://www.icar.beniculturali.it/index.php?id=338>

¹⁸⁰ Le entità prese in considerazione nei tracciati (complessi archivistici, unità/sottounità archivistiche, unità documentarie, soggetti produttori, soggetti conservatori, strumenti di ricerca, entità/voci d'indice) sono quelle che tipicamente caratterizzano i sistemi descrittivi sviluppati nel nostro paese.

¹⁸¹ vedi ICAR-IMPORT [cap. 2.6.1](#)

2.9 Riepilogo degli standard nazionali e internazionali

La tabella seguente riassume le caratteristiche principali degli standard di metadati precedentemente descritti.

Standard (ordine alfabetico)	Internazionale	Nazionale / esiste vers. ita	Per i beni culturali	Presente su Mèmora
BIBO	X		X	
CAT-SAN		X	X	X
CIDOC-CRM		X	X	X
Dublin Core	X		X	
EAC-CPF	X		X	X
EAD	X		X	X
EDM	X		X	
ICP dell'IFLA	X			
ICCD		X	X	X
ICCU		X		
ISAAR (CPF)	X	X	X	X
ISAD (G)	X	X	X	X
ISBD	X			
ISNI	X			
MAG		X	X	X
MARC	X			
METS e METS Right	X		X	X

METS-SAN		X	X	X
MIX	X			X
MODS	X		X	
NIERA (EPF)		X	X	X
PICO		X	X	X
SCONS		X	X	X
RDA	X	X		
VRA	X		X	

Tab. 2.3 Riepilogo degli standard di metadati

3. Analisi dello stato dell'arte sui *Linked Open Data*

Il bisogno di una rappresentazione semantica dei dati è considerato sempre di più una delle prossime, maggiori sfide nel campo delle applicazioni LOD. Per questo motivo sempre più istituzioni stanno investendo nella pubblicazione dei propri dati secondo il paradigma *Linked Data*, attività descritta nel dettaglio nei capitoli precedenti.

La letteratura sul ricorso al LOD è molto ricca e copre domini applicativi molto vasti, dalla pubblica amministrazione, all'istruzione e ai beni culturali.

Questo capitolo si concentra in particolar modo sull'analisi, in un primo momento, dello stato degli *open data* in Italia per stabilire *quanti* dei dati aperti sul web sono effettivamente collegati.

Una volta inquadrata la dimensione dei *Linked Data* italiani, si procede all'analisi di casi simili al lavoro in oggetto di questa tesi, ovvero la pubblicazione di dati secondo il paradigma LOD.

3.1 Lo stato degli Open Data in Italia

3.1.1 La posizione del nostro paese nella classifica europea

Il Portale Europeo dei dati, l'*Open Data Maturity*, si occupa di analizzare la "maturità" degli *open data* dei paesi europei: i Report mostrano che dal 2017¹⁸² fino ad oggi, l'Italia si è posizionata tra i "trend-setters", salendo di livello rispetto ai "followers" in cui si trovava nel 2016.

¹⁸² che abbraccia il periodo luglio 2016 – giugno 2017

L'Open Data Maturity Report infatti raggruppa i paesi analizzati in quattro diversi cluster: "beginners", "followers", "fast-trackers" e "trend-setters".

Un paese "trend-setter" ha una politica in atto di dati aperti avanzata con un forte coordinamento delle attività di Open Data in tutto il paese. Il portale nazionale offre una vasta gamma di funzionalità e soddisfa le esigenze degli utenti avanzati e degli editori. Il livello di qualità degli Open Data nel paese è molto buono, con varie iniziative in atto per garantire la pubblicazione di dati di alta qualità e la conformità con gli standard DCAT-AP. Esistono diversi ecosistemi di Open Data sviluppati attorno a domini di dati, con un alto livello di interazione e riutilizzo all'interno di questi domini. Le attività per misurare il riutilizzo sono condotte, con metodologie già in atto per valutare l'impatto in diversi domini. Sono osservabili poche o nessuna limitazione alla pubblicazione o al riutilizzo.¹⁸³

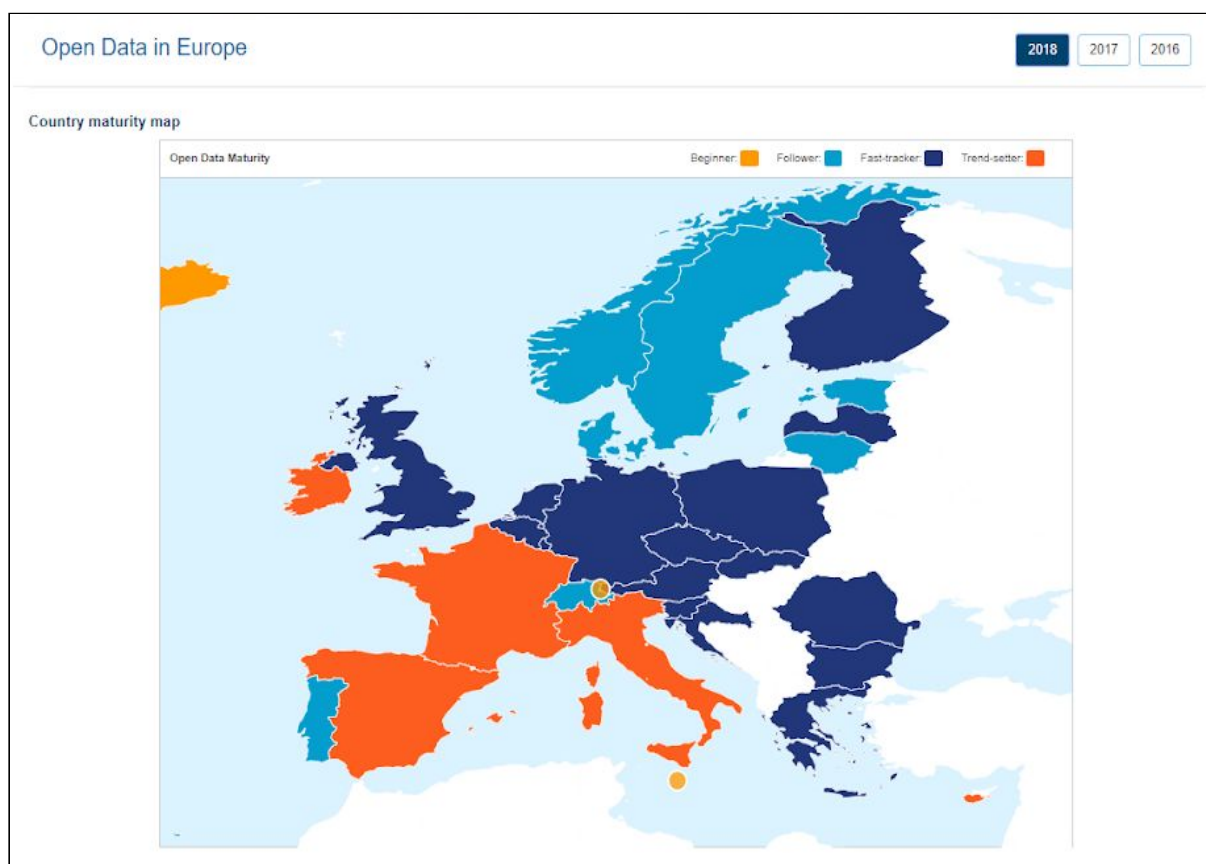


Fig. 3.1 Portale Europeo dei Dati¹⁸⁴: nel 2018 l'Italia è tra i "trend-setters".

¹⁸³ Report 2018, pag. 97

¹⁸⁴ <https://www.europeandataportal.eu/it/dashboard#2018>

Questo significa che il nostro paese è tra quelli che sono stati in grado di implementare una politica di *open data* avanzata e ben coordinata tra tutti gli attori che operano sul territorio nazionale e che dispongono di un portale nazionale dei dati aperti con funzionalità avanzate.

Sebbene le iniziative di *open data* italiane e il numero di dataset rilasciati non siano ancora paragonabili a contesti più maturi come quello statunitense o britannico, il movimento italiano sta lentamente crescendo.¹⁸⁵

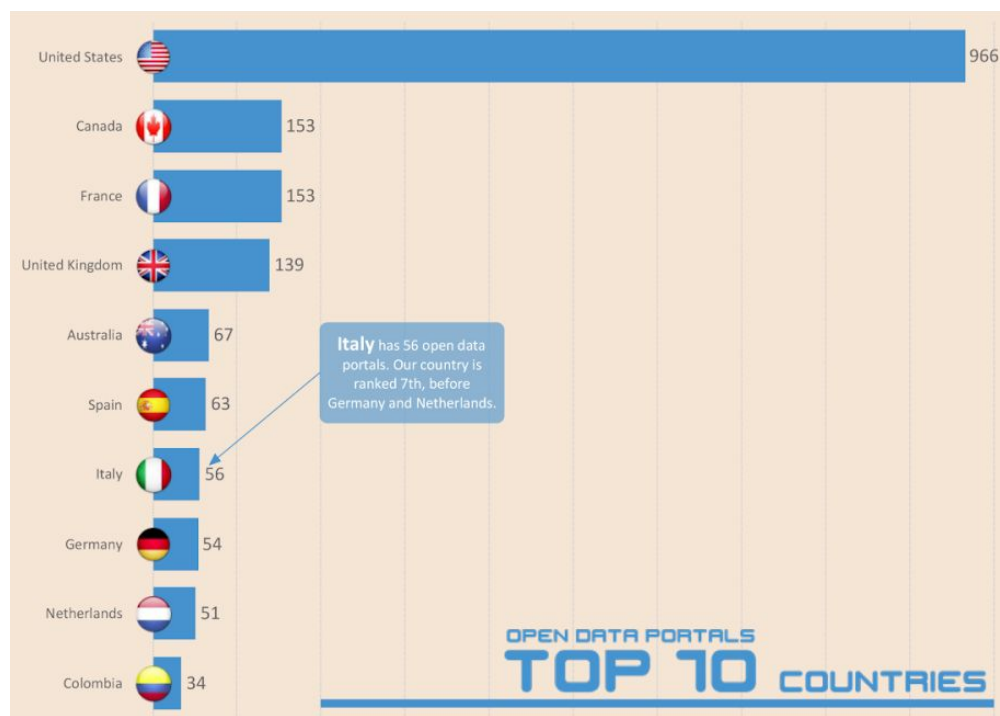


Fig. 3.2 *Open data portals: top 10 countries*

Il nostro paese, seppur lontano da Stati Uniti e Regno Unito, presenta ben 56 portali¹⁸⁶ di dati aperti: un numero superiore a paesi come Germania e Olanda.

¹⁸⁵ F. Mastroianni, *Nel mondo 2700 portali Open Data. La rimonta dell'Italia*, in *InfoData, Il Sole 24 Ore*, 2018

¹⁸⁶ Elenco dei 56 portali italiani:

<https://www.opendatasoft.com/a-comprehensive-list-of-all-open-data-portals-around-the-world/#/italy>



Fig. 3.3 La collocazione geografica dei portali italiani¹⁸⁷

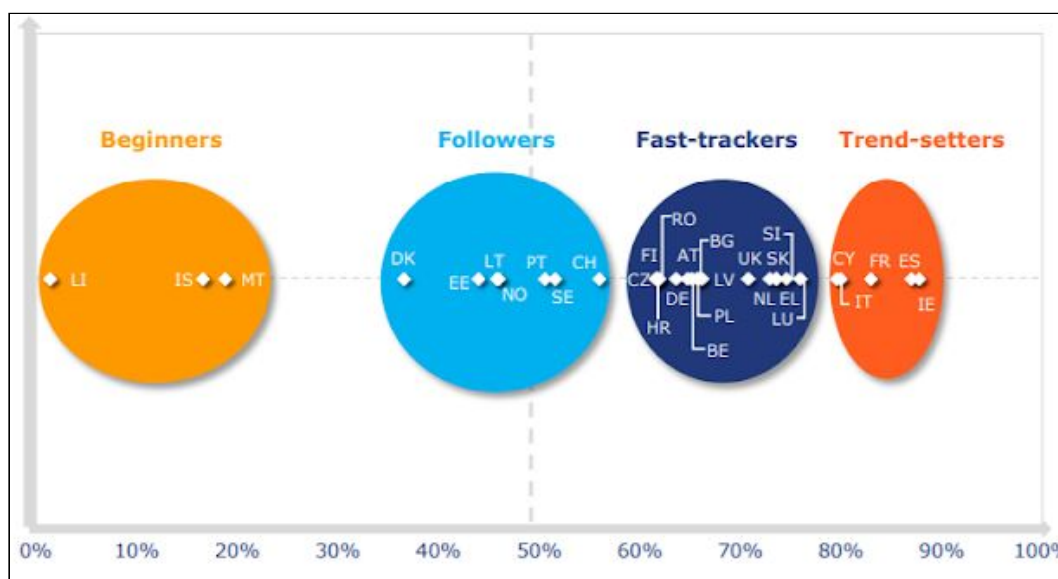
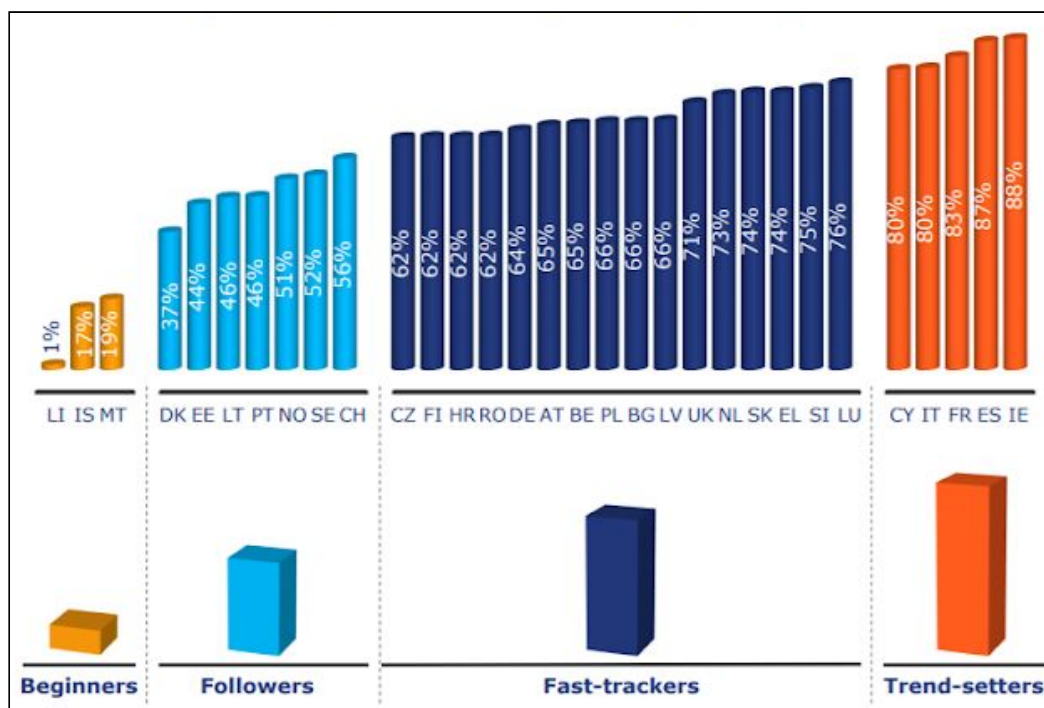
3.1.2 Punti di forza e punti deboli dell'open data italiano

Il Report del 2018, parla di un' "Europa a diverse velocità"¹⁸⁸ e di "eterogeneità nell'impegno" dei vari paesi riguardo al loro contributo per i dati aperti.

In 2018, only five European countries qualified as 'trend-setters': Ireland, Spain, France and Italy provided solid evidence of their sustained Open Data efforts, with scores of 80% and above.

¹⁸⁷ <https://opendatainception.io/>

¹⁸⁸ G. Cecconi, C. Radu, *Open Data Maturity in Europe. Report 2018*, pag. 6
https://www.europeandataportal.eu/sites/default/files/edp_landscaping_insight_report_n4_2018.pdf



Figg. 3.4 e 3.5 : La posizione dell'Italia tra i "trend-setters" nell'Open Data Maturity clustering. Si classifica al 4° posto con un livello di maturità del 80%.¹⁸⁹

Sempre il Report 2018, propone un intero paragrafo intitolato *Promuovere la cultura dell'"open" - Una storia di successo dall'Italia*¹⁹⁰, in cui vengono rimarcate non solo le iniziative italiane a livello nazionale, ma anche quelle locali di regioni e comuni

¹⁸⁹ Era al 8° posto nel 2017 e al 19° nel 2016

https://www.europeandataportal.eu/sites/default/files/country-factsheet_italy_2018.pdf

¹⁹⁰ Report 2018, pag. 19

(citate ad es. la Sardegna, Genova e Milano) e i loro sforzi di sensibilizzazione e incoraggiamento allo sviluppo di servizi innovativi basati sul riutilizzo di dati aperti.

Inoltre, l'Italia (insieme alla Francia) si classifica tra le performance più alte a livello europeo nei punteggi sulla policy (quadro normativo, licenze e coordinamento nazionale)¹⁹¹.

L'appartenenza a domini eterogenei (pubblica amministrazione, ambiente, infrastrutture, turismo, cultura, sanità, ...), rendono gli *open data* italiani "versatili" e "dalle molteplici possibilità di riutilizzo"¹⁹², nonché aventi un impatto di livello "alto" sull'*ambiente* (es. sostenibilità, tutela del territorio e del mare), sulla *società* (es. iniziative di volontariato o di ricerca)¹⁹³, e sulla *politica*, classificando il paese al terzo posto in questi tre settori.

Per quanto riguarda la *qualità* dei dati aperti, i parametri di giudizio sono l'automazione (leggibilità automatica) la circolazione e il riutilizzo di dati e metadati e la conformità con le specifiche DCAT-AP¹⁹⁴: l'Italia raggiunge anche qui punteggi molto alti, classificandosi, in particolare, tra i 17 paesi europei a fornire oltre il 90% dei propri dati in formato *machine readable*.

3.1.3 Il portale nazionale italiano dati.gov.it

Il portale nazionale dati.gov.it ha un grado di *machine-readability* del 96%, classificandosi decimo tra i portali delle nazioni europee¹⁹⁵, e con una percentuale di metadati conformi alle specifiche DCAT-AP del 100%¹⁹⁶.

¹⁹¹ *Report 2018*, pp. 25, 26

¹⁹² *Report 2018*, pag. 36

¹⁹³ *Report 2018*, pp. 54-56

¹⁹⁴ <https://joinup.ec.europa.eu/solution/dcat-application-profile-data-portals-europe>

¹⁹⁵ *Report 2018*, pag. 69

¹⁹⁶ <https://www.europeandataportal.eu/mqa/dati-gov-it>

Paese	Nome	Accessibilità degli URL d'accesso	Leggibilità automatica	Conformità alla specifica DCAT AP	Licenze conosciute	Accessibilità degli URL di download
	dati.gov.it (IT)	1 %	92 %	100 %	n/a	n/a

Fig. 3.6: controllo della qualità dei metadati del portale nazionale italiano dati.gov.it. ¹⁹⁷

3.1.4 Confronto sulla qualità dei metadati tra il portale dati.gov.it e gli altri cataloghi *open* europei.

Metadata Quality Assurance del Portale Europeo ha lo scopo di aiutare i fornitori di dati e i portali a controllare i loro record in base a quattro criteri: accessibilità delle distribuzioni, leggibilità delle macchine, conformità dei metadati relativi alle specifiche DCAT-AP e utilizzo di una licenza nota al Portale Europeo dei dati. Per ogni catalogo, tutti i record e le loro distribuzioni associate sono controllati.

¹⁹⁷ Per la percentuale di licenze conosciute al “Portale europeo dei dati” utilizzate: poiché la licenza non è obbligatoria, non è possibile controllare se mancano le informazioni. Per la percentuale di URL di download disponibili: l'accessibilità è determinata utilizzando un metodo HTTP GET. Poiché il “DownloadURL” non è un'informazione obbligatoria, non è possibile effettuare un controllo in caso di informazioni mancanti.
<https://www.europeandataportal.eu/mqa/>

Confronto sul parametro “accessibilità”.

La percentuale di AccessURLs accessibili¹⁹⁸ è del 1%, e non esistono URL di download, cioè quelli che contengono il link diretto al file scaricabile (ma va notato che l'URL di download è *opzionale* secondo il DCAT-AP).

Gli errori sono tutti del tipo “400 Bad Request” (Richiesta non valida).



Fig. 3.7 Accessibilità: portale dati.gov.it

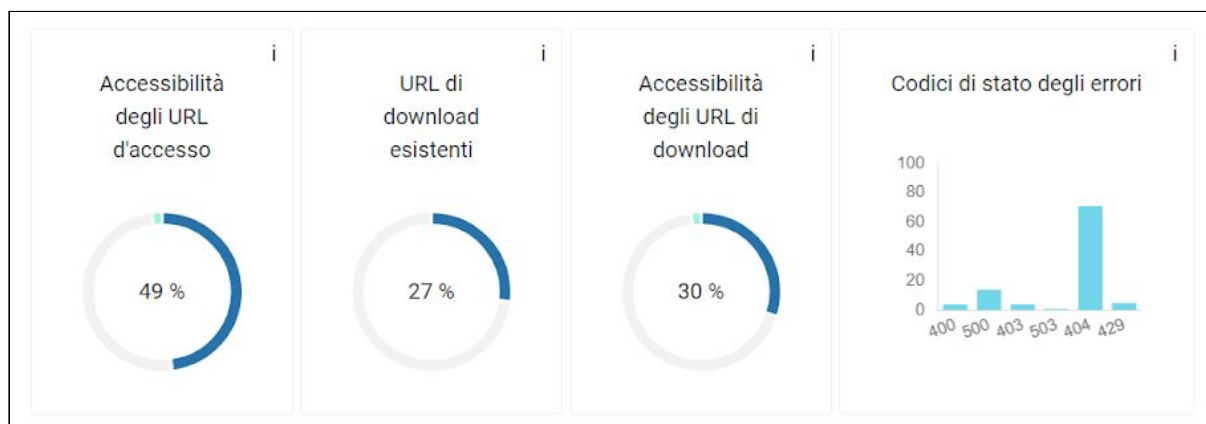


Fig. 3.8 Accessibilità: panoramica generale di tutti i metadati disponibili per il portale europeo di dati (tutti i cataloghi).

¹⁹⁸ Un URL d'accesso mette in collegamento con una pagina che dà accesso al set di dati ed è obbligatorio secondo la specifica DCAT-AP. L'accessibilità è determinata utilizzando un metodo HTTP GET.

Confronto sul parametro “leggibilità automatica”.¹⁹⁹

Il formato RDF è utilizzato solo al 5%.

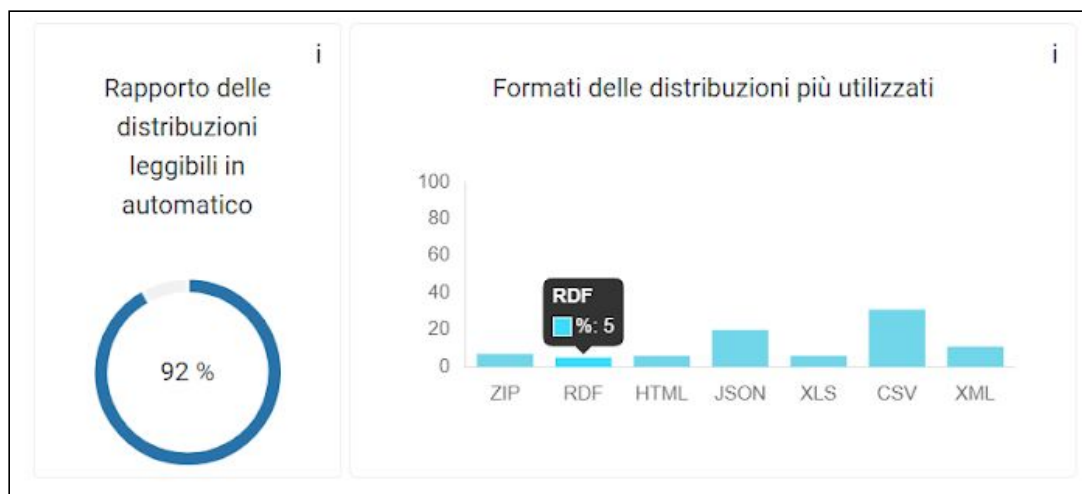


Fig. 3.9 Leggibilità automatica: portale dati.gov.it

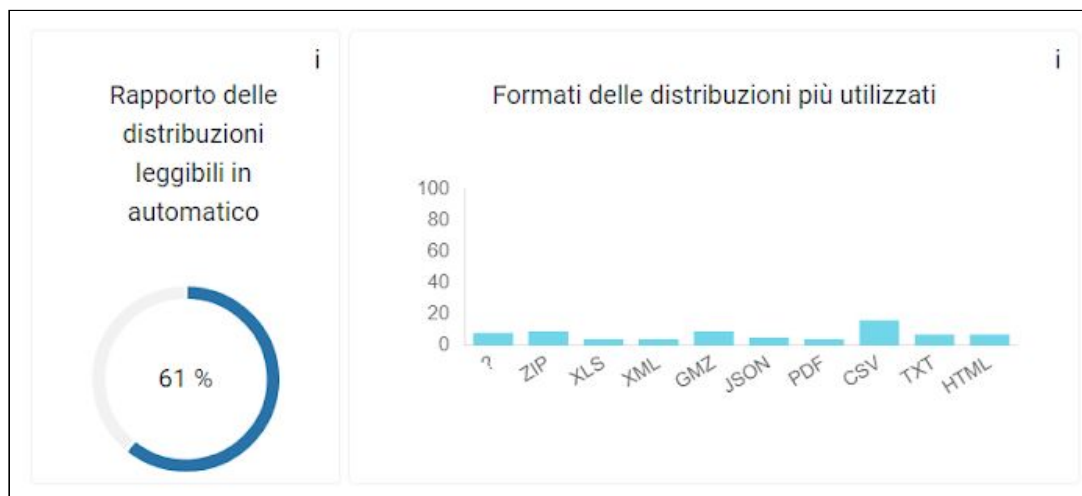


Fig. 3.10: Leggibilità automatica: panoramica generale di tutti i metadati disponibili per il portale europeo di dati (tutti i cataloghi)

¹⁹⁹ Lista dei formati leggibili automaticamente:
https://github.com/opendatamonitor/odm.restapi/blob/master/odmapi/def_formatLists.py#L5-L39

3.1.5 Riepilogo sugli open data in Italia

Complessivamente, la situazione dell'open data italiano è molto positiva.

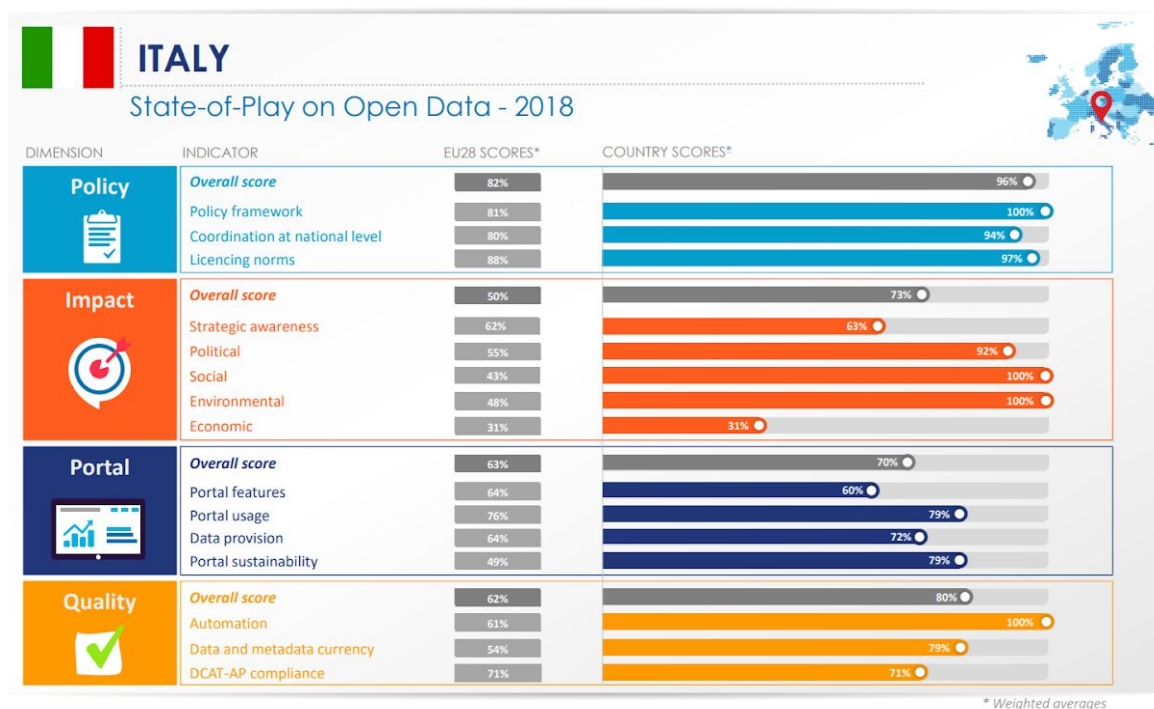


Fig. 3.11 Lo stato dell'Open Data italiano nel 2018

I punti su cui lavorare e che, secondo il Portale Europeo necessitano di un miglioramento, sono²⁰⁰:

- La mancanza di risorse per costruire un programma nazionale
- Mancanza di politiche esplicite sulla condivisione dei dati aperti
- Scarsa consapevolezza dei livelli decisionali / politici nei confronti degli Open Data
- Dubbi sulla protezione dei dati e sulla riservatezza
- Mancanza di una figura riconosciuta di data manager
- La documentazione, la qualità dei dati e l'automazione sono insufficienti
- *Mancanza di riconoscimento del valore del patrimonio informativo*
- *Condivisione della conoscenza insufficiente*
- *Gli esempi di riutilizzo sono ancora scarsi*
- *Difficoltà nella comprensione del potenziale dei dati aperti*

²⁰⁰ Italy - State-of-Play on Open Data - 2018, pag. 3

https://www.europeandataportal.eu/sites/default/files/country-factsheet_italy_2018.pdf

Ecco perché è di fondamentale importanza insistere sull'utilizzo del paradigma *Linked Data*.

3.1.6 Quanti *open data* sono anche *Linked Data*?

[...] *la tendenza prevalente sembra concentrarsi più sull'apertura di dati pubblici sotto forma di open data piuttosto che sul ricorso ai LOD*²⁰¹.

Laura Moro, direttore dell'ICCD, durante la conferenza *Patrimoni materiali e immateriali: innovazioni e resilienze degli standard italiani di catalogazione*, tenutasi ad aprile 2019 a Pisa, afferma che *pochissimi* dei dati aperti italiani disponibili sul web sono effettivamente dati collegati.

La scarsità dei *Linked Open Data* in Italia fa sì che vengano promosse iniziative per produrli e pubblicarli, da parte di varie istituzioni e da domini diversi tra loro.

Sempre Laura Moro, aggiunge che il LOD relativo ai beni culturali in Italia è *un campo ancora da conquistare*, e incentiva le istituzioni culturali e la ricerca all'adozione di questo paradigma, definendolo *la prateria delle nuove generazioni*.

Dopo questa panoramica generale sull'*open data* e i *Linked Data* in Italia, il capitolo prosegue con la selezione e l'analisi di alcuni casi di pubblicazione di dati secondo il paradigma LOD, aggiungendo un confronto con il lavoro svolto in questa tesi.

3.2 Lavori simili

La letteratura sull'adozione del paradigma LOD delle varie istituzioni per la pubblicazione dei propri dati è molto vasta. Qui sono riportati e analizzati solo alcuni dei casi più significativi.²⁰²

²⁰¹ Agenzia per l'Italia Digitale (AGID), *Linee guida per l'interoperabilità semantica attraverso i Linked Open Data*, 2012, pag. 46

²⁰² Per approfondire, vengono segnalate altre iniziative LOD interessanti:

- Museo del paesaggio sonoro:
<https://synapta.it/progetti/museo-del-paesaggio-sonoro/>

3.2.1 Amsterdam Museum Linked Open Data

Il lavoro presentato da Amsterdam Museum²⁰³ è probabilmente quello che presenta maggiori analogie, non solo dal punto di vista del dominio di applicazione (patrimonio culturale), ma anche da quello delle tecniche di cui si avvale (standard di metadati utilizzati).

Il museo olandese ospita oggetti relativi alla storia e alla cultura della città di Amsterdam e ai suoi cittadini: la natura di questi oggetti è eterogenea (dipinti, stampe, disegni, oggetti in vetro e argento, mobili, libri, abiti, ...).

Nel 2010 l'intera collezione è stata resa disponibile online²⁰⁴ e navigabile tramite un'interfaccia web²⁰⁵; in seguito per la fruizione non umana è stata resa disponibile una API per raccogliere tutti i metadati della collezione e/o recuperarli e restituirli in caso di una ricerca dell'utente (*search and retrieval*).

Essendo una istituzione più piccola, il museo non ha le risorse per produrre i propri Linked Data (al contrario di istituzioni più grandi come per esempio la *German National Library*²⁰⁶ o la *British National Library*²⁰⁷): queste istituzioni minori spesso dipendono da aggregatori di grandi dimensioni, come Europeana.

-
- Piattaforma ERCOLE:
<https://books.google.it/books?id=AwvNCwAAQBAJ&pg=PA39&lpg=PA39#v=onepage&q&f=false>
 - LOD della British National Library:
<http://www.theeuropeanlibrary.org/tel4/newsitem/5200>
 - Linked Data at the Guardian:
<https://www.theguardian.com/open-platform/blog/linked-data-open-platform>
 - SBN verso i LOD: sperimentazioni ed esperienze
<https://www.aib.it/aib/sezioni/emr/bibtime/num-xx-1-2-3/bozzarelli.htm>
 - Open data nella pubblica amministrazione italiana: l'esperienza del Comune di Firenze
<https://www.jlis.it/article/view/5500/7926>

²⁰³ <https://www.amsterdammuseum.nl/>

²⁰⁴ Software che gestisce gli oggetti digitali del museo: <http://www.adlibsoft.com/>

²⁰⁵ <https://hart.amsterdam/nl/page/232315/beeldcollectie>

²⁰⁶ https://www.dnb.de/EN/Home/home_node.html

²⁰⁷ <https://www.bl.uk/>

Il dataset conserva la complessità dei dati originali e al tempo stesso raggiunge l'interoperabilità attraverso la mappatura ad un modello standard, in questo caso l'EDM.

I record dei metadati sono stati quindi strutturati appositamente per adattarsi all'Europeana Data Model (EDM)²⁰⁸ e pubblicati sul server di Europeana come *Linked Data Five Star*.²⁰⁹

Nell'ambito del patrimonio culturale gli schemi di metadati più usati per i musei sono DC, VRA, CIDOC-CRM; EDM non è costruito su alcun particolare standard ma piuttosto adotta un framework aperto, *cross-domain* e basato sul Web Semantico che può gestire la gamma e la ricchezza di tutti questi schemi. È stato testato per la compatibilità con altri standard di comunità come EAD e METS.

L'EDM principalmente riutilizza o trae ispirazione da elementi appartenenti ad altri standard.

DC, CIDOC-CRM, SKOS sono utilizzati per i metadati "descrittivi", mentre per i metadati relativi a persone, EDM utilizza lo standard RDA-group2.

I concetti del vocabolario sono mappati a GeoNames e DBpedia.

Il dataset LOD "cinque stelle" pubblicato comprende le descrizioni di più di 70.000 oggetti del museo ed è stato incluso in *Europeana Thoughtlab*²¹⁰, un insieme di tecnologie e strumenti innovativi che aprono la strada agli sviluppi relativi a Europeana.

Il processo di conversione dei dati in *Linked Open Data* da parte dell'Amsterdam Museum si articola nelle fasi seguenti:

- Un insieme record di metadati per 73.447 oggetti, un thesaurus composto da 28.000 concetti utilizzati nei record di metadati e un *authority file* composto

²⁰⁸ <https://pro.europeana.eu/resources/standardization-tools/edm-documentation>

²⁰⁹ Anche se questo approccio garantisce un livello di coerenza e interoperabilità tra i set di dati di diverse istituzioni, la ristrutturazione crea una disconnessione tra il modello di metadati originale dell'istituto e quello dei Linked Data pubblicati. (V. de Boer et al., *Amsterdam Museum Linked Open Data*, in *Semantic web Journal*, pag. 2)

²¹⁰ <https://pro.europeana.eu/web/quest/thoughtlab/linked-open-data>

da 66.966 persone legate agli oggetti o ai metadati, sono stati tutti raccolti tramite un'interfaccia OAI-PMH.

- L'XML risultante è stato prima convertito in RDF grezzo e successivamente ristrutturato utilizzando regole di riscrittura interattiva, con il lo strumento XMLRDF.
- Alle risorse (oggetti, concetti, persone, ...) sono stati assegnati URI; per il basenome URI, è stato usato <http://purl.org/collections/nl/am/>.²¹¹ Gli URI terminano con un identificatore univoco del Museo di Amsterdam (il "preref").
- I collegamenti impliciti (ad esempio tra gli oggetti e i concetti del thesaurus) sono stati resi espliciti come relazioni RDF, e i valori letterali rappresentanti risorse sono stati sostituiti con le risorse stesse.
- Rendere interoperabili i metadati del museo con l'EDM sono state mappate le classi e le proprietà specifiche del museo a quelle dell'EDM (ovvero a elementi specifici di DC, SKOS, RDA e EDM) utilizzando le relazioni `rdf:subPropertyOf` o `rdfs:subClassOf`.

Il dataset LOD risultante è composto dalle seguenti parti:

- *Link interni* (tra le risorse del dataset): in totale 558,161.
- *Link esterni* (ad altre fonti di dati): stabiliti manualmente utilizzando la piattaforma di allineamento di *Amalgame*²¹².
- *Object metadata*: in totale sono 5.700.371 triple RDF.

²¹¹ Al momento della conversione, non si aveva ancora accesso ai server web del Museo di Amsterdam, quindi non era possibile avere URI nello spazio dei nomi di Amsterdam Museum reindirizzati al server semantico in modo che potessero essere risolti. Per questo è stato usato purl.org. (V. de Boer et al., *Amsterdam Museum Linked Open Data*, pag. 2)

²¹² AMALGAME, abbreviazione di AMsterdam ALignment GenerAtion MEtatool, è uno strumento DEL Web Semantico per la ricerca, la valutazione e la gestione di allineamenti di vocabolario.

<https://github.com/jrvosse/amalgame>

<https://www.cwi.nl/innovation/software/amalgame-managing-alignments-in-large-vocabularies>

- *Thesaurus*: 28.000 concetti rappresentati in SKOS, con un filtro che li divide in “geografici” (collegati a GeoNames²¹³) e “non geografici”.
- *Person authority file*: 66.966 istanze (autori, possessori di oggetti artistici, annotatori, persone dipinte, ...). In totale nel dataset ci sono 301,143 triple RDF di persone, collegate a ULAN²¹⁴ e DBpedia²¹⁵ (in misura minore perchè *not well-known enough to appear in DBPedia*²¹⁶).

<i>Data part</i>	<i>Resources</i>	<i>Predicates used</i>	<i>RDF triples</i>
Object metadata	73,447 (proxies)	100	5,700,371
Thesaurus	28,000 (concepts)	13	601,819
Person auth. list	66.966 (persons)	21	301,143
Total	168,413	134	6,603,333

Fig. 3.12 parti costitutive del dataset LOD dell'Amsterdam Museum

I dati del museo di Amsterdam (set di dati convertiti e file di mapping dello schema) sono disponibili come *Linked Data*²¹⁷ su *Europeana Semantic Layer* (ESL)²¹⁸, un'istanza in esecuzione del server semantico ClioPatria²¹⁹ che ospita dataset che sono stati mappati su EDM. Le URI PURL del museo sono reindirizzate verso questo server.

I dati sono disponibili sia in RDF/XML che in RDF/Turtle.²²⁰

Come applicazioni finali di utilizzo, il dataset del museo è stato ampiamente utilizzato per un certo numero di applicazioni di *mashup*, ed è centrale per lo sviluppo futuro (già in corso) di una serie di applicazioni, sia web che *mobile*, che combinano i dati del museo con altri dataset. Un esempio è una guida turistica culturale della città di Amsterdam attualmente in via di progettazione, pensata per smartphone, in cui il dataset LOD sarà una fonte di dati di fondamentale importanza.

²¹³ <https://www.geonames.org/>

²¹⁴ <https://old.datahub.io/dataset/getty-ulan>
<http://www.getty.edu/research/tools/vocabularies/ulan>

²¹⁵ <https://wiki.dbpedia.org/>

²¹⁶ V. de Boer et al., *Amsterdam Museum Linked Open Data*, pag. 5

²¹⁷ <https://semanticweb.cs.vu.nl/lod/am/>

²¹⁸ <http://www.europeanaconnect.eu/results-and-resources.php?page=1>

²¹⁹ <http://www.swi-prolog.org/web/ClioPatria/>

²²⁰ Il dataset del museo può essere anche consultato separatamente attraverso un *repository* GIT: <http://eculture.cs.vu.nl/git/public/?p=econnect/metadata/AHM.git;a=summary>

3.2.2 SHARE Catalogue

SHARE (*Scholarly Heritage and Access to Research*) consiste in un progetto di cooperazione e di condivisione di servizi tra le biblioteche campane, lucane e salentine che include la possibilità di accesso integrato all'informazione bibliografica e documentale relativa alle collezioni accessibili online sui siti dei fornitori, con indirizzamento degli utenti ai *full text*. Tale obiettivo è stato realizzato utilizzando sia un catalogo collettivo in *Linked Open Data* (SHARE Catalogue).

Scopo del progetto è di "integrare la "base di conoscenza" rappresentata dai differenti cataloghi di authority e bibliografici, per arricchirla con quella "nuova e in divenire" generata dal web e costituire un sistema informativo integrato in grado di offrire agli utenti uno strumento unico di accesso ai vari OPAC delle biblioteche.

Gli obiettivi del progetto comprendono, tra le altre cose:

- La conversione dei dati secondo il modello RDF
- L'arricchimento di questi attraverso l'integrazione di informazioni provenienti da diverse fonti, possibile tramite collegamenti a progetti esterni, quali per esempio:
 - VIAF, Virtual International Authority File
 - Library of Congress Name Authority File
 - ISNI – International Standard Name Identifier
 - Wikipedia e Wikidata
- La pubblicazione del dataset sui siti preposti ad accogliere dati LOD (ad es. CKAN), affinché altri utilizzino i risultati della conversione in RDF per estendere il progetto o per realizzarne di nuovi; il dataset sarà reso disponibile per uso pubblico come definito dalla licenza Creative Commons.
- Un'applicazione finale sotto forma di portale web navigabile dagli utenti. Il portale di consultazione è dotato di un'interfaccia di navigazione proposta in una modalità intuitiva per facilitare la fruizione dei dati da parte dell'utenza finale, sul modello di dati BIBFRAME (*Bibliographic Framework Initiative*)²²¹.

²²¹ Il documento *Bibliographic Framework as a Web of Data: Linked Data Model and Supporting Services*, pubblicato dalla Library of Congress il 21 Novembre 2012, espone un nuovo modello di dati pensato come evoluzione, in linked open data, del formato Marc 21.

I dati sono disponibili in RDF (*Linked Open Data*), memorizzati in uno SPARQL Endpoint e resi fruibili da opportune funzioni di ricerca e presentazione. I vari livelli informativi sono arricchiti dai dati provenienti da fonti esterne ai cataloghi delle biblioteche con la finalità di estendere il potenziale di ricerca.

3.2.3 Progetto ITACH@

Il progetto ITACH@ (*Innovative Technologies And Cultural Heritage Aggregation*) ha come campo di applicazione l'intero complesso informativo prodotto da enti e istituzioni che si riferiscono all'ambito dei beni culturali (biblioteche, archivi, musei ed enti turistici), e nasce da una riflessione circa:

- l'esigenza poco riconosciuta e non soddisfatta di un accesso integrato ai dati, al di là della loro eterogeneità, quantità, distribuzione, proprietà;
- la necessità di condivisione e utilizzo (e riutilizzo) dei dati: organizzazioni e individui che scelgono di condividere i dati e che traggono vantaggio dalla creazione di "ecosistemi" organizzati e fruibili;
- la necessità di rispondere ai seguenti quesiti:
 - qual è il modo migliore per fornire l'accesso ai dati così che possano essere facilmente riutilizzati?
 - come rendere possibile la scoperta di dati pertinenti all'interno della moltitudine di insiemi di dati disponibili?
 - come consentire alle applicazioni di integrare dati provenienti da fonti eterogenee e sconosciute?

Così il progetto si pone all'interno della più ampia tematica del web semantico: la pubblicazione dei dati secondo gli standard e le buone pratiche previste da questo ambito, come i *Linked Data*.

Una delle componenti tecnologiche del sistema è l'OpLiDaF, un framework per la creazione, strutturazione e visualizzazione di dati in formato RDF/XML, ideato per il trattamento (inteso come mappatura, conversione, pulizia e pubblicazione) di *Linked Data* a partire da dati in formati eterogenei (sia in termini di contenuti che di formati).

Le funzioni principali previste nella piattaforma OpLiDaF sono:

- selezione di ontologie;
- mappatura tra i dati di origine e l'ontologia o le ontologie selezionate;
- creazione di specifiche ontologie a partire da un set di dati;
- produzione di file RDF/XML;
- bonifica del dato (*data cleaning*).

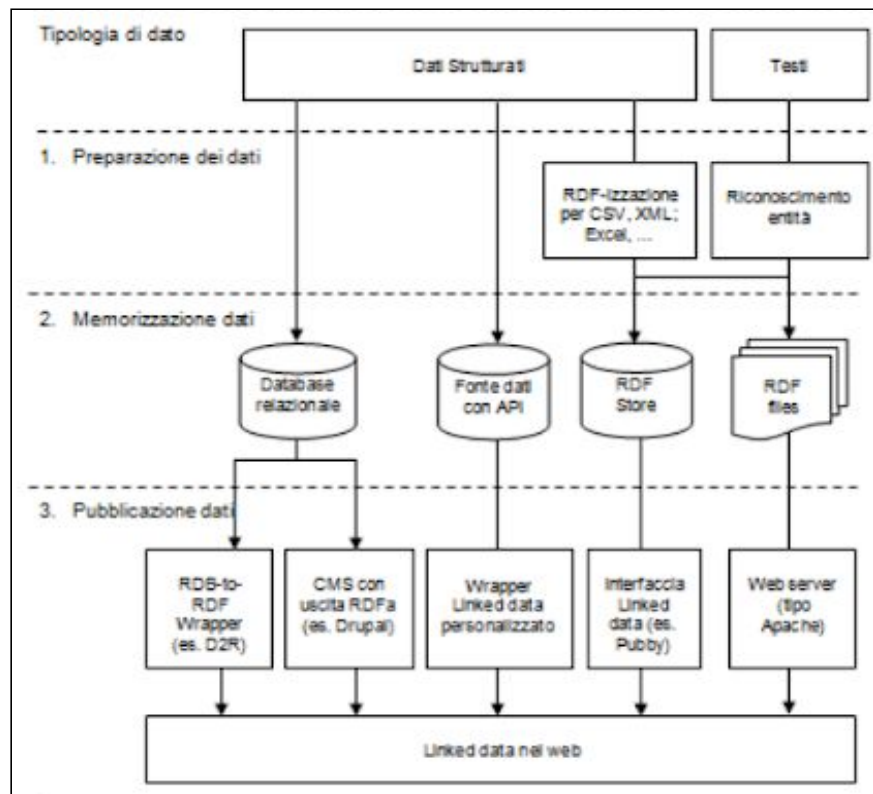


Fig. 3.13 Workflow di pubblicazione di dati eterogenei in Linked Data nella piattaforma OpLiDaF.

Il primo passo per pubblicare i dati nel web semantico è stato la loro conversione dal MARC21 (formato prediletto dalle biblioteche) al RDF/XML.

Molti dati bibliografici strutturati in MARC21 sono memorizzati in database relazionali.

Sono state quindi adottate tecniche di trattamento e traduzione di dati da database relazionali in *Linked Data*.

I dati conservati in database relazionali possono essere facilmente pubblicati come *Linked Data* attraverso l'uso di un insieme di tool disponibili, che partono da

processi di mappatura dai database relazionali in grafi RDF, poi pubblicati sul web secondo i principi dei LOD.

I *Linked Open Data* del progetto sono stati quindi pubblicati a partire da database relazionali.

Il W3C RDB2RDF Working Group sta lavorando all'elaborazione di linguaggi standard per la mappatura di dati relazionali e di schemi di database relazionali in RDF e OWL: tra i principali linguaggi ad oggi disponibili esiste il R2RML (RDB2RDF *Mapping Language*).²²²

Uno dei tool più diffusi e utilizzati per la pubblicazione di database relazionali nel web semantico è il D2R Server, che consente a browser RDF e HTML di navigare i contenuti del database utilizzando SPARQL come linguaggio di ricerca. Si tratta di standard e tecnologie ampiamente riconosciute nel *semantic web*, ma l'obiettivo voleva essere quello di mettere in evidenza il potenziale di un altro linguaggio di mappatura tra schemi di database relazionali e ontologie implementate in RDF(S) o OWL, utilizzato nella piattaforma OpLiDaF: l'R2O (*Relational to Ontology*).

L'R2O è un linguaggio ad alto livello che opera con database che utilizzano l'SQL standard; è ispirato al D2R, ma cerca di superarne i principali limiti: è più potente ed elastico, e dunque più adatto allo sviluppo di mappature complesse, colmando un certo grado di espressività che manca al D2R.²²³

A seguire, i passi del flusso di generazione di dataset RDF a partire da un database relazionale, all'interno della piattaforma OpLiDaF.²²⁴

1. mappatura tra il database relazionale e l'ontologia;
2. produzione del file R2O, che è l'XML che descrive la mappatura grafica tra database e ontologia e che serve poi per generare l'RDF;

²²² vedi <http://linkeddatabook.com/editions/1.0/#htoc69>

²²³ Un presupposto di utilizzo dell'R2O è che il database e l'ontologia (implementata in OWL/RDF) condividano un elevato grado di somiglianza nella struttura, presupponendo che sia il database sia l'ontologia utilizzata siano pre-esistenti e non necessitino di modifiche per essere utilizzate.

(T. Possemato, *OpLiDaF Open Linked Data Framework: una piattaforma per la creazione e la pubblicazione di linked data*, in *Rivista Italiana di Biblioteconomia, Archivistica e Scienza dell'informazione*, vol. 4, n°1, 2013, pag. 8)

²²⁴ utilizzate ontologie Bibtex e BIBO

3. interrogazione della base dati da parte del sistema, estrazione dei record e mappatura nel formato RDF, secondo gli schemi impostati nelle fasi precedenti.

La successiva fase è quella di *data cleaning* (bonifica dei dati) e prevede, oltre all'utilizzo di tool di pulizia per l'eliminazione di caratteri "sporchi" (come ad esempio i codici di sottocampo dei tag MARC21), la creazione automatica di entità identificabili da URI, nella maggior parte dei casi già presenti nei database relazionali, o creati secondo criteri stabiliti.

Il ciclo di vita dei *Linked Data*, all'interno di questo progetto, è stato pensato come suddiviso in sette passi²²⁵:

1. identificazione della fonte dati;
2. modellizzazione del vocabolario;
3. generazione dei dati in formato RDF, tramite i diversi linguaggi di mappatura disponibili;
4. pubblicazione dei dati in RDF;
5. bonifica dei dati prodotti;
6. creazione di collegamenti tra dataset differenti
7. messa a disposizione dei dati, con differenti passi, tra cui la pubblicazione del dataset ottenuto dal processo sul CKAN Registry (*Comprehensive Knowledge Archive Network*).

I passi da 2 a 5 sono pienamente soddisfatti dalla piattaforma, che *si propone come uno strumento utile per chiunque voglia arrivare alla produzione di Linked Data (a prescindere dal sistema di gestione, dal formato dei dati, dalla grandezza del dataset e dalla modalità e ciclicità di aggiornamento), risolvendo una parte di ostacoli e problematiche che il passaggio dal web tradizionale al web semantico pone.*²²⁶

²²⁵ B. Villazón-Terrazas, L. M. Vilches-Blázquez, O. Corcho, A. Gómez-Pérez, *Methodological Guidelines for Publishing Government Linked Data*, in D. Wood (ed.), *Linking Government Data*, 2011

https://link.springer.com/chapter/10.1007/978-1-4614-1767-5_2

²²⁶ T. Possemato, 2013, pag. 18

3.2.4 OpenERCH

Il progetto OpenERCH (*Open Emilia Romagna Cultural Heritage*) ha come scopo primario di convertire dati relativi all'ambito artistico e culturale, riferiti nello specifico a musei ed edifici storici presenti nella regione Emilia Romagna.

Il lavoro svolto da consiste principalmente nella creazione di un dataset²²⁷, rispondente alle caratteristiche tipiche del *Semantic Web*, ottenuto dalla conversione di dati di vario genere relativi a musei e altri edifici e contenitori dislocati nel territorio della regione, facendo riferimento ad uno specifico modello ontologico e sfruttando i principi e le tecnologie del *Linked Data* per ottenere un insieme di informazioni il più possibile correlate e arricchite.

I dati presenti all'interno del catalogo sono essenzialmente di due tipi: edifici ed oggetti contenitori (quali musei, teatri storici, luoghi d'arte contemporanea, ecc.) e beni di catalogo veri e propri in essi contenuti (opere ed oggetti d'arte, reperti archeologici, fotografie, ecc.).

Il progetto fa riferimento principalmente alla prima tipologia di dati, basandosi su un dataset in cui sono contenuti dati relativi ad oltre mille siti tra musei, teatri e altri luoghi di interesse artistico.

L'attività di modellazione e conversione semantica dei dati ha tenuto conto dei principi proposti dal paradigma LOD, esprimendo il significato dei dati tramite una loro rappresentazione basata sul modello ontologico proposto da CIDOC-CRM. Il CIDOC-CRM è quindi l'ontologia alla base del progetto, e fornisce un linguaggio altamente specifico per la rappresentazione di informazioni sui beni culturali relative anche ad ambiti spaziali e temporali.

Ogni risorsa in OpenERCH è individuata univocamente da specifici URI. Più precisamente, tutti gli URI che identificano delle risorse sono creati tramite uno

²²⁷L'IBC (Istituto per i Beni artistici, Culturali e naturali della regione Emilia Romagna) ha fornito i dati di interesse, oggetto del lavoro di conversione, che oggi sono presenti nel portale *Samira*: un sistema informativo che contiene il catalogo del patrimonio culturale della regione.
<http://bbcc.ibc.regione.emilia-romagna.it/>

specifico schema ottenuto tramite l'utilizzo di un prefisso del tipo

`http://dominioIBC/resource/`.²²⁸

I dati iniziali, oggetto di conversione, sono presenti all'interno di un database in cui sono conservati prevalentemente sotto forma di schede di catalogazione: ad ogni oggetto è assegnata una lista di proprietà che ne descrivono le varie caratteristiche. La descrizione degli oggetti è pertanto effettuata, all'interno di tale base di dati, prevalentemente tramite un elenco di valori testuali associati a ciascuno di essi.

L'attività di conversione di metadati in formati CSV²²⁹ e XML è stata effettuata da una componente software che si occupa, a partire dai file di dati appena indicati estratti dal database, di leggere le singole informazioni e creare a partire da esse degli statement RDF che le descrivano in base al modello ontologico proposto da CIDOC-CRM.

Al fine di dimostrare eventuali utilizzi di questo genere, nell'ambito del progetto è stato sviluppato anche l'applicativo ERCH-Mashup che permette di effettuare ricerche tra i metadati relativi ai musei e agli altri oggetti contenitori, aggiungendo, ove possibile, arricchimenti ottenuti da informazioni recuperate da altri dataset esterni, come DBpedia, Freebase, GeoNames e LinkedGeoData.

3.2.5 Archivio Fotografico della Fondazione Federico Zeri

La *Fondazione Federico Zeri*²³⁰ preserva una delle più prestigiose collezioni di fotografie di opere d'arte d'Europa (290.000 esemplari), una ricca biblioteca di storia dell'arte (46.000 volumi), una consistente raccolta di cataloghi d'asta (37.000),

²²⁸ In OpenERCH esiste una netta distinzione tra le risorse vere e proprie e i documenti che le descrivono. Ad ogni risorsa sono pertanto associati diversi URI, che indicano specifiche rappresentazioni documentali (G. De Marco, *Conversione per il Semantic Web e pubblicazione nel Linked Data di dati relativi a beni artistici e culturali della regione Emilia Romagna*, Università Di Bologna, 2011, pag. 51)

²²⁹ Il CSV (*comma-separated values*) è un formato basato su file di testo utilizzato per l'importazione ed esportazione (ad esempio da fogli elettronici o database) di una tabella di dati. In questo formato le informazioni relative ai diversi oggetti sono presentate su righe diverse (ad esempio una riga per ogni record di un database), che a loro volta sono divise in campi (ad esempio le singole colonne di una record di un database) separati da un apposito carattere separatore, in genere rappresentato dalla virgola.

²³⁰ <http://www.fondazionezeri.unibo.it/it/fondazione/chi-siamo>

nonché 15.000 fogli con preziosi appunti del suo creatore Federico Zeri: la più grande collezione privata di documentazione di storia dell'arte presente in Italia.²³¹

E' stato creato un database all'interno del quale far confluire le schede catalografiche dei beni museali, archivistici e bibliografici, oltre alla digitalizzazione delle immagini. Sono stati utilizzati gli standard ICCD per la catalogazione di fotografie e opere d'arte e altri standard e authorities per i materiali documentali.

Il progetto Zeri & LODE²³², nato a Bologna per volere della Fondazione aveva i seguenti scopi:

- riprodurre in *Linked Open Data* i dati di un patrimonio culturale eterogeneo come quello della Fototeca;
- tradurre in un nuovo modello di rappresentazione (RDF), gli standard di catalogazione ministeriali: Scheda OA, per la descrizione di opere d'arte, e Scheda F per la descrizione della fotografia, nonché le altre normative dell'ICCD per la redazione di record bibliografici inerenti la documentazione allegata, la descrizione archivistica e gli authority file degli artisti e fotografi.

Il progetto complessivo ambisce a riunire le immagini possedute da tutti i partner e i relativi metadati in un unico ambiente, inteso come spazio di ricerca, aggregazione e condivisione per la storia dell'arte, liberamente accessibile online e multilingue, che includa e valorizzi immagini, dati, documentazione di corredo e che sia anche capace di utilizzare soluzioni tecnologiche innovative (*image recognition applications, text and image annotation, etc.*).

*Si vuole superare i limiti delle singole banche dati per evidenziare similitudini tra le opere d'arte e avere elementi utili per la datazione. L'aggregazione dei dati provenienti da domini differenti ma affini e la valorizzazione delle loro relazioni permetterà la creazione di nuova conoscenza.*²³³

²³¹ <http://www.fondazionezeri.unibo.it/it/fototeca/fototeca-zeri>

²³² <http://data.fondazionezeri.unibo.it/>

²³³ O. Bozzarelli, V. Mandrile, E. Marangoni, *Quale futuro per i dati di biblioteche e musei? Linked Open Data e Open Data protagonisti al Pubblico dominio #open festival di Torino*, in *Bibliotime*, anno XX, nn° 1-2-3, 2017

Il progetto intende creare un modello ontologico, riutilizzabile da altri *partners*, che restituisca la complessità dell'oggetto fotografico (aspetti di unicità /serialità in diversi archivi, monumento/documento), traduca la specificità di una fotografia di opera d'arte in un sistema archivio, valorizzi e rafforzi la ricchezza semantica del DB Zeri, rispecchi l'analiticità delle schede di metadati ministeriali OA (Opera d'Arte) e F (Fotografia), e supplisca alle lacune del modello concettuale CIDOC-CRM tramite l'integrazione con altre ontologie consolidate.

Le ontologie create, *OA Entry Ontology* e *F Entry Ontology*, e i mapping degli standard OA/F a RDF, hanno richiesto come presupposto la disamina delle proposte di rappresentazione formale in uso nell'ambito catalogafico e archivistico. Per trovare il compromesso tra esigenze di preservazione e di studio/arricchimento dei dati è stata propedeutica l'analisi di alcuni tra i modelli attualmente in uso nei settori del *Cultural Heritage* e del *Semantic Publishing*: CIDOC-CRM (standard per l'interoperabilità semantica tra istituzioni museali), le SPAR Ontologies²³⁴ (una suite di ontologie definite nel dominio dell'editoria per la descrizione) e un modello creato *ad hoc*, HiCO²³⁵, in grado di rappresentare l'atto dell'attribuzione.

Per perseguire l'obiettivo di definizione di un evento, come può essere la creazione di un'opera, e quindi definire formalmente questo scenario complesso e rendere possibile la rappresentazione di tutte le esigenze descrittive²³⁶ i modelli più adatti ed effettivamente scelti per rappresentare gli oggetti dell'archivio sono:

- FRBR che consente di organizzare i metadati amministrativi, gestionali e descrittivi secondo una scomposizione delle caratteristiche estrinseche ed intrinseche dell'oggetto culturale grazie a quattro livelli descrittivi: opera, espressione, manifestazione ed esemplare.

²³⁴ <http://www.sparontologies.net/>

²³⁵ <http://purl.org/emmedi/hico>

²³⁶ ovvero consentire la definizione di una rete di eventi inerente il ciclo di vita dell'oggetto culturale catalogato, al fine di preservare un set finito di informazioni, mantenendo però aperte le porte a mutamenti, cambiamenti e contraddizioni nel corso del tempo, contestualizzando le informazioni nello spazio e nel tempo, e documentando le problematiche riscontrate nel definire tutte le implicazioni che hanno portato l'oggetto culturale ad essere al centro di una rete di relazioni di dipendenza con altri oggetti culturali

- CIDOC-CRM, che nel settore museale si impone come standard *de facto* ma non era stato mai applicato a fotografie del patrimonio storico-artistico. Esso mantiene solo parzialmente tale infrastruttura teorica nella definizione dell'oggetto culturale: riconduce i quattro livelli descrittivi proposti da FRBR a due fondamentali entità per descrivere l'oggetto culturale:
 - una entità prettamente concettuale che assimila il concetto di opera alle sue espressioni
 - una entità che rappresenta la concretezza dell'oggetto culturale assumendo le qualità delle sue possibili manifestazioni.

A differenza del precedente modello, la descrizione in CIDOC-CRM si focalizza sulla rete di eventi collegati all'oggetto, come il processo di produzione dell'opera e il suo ciclo di vita all'interno dell'istituzione. Nel modello si ritrova la rappresentazione di una serie di eventi, collegati alla rispettiva entità di riferimento, rappresentante un aspetto (concettuale o concreto) dell'oggetto, tramite relazioni dirette.

La proposta di estensione e allineamento di CIDOC-CRM a FRBRoo si propone di appianare le divergenze in termini di descrizione dell'oggetto culturale tra FRBR e CIDOC-CRM, ristabilisce la definizione dei quattro livelli descrittivi proposti da FRBR.

La Fototeca Zeri è uno tra i primi archivi fotografici di opere d'arte a offrire i propri dati in LOD: prima in Italia e seguita da pochi casi di rilievo a livello internazionale.

Oltre alla mappatura dei dati, alla successiva conversione dei dati in RDF e alla loro pubblicazione come LOD, visualizzabili sul web attraverso il browser RDF LODView²³⁷ e interrogabili attraverso un endpoint SPARQL, il progetto mira ad arricchire il dataset Zeri con link ad altri dataset e authorities (VIAF, DBpedia, Wikidata, Geonames, ecc...) e trasformare Zeri in un Authority (artisti, fotografi, attribuzioni, ecc.).

Il dataset pubblicato comprende un sottoinsieme di dati relativi alla pittura italiana del XVI secolo, già disponibili sul sito web della Fondazione Zeri. Circa 19.000 opere d'arte e 31.000 fotografie, pari a circa 50.000 schede catalografiche, oltre a

²³⁷ <https://www.w3.org/2001/sw/wiki/LodView>

informazioni su artisti, fotografi, bibliografia e relativa documentazione, sono descritti da più di 11 milioni di triple RDF.

Il completamento del progetto in corso prevede di terminare il *mapping* OA/F to RDF, pubblicare di tutte le schede catalografiche in RDF, integrare i dati RDF nel catalogo attuale online, consentire l'arricchimento della ricerca basata su LOD, integrare i dati con altri partner.

3.2.6 L'esperienza LOD del Museo Galileo: il progetto MINERV@

Il progetto MINERV@ realizza un dataset secondo i principi dei *Linked Open Data* dal titolo "Banca dati Museo Galileo: strumenti, libri, fotografie, documenti" per la creazione di un sistema informativo adeguato per la raccolta e il miglioramento delle diverse tipologie di risorse digitali (immagini, testi, video, suoni, animazioni, ecc...). Lo scopo è quello di connettere il Museo Galileo a una rete di operatori culturali in grado di dare nuovo slancio all'economia digitale e offrire risorse all'industria culturale e creativa.

Il modello precedente, fortemente orientato alla digitalizzazione di libri e manoscritti, risultava inadeguato alla realtà culturale del Museo Galileo, un'istituzione dotata di un patrimonio documentario vario e eterogeneo.

Per rendere fruibile una raccolta documentaria composta da materiale librario, strumenti scientifici, fotografie e risorse multimediali, la biblioteca digitale doveva necessariamente adottare un modello diverso da quello iniziale e ideare un sistema informativo per raccogliere e integrare l'insieme delle collezioni digitali tematiche.

Il percorso di sviluppo della Biblioteca digitale, contraddistinto da un costante e progressivo processo di integrazione e condivisione di dati e da un metodo di lavoro formatosi intorno ai concetti di cooperazione e interoperabilità, sembra naturalmente orientato verso l'universo in espansione dei LOD.

Per rafforzare e completare questa propensione era indispensabile il passaggio ai metadati semantici e alla loro elaborazione in un formato adatto alla logica delle macchine.

A questo si aggiunge il posizionamento storico del Museo Galileo all'interno di una rete di collegamenti istituzionali internazionali con altri soggetti culturali.

I dati selezionati per la conversione in LOD (e dunque in RDF) fanno parte della Banca Dati Cumulativa del Museo Galileo.

I dati selezionati per il progetto sono eterogenei, provenienti da differenti fonti, e molto strutturati (in MARC 21 come formato di origine)²³⁸; di grande interesse, dunque, per un progetto di condivisione delle informazioni attraverso la conversione e la pubblicazione in *Linked Open Data*.

Data la natura dei dati originari e le finalità progettuali, le attività eseguite sono state:

- Individuazione e selezione dei dati, realizzata tra i circa 250.000 record bibliografici e di authority provenienti dal Museo Galileo, in formato MARC21.
- Scelta dei singoli elementi descrittivi (tag MARC) inclusi nel record da utilizzare nella conversione in RDF (tutte le informazioni utili all'utenza finale (ricercatori), come il titolo, l'autore, l'editore, l'anno di pubblicazione, le informazioni aggiuntive identificative delle risorse).
- Analisi e modellazione, selezione delle ontologie utilizzate per la mappatura e conversione dei record in RDF. Tra le ontologie analizzate e poi utilizzate ci sono: DCMI Metadata Terms, RDF Schema, BIBO, FRBR, RDA, FOAF.
- Produzione dell'output e sua validazione: la mappatura dei dati è stata utilizzata per l'implementazione delle regole di conversione, che sono confluite nel framework di conversione dei dati in RDF/XML. Il file RDF prodotto è stato poi analizzato per verificare la correttezza formale e sintattica dell'output, e la corretta implementazione delle regole di applicazione delle ontologie selezionate.

²³⁸ Il processo di creazione e pubblicazione di Linked Open Data è avvenuto indipendentemente dal formato di origine dei dati.

- Pubblicazione: il dataset è stato pubblicato in *DataHub*²³⁹ (il portale della *OpenKnowledge Foundation*, un network mondiale no-profit dedicato all'apertura delle informazioni) e nel Portale OpenData della Regione Toscana²⁴⁰. Con questo ultimo passaggio i dati sono diventati accessibili, aperti, riutilizzabili.

3.2.7 LinkedUp Data Catalogue

LinkedUp Data Catalogue (anche conosciuto come *Linked Education Cloud*²⁴¹) è un meta-dataset dedicato al supporto di persone e applicazioni nella scoperta, esplorazione e utilizzo di dati nel web a scopo di servizi di istruzione. È un set di dati in evoluzione, con la maggior parte del suo contenuto che viene fornito estraendo automaticamente le informazioni rilevanti dalle descrizioni esterne e dagli stessi set di dati inclusi.

Se la scoperta dei dati e la facilità di accesso ad essi sono fondamentali in molti settori, nel campo dell'istruzione questo è particolarmente interessante; tuttavia l'adozione di pratiche di dati aperti e collegati, in questo ambito, è ancora molto limitata .

LinkedUp²⁴² è un'azione di supporto europea che promuove l'adozione di *Web Data* nell'educazione, e in particolare il loro uso in servizi di istruzione.

E' stato sviluppato un catalogo di dati rilevanti per l'istruzione, a cui è possibile partecipare e accedere ai dati di utilizzo per le proprie applicazioni. E' costruito per interagire con i cataloghi di dati esistenti (in particolare CKAN) in modo da estrarre, rappresentare e pubblicare informazioni sui dataset rilevanti pubblicati come dati collegati, con il catalogo stesso basato sui principi dei *Linked Data*.

Il catalogo contiene attualmente descrizioni di 50 set di dati. Ognuno di questi è accessibile almeno attraverso un endpoint SPARQL.

²³⁹ <https://datahub.io/>

²⁴⁰ <http://dati.toscana.it/>

²⁴¹ <https://old.datahub.io/organization/linked-education>

²⁴² <http://linkedup-project.eu/>

Il progetto ha eseguito una serie di trasformazioni di dataset esistenti, non basati su RDF, in dati RDF e collegati, in modo che possano essere forniti attraverso un endpoint SPARQL e inclusi nel catalogo.

L'applicativo principale del dataset è l'interfaccia²⁴³ del catalogo.

I set di dati in questo processo vengono identificati per la prima volta tramite l'uso di *Datahub.io*, un catalogo basato su CKAN di set di dati disponibili sul Web.

Da questi, l'API CKAN viene utilizzata per estrarre i metadati di base su ciascun set di dati (titolo, descrizione, licenza) e informazioni sul loro punto di accesso (endpoint SPARQL). Queste informazioni iniziali sono presentate usando VOID e Dublin Core. Dopo aver ricevuto le informazioni sull'endpoint SPARQL per ogni set di dati, vengono interrogate per ottenere l'elenco di grafici, classi e proprietà incluse, rappresentati come sottoinsiemi, partizioni di classe e partizioni di proprietà in VOID.

Come aggiunta al flusso di lavoro iniziale di catalogazione dei dati, abbiamo incluso una fase di mappatura del vocabolario nel flusso di lavoro. Questa fase è implementata utilizzando strategie di base di allineamento ontologico che ci permettono di formalizzare un grado di somiglianza semantica tra due termini provenienti da vocabolari distinti.

Oltre al processo di mappatura del vocabolario, che è semiautomatico, e all'input iniziale dei metadati in *Datahub.io*, l'intero flusso di lavoro è automatico e viene eseguito regolarmente, quando viene incluso un nuovo set di dati o semplicemente per aggiornare il dataset del catalogo.

Il catalogo LinkedUp di set di dati Web per l'istruzione è molto diverso, per scopo e forma, per molti altri set di dati nel *Linked Data Web*. È un metadata-set dedicato al supporto di persone e applicazioni per la scoperta, l'esplorazione e l'utilizzo di dati web per lo scopo di servizi educativi innovativi.

²⁴³ <http://data.linkededucation.org/linkedup/catalog/browse/>

3.2.8 Progetto LOD del COBIS

Il CoBiS (Coordinamento delle Biblioteche Speciali e Specialistiche di Torino) è una rete di 65 biblioteche dell'area torinese il cui profilo è eterogeneo per diversi aspetti: materiale posseduto, sistemi di catalogazione, modelli di rappresentazione del dominio e OPAC.

Il progetto *Linked Open Data* del CoBiS (finanziato dalla Regione Piemonte e dal Comitato ICT) ha lo scopo di fornire un unico punto di accesso alle collezioni delle biblioteche aderenti al progetto²⁴⁴.

Il progetto utilizza i *Linked Data* per due scopi principali:

- integrare i cataloghi eterogenei delle 12 biblioteche (generati a partire da 5 software gestionali diversi);
- arricchire il catalogo con informazioni aggiuntive sugli autori, provenienti da fonti esterne connesse alla *Linked Open Data cloud*.

I dati sono estratti dai database locali delle biblioteche, mappati sulle ontologie di riferimento e convertiti in un grafo RDF che si aggiorna periodicamente. Qui, grazie a un insieme di procedure di *entity matching*, annotiamo gli *interlink* tra le entità del CoBiS e quelle presenti in fonti esterne: i dati vengono arricchiti e collegati alla *Linked Open Data cloud*.

Infine il catalogo aumentato è pubblicato tramite un endpoint SPARQL che a sua volta alimenta un sito web.

²⁴⁴ Deputazione Subalpina di Storia Patria, Associazione Archivio Storico Olivetti, INAF Istituto Nazionale di Astrofisica, INRIM Istituto Nazionale di Ricerca Metrologica, Biblioteca Nazionale del Club Alpino Italiano, Accademia delle Scienze di Torino, Fondazione Giorgio Amendola, Biblioteca civica musicale Andrea Della Corte, Accademia di Medicina, Istituto per i Beni Musicali in Piemonte, Istituto di studi storici Gaetano Salvemini, Centro internazionale di studi Primo Levi.

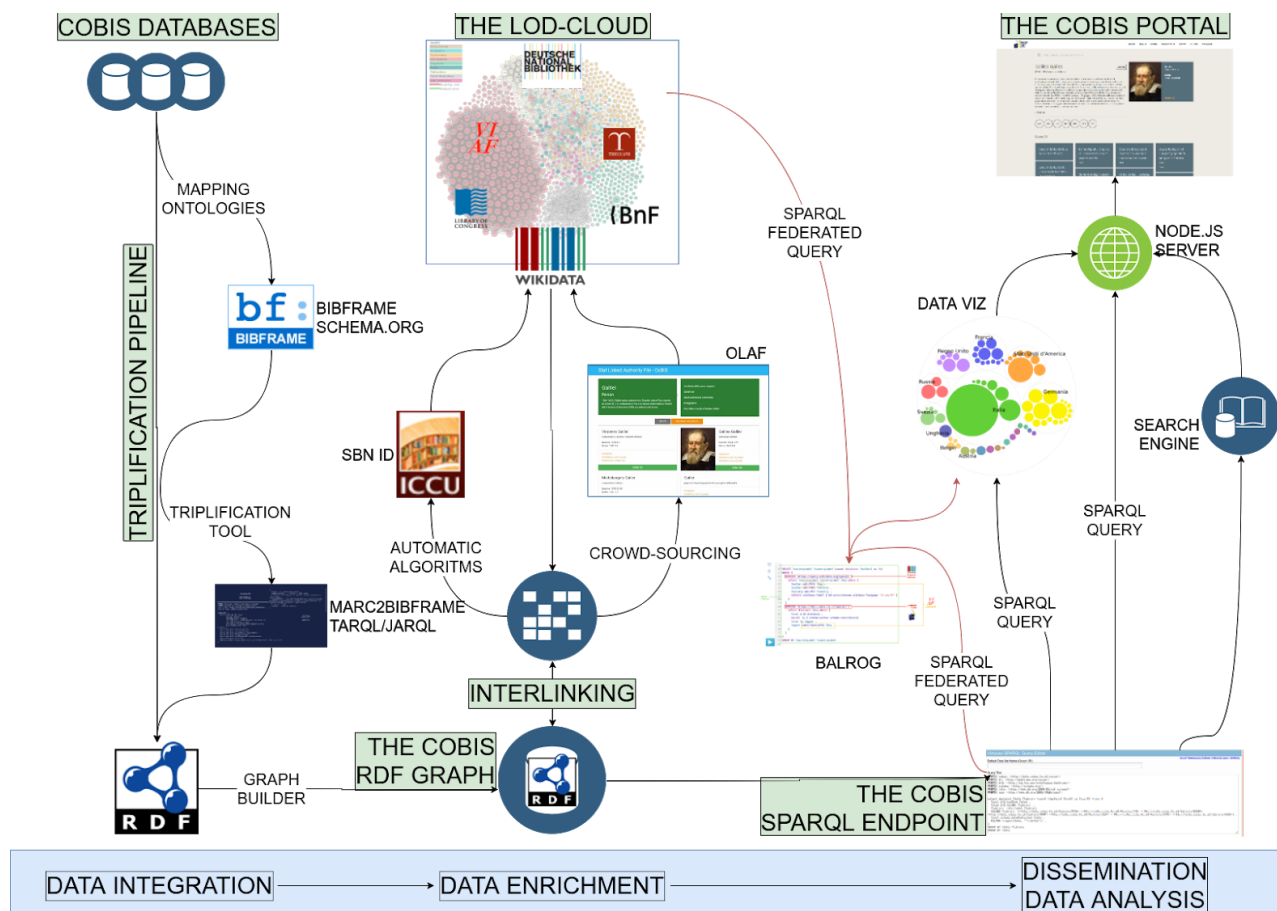


Fig. 3.14 Architettura del progetto LOD del CoBiS

I dati bibliografici sono convertiti dai formati utilizzati nei cataloghi locali (MARC, UNIMARC, Dublin Core, CSV custom) in triple RDF modellate secondo le ontologie BIBFRAME 2.0 e schema.org.

Il processo di triplificazione si avvale di due principali software *open source*:

- TARQL, converte i dati dal formato CSV a RDF, utilizzando per la mappatura la stessa sintassi SPARQL;
- JARQL, software realizzato durante la fase-pilota del progetto, che consente di prendere come input dati strutturati in forma di albero (come XML e JSON), i più frequenti tra le fonti dati del CoBiS.

La *pipeline* automatica di triplificazione è concepita in modo tale da essere facilmente replicabile, aggiornabile e apribile a nuovi partners. Qualsiasi catalogo che può essere esportato in MARC, UNIMARC o OAI-PMH Dublin Core, o è pubblicato nell'OPAC di SBN, può partecipare al progetto in modo semplice.

Il grafo RDF del CoBiS rappresenta un unico punto di accesso ai cataloghi delle biblioteche aderenti al progetto. Le tecnologie *Linked Data* hanno consentito di lavorare su un unico *authority file* all'interno del quale è stato avviato un processo di allineamento tra le diverse entità di autori presenti nei database di partenza. Grazie ai *Linked Open Data* esposti da Wikidata e VIAF le voci degli autori integrano ora nuove informazioni provenienti da fonti esterne disponibili online; per stabilire gli *interlink* che consentono questo flusso di dati sono stati usati processi automatici di *entity matching*.

Il Portale del progetto LOD del CoBiS implementa le best practices della comunità *Linked Open Data*.

Un endpoint SPARQL pubblico alimenta direttamente il sito web²⁴⁵ i cui contenuti sono generati da query SPARQL live. In questo modo i *Linked Open Data*, opportunamente organizzati, possono essere consultati da tutti gli utenti.

²⁴⁵ <https://dati.cobis.to.it/>

Galileo Galilei
(1564 - 1642) scienziato italiano

[Dati RDF](#)

Nascita:
1564-2-25, Pisa

Morte:
1642-1-8, Arcetri

[WIKIPEDIA](#)

Il suo nome è associato a importanti contributi in dinamica e in astronomia (legati al perfezionamento del telescopio, che gli permise importanti osservazioni astronomiche) oltre all'introduzione del metodo scientifico (detto spesso metodo galileiano o metodo scientifico sperimentale). Di primaria importanza fu anche il suo ruolo nella rivoluzione astronomica, con il sostegno al sistema eliocentrico e alla teoria copernicana. Sospettato di eresia e accusato di voler sovvertire la filosofia naturale aristotelica e le Sacre Scritture, Galileo fu processato e condannato dal Sant'Uffizio, nonché costretto, il 22 giugno 1633, all'abiura delle sue concezioni astronomiche e al confino nella propria villa di Arcetri. Solo 359 anni dopo, il 31 ottobre 1992, papa Giovanni Paolo II, alla sessione plenaria della Pontificia accademia delle scienze, ha dichiarato riconosciuti "gli errori commessi" sancendo la conclusione dei lavori di un'apposita commissione di studio da lui istituita nel 1981.

Interlink

VIAF WD LoC GND BNF SBN DBI

Opere (97)

Opere di Galileo Galilei a cura di Franz Brunetti	Ioannis Kepleri... Dioptrice seu demonstratio eorum quae visui & visi... 1611	Discorsi e dimostrazioni matematiche, intorno à due nuoue scienze atte... 1638	Vsus et fabrica circini cuiusdam proportionis, per quem omnia fere tum... 1655
---	--	---	---

Fig. 3.15 sito web del progetto LOD del CoBiS.

L'immagine di esempio è tratta dalla pagina "autori", dove il nome dell'autore stesso proviene dall'*authority file* LOD del CoBiS, la didascalia da Wikipedia, la biografia da SBN o, in alternativa, da Wikipedia, i metadati biografici da Wikidata e l'immagine da Wikidata o Wikipedia; il bottone "dati RDF" mostra la porzione di grafo che sottende la creazione della pagina. A metà della pagina è presente la lista delle risorse interconnesse, che rimanda alla consultazione di differenti repositories biografici e bibliografici (VIAF, Library of Congress, Deutsche National Bibliothek GND, Istituto Treccani ecc.). In fondo alla pagina c'è la lista dei libri dell'autore presenti nel database del CoBiS. Cliccando sopra i box, si viene rimandati all'informazione bibliografica e a chi possiede il libro. Le tre sezioni principali del sito (autori, opere e anagrafiche delle biblioteche) utilizzano lo stesso sistema di gestione dei contenuti.

3.3 Riepilogo dello stato dell'arte

La tabella seguente riporta un riepilogo delle caratteristiche principali dei progetti fin qui esaminati, sottolineandone i punti di contatto con il lavoro oggetto di questa tesi.

Progetto	LOD 5 star	beni culturali	ontologie / modelli di riferimento	contenuto eterogeneo	dataset esterni in comune	pubblicato su CKAN	altri elementi in comune
Amsterdam Museum	SI	SI	EDM	SI	GND DBpedia		
SHARE Catalogue		NO (biblioteche)	BIBFRAME	NO	VIAF Wikidata	SI	
ITAC@		SI				SI	Database relazionale e uso di D2RQ
OpenERCH		SI	CIDOC-CRM	SI	DBpedia		
Fototeca Zeri		SI	create sulla base di CIDOC-CRM FRBR	SI	DBpedia VIAF Wikidata	SI	scheda OA dell'ICCD come standard di partenza
MINERV@		SI	DCMI RDF Schema BIBO FRBR RDA FOAF.	SI			
LinkedUp			VoID DC				
CoBiS					VIAF GND Wikidata		finanziato dalla regione Piemonte come Mèmora

Tab. 3.1 Riepilogo dello stato dell'arte

4. Pubblicare dati secondo il paradigma LOD

Nei capitoli precedenti è stata presentata un'analisi delle principali caratteristiche e problematiche legate all'attività di divulgazione dei dati riguardanti beni artistici e culturali. Tale analisi ha portato all'individuazione delle tecnologie legate al *Semantic Web* e, più nello specifico, ai *Linked Open Data*, come gli strumenti attualmente più adatti per far fronte alla pubblicazione e rappresentazione della conoscenza relativa a tali dati.

Il progetto oggetto di questo lavoro di tesi ha proprio lo scopo primario di convertire dati relativi all'ambito artistico e culturale, in particolare riferiti alla regione Piemonte, nei formati e nei modi proposti dal *Semantic Web* e precedentemente descritti. Tale attività di conversione ha tenuto conto delle regole e dei principi dei *Linked Data*, permettendo ai dati di esprimere formalmente il proprio significato semantico, aderendo ad una specifica ontologia, e consentendo la creazione di collegamenti in uscita verso altri dataset presenti nel *web of data*.

I dati di riferimento, oggetto del lavoro di conversione, sono presenti nel portale Mèmora (cap. 4.5) consultabile liberamente tramite un apposito sito web²⁴⁶ che permette attività di ricerca, navigazione e consultazione delle informazioni sui beni. Le descrizioni dei vari oggetti sono presentate essenzialmente nella forma tipica delle schede di catalogazione: essa è infatti affidata a schede informative di vario genere che ne espongono, prevalentemente tramite contenuti testuali, le varie caratteristiche, come mostrato in figura.

²⁴⁶ <http://www.memora.piemonte.it/>

4.1 Linking Open Data Project

Il progetto L3²⁴⁷ (*Linking Open Data*) del W3C è uno sforzo collettivo di base fondato nel gennaio 2007 con l'obiettivo (che ha generato la crescente comunità dei *Linked Data*) di mandare avanti il *Semantic Web* identificando i set di dati esistenti disponibili con le licenze aperte, convertirli in RDF secondo i principi dei *Linked Data* e pubblicarli sul Web.

Il progetto è sempre stato aperto a chiunque pubblichi i dati secondo i principi dei *Linked Data*. Questa apertura è stata sicuramente un fattore rilevante per il successo del progetto.

4.1.1 Evoluzione del progetto

L'evoluzione del progetto L3 è facilmente e intuitivamente visibile grazie al modello grafico del *Linked Open Data Cloud*²⁴⁸, un diagramma che mostra i dataset collegati pubblicamente disponibili. Il diagramma è aggiornato regolarmente e gestito da *Insight Center for Data Analytics*²⁴⁹.

Lo schema del LOD Cloud è pubblicato sotto la licenza Creative Commons Attribution²⁵⁰ ed è, quindi, liberamente utilizzabile.

*The dataset currently contains 1,234 datasets with 16,136 links (as of June 2018).*²⁵¹

Il crescente ricorso alle tecnologie dei *Linked Data* da parte delle istituzioni e il conseguente aumento del numero di dataset aperti collegati sono visibili chiaramente nell'aumento della dimensione del *Cloud* nel corso degli anni:

²⁴⁷

https://www.w3.org/wiki/SweoIG/TaskForces/CommunityProjects/LinkingOpenData#Project_Pages

²⁴⁸ LOD Cloud: <https://lod-cloud.net/>

²⁴⁹ <https://www.insight-centre.org/>

²⁵⁰ <https://creativecommons.org/licenses/by/4.0/>

²⁵¹ lodcloud.net

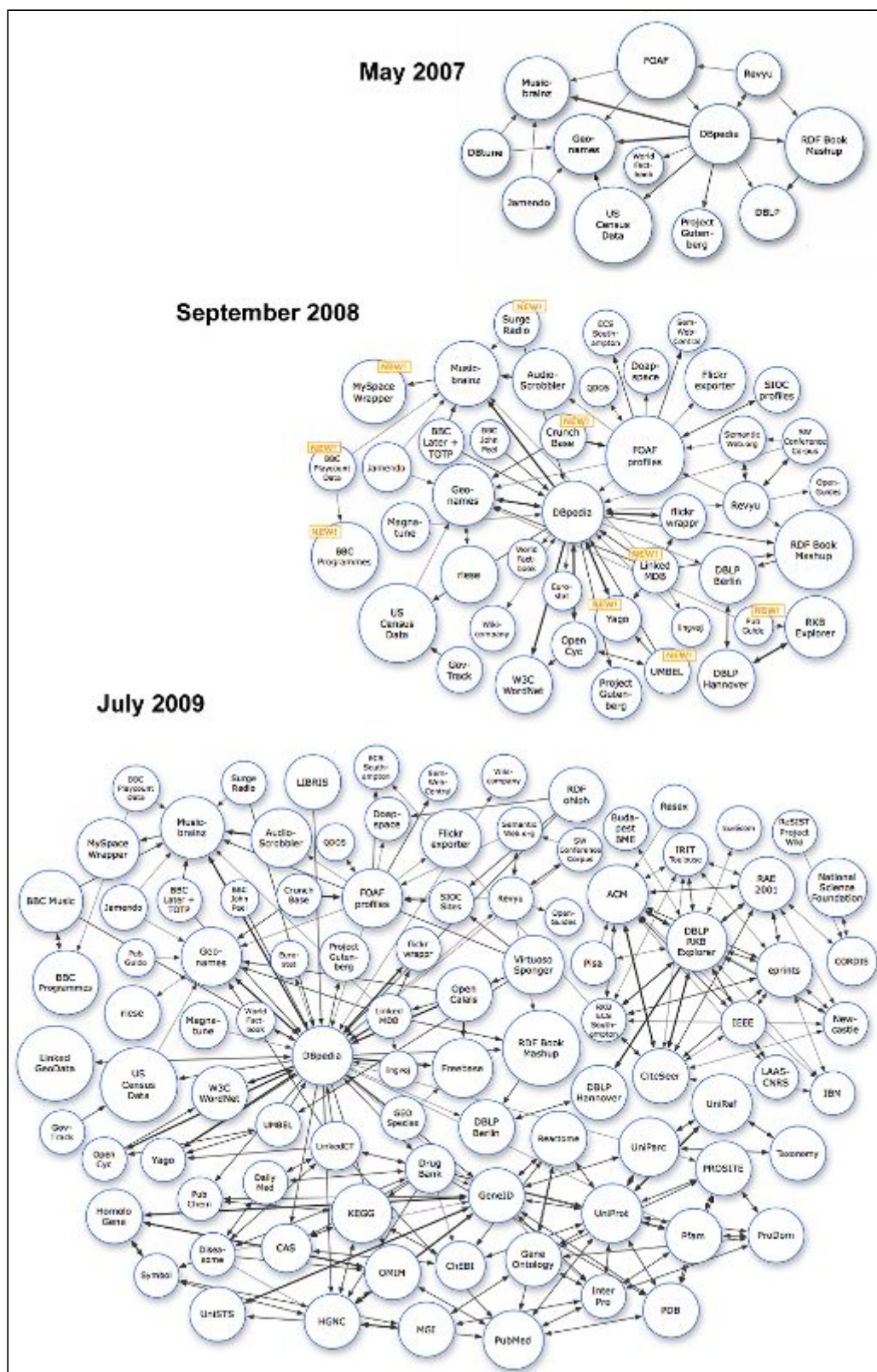


Fig. 4.2 Crescita del numero di serie di dati pubblicati sul Web come *Linked Data*.

Nel LOD Cloud, i colori classificano i set di dati per dominio tematico.

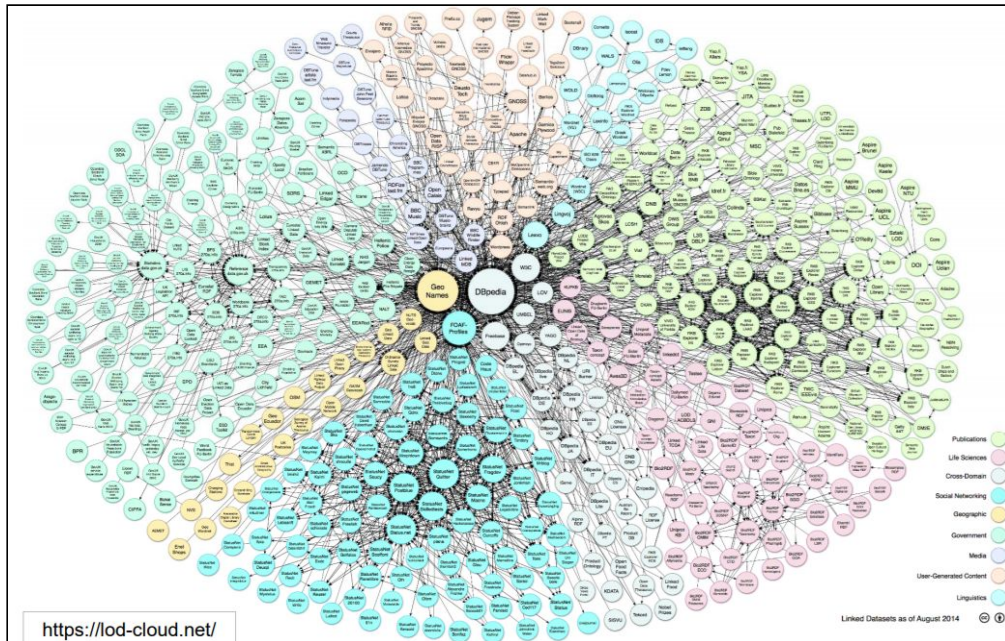


Fig. 4.3 LOD Cloud: agosto 2014

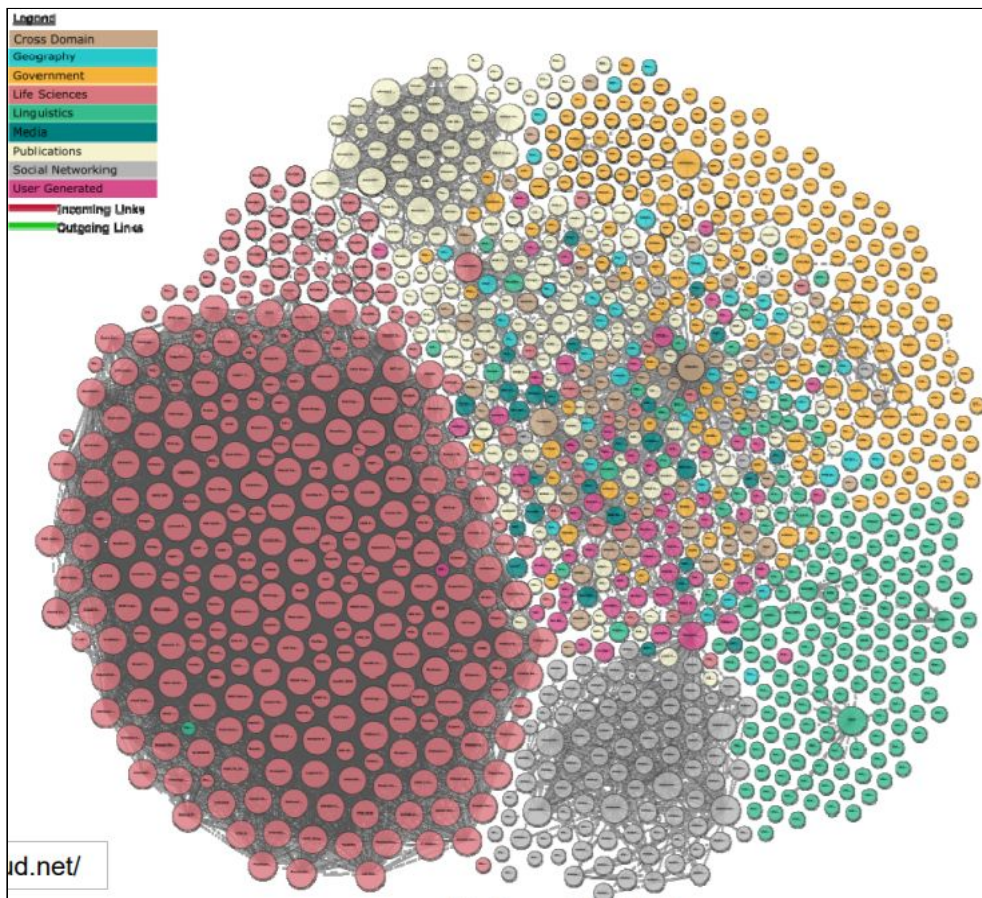


Fig. 4.4 LOD Cloud: settembre 2017

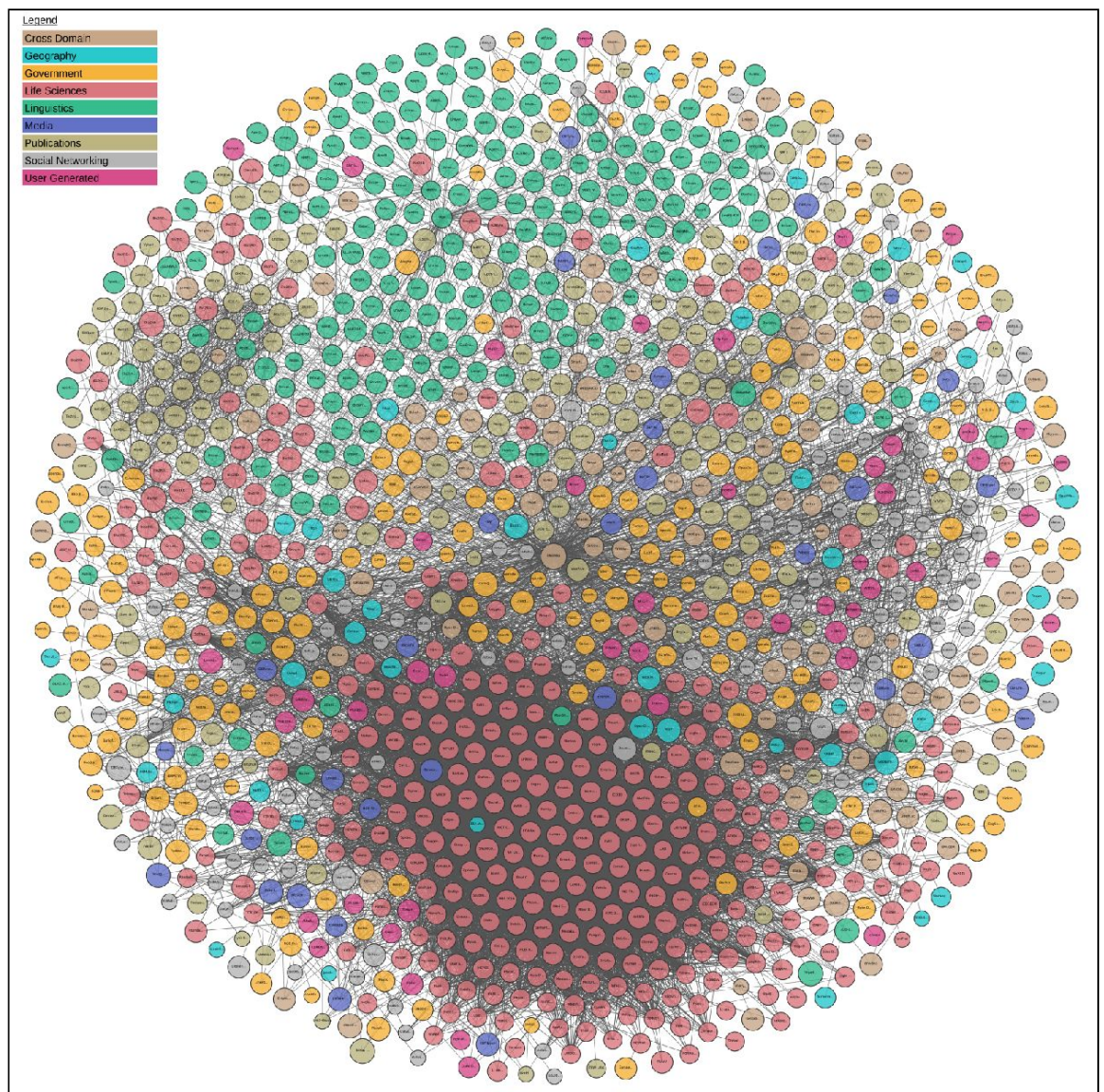


Fig. 4.5 LOD Cloud: marzo 2019

4.2 Contribuire al *Linking Data*

Negli ultimi anni, un numero sempre maggiore di individui e organizzazioni stanno adottando il LOD come un modo per pubblicare i propri dati nel web. Tutti questi “fornitori di dati” hanno contribuito alla crescita del *web of data*, contenente ormai miliardi di triple RDF provenienti da tantissime fonti diverse, che coprono tutti i tipi di

argomenti (località geografiche, persone, aziende, libri, film, musica, programmi televisivi e radiofonici, dati statistici, comunità online e recensioni, ecc...).

La forza e la diversità degli ecosistemi che si sono evoluti recentemente dimostra una richiesta di accesso ai dati mai riscontrata in precedenza: le organizzazioni e gli individui che scelgono di condividere i dati trarranno sicuramente beneficio dall'emergere di questi ecosistemi²⁵².

Vengono poste quindi tre domande chiave:

- Qual'è il modo migliore per fornire l'accesso ai dati in modo che possano essere facilmente riutilizzati?
- Come abilitare la scoperta di dati rilevanti all'interno della moltitudine di set di dati disponibili?
- Come integrare i dati da un gran numero di fonti precedentemente sconosciute?

Secondo la prospettiva qui proposta, i *Linked Open Data* (nel contesto del web semantico) sono il modo migliore per consentire la condivisione e il riutilizzo dei dati su vasta scala: essi, sfruttando l'infrastruttura del web, rivoluzionano il modo in cui *scopriamo, accediamo, integriamo e usiamo* i dati.

“Far conoscere” e “fare comunicazione” circa il patrimonio culturale italiano significa, in questo ambito, offrire le proprie conoscenze al resto del mondo (del web), ovvero pubblicare dati che seguano i principi dei LOD, e quindi:

- codificarli in RDF
- utilizzare ontologie e vocabolari standard per classi e proprietà
- utilizzare URI noti per denominare entità (persone, luoghi, concetti, eventi, ...)
- fornire agli URI le proprie risorse in modo che le altre istituzioni possano utilizzarle

I dataset devono essere collegati ad altre risorse sul web (link esterni): più sono numerosi i link in uscita, più attireranno link in entrata.

²⁵² Heath & Bizer, 2011, cap. 1.1

L'esistenza di una vera e propria rete di dati aperti molto ampia e già fruibile, quale quella offerta dal Linking Open Data Project, con i quali intrecciare le informazioni culturali disponibili, rappresenta un contesto ideale per l'esposizione dei contenuti presi in considerazione da Mèmore.

4.3 Modelli di pubblicazione di Linked Data

Publicare i *Linked Data* richiede l'adozione dei principi di base delineati al [capitolo 1](#). La conformità con gli standard e le *best practice* che sono alla base di questi principi è ciò che consente ai *Linked Data* di ottimizzare l'interoperabilità e il riutilizzo dei dati sul Web. Tuttavia, la conformità ai principi dei dati collegati non comporta l'abbandono dei sistemi di gestione dei dati ma semplicemente l'aggiunta di un ulteriore strato tecnico per collegarli nel Web dei dati. Mentre esiste un numero molto grande di sistemi tecnici che possono essere connessi al *Web of Data*, i meccanismi per farlo rientrano in un piccolo numero di *pattern di pubblicazione* di Linked Data. In questo capitolo viene fornita una panoramica di questi modelli.

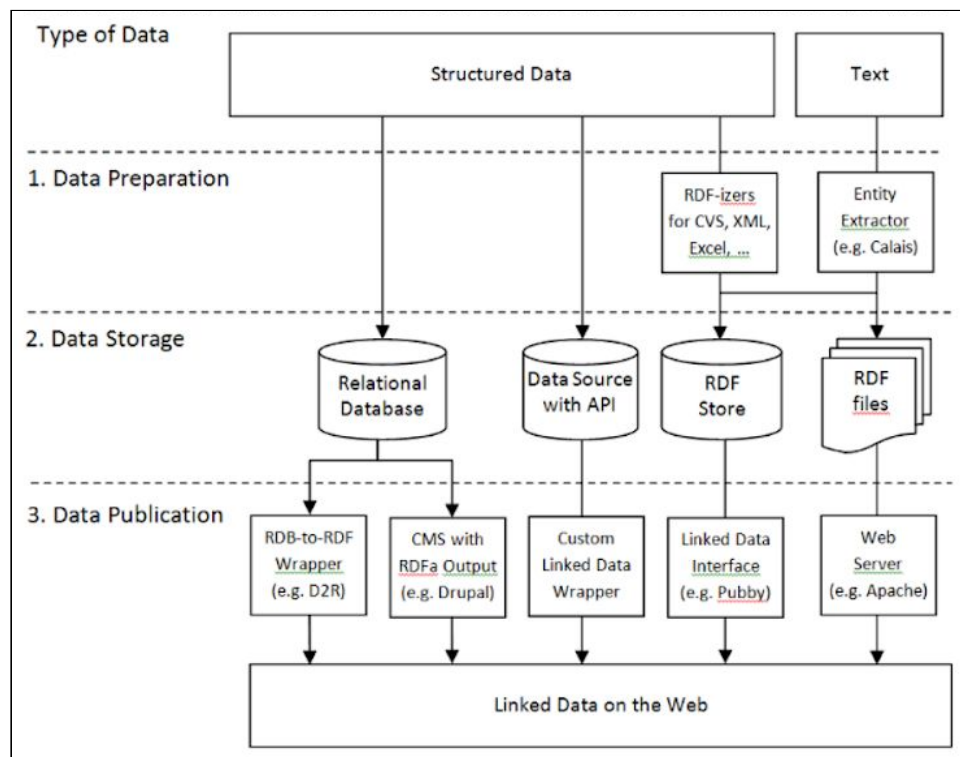


Fig. 4.6 Pattern di pubblicazione di Linked Data più comuni sotto forma di flussi di lavoro, da dati strutturati o contenuti testuali fino a dati collegati pubblicati sul Web. ²⁵³

²⁵³ <http://linkeddatabook.com/editions/1.0/#htoc69>

La considerazione primaria nella selezione di un flusso di lavoro per la pubblicazione di dati collegati riguarda la natura dei dati di input.

- Da dati strutturati interrogabili a dati collegati.
I set di dati memorizzati nei database relazionali possono essere pubblicati in modo relativamente semplice come dati collegati mediante l'uso del database relazionale ai wrapper RDF. Questi strumenti consentono all'editore di dati di definire i mapping dalle strutture di database relazionali ai grafici RDF pubblicati sul Web in base ai principi dei dati collegati. Uno strumento ampiamente utilizzato progettato per questo scopo è D2R Server, descritto al [cap. 5.5.1](#) .
- Da dati statici strutturati a dati collegati.
I dati di input statici possono essere costituiti da file CSV, fogli di calcolo Excel, file XML o dump di database. Per servirli come dati collegati sul Web, devono essere sottoposti a un processo di conversione che genera file RDF statici o carica i dati convertiti direttamente in un archivio RDF.²⁵⁴ Laddove i file statici sono già in formato RDF e seguono i principi dei dati collegati, possono essere semplicemente pubblicati sul Web utilizzando un server Web classico o caricati in un archivio RDF che disponga di un'interfaccia di dati collegati appropriata.
- Da documenti di testo a dati collegati.
Laddove l'input per un flusso di lavoro di pubblicazione di dati collegati comprende documenti testuali in linguaggio naturale è possibile passare questi documenti tramite un estrattore di entità di dati collegati come Calais²⁵⁵, Ontos²⁵⁶ o DBpedia Spotlight²⁵⁷ che annota i documenti con gli URI dei dati collegati delle entità a cui si fa riferimento nei documenti. La pubblicazione di queste annotazioni insieme ai documenti aumenta la rilevabilità dei documenti e consente alle applicazioni di utilizzare le origini dei dati collegati come conoscenza di base per visualizzare informazioni

²⁵⁴ Elenco degli strumenti RDFizing che possono essere utilizzati per questa conversione: <https://www.w3.org/wiki/ConverterToRdf>

²⁵⁵ <http://www.opencalais.com/>

²⁵⁶ <https://ontos.com/>

²⁵⁷ <https://www.dbpedia-spotlight.org/demo/>

complementari sulle pagine Web o per migliorare le attività di recupero delle informazioni, ad esempio offrire la navigazione sfaccettata anziché semplice completa ricerca di testo.

Relativamente a questa tesi, il lavoro fa riferimento al primo caso.

Generalmente, i dati strutturati sono quelli archiviati in database relazionali (SQL) o in file tabulari (CSV, XML, ecc...). Essi trasportano l'informazione di entità, per esempio trasportano informazioni associate ad un documento (entità) e per ogni entità trasportano le informazioni relative alla sua consistenza fisica, lingua, ecc...; queste ultime sono attributi dell'entità. In generale, un dato strutturato è facilmente interpretabile da un'applicazione e questo rappresenta un passo fondamentale per l'interoperabilità.

Prima della pubblicazione il dato, seppur strutturato, deve essere convertito.

Il processo di conversione e pubblicazione di dati presi da database SQL si effettua partendo da strumenti chiamati RDBtoRDF (*Relational Database to RDF*), come ad esempio D2R ([cap. 5.5.1](#)).

Alla fine del processo di conversione, il dato può essere pubblicato sul web usando sia lo schema di presentazione del dato raw (mediante SPARQL endpoint) sia mediante un'interfaccia web.

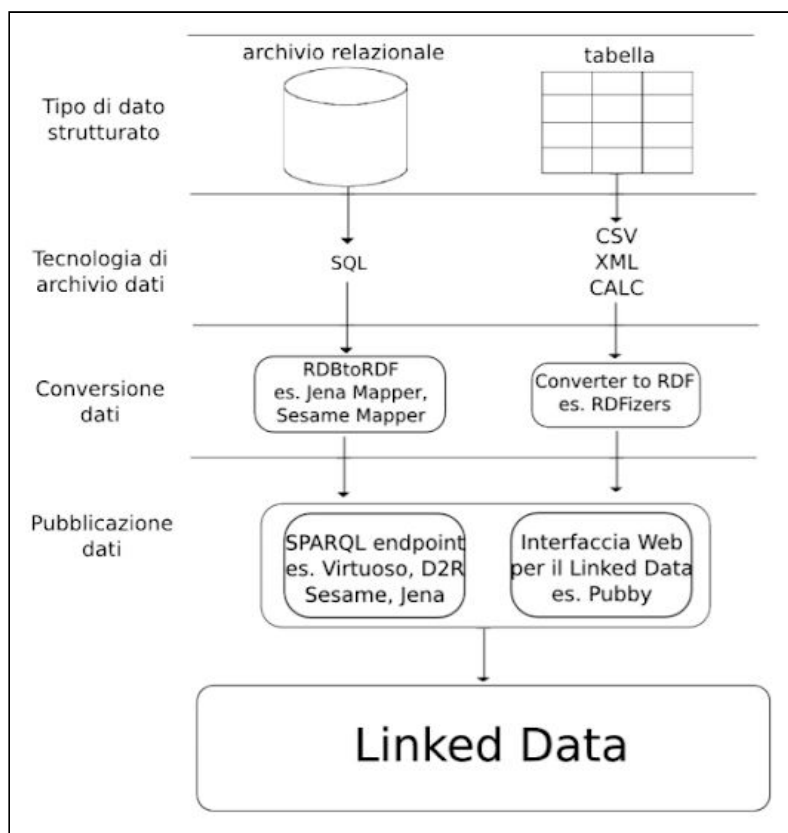


Fig. 4.7 riassunto dell'intero processo di pubblicazione, partendo dai dati strutturati fino ai dati *linkable*.²⁵⁸ Tutte le fasi del processo sono spiegate nel dettaglio nel capitolo seguente.

4.4 Fasi del processo di conversione e pubblicazione dei dati

Generalmente, le fasi in cui si articola il *workflow* di lavorazione per la conversione e pubblicazione di dati LOD, a prescindere dall'ordine esatto in cui vengono svolte, sono riconducibili alle seguenti attività:²⁵⁹

- I. Identificazione e analisi della fonte dati; disegno dell'URI; definizione delle licenze di distribuzione e uso.

²⁵⁸ G. Rizzo, F. Morando, J. C. De Martin, *Open Data: la piattaforma di dati aperti per il Linked Data*, in *Informatica e diritto*, XXXVII Annata, Vol. XX, 2011, n. 1-2, pp. 493-511

²⁵⁹ M. Guerrini, T. Possemato, *Linked Data, un nuovo alfabeto del web semantico*, in *Rivista italiana di biblioteconomia, archivistica e scienza dell'informazione*, vol.4, n°1, 2013, pp. 20-21: *Ciclo di vita dei Linked Data*.

- II. Modellizzazione: scelta o la creazione di ontologie e vocabolari controllati. In questo caso: adozione di ontologie esistenti, espresse in OWL, Web Ontology Language, o RDF(S) o con la creazione (più complessa) di nuove ontologie.
- III. Generazione dei dati in formato RDF, tramite diversi linguaggi di mappatura disponibili, anche in relazione al formato di origine del dato. In questa fase l'operazione più delicata è la creazione di URI, poiché essi sono la chiave per allineare risorse eterogenee provenienti da fonti differenti.
- IV. Bonifica dei dati prodotti, per individuare eventuali e possibili errori di conversione e rendere il dato qualitativamente usabile.
- V. Creazione di collegamenti tra dataset differenti, con l'identificazione di dataset di interesse che possano diventare *linking target*, identificando relazioni tra i singoli dati, validando le relazioni individuate.
- VI. Pubblicazione dei dati in RDF; rendere concreto l'utilizzo dei dati, con differenti passi, tra cui la pubblicazione del dataset ottenuto dal processo su uno o più portali o piattaforme (es: CKAN, Comprehensive Knowledge Archive Network, Socrata, ...).
- VII. Valorizzazione dei dati tramite la costruzione di strumenti semplici ed efficaci di utilizzo del dataset; per esempio, tramite la costruzione di interfacce per la fruizione.

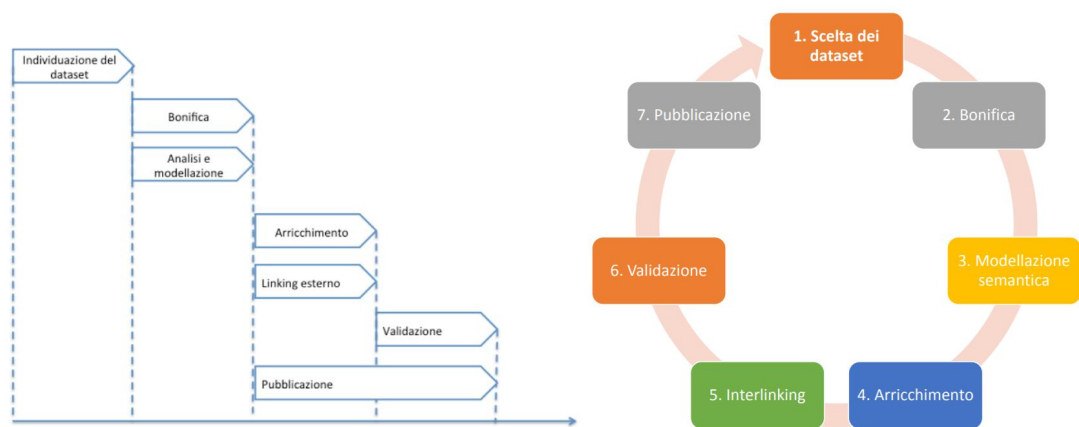


Fig. 4.8 Un altro modello per il workflow di pubblicazione dei *Linked Data*. A sinistra lo schema proposto dall’Agenzia per l’Italia Digitale (AGID) nel 2012²⁶⁰ e a destra quello di A. Iacono nel 2016²⁶¹.

Nel lavoro oggetto di questa tesi, l’intero processo si è svolto secondo le fasi riportate, in ordine d’esecuzione, al capitolo successivo.

Vengono descritti inoltre, nel dettaglio, alcuni aspetti tecnici e implementativi, indicando in che modo i dati sono stati modellati e rappresentati secondo le ontologie e fornendo alcuni dettagli sull’attività relativa alla conversione dei dati e al loro collegamento con i dataset esterni.

4.5 Mèmora

4.5.1 Descrizione del portale

Mèmora²⁶² è una piattaforma digitale pubblica, libera e gratuita per la descrizione del patrimonio culturale di archivi, istituti culturali e musei. Il progetto ha visto la migrazione in un unico applicativo di migliaia di oggetti digitali, schede e inventari di centinaia di enti.

²⁶⁰ <https://www.istat.it/it/files//2013/01/interoperabilitasemopendata.pdf> pag. 50

²⁶¹ in *Introduzione a Linked Open data e Web semantico*, pag. 119

²⁶² <http://www.memora.piemonte.it/>

Realizzato dal CSI Piemonte²⁶³, è basato su tecnologie open e gestisce attraverso un'unica interfaccia web beni culturali di tipologie diverse, integrando i molteplici software fino a ieri utilizzati. Mèmora supera così la divisione tra gli ambiti archivistico e museale, in favore di una visione complessiva del patrimonio, garantendone la conservazione e la fruibilità nel tempo.

Questo strumento digitale raccoglie e rende disponibili in un unico applicativo migliaia di documenti, fotografie, manifesti, video di centinaia di enti. Un ecosistema condiviso a livello regionale che a oggi conta 70.000 oggetti digitali, 500.000 schede descrittive di beni artistici e 500 inventari di archivi storici: un patrimonio destinato a crescere, giorno per giorno, per metterne in luce quantità, qualità e varietà.

Dal punto di vista tecnico Mèmora si basa sul software open source CollectiveAccess²⁶⁴, utilizzato da musei, archivi e biblioteche in tutto il mondo.

²⁶³ <http://www.csipiemonte.it/web/it/>

²⁶⁴ <https://www.collectiveaccess.org/>

5. Il Fondo antico del Comune di Novi Ligure

Prima di iniziare il lavoro sui dati, è stato necessario scegliere un archivio su cui procedere, tra quelli contenuti nel sistema Mèmora.

L'attività di individuazione dei dati infatti è il punto di partenza del processo di apertura e *linking*.

In questa fase emerge sia l'esigenza di selezionare un sottoinsieme dei dati che ha senso aprire e trasformare in *Linked Data* che la conseguente necessità di organizzare il lavoro di selezione seguendo dei criteri che *consentano di mediare tra la domanda di dati (interna o esterna) e lo sforzo richiesto per trattarli o la perdita di controllo che ne può derivare.*²⁶⁵

E' fondamentale anche considerare e i vincoli all'apertura dei dati presenti, sia da un punto di vista normativo che organizzativo, e valutare attentamente gli aspetti legati a privacy, proprietà intellettuale, licenza di rilascio ed eventuali limiti alla pubblicazione.²⁶⁶

E' inoltre fortemente raccomandato effettuare una fase di pre-analisi del dominio di riferimento in modo da acquisire informazioni preliminari utili per le fasi successive di trasformazione e identificare altri elementi-chiave per la misura della complessità.

Per ragioni di contenuto, di complessità e quantità dei dati, e di permessi relativi al download e alla pubblicazione dei dati, è stato selezionato dalla piattaforma Mèmora il Fondo antico del Comune di Novi Ligure.

5.1 Descrizione del Fondo

All'interno del sistema Mèmora, il fondo antico appartiene all'"Archivio storico del comune", nella cartella contenitore "Archivi e patrimonio documentale" dell'ente "Comune di novi Ligure".

²⁶⁵ <https://www.istat.it/it/files//2013/01/interoperabilitasemopendata.pdf> pag. 50

²⁶⁶ <https://docs.italia.it/italia/daf/lq-patrimonio-pubblico/it/bozza/aspecttiorg.html#linea-1-dati-nativi>

Contiene 3 sezioni, 70 serie, 83 sottoserie e un totale di 1641 unità (schede documentarie) che coprono un periodo che va dagli inizi del 1500 fino al 1870.

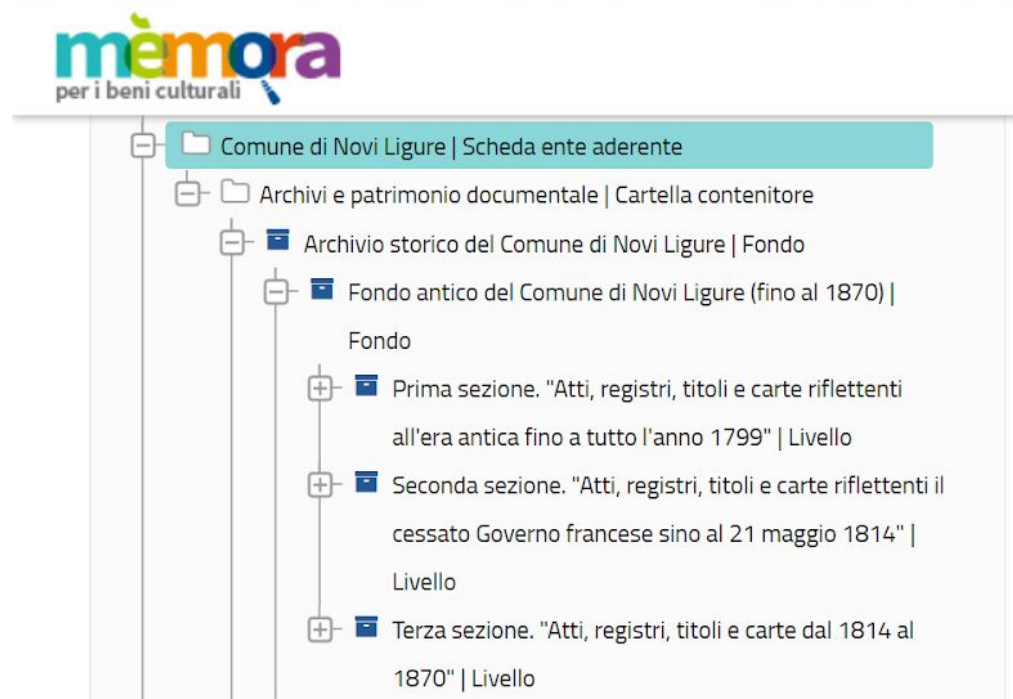


Fig. 5.1 Il Fondo antico del Comune di Novi Ligure nella gerarchia di Mémora.

Il fondo è suddiviso in tre livelli:

- Prima sezione. "Atti, registri, titoli e carte riflettenti all'era antica fino a tutto l'anno 1799"
- Seconda sezione. "Atti, registri, titoli e carte riflettenti il cessato Governo francese sino al 21 maggio 1814"
- Terza sezione. "Atti, registri, titoli e carte dal 1814 al 1870"

Il lavoro di conversione dei dati in LOD è stato eseguito sulla prima sezione del fondo, che contiene 18 serie e 223 unità che coprono il periodo 1135/1495 - 1799.

Questa prima sezione raccoglie al suo interno tutta la documentazione riferibile all'attività della Comunità novese fino al periodo della dominazione francese. In particolare sono presenti documenti in copia a partire dal 1135 e originali dal 1495 fino a tutto l'anno 1799, ma anche altre carte dal 1800 al 1811 poiché conservate in fascicoli prevalentemente riferibili a questa prima sezione.

La cesura fra la prima e la seconda sezione è determinata dall'adozione a metà Ottocento, come riferito nel resoconto delle vicende storiche dell'archivio comunale, del repertorio di ordinamento degli archivi allegato alle Istruzioni per l'amministrazione dei Comuni approvate da Sua Maestà il 1° aprile 1838, che prevede la suddivisione delle raccolte documentarie in tre serie sulla base del cambiamento di Governo (ovvero quello decorso prima dell'invasione francese; quello durante il governo francese; quello a cominciare dall'epoca della Restaurazione).


La documentazione, formata da fascicoli e registri di piccole e grandi dimensioni, è organizzata in 18 serie di ordinamento, per un totale di 223 schede di descrizione.

- La serie 1 contiene due copie degli statuti stampati in Alessandria nel 1605; le due copie si differenziano per legatura e formato e per la presenza sulla copia non rifilata nella rilegatura di numerose annotazioni a margine.
- La serie 2 contiene alcuni degli atti più interessanti e antichi per la storia della città; si tratta delle convenzioni con Genova del 1135, 1157, 1447 e 1756 e degli atti riguardanti le gabelle della macina e della carne, i mulini, gli antichi privilegi e le fiere cittadine; comprende anche l'importante saggio storico di Ercole Spinola.
- La serie 3 raccoglie i tre volumi in cui dalla seconda metà del Seicento alla fine del Settecento sono stati trascritti i decreti, le sentenze e gli atti più significativi riguardanti Nove e il suo territorio.
- La serie 4 contiene i registri degli atti consolari dal 1495 al 1800 e benché siano presenti numerose e vaste lacune (1519-1523, 1531-1561, 1595-1676, 1686-1745, 1749-1751), questa è una delle raccolte documentarie più importanti e meritevoli di studio.
- La serie 5 contiene gli ordini del Governo di carattere civile e militare.
- La serie 6 contiene la corrispondenza con le autorità dal 1795 e una piccola raccolta di lettere diverse di fine Cinquecento.
- La serie 7 contiene i libri mastri e dei conti dal 1535.
- La serie 8 contiene il carteggio e i registri riguardanti le imposte e le tasse.
- La serie 9 contiene una raccolta eterogenea di documenti riguardanti gli amministratori, le feste nazionali e le attività economiche, fra cui alcune grida del Cinquecento e del Seicento di notevole rilievo.

- La serie 10 contiene carteggio di fine Settecento riguardante la sanità e l'istruzione.
- La serie 11 contiene i materiali di carattere religioso fra cui il carteggio riguardante la soppressione dei monasteri; in questa serie sono stati inseriti anche due documenti in fotocopia, conservati in originale nelle Parrocchie della città, poiché ritenuti di eccezionale interesse e in modo da garantirne la consultazione.
- La serie 12 contiene i processi civili e criminali dal 1558 al 1696.
- La serie 13 contiene le contravvenzioni ai bandi campestri a partire dal 1602.
- La serie 14 contiene la documentazione sulle somministrazioni militari alle truppe francesi, austriache e russe e le note del loro passaggio e stazionamento.
- La serie 15 contiene il raro carteggio ancora esistente sui lavori pubblici alle strade e agli edifici.
- La serie 16 contiene alcuni registri e vari atti che tramandano le vicende del Monte di pietà, istituito in Novi nel 1611.
- La serie 17 contiene alcuni documenti privati di cittadini novesi.
- La serie 18 contiene quanto è rimasto del fondo notarile di cui si è trattato nella relazione storica all'archivio.

In sintesi, la Prima sezione dell'archivio storico comunale contiene quanto è rimasto della documentazione antica, che fino a metà Ottocento comprendeva molte pergamene e moltissimi altri documenti, la cui memoria ci è stata tramandata dalle relazioni di Omobono Sarti, Giovanni Barbieri, Ettore Rasponi e Vittorio Bozzola che in passato hanno avuto cura delle raccolte.

Il contenuto del fondo è eterogeneo: le risorse variano nella loro tipologia (volumi, fascicoli, quaderni, opuscoli, filze, rubriche, fogli, ecc...).



Convenzione con la città di Genova del 1135

SEGNATURA DEFINITIVA	A.S.N.L.3
UNITÀ DI CONSERVAZIONE	Numero: 2.2
IDENTIFICATIVO D'ORIGINE	221
SOGGETTO CONSERVATORE	Novi Ligure. Archivio storico
CRONOLOGIA E ALTRE DATAZIONI	Definizione cronologica: 1702
	Datazione: 1702
DATA TOPICA	Genova

*Copia della convenzione seguita mediante instrumento a rogito Pellegrino Bono fra i rappresentanti la Città ed il Popolo di Novi, ed il Legato del Popolo di Genova Bono Giovanni Capriato, che egli convocava nella Basilica del Beato Nicolò allora di recente edificata, colla quale convenzione, essi rappresentanti, confermarono e si obligarono di osservare i patti, già anteriormente stabiliti coi genovesi relativamente alla cessione del Castello di Novi in proprietà alla Chiesa di San Siro di Pavia, ed alla Repubblica di Genova, di conservarlo fedelmente per conto della medesima, di pagarne ad entrambe un annuo censo, di far con esse guerra contro i Tortonesi, e di non concluder pace, o far guerra con essi senza il consentimento dei Consoli di Genova e di Pavia, di coadiuvare finalmente con tutta la loro gente i Pavesi e i Genovesi, ma non però contro il Comune di Marinis, né contro la terra del marchese Rainerio, o del marchese Anselmo del Bosco" (copia dell'atto datato 5 gennaio 1135 estratto dall'Archivio Segreto di Genova il 20 marzo 1702)

DESCRIZIONE	DESCRIZIONE ESTRINSECA	1 fascicolo, lingua Latino, stato di conservazione buono
	Consistenza:	1
	Tipologia:	fascicolo
	LINGUA	LAT Latino
	STATO DI CONSERVAZIONE	Stato di conservazione: buono
SITO WEB		
SOGGETTI (ENTI, PERSONE, FAMIGLIE) COLLEGATE		
CONDIZIONI DI ACCESSO E RESTRIZIONI		
PROFILO DI ACCESSO		
IDENTIFICATIVO D'ORIGINE		
IDENTIFICATIVO DELLA SCHEDA		
COMPILAZIONE E CONTROLLO		

Fig. 5.2 esempio di una scheda di catalogo (unità documentaria) del portale Mèmora

5.2 Esportazione e pulizia dei dati da Mèmora

5.2.1 Individuazione delle entità

In primo luogo è stato necessario individuare delle entità di interesse su cui lavorare: la scelta si è basata considerando quali entità sono definite nelle unità documentarie dell'archivio, in quale quantità e frequenza ricorrono e quante di esse possono essere successivamente recuperate da dataset esterni.

Le entità individuate sono:

- Oggetto (la singola unità documentaria)
- Consistenza fisica (es. registro, fogli, fascicolo, quaderno, ...), relativa all'oggetto
- Lingua (italiano, latino, francese, ...), relativa all'oggetto
- Persone
- Date
- Luoghi

Sulla base di queste entità si sono delineati fin da subito i dataset esterni a cui collegarle: per esempio, un dataset authority per le persone è sicuramente VIAF, mentre le date possono essere recuperate da DBpedia.

I dataset effettivamente selezionati e utilizzati si trovano elencati al cap. [5.4.2](#)

5.2.2 Recupero dei dati: esportazione dal portale

Una volta definiti questi parametri, il lavoro pratico vero e proprio è iniziato partendo dall'esportazione delle unità documentarie della prima sezione dell'archivio.

La scelta del formato adatto per l'esportazione è una premessa necessaria: tra i 12 esportatori disponibili su Mèmora (vedi [cap. 2.8.1](#)) si è proceduto con una valutazione di ciascuno, considerando quali di essi riportano le informazioni (le entità) d'interesse su cui lavorare.

In ogni caso, il documento scaricato è sempre un file XML, strutturato a seconda del formato di esportazione selezionato.

I formati scelti sono:

- MINI-icar-san, che mappa e rende interoperabili al suo interno altri tre standard fondamentali per la definizione di metadati archivistici:
 - EAD3, per informazioni sui livelli gerarchici e relazioni degli oggetti (es. la consistenza fisica).
 - EAC-CPF per la descrizione dei soggetti produttori
 - SCONS per la descrizione dei soggetti conservatori

- scheda OA²⁶⁷, l'unico formato in cui è rintracciabile l'entità "luogo".

²⁶⁷ <http://docenti.unimc.it/pierluigi.feliciati/teaching/2014/12730/files/scheda-oa-obbligatorieta>

```

1 | <?xml version="1.0" encoding="utf-8"?>
2 | <icar-import xmlns:xlink="http://www.w3.org/1999/xlink" xmlns:icar-import="http://www.sar
3 | <header>
4 |   <systemId>Memora</systemId>
5 |   <systemTitle>La piattaforma digitale per i beni culturali piemontesi</systemTitle>
6 |   <contact>
7 |     <name>Csi Piemonte - Assistenza Cultura Gruppo</name>
8 |     <url>http://www.csipiemonte.it/web/it/</url>
9 |     <mail>assistenza.cultura@csi.it</mail>
10 |    <phone>telefono gruppo assistenza cultura</phone>
11 |   </contact>
12 |   <event eventType="other" eventDate="13 Apr 2019">
13 |     <agent id="export" agentType="software">Memora</agent>
14 |   </event>
15 |   <fileDesc>
16 |     <title>Esportazione dati per tracciato icar</title>
17 |     <abstract>Questo file xml contiene i dati per il tracciato icar</abstract>
18 |     <date>13 Apr 2019</date>
19 |   </fileDesc>
20 | </header>
21 | <listRecords>
22 |   <record>
23 |     <recordHeader action="insert" cascade="true" groupEad="single" type="ead3">
24 |       <id>ICAR-EAD-001</id>
25 |       <lastUpdate/>
26 |       <recordBody>
1028 |
1029 |     </recordHeader>
1030 |     <record action="insert" cascade="true" type="eac-cpf">
1031 |       <recordHeader/>
1032 |       <id>ICAR-EAC-CPF-001</id>
1033 |       <lastUpdate/>
1034 |       <recordBody>
4118 |
4119 |     </record>
4120 |     <record action="insert" cascade="true" type="scons">
4121 |       <recordHeader/>
4122 |       <id>ICAR-SCONS-001</id>
4123 |       <lastUpdate/>
4124 |       <recordBody>
5138 |
5139 |     </record>
5140 |   </listRecords>
5141 | </icar-import>

```

Fig. 5.3 Esempio di un'unità documentaria esportata nel formato MINI-icar-san (file XML): il tracciato raggruppa e rende interoperabili gli standard EAD, EAC.CPF e SCONS.

La seguente tabella riporta quanti e quali tag fanno riferimento alle entità di interesse e in quali dei formati di esportazione sono stati identificati, motivando così la scelta dei formati stessi per il download.

Entità	Tracciato da cui è stata estratta	Metadato	Tag
Nome dell'oggetto	MINI-icar-san	Titolo o denominazione	unittitle
Consistenza fisica	MINI-icar-san	Consistenza fisica - tipologia	unittype
Lingua	MINI-icar-san	estratto da: descrizione estrinseca	physdesc
Persone	MINI-icar-san		Estrate dal titolo con un tool di NER
Data	MINI-icar-san	Datazione / definizione cronologica	Estremi cronologici datesingle e note alla data datesingle - localtype
Luogo	scheda OA	Data topica	ADT ("altre datazioni") ²⁶⁸

Tab. 5.1: entità, formati e tag corrispondenti.

L'esportazione manuale delle 223 unità rispecchia, nella suddivisione in cartelle dell'output, la gerarchia delle 18 serie della sezione.

Oltre alle schede descrittive delle unità, sono state esportate anche eventuali immagini dell'oggetto materiale (figura sotto) e le schede descrittive dei livelli (i.e. delle serie).

La struttura del file XML di ognuna delle serie comprende anche la descrizione delle singole unità contenute, ciascuna racchiusa in un tag `<c0 level="Unità archivistica">`, nella parte di documento destinata alla definizione del tracciato EAD.

²⁶⁸ <http://www.iccd.beniculturali.it/getFile.php?id=3319> pag. 8



Fig. 5.4 esempi di immagini digitali degli oggetti materiali dell'archivio. In particolare queste unità documentarie appartengono tutte alla serie 2.

5.2.3 Conversione e pulizia dei dati

L'attività di conversione vera e propria, a partire dai file di dati estratti dal database, consiste nel recuperare le singole informazioni (entità) per poter creare a partire da esse, nella fase successiva di modellazione, degli *statement* RDF che le descrivano in base al modello ontologico prescelto.

Il primo passo è stato quello di marcare le entità d'interesse dai file XML estratti e memorizzarle in un file Excel.

Ciò è stato fatto trasformando i file dall'XML al CSV con Google Refine e aggiustando manualmente eventuali imprecisioni.

Google Refine

Anche conosciuto come Open Refine, è uno strumento *open source* per la pulizia, l'analisi e l'elaborazione di dati. Può operare su diversi tipi di dati in ingresso e offre un piccolo supporto alla funzione di *record linkage* grazie a funzionalità di espansione semantica.

Per la pubblicazione di dati LOD è necessario lavorare con un *plugin*²⁶⁹ per Refine che permette di fare il link automatico delle risorse a nodi esterni (estensione RDF).

²⁶⁹ <https://github.com/fadmaa/grefine-rdf-extension>

Oltre ai collegamenti con dataset esterni ([cap. 5.4.2](#)), Refine è stato utilizzato anche in questa fase di conversione di dati dall'XML al CSV: selezionando manualmente i nodi (tag) d'interesse, il programma crea in maniera automatica il formato tabulare su cui lavorare successivamente.

Sulla base delle entità marcate è stato creato un unico file Excel (vedi fig. sotto), relativo all'intera sezione, contenente informazioni rilevanti (es. la serie d'appartenenza, il titolo, la lingua, la datazione, ecc...) per tutte le 223 unità (ciascuna marcata da un identificatore univoco) e le relative entità.

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	id_ogg	serie	titolo	consistenza_fisica	lingua	esterni_cronologici	note_alla_data	luogo	ner_persone	ner_persone2	ner_persone3	ner_luoghi	ner_luoghi2
2	1	1	Statutorum civitum insignis oppidi Novarum domini ge volume		Latino, Italiano	1605							
3	2	1	Statutorum civitum insignis oppidi Novarum domini ge volume		Latino, Italiano	1605							
4	3	2	Convenzione con la città di Genova del 1135	fascicolo	Latino	1702		Genova					
5	4	2	Convenzione con la città di Genova del 1157	fascicolo	Latino	1702		Genova					
6	5	2	Convenzione con la città di Genova del 1447	fascicolo	Latino	1601 - 1800	XVII sec - XVIII sec.	Genova					
7	6	2	Molino di Busseto	fascicolo	Latino, Italiano	1500 - 1510	1500. Inizio XVI sec.					Busseto	
8	7	2	Disposizioni genovesi per l'amministrazione cittadina	registro		1527 - 1756	1527 - 1756						
9	8	2	Orazione di M. Lorenzo Capelloni al Serenissimo Prin	quaderno		1549	1549	Lorenzo Capelloni				Spagna	
10	9	2	Raccolta di atti antichi	quaderno	Latino, Italiano	6 June 1604	1604 giugno 06						
11	10	2	Vertenza della Terra di Novi riguardante le gabelle di ri	registro		30 August 1640	1640 agosto 30						
12	11	2	Cabella della macina e della carne	opuscolo		1667	1667						
13	12	2	Saggio storico della città di Nove del dottore Ercole Sp			1701 - 1800	XVIII sec.	Ercole Spinola					
14	13	2	Dritti e privilegi del Comune di Novi, e cambiamenti di	registro		23 June 1756	1756 giugno 23						
15	14	2	Convenzione con la città di Genova del 1756	fascicolo	Latino	13 September 1756	1756 settembre 13					Genova	
16	15	2	Acque	fascicolo		1774	1774						
17	16	2	Acque. Mulini del Capitolo del Comune	fascicolo		1779	1779						
18	17	2	Catastro. Memoria senza data	fascicolo		1801 - 1810	Inizio XIX sec.						
19	18	2	Storia dei molini per la Scivia dal 1356 al 1779	fascicolo		1801 - 1899	XIX sec.						
20	19	3	Libro dei decreti	registro		1641 - 1660	Metà circa XVI sec.						
21	20	3	Libro de decreti. Trascrizione di atti dal 1388 al 1796	registro	Latino, Italiano	1751 - 1800	Seconda metà XVIII sec.						
22	21	3	Sentenze decreti dei duchi di Milano e dei supremi ma	registro	Latino, Italiano	1752 - 1800	Seconda metà XVIII sec.	duchi di Milano					
23	22	4	Registro degli atti consulari dal 8 dicembre 1495 a tutt	registro	Latino	8 December 1495 - 1518	1495 Dicembre 08 - 1518						
24	23	4	Registro degli atti consulari dal 24 ottobre 1524 al 1531	registro	Latino	24 October 1524 - 1530	1524 Ottobre 24 - 1530						
25	24	4	Registro degli atti consulari dal 12 febbraio 1562 a tutt	registro	Latino	12 February 1562 - 1579	1562 Febbraio 12 - 1579						
26	25	4	Registro degli atti consulari dal 17 agosto 1579 a tutto	registro	Latino	17 August 1579 - 1594	1579 Agosto 17 - 1594						
27	26	4	Registro degli atti consulari da 25 luglio 1677 a tutto	registro		25 June 1677 - 1685	1677 Giugno 25 - 1685						
28	27	4	Deliberationum Magnifici Consilii et Magnifici Deput	registro		4 May 1746 - 1748	1746 Maggio 04 - 1748						
29	28	4	Registro degli atti consulari dal 9 luglio 1752 a tutto	registro		9 July 1752 - 1756	1752 Luglio 09 - 1756						

Fig. 5.5 File excel contenente tutte le informazioni rilevanti sulle unità documentarie e le entità d'interesse, creato come descritto al [cap. 5.2.3](#).

Ogni unità è identificata da un identificatore univoco, per un totale di 223 id.

Le colonne che riportano la dicitura "ner" nell'intestazione contengono le entità estratte dal tool di *Named Entity Recognition* (paragrafo seguente).

5.3 Arricchimento con entità estratte dal titolo delle opere attraverso un tool di Named Entity Recognition

Alcune entità, come per esempio le persone, sono state ricavate precedentemente estraendole dai titoli degli oggetti tramite un *tool online* di *Named Entity Recognition*, Dandelion API²⁷⁰ Entity Extraction.

Questo strumento è stato usato per estrarre i nomi delle persone dai titoli degli oggetti, nonché per estrarre altre informazioni e concetti interessanti circa gli oggetti stessi.

In alcuni casi dove l'entità "luogo" non era stata definita per un oggetto, essa è stata estratta sempre dal titolo tramite questo tool.

Dando in input al programma il titolo dell'oggetto, esso restituisce come output le entità riconosciute.

Il programma è in grado di riconoscere e classificare entità quali persone, opere, organizzazioni, luoghi, eventi e concetti.

²⁷⁰ <https://dandelion.eu/semantic-text/entity-extraction-demo/>

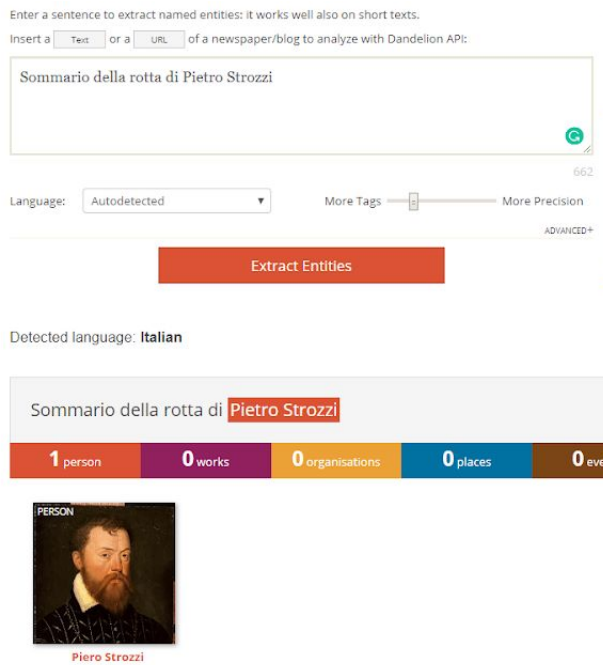
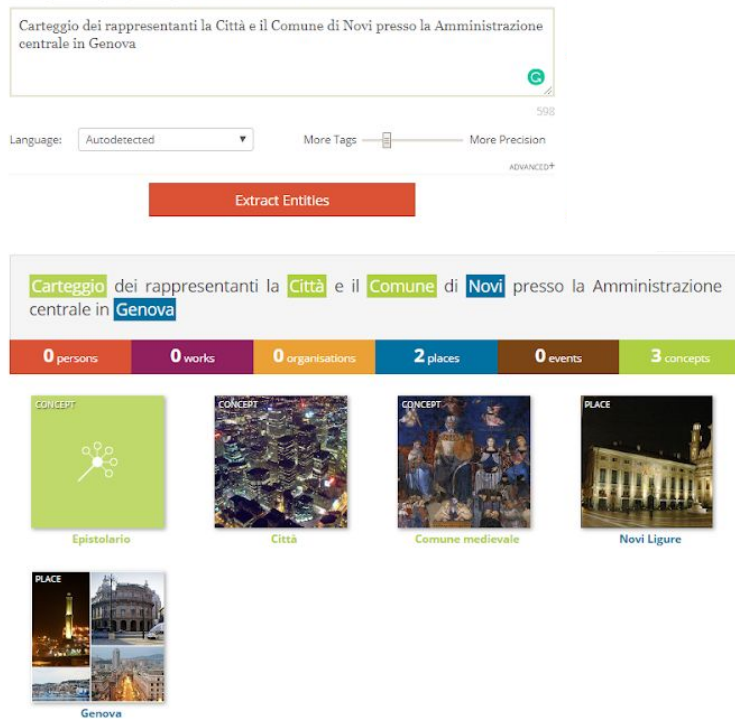


Fig. 5.6 e 5.7 due esempi di Named Entity Recognition con Dandelion API.



E' possibile regolare la quantità di tag estratti, a scapito della precisione nel riconoscimento delle entità, o viceversa.

Naturalmente, se si desidera un riconoscimento meno preciso sarà fondamentale un lavoro manuale di disambiguazione anche in questa fase.

5.4 Costruzione delle tabelle e import in un database relazionale

5.4.1 L'ambiente di sviluppo

L'ambiente di sviluppo utilizzato per la realizzazione del database è XAMPP²⁷¹, una piattaforma software che mette a disposizione un Server Apache HTTP²⁷² e tutti gli strumenti necessari per utilizzare i linguaggi di programmazione PHP e Perl.

Per gestire e memorizzare i dati raccolti è stato usato il DBMS (Database Management System) MySQL²⁷³, in quanto offre prestazioni veloci, elevata affidabilità, facilità d'uso ed è la soluzione standard adottata da XAMPP.

E' stato creato un primo database in MySQL denominato `fondo_antico_novi_ligure`, ed è stato organizzato, sulla base delle entità d'interesse, secondo le seguenti tabelle:

- persone
- luoghi
- oggetti
- date

Ogni persona, luogo, oggetto e data ha un identificatore univoco, necessario per costruire successivamente le tabelle mapping (cap. 5.4).

²⁷¹ <http://www.apachefriends.org/>

²⁷² <http://httpd.apache.org/>

²⁷³ <http://www.mysql.it/>

5.4.2 Scelta dei dataset esterni

La scelta dei dataset verso cui indirizzare i link esterni del Fondo dipende dal tipo di informazioni che è necessario rappresentare.

La tabella seguente riporta i dataset selezionati a seconda del tipo di entità da descrivere:

Entità	Dataset esterno
Persone	GND
	DBpedia (it)
	VIAF
Luoghi	GND
	DBpedia (it)
	Wikidata (en)
Date	DBpedia (it)
Oggetti	DBpedia (it)

Tab. 5.2 Dataset esterni da cui sono state recuperate le entità

DBpedia

Il dataset di DBpedia²⁷⁴, nato nel 2007, rappresenta uno dei punti di interconnessione più importanti del *Web of Data* e consiste in una gigantesca sorgente di dati ottenuti estraendo informazioni strutturate dalla più grande enciclopedia esistente, ovvero Wikipedia, e rendendole accessibili nel Web.

L'idea alla base di DBpedia è quindi quella di estrarre automaticamente informazioni da Wikipedia e di strutturarle in RDF. Per fare ciò, DBpedia sfrutta la conoscenza contenuta nelle tabelle di Wikipedia, definite utilizzando particolari template che permettono facilmente l'estrazione automatica dei dati.

²⁷⁴ <https://wiki.dbpedia.org/>

Alle risorse che sono oggetto di un articolo di Wikipedia viene automaticamente assegnato un URI DBpedia, basato sull'URI di quell'articolo di Wikipedia.

Il dataset raccoglie dati relativi a milioni di oggetti di vario tipo (persone, luoghi, album musicali, malattie, ecc.) descritte secondo una specifica ontologia tramite quasi 700 milioni di triple RDF estratte dalle varie edizioni, in diverse lingue, di Wikipedia.

Il *knowledgebase* messo a disposizione da DBpedia copre, pertanto, diversi domini, raccoglie informazioni in diverse lingue e viene costantemente aggiornato in modo che si evolva automaticamente in base ai cambiamenti di Wikipedia.

La conoscenza contenuta di DBpedia quindi è di tipo "non settoriale" (*cross-domain*) e proprio questo gli ha permesso di diventare l'*hub* del *Linked Open Data cloud*,²⁷⁵ in quanto altre sorgenti possono riutilizzare e linkare i concetti di senso in essa contenuti.

La copertura tra domini è fondamentale per aiutare a connettere set di dati specifici per un certo dominio in uno spazio dati interconnesso, evitando in tal modo la frammentazione del Web dei dati in "isole" di dati.

Data la vasta copertura di domini e l'elevato grado di sovrapposizione delle informazioni con altri dataset già presenti nel web, un numero sempre maggiore di sorgenti di dati hanno iniziato a creare dei collegamenti, tramite link RDF, con il dataset di DBpedia, rendendolo uno dei punti centrali di intercollegamento all'interno del Web of Data.

L'utilizzo delle risorse definite in DBpedia consente automaticamente di linkare il proprio dataset ad una miriade di dati già presenti nella LOD cloud.

I dati di DBpedia possono essere interrogati tramite query SPARQL facendo riferimento ad uno specifico endpoint che permette l'interrogazione sia tramite librerie di programmazione che tramite applicazioni web messe a disposizione per gli utenti.

²⁷⁵ DBpedia ha funzionato come hub all'interno del *Web of Data* fin dalle prime fasi del progetto *Linking Open Data*.

Nel dataset LOD oggetto di questo lavoro di tesi, per la creazione di link esterni è stata usata in particolare DBpedia-it²⁷⁶, ovvero la controparte italiana di DBpedia contenente i dati estratti da Wikipedia Italia.

VIAF

Virtual International Authority File (VIAF)²⁷⁷ costituisce una base dati di voci di autorità controllate (cap. 2.3.1) provenienti da diversi cataloghi nazionali. L'obiettivo del progetto è unire le singole voci di autorità nazionali in un'unica voce virtuale e collegare così record coincidenti provenienti dai diversi dataset.

VIAF
Virtual International Authority File

Suchen

Feld auswählen: Index auswählen: Suchbegriffe: **Suchen**

Cassirer, Ernst, 1874-1945

Cassirer, Ernst

1945-1874

Cassirer, Ernst German philosopher, 1874-1945

VIAF ID: 106965171 (Personal)
Permalink: <http://viaf.org/viaf/106965171>

Vorzugsbezeichnungen

- 100 1 _ [ta_Cassirer_Ernst_td_1874-1945](#)
- 100 1 0 [ta_Cassirer_Ernst](#)
- 100 1 _ [ta_Cassirer_Ernst_td_1874-1945](#)
- 100 1 _ [ta_Cassirer_Ernst](#)
- 200 _ | [ta_Cassirer_tb_Ernst_tf_1874-1945](#)
- 100 1 _ [ta_Cassirer_Ernst_td_1874-1945](#)
- 100 1 _ [ta_Cassirer_Ernst_td_\(1874-1945\)](#)
- 100 1 _ [ta_کاسیرر، ارنست، td_1874-1945](#)
- 100 1 [ta_Cassirer_Ernst_td_1874-1945](#)

Fig. 5.8 esempio della voce “Ernst Cassirer” (entità-persona) in VIAF: immagine tratta dalla rispettiva voce Wikipedia²⁷⁸

²⁷⁶ <https://it.dbpedia.org/sparql/>

²⁷⁷ <http://viaf.org/viaf/>

²⁷⁸

https://it.wikipedia.org/wiki/Virtual_International_Authority_File#/media/File:VIAF_Screenshot_2012.png

Nel VIAF così come, per esempio, nel progetto ISNI, non esiste una forma unica, universalmente riconosciuta come “autorizzata” o “uniforme” del nome di un’entità, né una forma che prevalga sulle altre, ma il criterio di identificazione dell’entità “persona” attraverso il nome e altri attributi è assolutamente democratico, con differenti forme equivalenti provenienti dalle varie agenzie nazionali che partecipano ai contenuti; tutte le forme attraverso le quali può essere rappresentato il dato ricercato, sono collegate tra loro e associate ad un medesimo identificativo.²⁷⁹

Wikidata

Wikidata²⁸⁰ è un database libero, collaborativo, multilingue e secondario che raccoglie dati strutturati per fornire supporto a Wikipedia, a Wikimedia Commons, agli altri progetti del movimento Wikimedia e a chiunque.

E’ stato utilizzato in questo caso per collegamenti sulle entità-luogo, essendo ritenuto molto utile in fase di arricchimento del dataset del Fondo antico di Novi Ligure: infatti, a partire da un luogo nel dataset, i link esterni a Wikidata permettono di raggiungere informazioni aggiuntive, non presenti in origine nel dataset (vedi cap. 5.4.2).

GND

Il *Gemeinsame Normdatei* (GND)²⁸¹ è un sistema di controllo di autorità (cap. 2.3.1) gestito dalla Biblioteca Nazionale Tedesca in collaborazione con altre biblioteche. Viene utilizzato principalmente dalle biblioteche, ma in maniera crescente anche da archivi, musei e per applicazioni in rete.

I dati del GND agevolano la procedura di catalogazione, consentono approcci di ricerca univoci e la possibilità di collegamento a differenti fonti di informazione. Inoltre fanno parte del sistema VIAF, nel quale il GND viene collegato virtualmente con altri controlli di autorità a livello internazionale.

Il GND pone gratuitamente a disposizione i dati in formato RDF con licenza CC0.

²⁷⁹ R. Delle Donne, T. Possemato, *SHARE Catalogue: un’esperienza di cooperazione*, in *Biblioteche Oggi*, vol. 35, 2017, pag. 24

²⁸⁰ <https://www.wikidata.org/wiki/>

²⁸¹ <https://lobid.org/gnd>

5.4.3 Collegamenti ai dataset esterni

Le tabelle persone, luoghi, date e oggetti sono state importate separatamente su Google Refine e arricchite con i link a dataset esterni.

E' stato possibile recuperare le entità dai dataset esterni grazie ai *reconciliation services* ("servizi di riconciliazione") aggiunti su Google Refine.

Sono questi servizi che permettono di estrarre i link relativi a ciascun record delle tabelle.

DBpedia è l'unico dataset ad avere un motore SPARQL, perciò l'aggiunta di questo servizio è stata semplice ed immediata.

Per gli altri dataset (Wikidata, VIAF, GND) è stato invece necessario recuperare dal web dei link a degli script seguenti per poterli utilizzare come *reconciliation services* su Google Refine.

- GND: <https://lobid.org/gnd/reconcile> ²⁸²
- Wikidata (en): <https://tools.wmflabs.org/openrefine-wikidata/en/api> ²⁸³
- DBpedia (it): <http://it.dbpedia.org/sparql/> (Virtuoso)²⁸⁴
- VIAF: <http://refine.codefork.com/reconcile/viaf> ²⁸⁵

La presenza di collegamenti esterni permette, ad esempio, di navigare da un dataset all'altro, ottenendo la descrizione di una stessa risorsa sotto diversi punti di vista, in quanto le diverse sorgenti di dati possono contenere informazioni completamente diverse e nuove relativamente ad uno specifico oggetto.

Ciò aiuta anche a garantire che i dati possano essere scoperti dai browser RDF e dai *crawler*, e supporta così la scoperta di nuova conoscenza.

²⁸² fonte: <http://blog.lobid.org/2018/08/27/openrefine.html>

²⁸³ fonte: <https://tools.wmflabs.org/openrefine-wikidata/en/api>

²⁸⁴ fonte: https://www.youtube.com/watch?v=99mRGIx4_QI

²⁸⁵ fonte: <https://libjohn.github.io/openrefine/hands-on-reconciliation.html>

Per fare un esempio, le risorse che fanno riferimento a dei luoghi presentano dei link in uscita verso le risorse che rappresentano gli stessi luoghi all'interno di altri dataset. Ciò permette, seguendo questo collegamento, di ottenere maggiori informazioni sul luogo stesso: in DBpedia, ad esempio, si potranno reperire informazioni generali come la descrizione piuttosto che la popolazione o ancora l'URI di altre risorse che rappresentano individui nati in quella città, mentre da Wikidata è possibile ottenere anche immagini, coordinate, ecc...

I collegamenti RDF consentono così di impostare riferimenti da un set di dati a entità descritte in un altro, che possono, a loro volta, avere descrizioni che fanno riferimento a entità in un terzo set di dati e così via. Pertanto, l'impostazione dei collegamenti RDF non solo collega un'origine dati a un'altra, ma abilita le connessioni in una rete potenzialmente infinita di dati che possono essere utilizzati collettivamente dalle applicazioni client.

La presenza di collegamenti esterni permette inoltre di arricchire il dataset tramite dati che originariamente non sono disponibili, lasciando agli utilizzatori di tali informazioni la possibilità di creare applicazioni o forme di utilizzo che individuino e sfruttino il valore aggiunto ottenibile tramite l'intreccio dei dati.

Questo è perfettamente in linea con l'approccio alla conoscenza nel web semantico e conforme ai principi dei *Linked Data*.

5.4.4 Disambiguazione

Una volta estratti i link è stato necessario verificarne la correttezza e, in caso necessario, apportare delle modifiche manualmente. Questa attività è detta di *disambiguazione*: consiste nell'aiutare manualmente il servizio di riconciliazione a *matchare* correttamente le entità e i rispettivi link esterni.

Il mapping ambiguo (riferito a collegamenti con più termini) viene separato da quello non ambiguo e viene corretto.

La riconciliazione automatica infatti può creare delle ambiguità sui link da associare ad una determinata entità: per esempio l'entità-luogo "Genova" può essere correlata, con lo stesso peso e nello stesso dataset, a più link, uno riferito alla risorsa "città di Genova", uno alla "Repubblica di Genova" e addirittura un terzo alla "provincia di Genova".

Per eliminare ambiguità come queste e selezionare solo il link corretto cui *matchare* l'entità, è necessario l'intervento manuale.



Fig. 5.9 un esempio di disambiguazione in Google Refine: più link DBpedia possono essere associati alla proprietà lingua "latino". L'intervento manuale è necessario per eliminare l'ambiguità e matchare il link corretto.

Google Refine permette di controllare la qualità dei *match* assegnando un punteggio di riconciliazione.

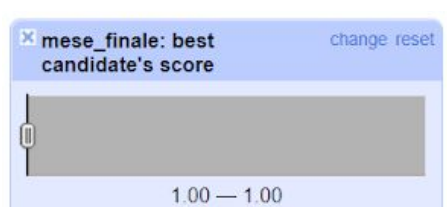


Fig. 5.10 esempio di punteggio di riconciliazione in Google Refine: in questo caso il match è perfetto e quindi il punteggio è 1. Refine permette comunque di controllare e/o modificare manualmente i *match* anche in questo caso.

Una volta terminata la riconciliazione per ogni tabella, i link esterni ottenuti, che costituiscono i collegamenti del nuovo dataset LOD, sono di tipo:

- *Identity links*, cioè collegamenti che puntano ad URI utilizzati da altre fonti di dati per identificare lo stesso oggetto del mondo reale o concetto astratto. Infatti, per monitorare i diversi fornitori di informazioni parlano della stessa entità, il paradigma LOD si basa sull'impostazione di collegamenti RDF tra alias URI. Di comune accordo, i *publisher* di *Linked Data* utilizzano la proprietà `owl:sameAs` per indicare che due alias URI si riferiscono alla stessa risorsa.

Esempio:

<u>Property</u>	<u>Value</u>
<code>foaf:name</code>	Paganino Capurri
<code>owl:sameAs</code>	<code><https://lobid.org/gnd/1151967343></code>
<code>owl:sameAs</code>	<code><https://viaf.org/viaf/6829151837996120520000></code>

- *Relationship links*, cioè collegamenti che puntano a risorse correlate in altri dataset.

Esempio:

<u>Property</u>	<u>Value</u>
<code>rdfs:label</code>	Convenzione con la città di Genova del 1756
<code>dc:language</code>	<code><http://it.dbpedia.org/property/latino></code>
<code>dcterms:medium</code>	<code><http://it.dbpedia.org/property/fascicolo></code>

- *Vocabulary links*, cioè collegamenti che puntano dai dati alle definizioni dei termini di vocabolario utilizzate per rappresentarli.

Esempio:

<u>Property</u>	<u>Value</u>
<code>rdfs:label</code>	Convenzione con la città di Genova del 1756
<code>rdf:type</code>	<code>pico:beni_archivistici</code>

Successivamente, per ogni colonna (in ogni tabella) contenente entità matchate con link esterni, è stata aggiunta con Google Refine una nuova colonna contenente gli URL dei link esterni.

Queste nuove colonne sono identificate dalla seguente sintassi:

`"datasetEsterno_link_nomeColonnaRiferimento"`

▼ anno	▼ DBpedia_link_anno	▼ mese	▼ DBpedia_link_mese
1555 Choose new match	http://it.dbpedia.org/resource/1555	Luglio Choose new match	http://it.dbpedia.org/resource/Luglio
1559 Choose new match	http://it.dbpedia.org/resource/1559	Settembre Choose new match	http://it.dbpedia.org/resource/Settembre

Fig. 5.11 esempio di colonne contenenti link. La colonna `anno` avrà la corrispondente colonna `DBpedia_link_anno` contenente tutti gli URL collegati alle entità.

I valori da inserire in queste nuove colonne sono ottenuti automaticamente scrivendo la seguente espressione in GREL (Google Refine Expression Language), che restituisce il valore del *match* (in questo caso proprio il valore dell'URI):

```
cell.recon.match.id
```

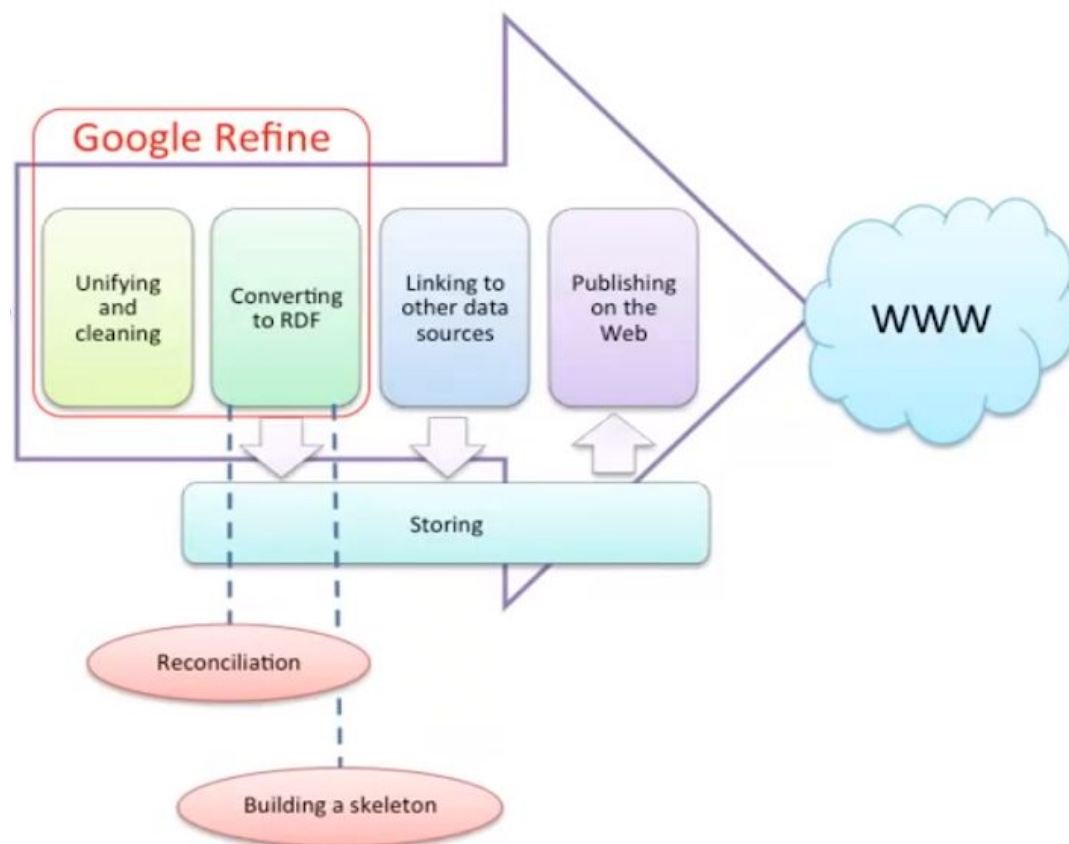


Fig. 5.12 Processo completo di creazione di *Linked Data* con Google Refine²⁸⁶. Nel caso di questo lavoro di tesi, la fase “Building a skeleton” è stata sostituita con le attività svolte tramite la piattaforma D2RQ (cap. 5.5.1).

5.4.5 Arricchimento del dataset: metadattazione e inferenza

In questa fase i dati sono stati quindi arricchiti attraverso l'esplicitazione di informazioni di contorno che ne semplificano il riutilizzo (*metadattazione*)²⁸⁷ e attraverso la derivazione di contenuto informativo aggiuntivo tramite l'utilizzo di tecniche di estrazione automatica dell'informazione o di ragionamento automatico (*inferenza*).

²⁸⁶ Youtube, *A complete example of how to link data, part 2*

https://www.youtube.com/watch?v=B5CsSE7_zeQ

²⁸⁷ I metadati arricchiscono il contenuto informativo dei dati esplicitandone delle proprietà che semplificano il processo di fruizione dei dati stessi facilitandone la ricerca, il recupero, la composizione e di conseguenza il riutilizzo.

In particolare, per quanto riguarda la metadattazione, le informazioni esplicitate sono relative alla semantica dei dati: le etichette utilizzate servono per chiarire il significato dei dati e facilitarne quindi l'interpretazione, ovvero servono per descrivere il significato di attributi, proprietà e classi di entità.

Circa l'inferenza, ovvero la possibilità di derivare nuovi dati a partire da quelli esistenti, anch'essa è resa possibile proprio dall'attività di arricchimento del dataset. Il modo tradizionale è usare la conoscenza già espressa per inferirne altra implicita; grazie ai dati noti è possibile inferire (per deduzione) nuove informazioni e nuova conoscenza.

L'esempio più banale è quello delle relazioni inverse²⁸⁸: se sappiamo che

```
<http://localhost:2020/resource/oggetti/125>  
<skos:related>  
<http://localhost:2020/resource/persona/14>
```

possiamo inferire automaticamente che:

```
<http://localhost:2020/resource/persona/14>  
<is skos:related of>289  
<http://localhost:2020/resource/oggetti/125 >
```

Un esempio più interessante è l'inferenza della classe di appartenenza (tipo) quando la relazione di un'ontologia è vincolata a quel tipo.

Per esempio, se sappiamo che

```
<foaf:based_near>
```

²⁸⁸ La piattaforma D2RQ sfrutta la relazione inversa e la mostra nella navigazione su browser ma nella forma passiva. In questo modo la relazione viene valorizzata.

<http://www.mokabyte.it/2012/11/d2rmapping-2/>

²⁸⁹ Scritta la relazione in forma passiva come accade in D2RQ.

ha come vincolo di dominio

`<http://xmlns.com/foaf/spec/#term_SpatialThing>`²⁹⁰

possiamo inferire che tutti i valori di quella proprietà saranno di tipo `SpatialThing`.

5.4.6 Creazione del database relazionale

Le nuove tabelle create con Google Refine, contenenti anche le colonne dei collegamenti ai dataset esterni, sono state esportate in formato CSV e importate in un nuovo database MySQL denominato `fondo_lod`.

Il nuovo database è organizzato come il precedente, con la stessa suddivisione in tabelle (persone, oggetti, luoghi, date), ma contiene le nuove tabelle contenenti anche i link esterni.

Sono state poi create, e aggiunte a questo database, le seguenti tabelle di mapping:

- persone-oggetti
- date-oggetti
- luoghi-oggetti

Le tabelle di mapping servono per costruire relazioni M-M (Molti-A-Molti) nel database, ovvero per costruire quelli che saranno i link interni all'archivio.²⁹¹

Ad ogni record di una delle due tabelle coinvolte possono corrispondere più record nell'altra tabella e viceversa: per es. tutti gli oggetti che fanno riferimento all'anno 1789, tutte le persone che fanno riferimento ad un oggetto, ecc...

²⁹⁰ Dal FOAF Vocabulary, Property: `foaf:based_near`
Domain: having this property implies being a Spatial Thing
Range: every value of this property is a Spatial Thing
http://xmlns.com/foaf/spec/#term_based_near

²⁹¹ <http://www.riochierego.it/docs/guida-SQL/AMySQL2.html>

In pratica per ogni oggetto ci saranno diversi record nella tabella `luoghi / date/ persone`, e ogni luogo, data o persona farà capo ad uno o più oggetti.

Di norma per realizzare una relazione di questo tipo si creano delle nuove tabelle, le tabelle di mapping, che faranno da “tramite”.

Le tabelle di mapping hanno il compito di contenere le *foreign key* (chiavi esterne)²⁹², che associano i rispettivi `id` tra le tabelle; essendo questo il loro unico scopo, e non essendo quindi necessarie ulteriori informazioni sui campi, le tabelle di mapping contengono come colonne solo gli identificatori delle entità da mappare.

Si crea così, per ciascuna tabella di mapping, una doppia relazione.

Per esempio, tra oggetti e luoghi:

- relazione 1-M tra oggetti (lato 1) e mapping-luogo-ogg (lato M), basata sull'uguaglianza del contenuto dei campi
`oggetti.id_ogg` e `mapping-luogo-ogg.id_ogg`
- relazione 1-M tra luoghi (lato 1) e mapping-luogo-ogg (lato M), basata sull'uguaglianza del contenuto dei campi
`luoghi.id_luogo` e `mapping-luogo-ogg.id_luogo`

La tabella `mapping-luogo-ogg` di fatto è la relazione di tipo M-M tra oggetti e luoghi.

292

<https://docs.microsoft.com/it-it/sql/relational-databases/tables/create-foreign-key-relationships?view=sql-server-2017>

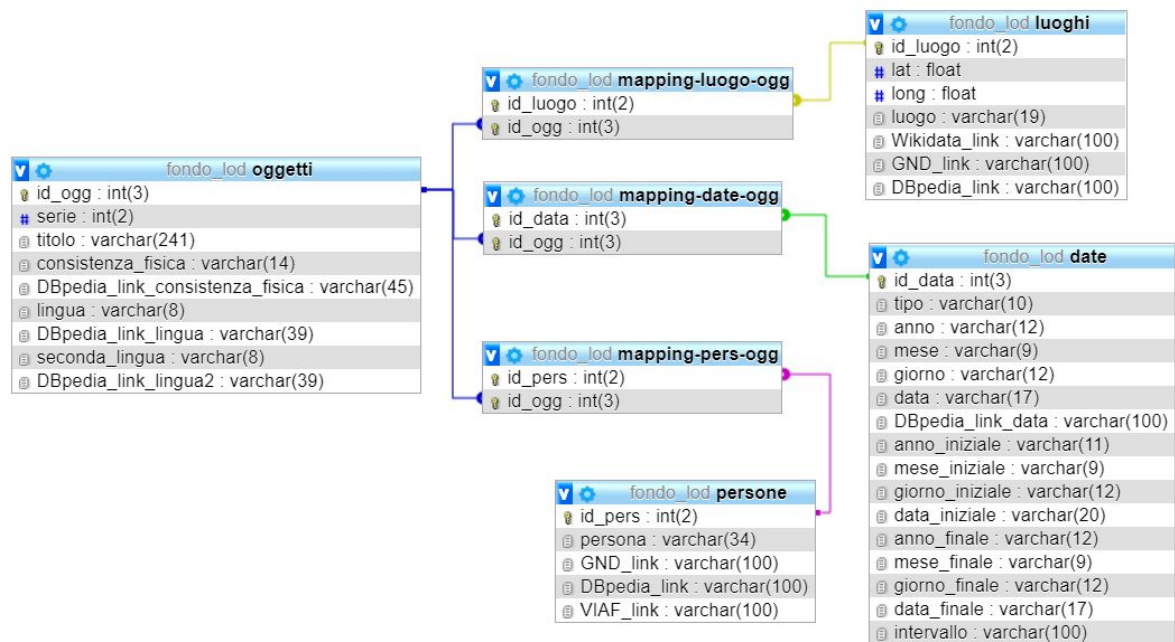


Fig. 5.13 Vista delle tabelle di mapping e delle relazioni all'interno del database.

```

243 # Mapping-date-ogg (n:m)
244 map:mapping-date-ogg_link a d2rq:PropertyBridge;
245   d2rq:belongsToClassMap map:oggetti;
246   d2rq:property edm:occurredAt;
247   d2rq:refersToClassMap map:date;
248   d2rq:join "mapping-date-ogg.id_data => date.id_data";
249   d2rq:join "mapping-date-ogg.id_ogg => oggetti.id_ogg";
250   .
251
252 # Mapping-luogo-ogg (n:m)
253 map:mapping-luogo-ogg_link a d2rq:PropertyBridge;
254   d2rq:belongsToClassMap map:oggetti;
255   d2rq:property edm:happenedAt;
256   d2rq:refersToClassMap map:luoghi;
257   d2rq:join "mapping-luogo-ogg.id_luogo => luoghi.id_luogo";
258   d2rq:join "mapping-luogo-ogg.id_ogg => oggetti.id_ogg";
259   .
260
261 # Mapping-pers-ogg (n:m)
262 map:mapping-pers-ogg_link a d2rq:PropertyBridge;
263   d2rq:belongsToClassMap map:oggetti;
264   d2rq:property skos:related;
265   d2rq:refersToClassMap map:persone;
266   d2rq:join "mapping-pers-ogg.id_ogg => oggetti.id_ogg";
267   d2rq:join "mapping-pers-ogg.id_pers => persone.id_pers";
268   .
269

```

Fig. 5.14 Rappresentazione delle relazioni M-M nel file di mapping del database secondo i modelli ontologici (cap. 5.5): appartengono agli oggetti e fanno riferimento a date / luoghi / persone.

Proprio le relazioni tra le tabelle del database sono ciò che costituisce i link interni all'archivio, ovvero collegamenti tra le entità / risorse dell'archivio stesso.

Esempio:

<u>Property</u>	<u>Value</u>
<code>rdfs:label</code>	Convenzione con la città di Genova del 1756
<code>edm:occurredAt</code>	<code><http://localhost:2020/resource/date/10></code>
<code>edm:happenedAt</code>	<code><http://localhost:2020/resource/luoghi/1></code>

Questi link interni si basano sulle relazioni tra un oggetto i rispettivi data e luogo. A partire dall'oggetto, collegamento punta alle risorse della data e del luogo corrispondenti tramite i loro `id`, e ciò è possibile proprio grazie alla struttura del database.

5.5 Scelta dell'ontologia per la rappresentazione dei dati

L'obiettivo principale di questa fase è dare una rappresentazione coerente con il modello ontologico del dataset di riferimento.

Si tratta di una vera e propria ristrutturazione logica e concettuale dei dati, un processo di *reingegnerizzazione* e *refactoring* della base informativa²⁹³.

Solo dopo aver ben chiari e pronti i dati che si vogliono rappresentare, si sceglie un modo (un'ontologia) per rappresentarli. (approccio *bottom up* descritto al [cap. 1.4.1](#)). La scelta delle ontologie e dei vocabolari parte sicuramente dall'analisi del dominio di appartenenza dei dati e delle risorse, in questo caso quello dei beni culturali.

Il primo criterio da seguire, quando si seleziona un'ontologia, è assicurarsi che questa riesca a rappresentare il dominio e gli argomenti relativi ai contenuti delle collezioni e dei dati che si vogliono convertire. Conoscere il settore di appartenenza

²⁹³ Agenzia per l'Italia Digitale (AGID), *Linee guida per l'interoperabilità semantica attraverso i Linked Open Data*, 2012, pag. 52

*e il significato dei dati risulta dunque cruciale. Solo così infatti si creano le premesse per scegliere in modo corretto l'ontologia più appropriata a rappresentare il proprio dominio.*²⁹⁴

Va valutata anche la sostenibilità del vocabolario o del modello. Spesso infatti alcuni di essi, normalmente accessibili nel web, vengono creati per supportare singoli progetti e, una volta raggiunto il risultato, non più aggiornati. In altri invece non è chiaro chi sia il proprietario e non si sa per quanto tempo il vocabolario resterà accessibile. Per questo è importante sapere se chi mantiene quel vocabolario è un'organizzazione stabile e affidabile.

Soprattutto quando si tratta di garantire l'accesso a collezioni di oggetti molto eterogenei o molto specifici, l'utilizzo di un solo schema non è sufficiente, poiché non riesce a rappresentare in modo analitico quella realtà di dati²⁹⁵. Una buona pratica, ampiamente utilizzata nelle comunità che sviluppano linked open data, è il riutilizzo di ontologie e vocabolari esistenti²⁹⁶ e per necessità particolari la creazione dei profili di applicazione.

Seguendo queste linee guida sono stati quindi scelti i seguenti modelli di rappresentazione:

- PICO²⁹⁷ (vedi [cap. 2.6.4](#)), Il Thesaurus del Portale della Cultura Italiana, un vocabolario controllato, progettato per la soggettazione e la classificazione di risorse molto eterogenee, provenienti da data-source culturali di ambiti estremamente differenti, compatibile con gli standard SKOS e RDF.

²⁹⁴ C. Fusetti, *Dati bibliografici e linked (open) data - Prospettive e criticità*, 2016, pag. 64

²⁹⁵ S. Peruginelli, *I profili di applicazione*, ICCU, Roma, 2001, www.iccu.sbn.it/upload/documenti/Peruginelli.ppt

²⁹⁶ D. Allemang, J. A. Hendler. *Semantic Web for the working ontologist: effective modeling in RDFS and OWL*, 2nd ed., Waltham: Morgan Kaufmann, 2011

²⁹⁷

http://www.culturaitalia.it/opencms/export/sites/culturaitalia/attachments/thesaurus/4.3/thesaurus_4.3.0.skos.xml

- EDM²⁹⁸, lo standard di metadati sviluppato da Europeana.
- Il *Basic Geo Vocabulary*, raccomandato dalle *guidelines*²⁹⁹ dell'EDM per la definizione delle proprietà relative ai luoghi (utilizzato in questo contesto per rappresentare coordinate spaziali -latitudine e longitudine-).

Inoltre sono stati inclusi ed utilizzati i vocabolari e le ontologie raccomandati dai principi LOD, quelli più utilizzati e condivisi all'interno della comunità del web semantico: SKOS, Dublin Core, FOAF, RDF Schema e OWL.

<i>namespace</i>	URI
skos	http://www.w3.org/2004/02/skos/core#
dc	http://purl.org/dc/elements/1.1/
dcterms	http://purl.org/dc/terms/
foaf	http://xmlns.com/foaf/0.1/
rdfs	http://www.w3.org/2000/01/rdf-schema#
owl	http://www.w3.org/2002/07/owl#
pico	http://www.culturaitalia.it/opencms/export/sites/culturaitalia/attachments/thesaurus/4.3/thesaurus_4.3.0.skos.xml#
edm	http://www.europeana.eu/schemas/edm/
geo	http://www.w3.org/2003/01/geo/wgs84_pos#

Tab. 5.3 Tabella riassuntiva dei vocabolari e delle ontologie utilizzati.

²⁹⁸

https://pro.europeana.eu/files/Europeana_Professional/Share_your_data/Technical_requirements/EDM_Documentation//EDM_Definition_v5.2.8_102017.pdf

²⁹⁹

https://pro.europeana.eu/files/Europeana_Professional/Share_your_data/Technical_requirements/EDM_Documentation/EDM_Mapping_Guidelines_v2.4_102017.pdf

5.5.1 D2RQ

Per generare una prima modellazione automatica del dataset in conformità di un modello logico è stata utilizzata la piattaforma D2RQ.

La piattaforma D2RQ³⁰⁰ è un sistema per accedere ai database relazionali come grafi RDF.

D2R (*Database to RDF*) è un framework che permette l'accesso al contenuto di un database relazionale come se si trattasse di un dataset RDF. Tramite un linguaggio dichiarativo³⁰¹ è possibile definire clausole che, in modo flessibile descrivono le relazioni tra le ontologie di riferimento e la struttura tabellare della base di dati. L'applicazione delle clausole consente, quindi, la produzione di un dataset RDF (anche chiamato "RDF dump") secondo le specifiche desiderate.

Oltre agli strumenti per la modellazione di ontologie risultano di particolare interesse il riuso di vocabolari ed ontologie già esistenti. Il riuso di basi di conoscenza già esistenti e riconosciute a livello internazionale permette di realizzare in maniera automatica l'*interlinking* dei dati, in quanto riutilizzando concetti già presenti e utilizzati in altre basi di conoscenza sarà possibile collegare varie sorgenti.

Il server D2R si basa su un mapping dichiarativo tra lo schema del database e i termini RDF di destinazione, fornito dal publisher dei dati. Sulla base di questa mappatura, il server D2R serve un punto finale SPARQL e le viste dei dati collegati del database.

³⁰⁰ <http://d2rq.org/>

³⁰¹ Linguaggio di mapping D2RQ è un linguaggio di mappatura dichiarativo per descrivere la relazione tra un'ontologia e un modello di dati relazionali. <http://d2rq.org/d2rq-language>

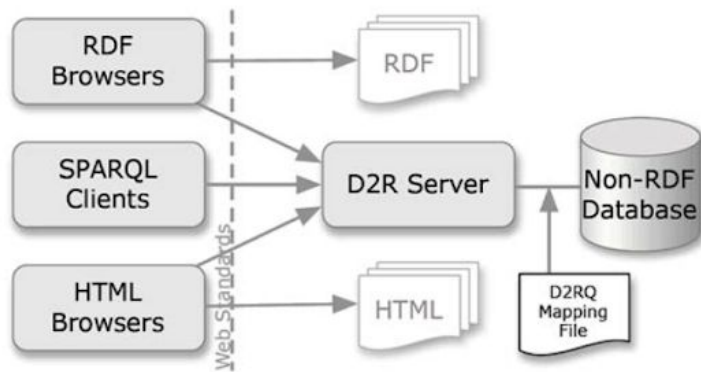


Fig. 5.15 Diagramma dell'architettura del server D2R.

Usando D2RQ è possibile:

- interrogare un database non RDF utilizzando SPARQL
- accedere al contenuto del database come Linked Data sul Web
- creare dump personalizzati del database in formati RDF per il caricamento in un archivio RDF
- accedere alle informazioni in un database non RDF utilizzando un'apposita API³⁰²

Il driver JDBC³⁰³ per MySQL è già incluso nella piattaforma D2RQ.

Un primo file di mapping è stato generato dal d2r-server, e successivamente è stato modificato manualmente per renderlo conforme all'ontologia precedentemente scelta su cui modellare i dati.

Il comando `generate-mapping`, analizzando lo schema del database esistente, crea un file di mapping, denominato "mapping di default", che associa ogni tabella ad una nuova classe RDFS basata sul nome della tabella, e ogni colonna ad una

³⁰² <http://incubator.apache.org/jena/>

³⁰³ JDBC (*Java DataBase Connectivity*) è un connettore (*driver*) per database che consente l'accesso e la gestione della persistenza dei dati sulle basi di dati da qualsiasi programma scritto con il linguaggio di programmazione Java, indipendentemente dal tipo di DBMS utilizzato. È costituito da un'API object oriented orientata ai database relazionali, raggruppata nel package `java.sql`, che serve ai client per connettersi a un database fornendo i metodi per interrogare e modificare i dati.
https://it.wikipedia.org/wiki/Java_DataBase_Connectivity

proprietà in base al nome della colonna. Questo file di mappatura può successivamente essere personalizzato.

Per generale il primo mapping³⁰⁴ sono stati eseguiti i seguenti comandi da terminale³⁰⁵.

```
cd "path della directory del server D2R"
```

```
generate-mapping -o mapping.ttl -d driver.class.name  
-u db-utente -p db-password jdbc: url: ...
```

mapping.ttl è il nome per il nuovo file di mappatura.

-d può essere saltato per MySQL.

L'URL di connessione JDBS per i database MySQL è il seguente:

```
jdbc:mysql://servername/databasename
```

Quindi, in questo caso, il comando diventa:

```
generate-mapping -o mapping.ttl -u root  
jdbc:mysql://localhost/fondo_lod
```

NB: se non sono state precedentemente settate le chiavi primarie per tutte le tabelle del database il comando non verrà eseguito e genera errore.

Avviare il server:

```
d2r-server mapping.ttl
```

Testare il file di mapping generato su browser all'indirizzo

```
http://localhost:2020.
```

Per rappresentare l'archivio secondo l'ontologia scelta si è proceduto con la personalizzazione (modifica manuale) del file di mapping di default prodotto in

³⁰⁴ <http://d2rq.org/getting-started>

³⁰⁵ dalla shall di XAMPP

maniera automatica dal server D2R: per ogni colonna delle tabelle sono state trovate le corrispondenze nell'ontologia EDM ed è stato modificato il file scrivendo quelle corrispondenze nel linguaggio di mapping dichiarativo di D2R.

Questa attività ha portato quindi alla creazione di diverse classi e proprietà secondo i modelli di rappresentazione prescelti.

Di particolare importanza in questa fase sono le triple che forniscono etichette *human readable* per le risorse che possono essere utilizzate all'interno delle applicazioni client. Predicati come `rdf:label`, `foaf:name`, `skos:prefLabel` e `dc:title` devono essere utilizzati a tale scopo poiché sono ampiamente supportati dalle applicazioni di *Linked Data*.

5.5.2 Creazione degli URI

Ogni risorsa del dataset ottenuto è innanzitutto individuata univocamente da specifici URI. Tali indirizzi fanno riferimento al nome associato all'indirizzo della macchina locale dove è stato implementato il progetto, ovvero `http://localhost:2020/`.³⁰⁶

Tutti gli URI che identificano le risorse (entità) sono creati tramite un preciso schema, ottenuto tramite l'utilizzo di un prefisso del tipo `http://localhost:2020/resource/` al quale segue l'indicazione specifica della tipologia di risorsa (entità) e l'identificatore, ottenendo quindi un suffisso del tipo `entità/id`.

Per garantire l'univocità degli URI, infatti, è spesso utile basarli su alcune chiavi primarie esistenti, come un id univoco in una tabella di database.

³⁰⁶ Dato che al momento di scrittura della presente tesi il dataset non è ancora stato ufficialmente pubblicato in un dominio web raggiungibile pubblicamente, gli URI utilizzano una struttura generica come quella indicata. Nel momento in cui i dati saranno raggiungibili nel Web e nel *Linked Data*, e quindi pubblicati su uno specifico dominio Web, la parte indicata come `localhost:2020` sarà sostituita con l'indirizzo di dominio assegnato.

Le principali tipologie di risorse presenti nel dataset sono le seguenti:

- Risorse relative a persone, identificate dal seguente URI:
`http://localhost:2020/resource/persona/`
- Risorse relative a date, identificate dal seguente URI:
`http://localhost:2020/resource/date/`
- Risorse relative a luoghi, identificate dal seguente URI:
`http://localhost:2020/resource/luoghi/`
- Risorse relative a oggetti, identificate dal seguente URI:
`http://localhost:2020/resource/oggetti/`

5.5.3 Dereferenziazione degli URI

L'atto di recuperare una rappresentazione di una risorsa identificata da un URI è noto come la dereferenziazione mediante tale URI.

Ad ogni risorsa sono associati diversi URI: ognuno indica specifiche rappresentazioni documentali.

Qualsiasi URI HTTP deve essere dereferenzabile, il che significa che i client HTTP possono cercare l'URI usando il protocollo HTTP e recuperare una descrizione della risorsa identificata dall'URI. Ciò si applica agli URI utilizzati per identificare i documenti HTML classici, nonché gli URI utilizzati nel contesto dei dati collegati per identificare oggetti del mondo reale e concetti astratti.

Deve infatti essere sempre possibile ottenere la descrizione di ogni risorsa presente nel dataset digitandone l'URI relativo all'interno di un web browser.

Laddove gli URI identificano risorse del mondo reale, è essenziale non confondere gli oggetti stessi con i documenti Web che li descrivono. Pertanto, è prassi comune utilizzare diversi URI per identificare l'oggetto del mondo reale e il documento che lo descrive, in modo da essere non ambigui. Questa pratica consente di fare dichiarazioni separate su un oggetto e su un documento che descrive quell'oggetto. Essere in grado di distinguere i due attraverso l'uso di diversi URI è fondamentale per la coerenza del *Web of Data*.

Dato che il paradigma LOD richiede che le risorse siano conformi al protocollo HTTP, cioè che sia possibile dereferenziarle, in questo progetto è stata utilizzata

una strategia per rendere degli URI dereferenziabili, detta *hash URI*: essa si basa sulla caratteristica che gli URI possono contenere una parte speciale separata dalla parte di base dell'URI da un simbolo di hash (#). Questa parte speciale è chiamata identificatore di frammento.

Quando un client desidera recuperare un URI hash, il protocollo HTTP richiede che la parte frammento venga rimossa prima di richiedere l'URI dal server. Ciò significa che un URI che include un hash non può essere recuperato direttamente e quindi non identifica necessariamente un documento Web. Ciò consente di utilizzare tali URI per identificare oggetti del mondo reale e concetti astratti, senza creare ambiguità.

Se prendiamo ad esempio una risorsa di tipo “data”, è necessario distinguere la data in sé dal documento / termine di vocabolario che la descrive.

Grazie alla strategia *hash URI* è possibile evitare questa ambiguità e recuperare, in maniera distinta dalla risorsa, la descrizione della risorsa stessa grazie al suo identificatore, preceduto dal simbolo #, a partire dall'URI del vocabolario.

URI + # + id

Es. `http://purl.org/dc/terms/#date`

L'attività appena descritta rappresenta una delle metodologie di base per l'accesso e la navigazione dei dati presenti all'interno di in una sorgente pubblicata come Linked Data.

5.5.4 Grafi esplicativi

Questa sezione ha come scopo quello di mostrare le rappresentazioni RDF ottenute, secondo il modello teorico descritto al [cap. 1.2](#).

Le tabelle seguenti schematizzano degli esempi di triple RDF tratti dalle risorse contenute nel fondo.

<i>Subject</i>	<i>Predicate</i>	<i>Object</i>
La persona con id 11	ha un nome	Pietro Strozzi
<code>http://localhost:2020/resource/persona/11</code>	<code>http://xmlns.com/foaf/0.1/name</code>	"Pietro Strozzi"

Tab. 5.4 un esempio di *tripla letterale*.

<i>Subject</i>	<i>Predicate</i>	<i>Object</i>
La persona con id 11	è correlato a	l'oggetto con id 84
<code>http://localhost:2020/resource/persona/11</code>	<code>http://www.w3.org/2004/02/skos/core#related</code>	<code>http://localhost:2020/resource/oggetti/84</code>

Tab 5.5 un esempio di *link RDF interno*.

<i>Subject</i>	<i>Predicate</i>	<i>Object</i>
Questo URI	si riferisce alla stessa cosa	di questo URI
<code>http://localhost:2020/resource/persona/11</code>	<code>http://www.w3.org/2002/07/owl#sameAs</code>	<code>https://lobid.org/gnd/121994724</code>

Tab. 5.6 un esempio di *link RDF esterno*.

Le seguenti figure mostrano altri esempi di risorse rappresentate in RDF ma sotto forma di grafi. In blu sono rappresentate le triple letterali, in giallo i link RDF esterni, in verde i link RDF interni.

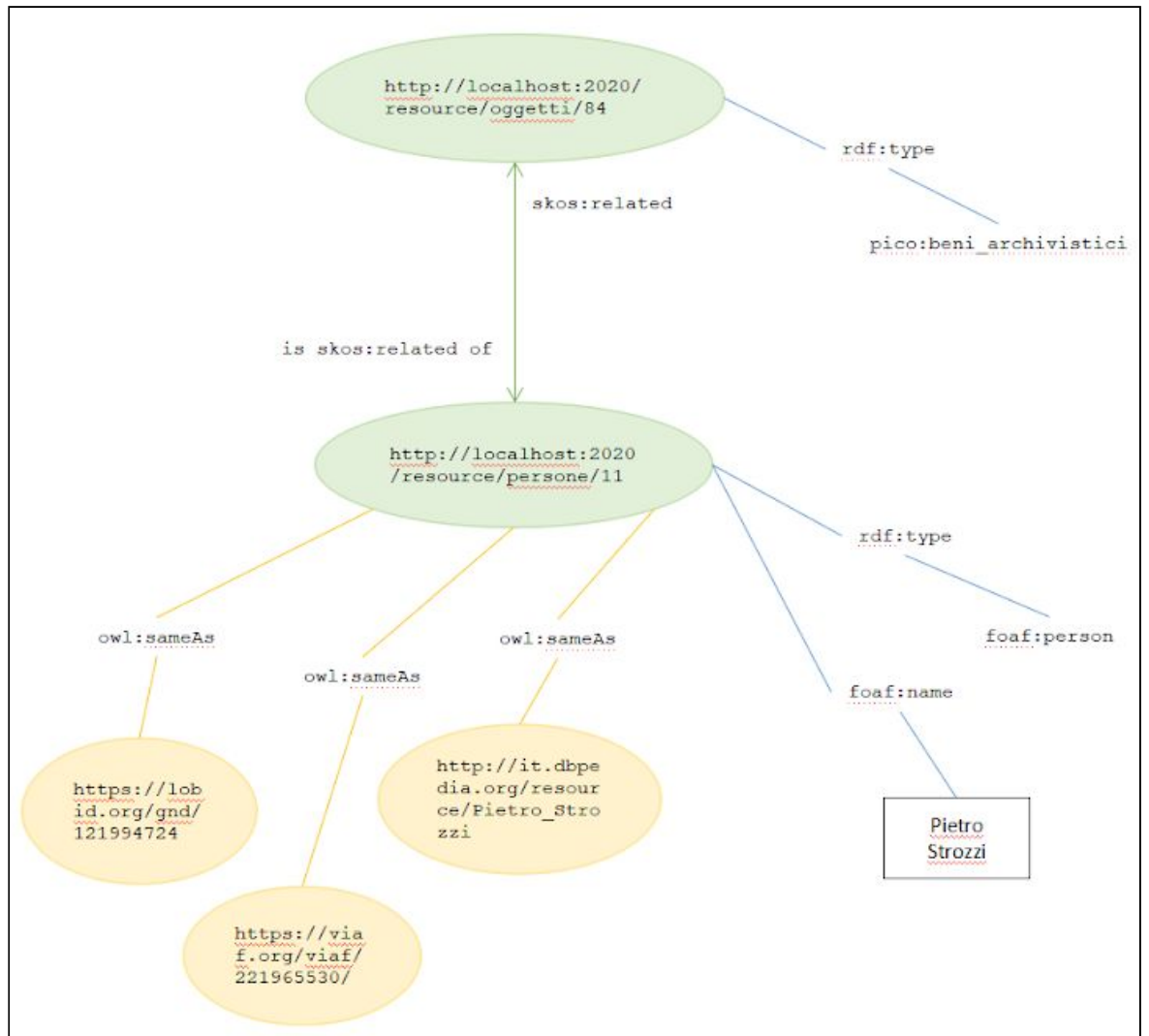


Fig. 5.16 Grafi RDF dell'archivio LOD.

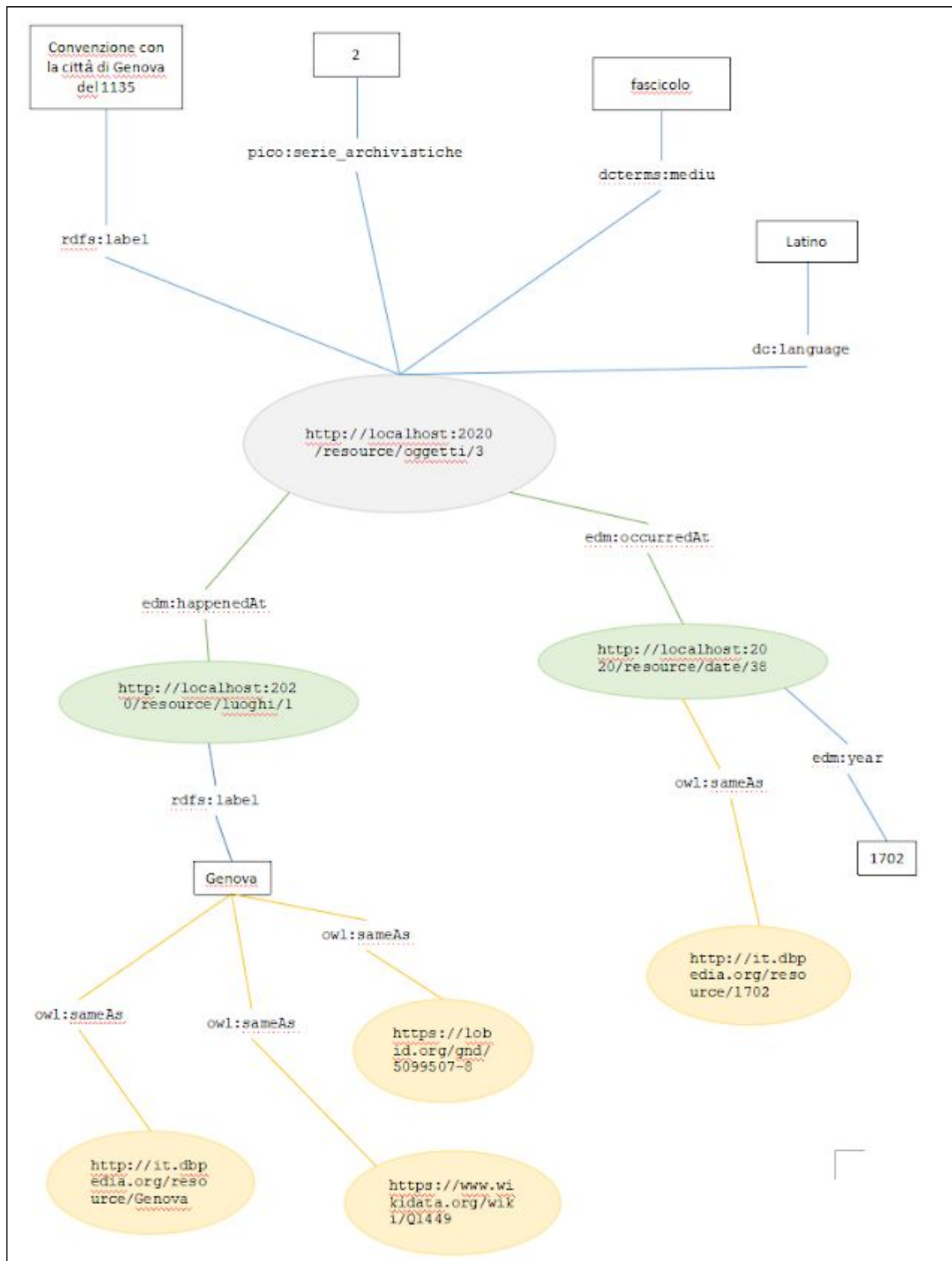


Fig. 5.17 Grafi RDF dell'archivio LOD.

5.6 Esportazione dell'archivio in Linked Open Data attraverso D2RQ

Al completamento di questa fase, il dataset risultante sarà rappresentato in RDF. Tutte le informazioni del dataset saranno rappresentate da risorse, che a loro volta sono descritte tramite relazioni con altre risorse, secondo il modello a grafo proposto da RDF. Questo mette in evidenza come tale modello di rappresentazione si discosti da quelli utilizzati dalle tradizionali schede di catalogazione, in quanto sposta l'attenzione dall'oggetto culturale alle relazioni che esso intrattiene con gli altri oggetti e concetti (principio che è alla base del funzionamento del *Knowledge Graph*).

Questo nuovo approccio supera uno dei principali limiti della rappresentazione tradizionale della conoscenza in ambito culturale.³⁰⁷

L'esportazione dell'archivio in *Linked Open Data* è stata possibile scaricando il contenuto dell'intero database in un singolo file RDF tramite D2RQ³⁰⁸ e sulla base del file di mapping definitivo. Quest'ultimo infatti servirà al comando che segue a tradurre il contenuto del database in RDF.

Il comando da eseguire per il dump del database è il seguente.

```
dump-rdf [-f format] [-b baseURI] [-o outfile.ttl]
mapping_file.ttl
```

In questo caso il `baseURI` è il seguente:

```
http://localhost:2020/resource/
```

e i formati in cui è stato esportato il database sono:

³⁰⁷ L. Moro, 2019.

³⁰⁸ <http://d2rq.org/dump-rdf>

- RDF/XML, sintassi standard del W3C, ampiamente utilizzata per pubblicare i dati LOD sul Web ma difficile da leggere per l'uomo.
- TURTLE, formato di testo semplice per serializzare i dati RDF di lettura più semplice per i fruitori umani.
- N-TRIPLE, sottoinsieme di Turtle ma senza le funzioni come prefissi e scorciatoie dei namespace. Il risultato è un formato di serializzazione con molta ridondanza, poiché tutti gli URI devono essere specificati per intero in ogni tripla. Di conseguenza, i file N-Triple possono essere piuttosto grandi rispetto a Turtle e persino a RDF / XML. Tuttavia, questa ridondanza è anche il vantaggio principale di N-Triple rispetto ad altri formati di serializzazione, poiché consente di analizzare i file una riga alla volta, rendendola ideale per il caricamento di file di dati di grandi dimensioni che non si adattano alla memoria principale. La ridondanza rende inoltre N-Triple molto suscettibile alla compressione, riducendo il traffico di rete durante lo scambio di file. Questi due fattori rendono N-triple lo standard *de facto* per lo scambio di grandi dump di dati collegati.

I dump sono scaricabili dal sito oppure accessibili tramite il relativo repository su GitHub³⁰⁹.

Nello stesso repository sono disponibili anche il codice del sito web e il file di mapping.

5.7 Costruzione del sito web per la navigazione delle informazioni sul fondo

Una volta ottenuto l'archivio *Linked Open Data* è stato costruito un sito web per l'accessibilità e la presentazione dei contenuti del fondo.

³⁰⁹ https://github.com/alod83/fondo_novi

Home Risorse Ontologia Dataset About

Per approfondire:

Città di Novi Ligure
Il fondo antico
Mèmora Piemonte

Fondo antico del Comune di Novi Ligure


Il Fondo raccoglie al suo interno tutta la documentazione archivistica riferibile all'attività della comunità novese dal 1495 fino al periodo della dominazione francese (fino a tutto l'anno 1799) e anche altre carte dal 1800 al 1811.

Le unità documentarie dell'archivio sono state estratte dal sistema Mèmora, una piattaforma digitale pubblica, per la descrizione del patrimonio culturale di archivi, istituti culturali e musei, realizzata da [CSI Piemonte](#).

Il Fondo appartiene all'archivio storico del Comune di Novi Ligure ed è classificato nella piattaforma Mèmora all'interno della cartella "Archivi e patrimonio documentale" del Comune.

In totale, il Fondo contiene 3 sezioni, 70 serie, 83 sottoserie e un totale di 1641 unità (schede documentarie) che coprono un periodo che va dagli inizi del 1500 fino al 1870.

Il lavoro di conversione dei dati in Linked Open Data è stato eseguito sulla prima sezione del fondo, intitolata *Atti, registri, titoli e carte riflettoni all'era antica fino a tutto l'anno 1799*, che contiene 18 serie e 223 unità che coprono il periodo 1135/1495 - 1799.



Il Fondo antico del comune di Novi Ligure

Home Risorse Ontologia Dataset About

Per approfondire:

Città di Novi Ligure
Il fondo antico
Mèmora Piemonte

Risorse

Dal Fondo sono state estratte le seguenti risorse:

- **Oggetti**, cioè le unità archivistiche.
- **Persone** relative ad un oggetto.
- **Luoghi** cui l'oggetto si riferisce.
- **Date** cui l'oggetto si riferisce, modellate anche come giorni, mesi e anni.

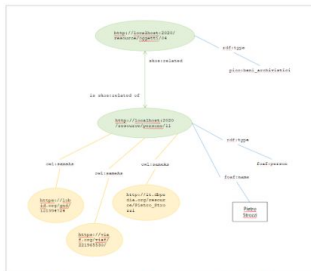
Collegamenti interni

Ogni oggetto può avere uno o più link interi a persone, luoghi e date correlate.

Collegamenti esterni

Il Fondo contiene collegamenti esterni ai seguenti dataset

- **Oggetti:** [DBpedia \(it\)](#)
- **Persone:** [DBpedia \(it\)](#), [GND](#) e [VIAF](#)
- **Luoghi:** [DBpedia \(it\)](#), [GND](#) e [Wikidata \(en\)](#)
- **Date:** [DBpedia \(it\)](#)



Esempi di risorse con relativi link interni (in verde) ed esterni (in giallo).

Home Risorse Ontologia Dataset About

Per approfondire:

Città di Novi Ligure
Il fondo antico
Mèmora Piemonte

Ontologia

Per la rappresentazione dei dati sono state utilizzate i seguenti vocabolari e ontologie:

- [Europeana Data Model ontology \(EDM\)](#)
- [PICO Thesaurus](#)
- [Basic Geo Vocabulary](#)
- [Web Ontology Language \(OWL\)](#)
- [SKOS](#)
- [FOAF](#)
- [Dublin Core](#)
- [RDF Schema](#)

Sono stati utilizzati le seguenti classi e proprietà:

Per le risorse "luoghi":

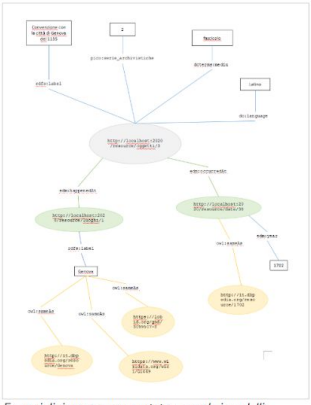
- **dc:coverage**, **edm:Place** come classi per definire i luoghi.
- **dcterms:spatial**, **geo:lat**, **geo:long**, per la rappresentazione di coordinate spaziali.
- **edm:happenedAt** per gli oggetti che fanno riferimento ad un luogo

Per le risorse "oggetti":

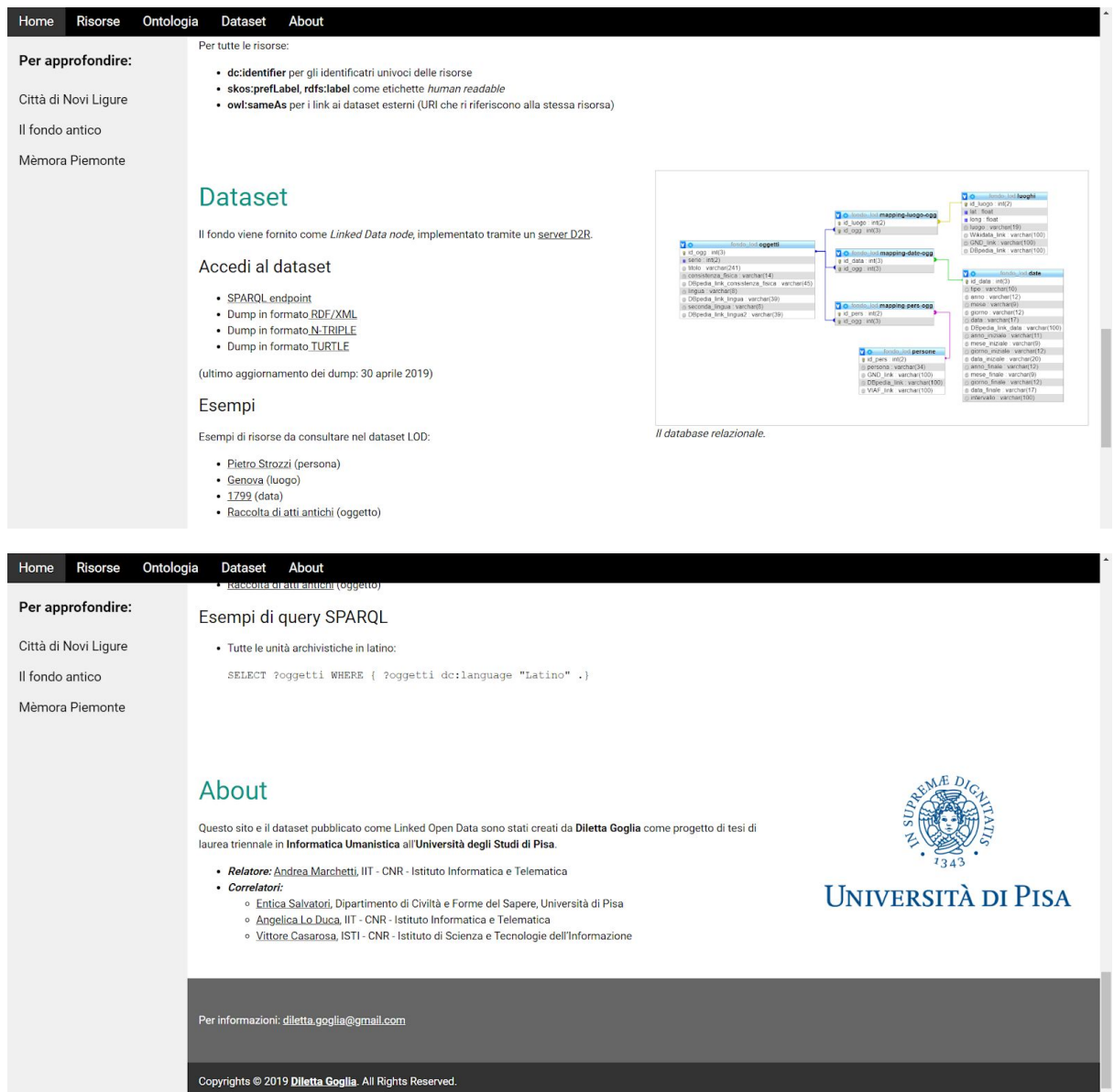
- **pico:beni_archivistici** come classe per gli oggetti
- **pico:serie_archivistiche** la proprietà "numero di serie d'appartenenza nell'archivio" degli oggetti
- **dcterms:medium** come proprietà "consistenza fisica" degli oggetti
- **dc:language** per la proprietà "lingua" degli oggetti

Per le risorse "persone":

- **foaf:person** come classe per le persone
- **foaf:name** per la proprietà "nome" delle persone



Esempi di risorse rappresentate secondo i modelli ontologici di riferimento.



Figg. 5.18-22 Il sito web per il Fondo Antico del Comune di Novi Ligure

Affinché i dati siano consultabili fornendo la propria rappresentazione documentale, il dataset è navigabile dal sito grazie alla piattaforma D2RQ, tramite il link che indirizza all'Endpoint SPARQL.

Il sito web permette infatti di accedere ai dati anche tramite un'interfaccia di interrogazioni SPARQL. E' messo a disposizione sul sito, tramite D2R, uno SPARQL Endpoint per interrogare il database, cioè un servizio conforme al protocollo proposto dal Semantic Web, che è in grado di accettare delle richieste di interrogazione in linguaggio SPARQL restituendone gli eventuali risultati, e permettendone la visualizzazione navigabile sempre in modalità *Linked Data*.



Fig. 5.23 Endpoint SPARQL

L'accesso ai dati tramite l'utilizzo di interrogazioni SPARQL rappresenta un metodo alternativo per visualizzare le informazioni di interesse rispetto a quello effettuato tramite la dereferenziazione degli URI prima descritto. Quest'ultimo può essere utilizzato, ad esempio, per ottenere la descrizione completa di una risorsa e per comprendere quindi come questa è rappresentata, ottenendo informazioni utili su come strutturare eventuali interrogazioni. La creazione di query richiede infatti una conoscenza precisa del modo in cui i dati sono descritti, e quindi dell'ontologia utilizzata, ma risulta un meccanismo molto potente per filtrare e individuare precisamente le informazioni di interesse.

Per permettere anche agli utenti meno esperti di effettuare delle query SPARQL, sono presentati sul sito degli esempi.

Il codice del sito web, così come i dump del database nei tre formati e il file di mapping con il modello ontologico sono disponibili nel repository GitHub accessibile al seguente indirizzo:

https://github.com/alod83/fondo_novi

6. Conclusioni e sviluppi futuri

6.1 Aderenza ai principi del Linked Data e scoperta informativa

L'obiettivo principale dell'attività svolta nell'ambito del progetto consiste nel rendere facilmente accessibili e utilizzabili i dati sui beni culturali relativi al Fondo antico di Novi Ligure utilizzando le tecniche proposte dal *Linked Data* che, come si è analizzato nel corso di questa tesi, risultano tra le più adatte per rendere tali informazioni raggiungibili e consumabili anche in un'ottica di relazione con altri dati collegati.

Per valutare qualitativamente l'attività svolta non esistono dei parametri oggettivi e facilmente determinabili³¹⁰; tuttavia, al fine di misurare la *bontà* del lavoro svolto si può fare riferimento ai principi (precedentemente descritti al [cap. 1](#)) su cui l'intero concetto di *Linked Data* è basato, valutando se ognuno di essi è stato rispettato e in che modalità.

- *Utilizzare gli URI per identificare le risorse.* Ogni risorsa del Fondo è identificata da specifici URI costruiti secondo delle regole ben precise.
- *Utilizzare gli URI in conformità al protocollo HTTP*, in modo che possano essere consultati. Anche tale principio è stato rispettato grazie alla dereferenziazione degli URI.

E' importante specificare che il dataset non è ancora stato formalmente caricato su un dominio web nel momento in cui si scrive, principalmente a causa della mancanza degli oggetti digitalizzati che solo il proprietario, ovvero il portale Mèmora da cui sono stati recuperati, può fornire. La costruzione di URI, secondo le regole LOD, sulla base di un nuovo nome di

³¹⁰ De Marco G., 2011

dominio pubblico, è presa in considerazione come uno degli sviluppi futuri del progetto ([cap. 6.2](#))

- *Utilizzare gli standard RDF e SPARQL* come modello per la rappresentazione dei dati. Il dataset costruito è completamente basato sul modello a triple di RDF e facilmente accessibile sia dal sito web che tramite interrogazioni gestite da un endpoint SPARQL. Tali standard sono pertanto stati utilizzati pienamente per la rappresentazione dei dati.
- *Includere link agli URI che identificano altre risorse*, in modo che possano essere scoperte nuove informazioni. L'attività di collegamento con risorse esterne ha rappresentato uno dei punti di maggiore attenzione durante tutto il lavoro, permettendo la creazione di percorsi di scoperta informativa.

Si può pertanto affermare che tutti i principi sono stati rispettati e adempiuti.

Oltre a quelli appena descritti si può fare riferimento anche all'interpretazione di tali principi proposta dal *Linking Open Data Project* ([cap. 4.1](#)), che fornisce una lista di regole da rispettare

per richiedere la pubblicazione di un proprio dataset all'interno del progetto.

Nello specifico, oltre a regole che fanno riferimento essenzialmente a quanto già richiesto dai principi precedenti, sono richieste le seguenti direttive, di cui si valuta l'aderenza:

- *Il dataset deve contenere almeno mille triple*. Il dataset contiene 16.889 triple che descrivono oltre 200 oggetti che a loro volta fanno riferimento a diversi altri concetti tra persone, luoghi, date, ecc...
- *Il dataset deve essere connesso con link RDF ad altri dataset già presenti nel Linking Open Data Project*, sono richiesti arbitrariamente almeno 50 collegamenti. La maggior parte dei collegamenti sono stati effettuati verso il dataset di DBpedia che fa parte del *Linking Open Data Project*; le triple che fanno riferimento a collegamenti di questo genere superano il numero richiesto di 50³¹¹.

³¹¹ Sono 56 in totale.

- *L'accesso all'intero dataset deve essere possibile tramite crawling RDF, dump RDF o tramite un endpoint SPARQL. Il crawling si riferisce alla possibilità di navigare tra le risorse visualizzando, per ognuna di esse, una rappresentazione accedendo ad uno specifico indirizzo. Tale attività è possibile grazie ai meccanismi realizzati per la dereferenziazione degli URI delle risorse. Il dump RDF è disponibile sul sito in tre formati, sito che dispone anche di un endpoint per le interrogazioni SPARQL. Tutti questi meccanismi di accesso sono pertanto stati implementati.*

La modellazione dei dati del progetto rispetta quindi tutti i principi proposti dal *Linked Data* e ciò permette di rispondere alle principali esigenze per le quali tale attività è stata svolta.

6.2 Proposte di sviluppo futuro

Esistono alcuni aspetti del progetto che possono essere migliorati e potenziati.

Una prima, importante, attività che potrà essere svolta consiste nel rendere effettivamente reperibile ed utilizzabile il dataset prodotto all'interno del *Linking Open Data Project*.

La pubblicazione all'interno di un dominio web è possibile senza necessità di particolari modifiche al lavoro svolto, che aderisce già completamente a tutti i requisiti richiesti per entrare a far parte del *Web of Data*.

Un auspicabile e imminente sviluppo futuro del progetto consiste pertanto nel rendere raggiungibile il dataset creato all'interno di un dominio web per chiederne la formale pubblicazione e ammissione all'interno della rete di dati che costituisce il *LOD Cloud*.

In questo modo le informazioni prodotte potranno non solo essere utilizzate molto più facilmente per i più disparati scopi applicativi (in relazione con le altre già presenti con

le quali presentano collegamenti e associazioni) ma anche essere oggetto di future correlazioni da parte di nuovi dataset inseriti nella rete.

Un'altra attività, strettamente correlata a quella appena descritta, consiste quindi nel coniare un nuovo nome di dominio, facilmente leggibile e memorabile, per pubblicare l'archivio LOD, nonché rilasciarlo secondo una licenza *open*.

Inoltre, è auspicabile l'estensione del progetto alle altre sezioni del Fondo, possibile tramite un lavoro analogo a quello finora svolto e descritto in questa tesi.

Nella stesura infatti, è stata riservata particolare attenzione nello spiegare in maniera chiara e completa tutti i dettagli di implementazione del progetto in modo da rendere possibile la riproducibilità, anche a chiunque voglia intraprendere un'iniziativa simile di conversione e pubblicazione di dati LOD.

Infine, si propone come ulteriore attività futura un'analisi mirata al testo degli oggetti per ricercare entità e risorse: l'attività di *entity extraction* e di *Named Entity Recognition* è stata infatti svolta, all'interno di questo progetto, limitatamente ai titoli degli oggetti e non al loro contenuto testuale.

6.3 Considerazioni finali

L'obiettivo dominante che ha portato alla realizzazione del lavoro presentato in questo documento di tesi si basa sulla convinzione che le tecnologie legate al Semantic Web rappresentino lo strumento ideale per la rappresentazione della conoscenza in ambito artistico e culturale.

Il modello a grafo tipico dell'RDF, l'utilizzo di modelli ontologici per la definizione formale delle informazioni e la possibilità di associare e arricchire le proprie sorgenti dati sfruttando carichi informativi esterni rappresentano sicuramente le caratteristiche principali che consentono di ottenere una modellazione dei dati relativi a beni artistici e culturali in grado di esprimere a pieno il valore informativo degli oggetti.

Il dataset prodotto tramite questo progetto, e il sito web realizzato per consentirne un accesso e un utilizzo conforme agli standard e ai principi del Linked Data, ne costituisce un esempio pratico e diretto.

L'attività svolta ha lo scopo di mostrare come sia possibile, sfruttando queste tecnologie, rendere disponibili i dati in un formato aperto e aderente agli standard,

consentendone un utilizzo il più possibile generico e non legato esclusivamente alla sola attività di consultazione.

Viene messo in risalto anche il nuovo modo di *ripensare la conoscenza* abbattendo le frontiere imposte dai singoli *silos* informativi, come Tim Bernes Lee stesso auspica e suggerisce.

E' importante, per concludere, osservare come nuovi interessanti sviluppi possono delinearsi indipendentemente da quanto è stato realizzato in quest'ambito.

Si è visto infatti come risultato importante, specialmente nel contesto del patrimonio culturale *Cultural Heritage* italiano, avere dei *Linked Data* che portino alla creazione e al raggiungimento di valore aggiunto al nostro patrimonio culturale e ai beni (materiali e immateriali) che lo costituiscono.

È stato inoltre ampiamente descritto come il Web of Data sia una rete in continua evoluzione, e come esistano diversi progetti che puntano ad utilizzare ed espandere sempre

di più la rete informativa da esso messa a disposizione. Gli sviluppi di cui si parla si riferiscono pertanto alla possibilità, in futuro, di correlare i dati del Fondo antico di Novi Ligure con nuove sorgenti informative di interesse, incrementando la quantità di collegamenti in uscita e dando la possibilità, di conseguenza, di creare arricchimenti ancora più utili e interessanti.

Più la rete dei dati collegati si espanderà, specie se in riferimento al campo artistico e culturale italiano, maggiori saranno le prospettive future di utilizzo di sorgenti dati come quella realizzata in questo lavoro di tesi.

Bibliografia

Libri

Linked Data e Web Semantico

Berners-Lee, T. *Weaving the Web: The original design and ultimate destiny of the World Wide Web (1st edition)*, Thomson Learning Emea, 2000 [1999], ISBN: 9780062515872, <https://www.w3.org/People/Berners-Lee/Weaving/>

Bizer C., Heath T., *Linked Data: Evolving the Web into a Global Data Space (1st edition)*, 2011, ISBN: 9781608454303 (versione cartacea), ISBN: 9781608454310 (ebook), versione HTML: <http://linkeddatabook.com/editions/1.0/>

Capitani P., *Il Knowledge Management*, in: *Associazione Italiana Formatori*, Franco Angeli Editore, Milano, 2006, ISBN: 8846472535

Guerrini M., Possemato T., *Linked Data per biblioteche, archivi e musei*, Editrice Bibliografica, 2015, ISBN: 9788870758306

Hooland S., Verborgh R., *Linked Data for Libraries, Archives and Museums: how to clean, link and publish your metadata*, Facet Publishing, Londra 2014, ISBN: 9781856049641

Standard di metadati

American Library Association, Canadian Library Association, CILIP, *Resource Description and Access*, 2014, traduzione italiana a cura dell'ICCU, 2015, versione PDF:

https://www.iccu.sbn.it/export/sites/iccu/documenti/2015/RDA_Traduzione_ICCU_5_Novembre_REV.pdf

Corti L., *Beni culturali, standards di rappresentazione, descrizione e vocabolario*, in: *Informatica e beni culturali*, Franco Cosimo Panini Editore, Modena, 1992, ISBN: 8876861939

ICA - International Council on Archives, *ISAAR (CPF): Standard internazionale per i record d'autorità archivistici di enti, persone, famiglie*, traduzione italiana di ISAAR (CPF): *International Standard Archival Authority Records for Corporate Bodies, Persons and Families*, a cura di Stefano Vitali, seconda edizione, 2004

<https://www.ica.org/en/isaar-cpf-international-standard-archival-authority-record-corporate-bodies-persons-and-families-2nd>

Versione in italiano:

https://www.ica.org/sites/default/files/CBPS_Guidelines_ISAAR_Second-edition_IT.pdf

ICA - International Council on Archives, *ISAD(G): General International Standard Archival Description, second edition*, 2000

https://www.ica.org/sites/default/files/CBPS_2000_Guidelines_ISAD%28G%29_Second-edition_EN.pdf

IFLA, *ISBD International Standard Bibliographic Description*, 2011, traduzione italiana a cura dell'ICCU, 2012, ISBN 9788871071312, versione PDF:

https://www.iccu.sbn.it/export/sites/iccu/documenti/2012/ISBD_NOV2012_online.pdf

IFLA, *Dichiarazione di Principi Internazionali di Catalogazione (ICP)*, 2016, traduzione italiana a cura del Gruppo di lavoro tecnico per la traduzione degli ICP,

2017, versione PDF: https://www.ifla.org/files/assets/cataloguing/icp/icp_2016-it.pdf

Ministero per i Beni e le Attività Culturali, *Norme italiane per l'elaborazione dei record di autorità archivistici di enti, persone, famiglie*, prima edizione, a cura di E. Fregni e R. Santolamazza, 2011

http://media.regesta.com/dm_0/ANAI/anaiCMS//ANAI/000/0111/ANAI.000.0111.0009.pdf

Articoli

Linked Data e Web Semantico

Adams B., Hitzler P., Janowicz K., Kolas D., Vardeman II C., *Five Stars of Linked Data Vocabulary Use*, in *Semantic Web Journal*, 2014

<http://www.semantic-web-journal.net/system/files/swj653.pdf>

Agenzia per l'Italia Digitale (AGID), *Linee guida per l'interoperabilità semantica attraverso i Linked Open Data*, 2012

<https://www.istat.it/it/files//2013/01/interoperabilitasemopendata.pdf>

Bacchi C., Bergamin G., *New ways of creating and sharing bibliographic information: an experiment of using the WikibaseData Model for UNIMARC data*, in *Rivista italiana di biblioteconomia, archivistica e scienza dell'informazione*, vol.9, n°3, 2018

<https://www.jlis.it/article/view/12458/11344>

Barbera M., *Linked data su larga scala: alcune sfide tecnologiche, ingegneristiche e sociali nell'ambito delle digital humanities, delle biblioteche digitali e dei beni culturali*, in *Rivista italiana di biblioteconomia, archivistica e scienza dell'informazione*, vol.4, n°1, 2013

<https://www.jlis.it/article/view/6333/7869>

Berners-Lee T., *Giant Global Graph*, in "timbl's blog" (webarchive.org), 2007

<https://web.archive.org/web/20160713021037/http://dig.csail.mit.edu/breadcrumbs/node/215>

Berners-Lee T., *Linked Data*, in *Design Issues (W3C)*, 2006

<https://www.w3.org/DesignIssues/LinkedData.html>

Bizer C., *The Emerging Web of Linked Data*, in *IEEE Intelligent Systems*, Vol. 24, n°5. pp. 87-92, IEEE 2009 <https://ieeexplore.ieee.org/document/5286174>

Bizer C., Heath T., Berners-Lee T., *Linked Data - The Story So Far*, 2009

<http://tomheath.com/papers/bizer-heath-berners-lee-ijswis-linked-data.pdf>

Borelli G., *Open Data e Linked Data: cosa sono e come pubblicarli*, in *Abstract*, 2014

<https://abstract.it/lab/articles/open-data-e-linked-data-cosa-sono-e-come-pubblicarli>

Byrne G., Goddard L., *The Strongest Link: Libraries and Linked Data*, in *D-Lib Magazine: The Magazine of Digital Library Research*, vol. 16, n°11-12, 2010

<http://www.dlib.org/dlib/november10/byrne/11byrne.html>

Canfora G., Di Fatta D., Pilato G., *Ontologie e Linguaggi Ontologici per il Web Semantico*, ICAR CNR (Consiglio Nazionale delle Ricerche - Istituto di Calcolo e Reti ad Alte Prestazioni), 2004,

<https://intranet.icar.cnr.it/wp-content/uploads/2016/11/TechReport-04-06.pdf>

Cecconi G., Radu C., *Open Data Maturity in Europe. Report 2018*

https://www.europeandataportal.eu/sites/default/files/edp_landscaping_insight_report_n4_2018.pdf

Coyle K., *Library linked data: an evolution*, in *Rivista italiana di biblioteconomia, archivistica e scienza dell'informazione*, vol.4, n°1, 2013

<https://www.jlis.it/article/viewFile/5443/7889>

Crupi G., *Oltre le colonne d'Ercole: Linked data e cultural heritage*, in *Rivista italiana di biblioteconomia, archivistica e scienza dell'informazione*, vol.4, n°1, 2013

<https://www.jlis.it/article/view/8587/7888>

Fusetti C., *Dati bibliografici e linked (open) data - Prospettive e criticità*, 2016, <http://dspace.unive.it/bitstream/handle/10579/9479/845606-1184567.pdf?sequence=2>

Gnoli C., *Connettere la conoscenza: quattro livelli di interoperabilità*, in *Biblioteche oggi*, vol.32, n°5, 2014, <http://www.bibliotecheoggi.it/rivista/article/view/111/395>

Guerrini M., Possemato T., *Linked Data, un nuovo alfabeto del web semantico*, in *Rivista italiana di biblioteconomia, archivistica e scienza dell'informazione*, vol.4, n°1, 2013
<https://www.jlis.it/article/view/6305/7892>

Hendler J., Berners-Lee T., *Publishing on the Semantic Web*, in *Nature*, 2001 pp. 1023-1025.
<https://www.nature.com/articles/35074206>

Hendler J., Berners-Lee T., Miller E., *Integrating Applications on the Semantic Web*, in *Journal of the Institute of Electrical Engineers of Japan*, vol. 122, n°10, 2002, p. 676-680
<https://www.w3.org/2002/07/swint>

Hendler J., Berners-Lee T., Lassila O., *The Semantic Web*, in *Scientific American*, 2001
<https://web.archive.org/web/20081204134900/http://www.sciam.com/article.cfm?id=the-semantic-web>
versione PDF:
https://www-sop.inria.fr/acacia/cours/essi2006/Scientific%20American_%20Feature%20Article_%20The%20Semantic%20Web_%20May%202001.pdf

Lamandini A., *Semantic Web*, in *Ricerche di Pedagogia e Didattica - Journal of Theories and Research in Education*, vol.6, n°1, 2011, pp. 3-5
<https://rpd.unibo.it/article/view/2248/1626>

Lo Castro V., *Web semantico e Linked Open Data: best practices, prospettive e criticità*, in *Nuovi Annali Della Scuola Speciale per Archivisti e Bibliotecari*, Firenze, Vol. 28, 2014, pp. 207-221
<https://search.proquest.com/docview/1651714733/fulltextPDF/4D4A6B18729E499EPQ/1?accountid=16773>

Lo Castro V., *Linked data nelle biblioteche digitali e di ricerca*, in *Biblioteche Oggi*, vol. 33, 2015
<http://www.bibliotecheoggi.it/rivista/article/view/5/247>

Lo Duca A., *Linked Data per i Beni Culturali: un approccio tecnico*, in *Digitalia*, vol.1, 2018
<http://digitalia.sbn.it/article/view/2046/1420>

Mastroianni F., *Nel mondo 2700 portali Open Data. La rimonta dell'Italia*, in *InfoData, Il Sole 24 Ore*, 2018

https://www.infodata.ilssole24ore.com/2018/09/20/nel-mondo-2700-portali-open-data-la-rimonta-dellitalia/?refresh_ce=1

Pace A. K., *Linked data in libraries: From disillusionment to productivity*, in OCLC.org, 2018

<http://www.oclc.org/blog/main/linked-data-in-libraries-from-disillusionment-to-productivity/>

Rizzo G., Morando F., De Martin J. C., *Open Data: la piattaforma di dati aperti per il Linked Data*, in *Informatica e diritto*, XXXVII Annata, Vol. XX, 2011, n. 1-2, pp. 493-511

http://www.ittig.cnr.it/EditoriaServizi/AttivitaEditoriale/InformaticaEDiritto/leD2011_1-2_RizzoEtAl.pdf

Shadbolt N., Hall W., Berners-Lee T., *The Semantic Web Revisited*, in *International Journal on Semantic Web and Information Systems*, IEEE 2006

https://eprints.soton.ac.uk/262614/1/Semantic_Web_Revisted.pdf

Signore O., Missikoff O., Moscati P., *La gestione della conoscenza in archeologia: modelli, linguaggi e strumenti di modellazione concettuale dall'XML al Semantic Web*, in *Archeologia e Calcolatori*, n°16, 2005, 291-319.

Solodovnik I., *Comunicazione e ricerca semantica di contenuti informativi tra Metadati, Linked open Data e Ontologie*, in *Bibliotime*, vol. XIV, n. 2, 2011

<http://eprints.rclis.org/15966/1/Metad-LOD-Ontologie.pdf>

Tammaro A. M., *Open Source, Open Access ed Open Content: verso sistemi aperti di condivisione della conoscenza*, 2005

Tammaro A. M., *Sviluppi internazionali dei linked open data (LOD) nelle biblioteche: sfide ed opportunità*, in *Digitalia*, 2015

Beni culturali

Agenzia per l'Italia Digitale (AGID), *datipubblici - linee guida cataloghi dati*, 2017

<https://www.dati.gov.it/sites/default/files/linee-guida-cataloghi-dati-profilo-dcat-ap-it-2.pdf>

Brunetti D., *Mèmore. La nuova piattaforma digitale per i beni culturali piemontesi*, in *Digitalia, rivista del digitale nei beni culturali*, vol.1, 2018,

<http://digitalia.sbn.it/article/view/2042/1416>

Cantone F., *Nuove prospettive di gestione e condivisione di dati in rete*, in *Archeologia e Calcolatori* n°16, 2005, 271-290.

Grew F., *From museum store to data warehouse: Archaeological archives for the twenty-first century*, in G. Lock, K. Brown (eds.), *On the Theory and Practice of Archaeological Computing*, Oxford University Committee for Archaeology, Oxford, 2000, pp.59-71.

Mèmore Piemonte, *Ecosistema beni culturali. Linee guida per i progetti di digitalizzazione*, 2018

http://www.memora.piemonte.it/dwd/EcosistemaBeniCulturali_LineeGuida_digitalizzazioneV01.pdf

McKenna G., *L'esperienza di linked heritage nell'ambito del linking heritage information*, in *Rivista Italiana di Biblioteconomia, Archivistica e Scienza dell'informazione*, vol.4, n°1, 2013

<https://www.jlis.it/article/view/6304/7931>

Moro L., *Mibact: "Così gli Open Data della cultura creano crescita e sviluppo"*, in *Agenda Digitale*, 2017

<https://www.agendadigitale.eu/cittadinanza-digitale/mibact-così-gli-open-data-della-cultura-creano-crescita-e-sviluppo/>

Negri A., *Catalogazione dei beni culturali in Italia: metodi, strumenti e cooperazione tra sistemi informativi per la gestione della conoscenza*, in *Revista CPC, São Paulo*, n. 21, pp. 99-118, 2016

<https://www.revistas.usp.br/cpc/article/download/111860/115882/>

Paolini P., Di Blas N., Alonzo F., *ICT per i beni culturali: esempi di applicazione in Mondo Digitale*, n°3, 2005

Scuola Normale Superiore, *Documento di sintesi del progetto tecnico-scientifico per il Portale della Cultura Italiana del Ministero per i Beni e le Attività Culturali*

http://www.culturaitalia.it/opencms/export/sites/culturaitalia/attachments/documenti/progetto/sintesi_progettotecnicoscientifico.pdf

Standard di metadati

Buttò S., *RDA: analisi, riflessioni e attività all'ICCU*, in *Rivista Italiana di Biblioteconomia, Archivistica e Scienza dell'informazione*, vol.7, n°2, 2016

<https://www.jlis.it/article/view/11798/10914>

De Robbio A., *Metadati per la comunicazione scientifica*, in *Biblioteche Oggi*, 2001

<http://www.bibliotecheoggi.it/2001/20011005401.pdf>

Doerr M., *The CIDOC CRM - an Ontological Approach to Semantic Interoperability of Metadata*, in *Research Gate*, 2003
https://www.researchgate.net/publication/240290981_The_CIDOC_CRM_-_an_Ontological_Approach_to_Semantic_Interoperability_of_Metadata

Genetasio G., *I Principi internazionali di catalogazione e il loro futuro*, in *Rivista Italiana di Biblioteconomia, Archivistica e Scienza dell'informazione*, vol.3, n°1, 2012
<https://www.jlis.it/article/view/4936/5355>

ICAR, SAN, *Interoperabilità fra sistemi archivistici: tracciati EAD3, EAC-CPF, SCONS2, ICAR-IMPORT*, 2018
http://www.icar.beniculturali.it/fileadmin/risorse/Accordi_e_convenzioni/Interoperabilita_sistemi_archivistici_tracciati_ICAR_20180925.pdf

Stato dell'arte

Abrescia A., *Verso i Linked Open Data Geografici*, in *Geomedia*, vol.16, n°4, 2012
<http://mediageo.it/ojs/index.php/GEOMedia/article/view/202/182>

Adamou A., D'Aquin M., Dietze S., Fetahu B., *The LinkedUp Data Catalogue: A Meta-Dataset of Linked Datasets in the Education Domain*, in *Semantic Web Journal*, IOS Press, 2014
<http://www.semantic-web-journal.net/system/files/swj860.pdf>

Boer V., Van Gent J., Hildebrand M., Isaac A., Oosterbroek M., Van Ossenbruggen J., Schreiber G., Wielemaker J., *Amsterdam Museum Linked Open Data*, in *Semantic web Journal*
http://www.semantic-web-journal.net/sites/default/files/swj293_2.pdf

Bozzarelli O., Mandrile V., Marangoni E., *Quale futuro per i dati di biblioteche e musei? Linked Open Data e Open Data protagonisti al Pubblico dominio #open festival di Torino*, in *Bibliotime*, anno XX, nn° 1-2-3, 2017
<https://www.aib.it/aib/sezioni/emr/bibtime/num-xx-1-2-3/bozzarelli.htm>

Casati S., Possemato T., *L'esperienza LOD (Linked Open Data) del Museo Galileo*, in *Digitalia*, vol. unico, 2015
<http://digitalia.sbn.it/article/view/1472/980>

Daquino M., Tomasi F., *Linked Cultural Objects: dagli standard di catalogazione ai modelli per il web of data. Spunti di riflessione dalla Fototeca Zeri*, in *Umanistica Digitale*, n°1, 2017
<https://umanisticadigitale.unibo.it/article/view/7195/7004>

De Marco G., *Conversione per il Semantic Web e pubblicazione nel Linked Data di dati relativi a beni artistici e culturali della regione Emilia Romagna*, Università Di Bologna, 2011

https://amslaurea.unibo.it/2360/1/demarco_giovanni_tesi.pdf

Delle Donne R., Possemato T., *SHARE Catalogue: un'esperienza di cooperazione*, in *Biblioteche Oggi*, vol. 35, 2017

<http://www.bibliotecheoggi.it/rivista/article/view/596/643>

Possemato T., *OpLiDaF Open Linked Data Framework: una piattaforma per la creazione e la pubblicazione di linked data*, in *Rivista Italiana di Biblioteconomia, Archivistica e Scienza dell'informazione*, vol. 4, n°1, 2013

<https://www.jlis.it/article/view/6313/7910>

Synapta, *Progetto LOD del CoBis*, 2017

<https://synapta.it/progetti/progetto-linked-open-data-cobis/>

Slide / altro

Linked Data e Web Semantico

Bizer C., Heath T., Hausenblas M., Hartig O., Cyganiak R., *How to Publish Linked Data on the Web*, 2008

http://videlectures.net/iswc08_heath_hpldw/

Bizer C., Heath T., Berners-Lee T., *Linked Data: Principles and State of the Art*, 17th International World Wide Web Conference, 2008

<https://www.w3.org/2008/Talks/WWW2008-W3CTrack-LOD.pdf>

Casarosa V., *I protocolli Z39.50 e OAI-PMH, Semantic Web e Linked Open Data (LOD)*, in *Biblioteche Digitali*, 2018

<http://nmis.isti.cnr.it/casarosa/BDG/>

Casarosa V., *LINKED DATA: - Come esportare il catalogo di una biblioteca nel web*, Roma, 2014,

<http://www.w3c.it/events/2014/lod2014/slides/paper55-slides.pdf>

Deng S., *Linked Data in the Library and OpenRefine*, UCF Libraries

European Data Portal, *Italy - State-of-Play on Open Data - 2018*

https://www.europeandataportal.eu/sites/default/files/country-factsheet_italy_2018.pdf

Felicetti A., *Pubblicare Linked Open Data*, 2013

<https://www.slideshare.net/culturaitalia/2013-11-29-pubblicare-linked-open-data>

Guerrini M., *Great tradition: Linked Data e SBN - Il futuro che c'è già*, Università di Firenze, 2016

https://e-l.unifi.it/pluginfile.php/245878/mod_resource/content/1/Firenze_SBN_18_ottobre_2016_Guerrini_15_10_2016_File_DEF_PDF.pdf

Iacono A., *Introduzione a Linked Open data e Web semantico*, 2016, Università La Sapienza, Roma

<https://www.slideshare.net/libriedocumenti/introduzione-a-linked-open-data-e-web-semantico-antonella-iacono>

Lo Duca A., *Linked Data: Introduction to the Semantic Web*

http://digitaltools.labcd.unipi.it/wp-content/uploads/2019/02/AngelicaLoDuca_LinkedData.pdf

Masci M. E., *I Linked Open Data nei Beni Culturali, alcuni progetti e casi di studio*, Roma, ICCU, 2013

<https://www.slideshare.net/culturaitalia/i-linked-open-data-nei-beni-culturali-alcuni-progetti-e-casi-di-studio>

Beni culturali

Buonazia I., Masci M. E., *Il PICO Application Profile. Un Dublin Core Application Profile per il Portale della Cultura Italiana*, Roma, 2007

<https://core.ac.uk/download/pdf/11876049.pdf>

Di Giorgio S., *Il portale Culturaitalia nel quadro delle collaborazioni europee*, Roma, 2016

<https://www.slideshare.net/libriedocumenti/il-portale-culturaitalia-nel-quadro-delle-collaborazioni-europee-sara-di-giorgio>

Mancinelli M. L., Veninata C., *Architettura della Conoscenza*, 2018

http://dati.beniculturali.it/wp-content/uploads/2018/01/MeetupArCoGennaio2018_1_intro.pdf

Masci M. E., *Linked Open Data (LOD): Un'Opportunità per il Patrimonio Culturale Digitale*, Roma, ICCU, 2013

<https://www.slideshare.net/culturaitalia/i-linked-open-data-nei-beni-culturali-alcuni-progetti-e-casi-di-studio>

Standard di metadati

Casarosa V,

Catalogazione - Schede bibliografiche, OPAC, MARC, Risorse nel Web - Metadati, Dublin Core, FRBR - Modello concettuale di scheda bibliografica, RDF e RDF Schema - Resource Description Framework, in Biblioteche Digitali, 2018
<http://nmis.isti.cnr.it/casarosa/BDG/>

ICCD, MiBACT, *Standard catalogafici - Principi generali - modulo 1*, 2016

<http://www.iccd.beniculturali.it/getFile.php?id=4890>

Sitografia

Linked Data e Web Semantico

All About Web, *Getting the Essence of the Semantic Web*

<http://paanchiweb.blogspot.com/2012/12/getting-essence-of-semantic-web.html> (24 marzo 2019)

Berners-Lee T., *Stop hugging your data*, 2009

https://www.youtube.com/watch?v=_J8H-ZVJwk8

CulturalItalia, *Linked Open Data*

http://www.culturalitalia.it/opencms/linked_open_data_it.jsp (18 marzo 2019)

FontiStoriche, *Linked Data: cosa sono e come si utilizzano*

<https://fontistoriche.org/linked-data/> (01 marzo 2019)

linkeddata.org (05 marzo 2019)

linkeddatatools.com (05 marzo 2019)

publicresource.org, *RFC 7-8 dicembre 2017* (14 marzo 2019)

Sapienza Università di Roma, Banca dati LOD della Biblioteca di Economia "Enrico Barone"

<https://web.uniroma1.it/bibbarone/archivionotizie/linked-open-data-lod> ,

Il web semantico oggi

<https://drive.google.com/file/d/0B2c4zjw1FuOgbE9CdkFmamIpajBlaExsVUR4ZXZVQ2NDeEtF/edit?pref=2&pli=1> (08 aprile 2019)

Casarosa V., *Knowledge representation*, in *Biblioteche Digitali*

http://nmis.isti.cnr.it/casarosa/BDG/lezioni/Know_LOD_IU-BDG_AA18-19_Casarosa.pdf (03 marzo 2019)

w3.org, *Dereferencing HTTP URIs*

<https://www.w3.org/2001/tag/doc/httpRange-14/2007-05-31/HttpRange-14> (01 maggio 2019)

w3.org, *Linked Data* <https://www.w3.org/standards/semanticweb/data> (05 marzo 2019)

w3.org, *Publications of Tim Berners-Lee,*

<https://www.w3.org/People/Berners-Lee/Publications.html> (04 aprile 2019)

Wikipedia, voce *Web Semantico*

https://it.wikipedia.org/wiki/Web_semantico (07 marzo 2019)

Wikipedia, voce *Linked Data*

https://it.wikipedia.org/wiki/Linked_data (05 marzo 2019)

Wikipedia, voce *Open Data*

https://it.wikipedia.org/wiki/Dati_aperti (05 marzo 2019)

Wikipedia, voce *Gartner*

<https://it.wikipedia.org/wiki/Gartner> (14 aprile 2019)

Wikipedia, voce *Giant Global Graph*

https://it.wikipedia.org/wiki/Giant_Global_Graph (07 marzo 2019)

Wikipedia, voce *Hype Cycle*

https://it.wikipedia.org/wiki/Hype_cycle (14 aprile 2019)

WordLift, *Linked Open Data Cloud*

<https://wordlift.io/blog/en/linked-open-data-cloud/> (01 marzo 2019)

WordLift, *Semantic Web*

<https://wordlift.io/blog/en/entity/semantic-web/> (01 marzo 2019)

Youtube, *Intro to the Semantic Web*

<https://www.youtube.com/watch?v=OGg8A2zfWKg> (08 marzo 2019)

Youtube, *What is Linked Data?*

https://www.youtube.com/watch?time_continue=1&v=4x_xzT5eF5Q (01 marzo 2019)

Youtube, *GoogleRefine How to link dataset to external sources,*

https://www.youtube.com/watch?v=99mRGlx4_QI (20 aprile 2019)

Youtube, *Google Refine with RDF extension: From Table to RDF*,
<https://www.youtube.com/watch?v=kn-YeDXR4R8> (20 aprile 2019)

Youtube, *A complete example of how to link data, part 1, 2, 3*
https://www.youtube.com/watch?v=B5CsSE7_zeQ (21 aprile 2019)

Youtube, *Reconciliation in OpenRefine, part 1, 2*
<https://www.youtube.com/watch?v=q8ffvdeyuNQ> (21 aprile 2019)

Youtube, *From Excel file to RDF with links to DBpedia and Europeana*
<https://www.youtube.com/watch?v=XdpzmGxA33U> (21 aprile 2019)

Beni culturali

CSI Piemonte, *Mèmora. La nuova piattaforma digitale per i beni culturali piemontesi*.
<http://www.csipiemonte.it/web/it/comunicazione/iniziative-ed-eventi/1346-memora-la-nuova-piattaforma-digitale-per-i-beni-culturali-piemontesi> (17 aprile 2019)

CSI Piemonte, *Mèmora. La nuova piattaforma open source per i beni culturali piemontesi*.
<http://www.csipiemonte.it/web/en/rss-csi-piemonte/16-area-media/comunicati-stampa/1344-memora-la-nuova-piattaforma-open-source-per-i-beni-culturali-piemontesi> (17 aprile 2019)

Culturaitalia, *Documentazione tecnica*
http://www.culturaitalia.it/opencms/documentazione_tecnica_it.jsp (15 marzo 2019)

Mèmora, *Manuale d'uso, versione 3, 2018*
http://www.memora.piemonte.it/dwd/manuale_memora_V03_2018.pdf (23 marzo 2019)

Regione Piemonte, *Mèmora*,
<https://www.regione.piemonte.it/web/temi/cultura-turismo-sport/cultura/ecosistema-digitale-dei-beni-culturali/memora> (17 aprile 2019)

Standard di metadati e ontologie

Accademia della Crusca, voce *Metadati/metadati*
<http://www.accademiadellacrusca.it/it/lingua-italiana/consulenza-linguistica/domande-risposte/metadati-metadati> (08 marzo 2019)

d2rq.org, *The D2RQ Mapping Language*,
<http://d2rq.org/d2rq-language>

dati.beniculturali.it, *Gli Open Data del MiBAC*

<http://dati.beniculturali.it/altre-ontologie-utilizzate/> (07 marzo 2019)

Europeana, *The Europeana Data Model for Cultural Heritage*,
https://pro.europeana.eu/files/Europeana_Professional/Share_your_data/Technical_requirements/EDM_Documentation/EDM_Factsheet.pdf (15 marzo 2019)

ICCD - MiBACT, *Standard Catalografici* su GitHub
<https://github.com/ICCD-MiBACT/Standard-catalografici> (06 marzo 2019)

ICCD, *Interoperabilità*
<http://www.iccd.beniculturali.it/it/per-condividere/interoperabilita> (15 marzo 2019)

IFLA, *MARCXML*,
<https://www.ifla.org/best-practice-for-national-bibliographic-agencies-in-a-digital-age/node/8858> (28 novembre 2018)

MokaByte, *D2R Server per mappare e importare ontologie da DB relazionali, parte I, II, III*
<http://www.mokabyte.it/2012/10/D2Rmapping-1/>

wikidata.org https://www.wikidata.org/wiki/Wikidata:Main_Page (28 aprile 2019)

Wikipedia, voce *Bibliographic Ontology*
https://en.wikipedia.org/wiki/Bibliographic_Ontology (18 marzo 2019)

Wikipedia, voce *Catalogazione dei beni culturali*
https://it.wikipedia.org/wiki/Catalogazione_dei_beni_culturali (15 marzo 2019)

Wikipedia, voce *CIDOC Conceptual Reference Model*,
https://en.wikipedia.org/wiki/CIDOC_Conceptual_Reference_Model (16 marzo 2019)

Wikipedia, voce *Dublin Core*
https://it.wikipedia.org/wiki/Dublin_Core (28 novembre 2018)

Wikipedia, voce *Encoded Archival Context*
https://en.wikipedia.org/wiki/Encoded_Archival_Context (16 marzo 2019)

Wikipedia, voce *Encoded Archival Description*
https://en.wikipedia.org/wiki/Encoded_Archival_Description (28 novembre 2018)

Wikipedia, voce *FOAF (ontology)*
[https://en.wikipedia.org/wiki/FOAF_\(ontology\)](https://en.wikipedia.org/wiki/FOAF_(ontology)) (07 marzo 2019)

Wikipedia, voce *Gemeinsame Normdatei*
https://it.wikipedia.org/wiki/Gemeinsame_Normdatei (28 aprile 2019)

Wikipedia, voce *International Standard Archival Authority Record*
https://en.wikipedia.org/wiki/International_Standard_Archival_Authority_Record (28 novembre 2018)

Wikipedia, voce *International Standard Bibliographic Description*
https://it.wikipedia.org/wiki/International_Standard_Bibliographic_Description (28 novembre 2018)

Wikipedia, voce *ISAD(G)*
[https://it.wikipedia.org/wiki/ISAD\(G\)](https://it.wikipedia.org/wiki/ISAD(G)) (28 novembre 2018)

Wikipedia, voce *MARC*
<https://it.wikipedia.org/wiki/MARC> (28 novembre 2018)

Wikipedia, voce *Metadata Encoding and Transmission Standard*
https://en.wikipedia.org/wiki/Metadata_Encoding_and_Transmission_Standard (28 novembre 2018)

Wikipedia, voce *Metadati Amministrativi Gestionali*
https://it.wikipedia.org/wiki/Metadati_Amministrativi_Gestionali (16 marzo 2019)

Wikipedia, voce *MODS*
<https://en.wikipedia.org/wiki/MODS> (28 novembre 2018)

Wikipedia, voce *NISO Metadata for Images in XML*
https://de.wikipedia.org/wiki/NISO_Metadata_for_Images_in_XML (28 novembre 2018)

Wikipedia, voce *Ontologia (informatica)*
[https://it.wikipedia.org/wiki/Ontologia_\(informatica\)](https://it.wikipedia.org/wiki/Ontologia_(informatica)) (07 marzo 2019)

Wikipedia, voce *RDF Schema*
https://it.wikipedia.org/wiki/RDF_Schema (07 marzo 2019)

Wikipedia, voce *Resource Description and Access*
https://it.wikipedia.org/wiki/Resource_Description_and_Access (17 marzo 2019)

Wikipedia, voce *Resource Description Framework*
https://it.wikipedia.org/wiki/Resource_Description_Framework (05 marzo 2019)

Wikipedia, voce *Semantically-Interlinked Online Communities*
https://en.wikipedia.org/wiki/Semantically-Interlinked_Online_Communities (07 marzo 2019)

Wikipedia, voce *Simple Knowledge Organization System*

https://en.wikipedia.org/wiki/Simple_Knowledge_Organization_System (07 marzo 2019)

Wikipedia, voce *Virtual International Authority File*
https://it.wikipedia.org/wiki/Virtual_International_Authority_File (28 aprile 2019)

Wikipedia, voce *Visual Resources Association*
https://en.wikipedia.org/wiki/Visual_Resources_Association (28 novembre 2018)

Wikipedia, voce *Web Ontology Language*
https://it.wikipedia.org/wiki/Web_Ontology_Language (07 marzo 2019)