

OntoPiA: la rete di ontologie e vocabolari controllati della PA

Aldo Gangemi^{1,2}, Raffaele Lillo³, Andrea Giovanni Nuzzolese^{1*}, Valentina Presutti¹

¹STLab, ISTC-CNR, Roma, Italia

²FICLIT, Università di Bologna, Bologna, Italia

³Team per la Trasformazione Digitale, Roma, Italia

{aldo.gangemi, andreagiovanni.nuzzolese, valentina.presutti}@cnr.it
raffaele@teamdigitale.governo.it

Abstract

OntoPiA è la rete di ontologie e vocabolari controllati modellati per offrire alla Pubblica Amministrazione un linguaggio condiviso per la pubblicazione degli open data utilizzando i principi Linked Data e FAIR. OntoPiA è il risultato di una collaborazione tra STLab, ISTC-CNR, AgID ed il Team per la Trasformazione Digitale e si colloca nel più ampio progetto denominato Data & Analytics Framework.

1 Introduzione

Il numero e la varietà di dataset rilasciati con licenze open e quindi come open data ha osservato una crescita esponenziale negli ultimi anni. In questo lavoro presentiamo la rete di ontologie e vocabolari controllati progettata e modellata per fornire alle varie PA un linguaggio comune per favorire l'interoperabilità ed il riuso degli open data attraverso le tecnologie semantiche ed i principi Linked Data [1] e FAIR [7]. Tale rete è denominata OntoPiA ed è parte di un più ampio progetto di interoperabilità semantica dell'AgID e del Team per la Trasformazione Digitale del Governo. OntoPiA definisce un insieme di ontologie e vocabolari controllati che forniscono concetti e proprietà con diversi livelli di granularità. Esse infatti spaziano da concettualizzazioni utili per modellare conoscenza di alto livello o di supporto come, ad esempio, la modellazione di ruoli, eventi, ecc..., a concettualizzazioni utili per modellare conoscenza di dominio come, ad esempio, gli eventi legati ai luoghi della cultura.

2 Definizione del problema

Il Piano triennale per l'informatica nella Pubblica amministrazione¹ identifica esplicitamente gli open data come strumento per porre in essere il processo di gestione trasparente della PA. Ciò è in ottemperanza con l'art. 1, 50, e 52 del Codice dell'Amministrazione Digitale (CAD) che definisce gli open data come elementi fondanti nel recepimento della

*Contact Author

¹https://docs.italia.it/italia/piano-triennale-ict/pianotriennale-ict-doc/it/stabile/doc/01_piano-triennale-per-informatica-nella-pa.html#premess

Direttiva europea sull'informazione nel settore pubblico². OntoPiA si colloca nel contesto del più ampio progetto denominato Data & Analytics Framework (DAF) che è il risultato di una collaborazione che vede come attore principale il Team per la Trasformazione Digitale. Il DAF ha come obiettivo quello di rendere più semplice la produzione, la pubblicazione, lo scambio ed il riuso degli open data nelle PA attraverso la standardizzazione e l'ottimizzazione di processi automatici per analisi e l'arricchimento automatico di dati. L'idea chiave del DAF è quella di aprire il mondo della Pubblica Amministrazione ai benefici offerti dalle moderne piattaforme per la gestione e l'analisi dei Big Data al fine di promuovere lo sviluppo di applicazioni *intelligenti*, nonché iniziative utili a promuovere attività di ricerca scientifica su tematiche applicative di interesse per la PA. Tuttavia, la capacità del DAF di fornire alle PA uno strumento per la gestione di tutti i processi associate alla produzione di open data integrati e pianamente fruibili è subordinata alla definizione di un formalismo che favorisca l'interscambio, il riuso ed, altresì introduca semantica nella modellazione dei dati. OntoPiA risponde proprio a questi requisiti. Inoltre, non solo favorisce la pubblicazione e l'interscambio di open data attraverso i principi dei Linked Data, ma abilita la generazione di quello che costituisce il *grafo di conoscenza* delle PA. Come accennato in precedenza OntoPiA è costituita da una rete di ontologie interconnesse progettate in OWL³ ed una serie di vocabolari controllati modellati in SKOS⁴.

3 Materiale e metodo

Le ontologie ed i vocabolari controllati di OntoPiA sono il risultato dell'analisi di una parte di quelle che sono definite come *basi di dati chiave* nell'ambito del Piano Triennale per l'Informatica nella PA. La metodologia di progettazione e modellazione delle network di ontologie adottata è basata su un'estensione di eXtreme Design (XD) [3]. XD è sia un metodo di progettazione agile di ontologie, che un tool sviluppato nel contesto del progetto EU FP6 NeOn e si basa sul riuso di Ontology Design Pattern [4] (ODP) al

²https://eur-lex.europa.eu/resource.html?uri=cellar:4e790e4c-4969-11e8-be1d-01aa75ed71a1.0010.02/DOC_1&format=PDF

³<https://www.w3.org/TR/2004/REC-owl-guide-20040210/>

⁴<https://www.w3.org/2004/02/skos/>

fine di risolvere problemi di modellazione ontologica ricorrenti. Il riuso degli ODP è ottenuto attraverso l'elicitazione dei requisiti rappresentati sotto forma di Competency Question (CQ) [5], ovvero vere e proprie domande a cui le ontologie devono fornire risposta. Le CQ emergono trasformando in domande le user story costruite analizzando le banche dati champion. Pertanto, le CQ identificano i requisiti di modellazione e, nel contesto della metodologia XD, vengono utilizzate anche come supporto per identificare potenziali ODP da riusare. Infatti, le CQ elicitate sono confrontate con i casi d'uso associati agli ODP al fine di riusare le soluzioni di modellazione proposte dagli ODP stessi. Inoltre, l'estensione di XD [6] utilizzata è focalizzata principalmente a definire e fornire linee guida chiare per il riuso di ODP e la loro organizzazione in una network. La strategia utilizzata per la progettazione della network è quella chiamata riuso indiretto degli ODP. Questo vuol dire che gli ODP sono usati esclusivamente come schemi di riferimento e replicati dalle strutture definite all'interno delle ontologie della network. Questo approccio garantisce autoconsistenza delle ontologie della network poiché non viene introdotta nessuna dipendenza con ontologie, vocabolari o pattern definiti esternamente. Tuttavia, il riuso indiretto prevede come buona pratica la definizione di assiomi di allineamento da mantenere separatamente alle ontologie della network al fine di garantire interoperabilità con le ontologie di riferimento come, ad esempio, FOAF, Schema.org, DOLCE, ecc. . .

4 Risultati e valutazione

OntoPiA conta allo stato attuale di 25 ontologie organizzate in una rete interconnessa di assiomi di tipo `owl:import` e 26 vocabolari controllati. Sia le ontologie che i vocabolari controllati sono pubblicati e versionati nel repository GitHub developers Italia⁵. Il namespace adottato per le ontologie e le loro entità è <https://w3id.org/italia/onto/>, mentre quello adottato per i vocabolari controllati e le loro entità associate è <https://w3id.org/italia/controlled-vocabulary/>. Entrambi i namespace sono definiti all'interno del dominio `w3id.org`. W3ID è il servizio gestito dal Permanent Identifier Community Group del W3C. Tale servizio è progettato per mantenere una redirectione delle URI sicura e permanente verso gli indirizzi finali. Questo vuol dire che le ontologie e i vocabolari controllati sono fisicamente salvati su un repository esterno a `w3id`, che nel caso specifico è il repository GitHub del progetto. Tuttavia, viene usato il meccanismo di redirect basato su `w3id` per la gestione permanente delle URI. Ciò permette di non cambiare gli identificativi di ontologie, classi e proprietà qualora il repository fisico dovesse cambiare in futuro. Oltre alle ontologie ed ai vocabolari controllati OntoPiA rilascia una piattaforma che fornisce una serie di strumenti per l'interrogazione e la visualizzazione orientati all'interazione utente. Tale piattaforma, denominata OntoPiA-UI, è rilasciata sotto forma di container Docker ed è anch'essa disponibile in un partizione del repository GitHub developers italia⁶.

⁵Il repository è disponibile all'indirizzo <https://github.com/italia/daf-ontologie-vocabolari-controllati>

⁶<https://github.com/italia/OntoPiA-UI>

OntoPiA-UI permette la visualizzazione delle ontologie in formato HTML attraverso gli strumenti LODE e WebVOWL, l'interrogazione della rete di ontologie e dei vocabolari controllati attraverso uno SPARQL endpoint basato su OpenLink Virtuoso, e la navigazione dell'entità ontologiche e dei termini dei vocabolari controllati attraverso LodView. OntoPiA è rilasciata con licenza CC-BY-4.0 e segue i principi FAIR, che intendono fornire delle linee guida per migliorare la reperibilità, l'accessibilità, l'interoperabilità ed il riuso di asset digitali⁷. La valutazione di OntoPiA si basa su una metodologia definita in [2]. Tale metodologia prevede tre fasi, ossia, (i) la verifica delle CQ, (ii) la verifica delle inferenze logiche e (iii) la propagazione degli errori. Le prime due fasi riguardano la corretta implementazione dei requisiti, mentre la terza fase riguarda l'analisi delle possibili vulnerabilità della rete di ontologie e dei vocabolari controllati. Quest'ultima analisi serve, non solo a identificare casi in cui OntoPiA fornisce modellazioni efficaci, ma anche a identificare potenziali casi di modellazione inefficace o erronea.

5 Conclusioni

In questo articolo è stata presentata OntoPiA, la rete di ontologie e vocabolari controllati che implementano il linguaggio per favorire la produzione del grafo di conoscenza delle PA attraverso le tecnologie semantiche ed i principi FAIR. OntoPiA nasce da una collaborazione tra STLab, ISTC-CNR, AgID ed il Team per la Trasformazione Digitale e si colloca nel più ampio progetto denominato Data & Analytics Framework.

Riconoscimenti. Giorgia Lodi di AgID ha contribuito alla progettazione e allo sviluppo delle ontologie e dei vocabolari di OntoPiA.

References

- [1] Christian Bizer, Tom Heath, and Tim Berners-Lee. "Linked Data - The Story so far". In: *International Journal on Semantic Web and Information Systems* 5.3 (2009), pp. 1–22.
- [2] Eva Blomqvist, Azam Seil Sepour, and Valentina Presutti. "Ontology testing-methodology and tool". In: *Proc. of EKAW 2012*. (Galway, Ireland). Springer, 2012, pp. 216–226.
- [3] Eva Blomqvist et al. "Experimenting with eXtreme Design". In: *Proc. of EKAW 2010*. (Lisbon, Portugal). Springer, 2010, pp. 120–134.
- [4] Aldo Gangemi and Valentina Presutti. "Ontology design patterns". In: *Handbook on ontologies*. Springer, 2009, pp. 221–243.
- [5] Michael Grüninger and Mark S Fox. "The role of competency questions in enterprise engineering". In: *Benchmarking - Theory and practice*. Springer, 1995, pp. 22–31.
- [6] Valentina Presutti et al. "The role of Ontology Design Patterns in Linked Data projects". In: *Proc. of ER 2016*. Springer, 2016, pp. 113–121.
- [7] Mark D Wilkinson et al. "The FAIR Guiding Principles for scientific data management and stewardship". In: *Scientific data* 3 (2016).

⁷FAIR è l'acronimo inglese per Findability, Accessibility, Interoperability, and Reuse.