
David Stuart

**Practical Ontologies
for Information
Professionals**

New York, Neal-Schuman
Publishers, 2016, 224 p.

Practical Ontologies for Information Professionals ha tre obiettivi principali: dimostrare ai professionisti dell'informazione l'importanza delle ontologie per la scoperta della conoscenza, evidenziare l'importante contributo che essi possono dare allo sviluppo di ontologie e infine fornire un'introduzione pratica allo sviluppo di ontologie. Il capitolo introduttivo è utile per documentare ontologie consolidate e ampiamente utilizzate e il ruolo dei professionisti dell'informazione in questo processo, idee che vengono argomentate e rafforzate in tutto il resto del libro. Le ontologie in realtà rappresentano solo uno dei vari sistemi di organizzazione della conoscenza

sviluppati dai professionisti dell'informazione per migliorare il reperimento delle risorse informative e vengono dette anche *tassonomie* o *vocabolari controllati* a seconda del settore in cui sono utilizzate (più nel settore commerciale il primo termine e culturale il secondo). Il libro privilegia la seconda definizione sia per la potenziale confusione causata dal duplice significato delle tassonomie sia per il profilo professionale dell'autore.

Il primo capitolo fa il punto sulla terminologia: i vocabolari controllati sono inevitabilmente soggettivi e riflettono il punto di vista del creatore in un particolare momento e risulta poi molto difficile rettificarne in futuro l'impostazione. Le intestazioni di soggetto sono un set controllato di termini che descrivono il soggetto o l'argomento di una risorsa (libro, articolo, data set) mentre gli *authority files* sono set di intestazioni di soggetto preferite. Un thesaurus, come una tassonomia nel senso più stretto del termine, fornisce relazioni gerarchiche (o di equivalenza o associative) tra i concetti.

La più usata definizione di ontologia è quella di Gruber³ "una specificazione esplicita di una concettualizzazione", molto criticata sia per la sua vastità sia per il focus su concetti soggettivi piuttosto che entità reali. Ma cosa può fare un'ontologia? Hedden⁴ identifica i tre scopi principali: supporto all'indicizzazione, supporto al recupero dell'informazione e supporto all'organizzazione e alla navigazione. Questo libro viene pubblicato in un momento cruciale della storia delle ontologie: da una parte il web e lo sviluppo delle tecnologie per il semantic web hanno fornito l'op-

portunità per le ontologie di essere adottate da più persone in più posti rispetto a prima, raccogliendo i dati a livello mondiale in un data set che può essere interrogato da chiunque, dall'altra gli ideali di un semantic web si sono dovuti adattare agli aspetti pratici delle abilità umane, riconoscendo l'importanza di pubblicare i dati anche se non accompagnati da robuste ontologie formali. I professionisti dell'informazione possono dare un grosso apporto allo sviluppo delle ontologie, nell'ottica del doppio esperto di progettazione di ontologie e di domini di soggetto. Per gli stessi tassonomisti ed ontologisti, per i quali lo sviluppo e il mantenimento di vocabolari controllati può essere un lavoro full-time, anche la conoscenza e l'esperienza di ontologie è necessaria come parte di un più ampio set di abilità nella catalogazione e nei metadati. Il focus delle ontologie in questo libro è su come sono state usate nel web semantico.

Il secondo capitolo fornisce un'introduzione al web semantico e al ruolo delle ontologie, e a come esse siano state sempre più adottate in una vasta gamma di biblioteche, così come in altre istituzioni culturali e organizzazioni commerciali. Il web ci dà una gran quantità di informazione ma questa è destrutturata e richiede molto sforzo per trovare e analizzare ciò che ci serve: se si ricerca un'informazione complessa si fa riferimento a siti che forniscono informazioni aggregate (piuttosto che costruire la ricerca da soli) ma non tutte le informazioni sono aggregabili o semplicemente sono informazioni mancanti. Ciò nonostante progressi sono stati fatti non solo stabi-

lendo nuove specifiche ma anche stabilendo un nuovo paradigma per la pubblicazione dei dati: i *linked data*. Questi danno priorità alla pubblicazione dei dati in un formato leggibile dalla macchina piuttosto che dei concetti sottostanti e la semplicità dell'approccio ha incoraggiato la pubblicazione di dati sul web semantico da parte di molti individui e organizzazioni. Il capitolo descrive il *Resource Description Framework* (RDF) un modello concettuale per fare *dichiarazioni* sulle risorse attraverso triplette RDF, ossia un modo di esprimere e relazionare le informazioni come dichiarazione in tre parti, strutturate in un semplice formato *soggetto-predicato-oggetto*. Per rendere esplicite le triplette possono essere usati gli URI per rappresentare ogni risorsa (nel web semantico ogni cosa rappresentabile è definita risorsa): una volta che soggetti, predicati e oggetti sono disambigui possono essere combinati in un singolo grafico che rappresenta la *query* per il computer. Dopo una panoramica su classi, sottoclassi, proprietà, identificatori e set di caratteri (URI e Unicode), si analizza la sintassi, l'XML (*Extensible Markup Language*), un linguaggio di marcatura leggibile sia dalla macchina sia dall'uomo. Sebbene RDF/XML non sia l'unico formato usato per condividere dati sul web, è di gran lunga il più consolidato e raccomandato. RDFS (*Resource Description Framework Schema*) fornisce il vocabolario necessario a creare il primo livello di vincoli, fornendo un linguaggio comune per la descrizione delle classi, delle sottoclassi e delle proprietà all'interno e tra di esse. Ma anche RDFS ha limiti, per esempio mentre è possibile sostenere che qualcosa è membro di una

certa classe non è possibile stabilire che sia membro anche di un'altra classe: per fare questo è necessario ricorrere a OWL (*Web Ontology Language*) che fornisce un linguaggio ontologico espressivo per il web semantico ed è diventato una raccomandazione del W3C nel 2004. L'inferenza non deve essere basata su molteplici relazioni ma può essere costruita su un semplice predicato. Rispetto a RDFS, OWL permette un'inferenza più complessa costruita su relazioni più espressive oltre a permettere *dichiarazioni* riguardo a una classe che è disgiunta da un'altra. In ambito bibliotecario c'è stato molto interesse nel pubblicare i dati secondo i principi dei *linked data* e ci sono stati eventi dedicati (come *IFLA 2014 Satellite Meeting Linked data for libraries* <http://ifla2014-sat-data.bnf.fr/program.html>) e diversi libri. Il focus primario della comunità bibliotecaria nella pubblicazione di *linked data* è la pubblicazione di risorse bibliografiche e molti dei dati che sono stati pubblicati hanno preso la forma di record catalografici o metadati bibliografici. La *British National Bibliography* (BNB)⁵ è stata resa disponibile online dalla British Library come *linked data* e ci sono molti altri esempi illustri (OCLC, *DDC system*). Rendere i metadati bibliografici disponibili come *linked data* ha molti vantaggi soprattutto se pubblicati con una appropriata licenza *open* che ne permette il riutilizzo (per esempio *Creative Commons* o *Open Data Commons Attribution License* di OCLC). I bibliotecari sono consapevoli che le risorse non sono solo disponibili ma vengono usate; "books are for use" è la prima legge della biblioteconomia di Ranganathan e "data is for use" può essere considerato una versione aggiorna-

ta della legge purché i ricercatori siano in grado di usare tali dati. Ciò significa assicurare che i dati (o più specificatamente le ontologie) siano appropriatamente formattati e abbiano i metadati necessari in modo che possano essere trovati, assicurino il riutilizzo di vocabolari esistenti, abbiano documentazione sufficiente da essere usata e siano pubblicati apertamente con una licenza appropriata in modo da essere riutilizzabili.

Il terzo capitolo considera alcune delle principali ontologie, comprese quelle utilizzate per rappresentare le ontologie stesse, quelle ampiamente adottate dalle biblioteche e quelle largamente utilizzate sul web, dal momento che sono davvero tante quelle pubblicate soprattutto in relazione a progetti di nicchia per specifici domini. Non vengono analizzate ontologie associate a progetti di ricerca specifici in quanto non ampiamente adottate, e ontologie che sono altamente specifiche per soggetto: unica eccezione la *Bible Ontology* poiché costituisce un esempio di trasformazione di testo narrativo in ontologia e fornisce un interessante confronto con le altre ontologie del capitolo, tutte - a eccezione di *DBpedia - element set*. Dopo una breve nota relativa alla documentazione sulle ontologie, il capitolo si divide in quattro parti: la prima copre le più importanti ontologie usate per costruire le ontologie stesse (RDF, RDFS, SKOS, OWL2). La parte successiva considera alcune delle ontologie associate con biblioteche e ruoli informativi più tradizionali, rappresentando libri e altri tipi di risorse. Questa parte è seguita da uno sguardo all'ontologia superiore *Basic Formal Ontology*⁶ e due

dei *data model* dominanti del patrimonio culturale *Europeana Data Model*⁷ e *CIDOC-Conceptual Reference Model*.⁸ Infine vengono considerate le ontologie più adottate nel web. Nel far uso di ontologie esistenti il professionista dell'informazione si può trovare di fronte a due problemi: la mancanza di documentazione sull'ontologia che si sta utilizzando o al contrario un eccesso di documentazione, dovuto in genere alla necessità di disambiguazione dell'ontologia. Laddove i diagrammi delle ontologie devono migliorare la chiarezza o la comprensione fanno uso dei *building blocks* di *VOWL* (*Visual Notation for OWL Ontologies*):⁹ cerchi o più spesso ellissi (per la lunghezza delle etichette) sono usati per rappresentare le classi mentre le linee rappresentano le relazioni di proprietà con frecce che puntano all'oggetto di una tripletta. Il primo esempio riportato per rappresentare le ontologie stesse è RDF e RDFS che formano la base del web semantico. *SKOS* (*Simple Knowledge Organization System*)¹⁰ è di particolare interesse per la struttura di molti dei sistemi di organizzazione della conoscenza esistenti in ambito culturale, come intestazioni di soggetto, schemi di classificazione, thesauri e altre tassonomie. Fornisce un semplice vocabolario per rappresentare le collezioni di concetti e permette relazioni tipo il thesaurus per esprimerle (tipicamente *skos:broader*, *skos:narrower*, *skos:related*). Sebbene sia molto adottato, ha anche notevoli limitazioni: ad esempio non può sistemare tutte le informazioni contenute negli *authority records* in formato MARC che sono ampiamente adottati nelle biblioteche. OWL, il *Web*

Ontology Language, diventato una raccomandazione del W3C nel 2004, è stato soppiantato da OWL2 nel 2009. È uno strumento che va molto al di là della creazione di semplici ontologie informali. Sebbene abbia molte potenzialità che non vengono utilizzate appieno, è necessario che i professionisti dell'informazione ne conoscano alcune di base in quanto possono tornare utili alla creazione di ontologie più semplici. Esistono diverse sintassi specifiche per OWL: le sintassi per descrivere le ontologie possono essere separate in due gruppi, quelle che descrivono il grafico che descrive l'ontologia e quelle che descrivono l'ontologia stessa. Per quanto riguarda le biblioteche il Dublin Core è l'*element set* che richiede meno spiegazioni. L'originale *Dublin Core Metadata Element Set* consiste di 15 proprietà generiche per descrivere le risorse online; nel tempo si è evoluto e ora consiste di ben 55 termini definiti come proprietà RDF. Sebbene Dublin Core sia criticato per la sua semplicità questa caratteristica l'ha reso ampiamente adottato rispetto ad altre ontologie e profili di applicazione. In alcuni casi è usato per includere metadati riguardo all'ontologia stessa e rappresenta lo standard di metadati di default per molte piattaforme di pubblicazione. La *Bibliographic Ontology*,¹¹ basata su standard di metadati esistenti, è stata specificamente progettata per descrivere citazioni e risorse bibliografiche nel web semantico. In aggiunta a molte classi e proprietà dei set esistenti sono stati definiti nuovi elementi per definire specifici tipi di documenti (manuali, manoscritti, brevetti, tesi) e le relazioni tra i documenti e

gli agenti (per es. direttore, editore, intervistatore). Un importante sviluppo nella rappresentazione degli oggetti bibliografici è stato l'elaborazione del modello FRBR (*Functional Requirements for Bibliographic Records*) che rappresenta gli oggetti bibliografici e le loro relazioni con l'entità responsabile associata. Un catalogo costruito su FRBR non descrive solo l'oggetto bibliografico che la biblioteca possiede ma le sue relazioni con altre cose nell'"universo bibliografico". Fornisce un modello pratico per comprendere le relazioni tra i diversi tipi di oggetto che possono essere correlati in differenti database bibliografici: distingue tra tre gruppi di entità, quelle che rappresentano il contenuto intellettuale, gli agenti associati con il contenuto e il soggetto del contenuto. FRBR dispone di vocabolari specifici per rappresentare un oggetto bibliografico cosicché differenti vocabolari possono essere usati da diverse comunità di utenti o per raggiungere diversi livelli di dettaglio pur aderendo al modello FRBR.

Con l'adozione dei social media e del web 2.0 molta informazione di valore circola nel web e pertanto alcune delle più famose ontologie sono state create per codificarne e condividerne i dati. L'ontologia FOAF (Friend Of A Friend) è una delle più usate nel semantic web. Progettata inizialmente per descrivere persone e organizzazioni sul web e collegarle insieme, molti dei termini sono stati adottati in molte altre ontologie e dispone di un generatore di pagine FOAF-a-Matic.¹² Lo stesso Tim Berners-Lee nel 2007 sosteneva "I express my network in a FOAF file, and this is a start of the revolution"¹³ anche

se la rivoluzione è stata lenta a venire. Schema.org è forse oggi il più importante *element set* del web: non solo è progettato specificamente per il web ma anche per i motori di ricerca. È stato lanciato nel 2011 da Google, Yahoo e Bing, progettato per creare vocabolari per ogni sorta di informazioni che si possono trovare nei siti web, specialmente laddove è importante avere un'informazione strutturata e aggregata. Infine uno sguardo a *Facebook Open Graph Protocol* (OGP) creato per rappresentare il contenuto web all'interno di grafici *social* attraverso l'aggiunta di metadati per tipi di contenuti quali video, audio e immagini come anche le stesse pagine web. L'importanza di Facebook sul web rende OGP uno dei vocabolari online più di successo.

Il riutilizzo di ontologie esistenti è importante sia per l'integrazione dei dati attraverso sistemi diversi sia per evitare di ripetere un lavoro già fatto. Il quarto capitolo considera gli strumenti disponibili per identificare ontologie esistenti, come ontologie (o loro componenti) che possono essere combinate nella creazione di profili di applicazione, e alcuni dei criteri che dovrebbero essere considerati quando si seleziona un'ontologia. Si apre sottolineando l'importanza del riuso e alcuni dei problemi che possono contribuire nella decisione di adottare o meno una certa ontologia. In ambito bibliotecario il sistema di riuso delle ontologie è ben consolidato: molte biblioteche trovano più appropriato adottare o estendere un sistema di classificazione o di catalogazione esistente piuttosto che crearne uno su misura (che ha i vantaggi di un sistema customizzato ma di gran lunga su-

perati dai costi di mantenimento). Le principali ragioni di riuso delle ontologie esistenti sono le stesse del riuso di sistemi di organizzazione della conoscenza già in uso: efficacia dei costi e interoperabilità dei dati. Lo sviluppo di un *element set* logicamente coerente con classi e proprietà chiaramente definite che possono coprire tutti i tipi di istanze non è un processo semplice mentre l'adozione di una stessa istanza su diverse piattaforme può facilitare il recupero dell'informazione da molteplici siti e fonti. Per semplicità comunque il libro considera l'efficacia dei costi in termini di ontologie con un vasto numero di istanze e il vantaggio di adottare *element set* esistenti per l'interoperabilità dei dati. La creazione di un'ontologia di alta qualità è estremamente laboriosa specialmente quando tratta di soggetti estesi e richiede anni per essere sviluppata. Anche la creazione di un *element set* può richiedere molto tempo specialmente se vi vanno incorporate molte classi o proprietà. La reale importanza di riusare un'ontologia esistente è comunque, specialmente per le ontologie meno formali, l'interoperabilità dei dati. L'adozione di un *element set* impiegato vastamente e la relativa informazione codificata garantisce una maggiore comprensione da parte dei servizi esterni (se l'informazione è resa pubblica) ed è più probabile la portabilità in altri software (si veda ad es. Dublin Core). Un *application profile* è uno schema che combina termini da uno o più *element set* per un'applicazione particolare ed è un approccio pratico per implementare ontologie: fare uso di ciò che è disponibile piuttosto che fissarsi su ciò che soddisfi

tutti i criteri di uno schema. Dopo tutto nella maggior parte delle istanze un'ontologia che copra esattamente tutti i requisiti di un implementatore non è disponibile. Anche se esiste un'ontologia creata per un particolare argomento, potrebbe non includere necessariamente le stesse proprietà o andar bene per un particolare livello di granularità. Ma come si identifica la giusta ontologia? Attualmente esistono a questo scopo *ontology libraries*, o si utilizza un motore di ricerca specifico o anche quelli generici. Una *ontology library* è un sistema web che fornisce accesso a una collezione estensibile di ontologie con lo scopo primario di mettere gli utenti nelle condizioni di trovare ed usare una o più ontologie da tale collezione.¹⁴ Il termine collezione sottende un elemento di cura rispetto a un'ontologia o un motore di ricerca basato su stringhe automatiche di web semantico. In realtà la distinzione tra i due non è sempre così chiara: una *ontology library* può far uso di una certa quantità di indicizzazione automatica e un motore di ricerca al contrario può avere molta accuratezza. Esistono online diverse *ontology libraries*, che coprono diverse aree e hanno diversi livelli di funzionalità. Il libro ne analizza quattro: *Linked Open Vocabularies* ospitata presso la OPEN Knowledge Foundation, *BARTOC.org* (*Based Register of Thesauri, Ontologies & Classifications* - www.bartoc.org) che fornisce un registro di vocabolari controllati e strutturati arricchiti con la DDC e intestazioni di soggetto da Eurovoc (<http://eurovoc.europa.eu>) e che prevede che non tutti i vocabolari usati per costruire ontologie siano classificati come ontologie (per

esempio Dublin Core e Bibframe sono classificati come vocabolari controllati), *Taxonomy Warehouse* (www.taxonomywarehouse.com) che ha funzionalità base per la consultazione di 666 vocabolari scorribili in ordine alfabetico o ricercabili attraverso un'interfaccia di ricerca ed infine *BioPortal* (<http://bioportal.bioontology.org>) che è un esempio di *ontology library* specializzata fornendo un *repository* di ontologie biomediche ed include anche strumenti di valutazione delle ontologie ed informazioni sui progetti che le stanno utilizzando. Rispetto alle ontologie i motori di ricerca semantici (*Swoogle, Watson, Falcons*) assicurano informazioni aggiornate ma generano ontologie di scarsa qualità e limitano i metadati che possono essere estratti automaticamente. L'uso invece di un motore di ricerca generico quale Google presenta più svantaggi che vantaggi: tra i primi l'uso incoerente del termine "ontologia", l'inadeguatezza delle ontologie nell'essere indicizzate da un motore di ricerca e l'inadeguatezza degli indici del motore per scovare le ontologie. D'altra parte ha la possibilità di scoprire ontologie che non sono indicizzate nel web semantico e nonostante le sue carenze Google è stato usato come base per la raccolta dati in uno studio di FOAF. È dunque necessario tener conto di alcuni criteri di selezione ed il libro ne suggerisce alcuni. In primis la *chiarezza* ossia la capacità di un'ontologia di trasmettere il significato dei suoi termini associati. Sfortunatamente le ontologie spesso necessitano della definizione di termini altamente tecnici o termini multipli con sottili differenze tra essi e poiché pochi creatori di

ontologie sono esperti lessicografi raramente le definizioni sono chiare per l'utente. La *coerenza* si riferisce alla congruenza dell'ontologia mentre l'*estendibilità* riguarda la possibilità per gli utenti di estendere un'ontologia per le proprie specifiche esigenze. L'*adattabilità* fa riferimento a come un'ontologia esistente (o un elemento) incontra i requisiti dell'ontologia richiesta ed è il fattore più importante. Il *formato* si riferisce al modo in cui l'ontologia è stata codificata.

Il quinto capitolo fornisce un'ampia metodologia per la costruzione di una ontologia ed una panoramica di alcuni degli strumenti disponibili, prima di guidare il lettore attraverso lo sviluppo di una semplice ontologia con Protégé, il software più popolare (e gratuito) per lo sviluppo di ontologie. Vengono distinti tre tipi di metodologie per lo sviluppo di ontologie: metodologia "auto-riflettente", un approccio collaborativo e metodologie empiriche. La prima è la più semplice: una singola persona è responsabile per lo sviluppo di un'ontologia basata sulla conoscenza che possiede e questo tipo va bene se si tratta di un settore altamente specializzato con una limitata condivisione di dati. Laddove c'è una storia più consolidata di condivisione di dati o il dominio è troppo esteso per una singola persona per comprendere diverse prospettive allora un approccio collaborativo può essere più appropriato. La metodologia empirica è basata sui dati contenuti nel documento ed ha il vantaggio di essere più aderente ai termini usati piuttosto che a quelli che *dovrebbero* essere usati. Viene dettagliata una metodologia composta di dodici step, basata sugli undici step del

progetto CENDARI¹⁵ un progetto di ricerca relativo al medioevo e alla prima guerra mondiale la cui metodologia è basata su quattro precedenti metodologie ognuna delle quali enfatizza aspetti differenti del processo di sviluppo dell'ontologia. Quella di Uschold e King's è una delle prime metodologie generali e consiste di quattro stadi: identificare lo scopo dell'ontologia, costruire l'ontologia, valutarla e documentarla. Questo in linea di massima l'approccio per qualsiasi ontologia a cui si possono eventualmente aggiungere ulteriori dettagli. Il capitolo si conclude con un esempio di ontologia costruita con il software Protégé secondo i dodici step sopra indicati. Il sesto capitolo fornisce una panoramica degli strumenti disponibili per interrogare ontologie del web semantico per acquisire nuove conoscenze, sia attraverso il *Simple Protocol* e il linguaggio di interrogazione di RDF (SPARQL) sia attraverso i motori di ricerca. Le ontologie possono essere interrogate da utenti diversi in modi differenti ed esiste un'ampia gamma di strumenti di varia complessità per interagire con differenti ontologie ma queste sono rappresentazioni formali della conoscenza e per estrarne il pieno valore richiedono *queries* formali. Le ontologie e i dati di istanza associati sono generalmente interrogati per tre ragioni: determinare se un'ontologia è adatta per il riuso; estrarre informazioni dall'ontologia e/o dal data set; raccogliere informazioni riguardo l'uso di un'ontologia. Il capitolo analizza queste ragioni e discute alcune delle tecnologie più appropriate per ognuna. Il modo migliore per indagare un'ontologia è leggere la documentazione relativa ed esplorarla in uno degli strumenti

per sviluppare o pubblicare ontologie menzionati nel capitolo 5 (TopBraid o Protégé). Tutte le ontologie sono formali ma alcune lo sono più di altre: da quelle con pochi vincoli a quelle con rigide regole. Il web semantico non è altro rispetto al web che si usa ogni giorno, anzi ne è una parte sempre in crescita. Milioni di persone interagiscono con contenuti semantici tramite i motori di ricerca senza rendersene conto. Tali applicazioni sono pensate per l'utente generico con semplici *queries* o interfacce a linguaggio naturale: questo ovviamente limita la complessità delle *queries* che possono essere create. Per *queries* complesse è necessario un linguaggio più formale, nello specifico SPARQL. Creare *queries* SPARQL è molto più complesso che inserire parole chiave in Google e si basa su una comprensione dell'*element set* dell'ontologia associato. Il libro fornisce solo una introduzione guidando il lettore da esempi più semplici a esempi più complessi, basandosi su tre ontologie pubbliche ognuna con la query SPARQL cosicché il lettore può inserire e adattare le *queries* alle sue esigenze. Esistono molte fonti di informazione su SPARQL compresi tutorial online e lo standard stesso.¹⁶ Il capitolo finale guarda al futuro delle ontologie e al ruolo dei professionisti dell'informazione nel loro sviluppo e utilizzo. Il futuro delle ontologie sarà senza dubbio una miscela di ontologie leggere e più formali, e il loro sviluppo sarà integrabile con altre tecnologie, come *Natural Language Processing* e potenziali flussi di lavoro *crowdsourcing*. Il primo passo che le biblioteche possono fare per impegnarsi nel campo delle ontologie è fare uso di ontologie esistenti piuttosto che crearne di nuove. Le ontologie do-

vrebbero essere inserite tra i molti ruoli dei professionisti dell'informazione al pari della catalogazione e della classificazione come pure la pubblicazione di dati e la raccolta di nuovi indicatori di impatto. Una *best practice* su quali dati debbano essere pubblicati ed il modo migliore di pubblicarli deve ancora emergere. C'è una gran varietà di modi di pubblicazione dei dati e non tutti necessitano di essere codificati in un'ontologia o *knowledge base* sebbene queste abbiano un importante ruolo per assicurare che i dati siano accessibili attraverso la creazione di metadati durante il processo di catalogazione. In molti casi comunque i dati sono adattabili per la pubblicazione come *linked data* ma mentre alcuni grandi progetti di ricerca hanno persone dedicate alla pubblicazione dei dati in altri casi viene considerato un aspetto secondario. I professionisti che operano all'interno di organizzazioni commerciali e biblioteche speciali hanno un ruolo chiave nello sviluppo di ontologie che rispondano alle necessità specifiche del loro ambito. Come dimostra *Facebook's Open Graph Protocol* una singola organizzazione commerciale può avere molto successo nella promozione di una particolare ontologia ma ciò non significa che questa sia particolarmente ben progettata e ampiamente applicabile. Tuttavia tale ontologia può essere largamente usata perché è disponibile e non richiede alcun ulteriore

sviluppo. Rimane aperta la questione del ruolo futuro delle ontologie e della loro natura (formale, informale, aperta o chiusa ad esempio). Quanto all'accesso, ossia se le ontologie debbano essere aperte o chiuse, ci si sta muovendo verso un prevalere di risorse ad accesso aperto e su queste per lo più si focalizza il libro anche se molto numerose sono quelle ad accesso chiuso per esempio quelle di imprese private che forniscono accesso a intranet private e risorse di vasto pubblico come la *Knowledge Vault* di Google. Comunque mentre c'è senza dubbio necessità di condividere ontologie pubbliche, c'è anche necessità di trovare modelli che rendano lo sviluppo delle ontologie sostenibile dal punto di vista finanziario. In realtà quelle proposte sono false dicotomie, non è una scelta di sviluppo di ontologie aperte o chiuse, formali o informali, centralizzate o distribuite: le ontologie continueranno ad avere una gran varietà di forme ed i professionisti dell'informazione avranno modo di lavorare nelle aree che più si confanno ai propri interessi ed abilità. Infatti il semantic web è progettato per supportare molteplici soluzioni.¹⁷

DANIELA CANALI

Dipartimento Sviluppo
Comune di Terni
daniela.canali@comune.terni

NOTE

³ THOMAS R. GRUBER *A Translation Approach to Portable Ontology Specification*, *Knowledge Acquisition*, 5 (1993), 2, p. 199-220.

⁴ HEATHER HEDDEN, *The Accidental Taxonomist*, Medford, Information Today Inc., 2010.

⁵ www.bl.uk/bibliographic/datafree.html#lod.

⁶ <http://ifomis.uni-saarland.de/bfo>.

⁷ <http://pro.europeana.eu/page/edm-documentation>.

⁸ <http://www.cidoc-crm.org>.

⁹ <http://vowl.visualdataweb.org/v2>.

¹⁰ www.w3.org/2004/02/skos.

¹¹ <http://bibliontology.com>.

¹² www.ldodds.com/foaf/foaf-a-matic.html.

¹³ BERNERS-LEE TIM, GIANT GLOBAL GRAPH, *Timbl's blog*, <http://dig.csail.mit.edu/breadcrumbs/node/215>.

¹⁴ NATALYA F. NOY, MATHIEU D'AQUIN, *Where to publish and find ontologies?: a survey of ontology libraries*, "Journal of Web Semantic science, services and agents on the world wide web", 11 (2012), p. 96-111.

¹⁵ Guidelines for Ontology Building, www.cendari.eu/sites/default/files/CENDARI%20_6.3%20Guidelines%20for%20Ontology%20Building.pdf.

¹⁶ *SPARQL 1.1 Query Language* (2013), www.w3.org/TR/sparql11-query.

¹⁷ DEAN ALLEMANG, JAMES HENDLER, *Semantic Web for the Working Ontologist: effective modeling in RDFS and OWL*, 2nd edition, Morgan Kaufman, 2011.

DOI: 10.3302/0392-8586-201805-058-1