

Linked data for cultural heritage

Edited by Ed Jones, Michele Seikel

London, Facet Publishing,
2016, 134 p.

Nell'introduzione a *Linked data for cultural heritage*, Ed Jones ricostruisce una breve tabella sui livelli di interoperabilità e condivisione dei dati disponibili nella rete internet, riassumendo una lettura del fenomeno proposta nel 2010 da Tim Berners-Lee.¹ Il World Wide Web Consortium, che stabilisce le regole e propone le nuove *realises* per il web e per internet, ha sviluppato modelli, principi, tool e componenti che fanno in modo che il linking sia più dinamico per il consumo e l'uso di identificatori universali di risorse (Uris). Le Uris possono essere accettate automaticamente per ottenere un più pertinente livello di informazione e i link possono essere recuperati o possono essere fatte inferenze che generino un'informazione ulteriore o più appropriata. Questa impostazione non è estranea ai bibliotecari, da quando a partire dagli anni '60 (almeno per la biblioteconomia angloamericana) hanno incominciato a collegare le informazioni senza i tool e i principi odierni. Avvalendosi dello standard MARC, la library community ha sviluppato lo standard per condividere record fra biblioteche e consorzi in modo da ridurre i costi della catalogazione e supportare la condivisione delle risorse; ha, inoltre, risposto sino ad oggi alle esigenze di un'utenza tradizionalmente rivolta alla ricerca a catalogo (OPAC) e alla bibliografia. Le applicazioni del modello RDF, che sottende al web semantico, rappresentano gli strumenti necessari a un miglioramento dello standard MARC in grado di garantire un globale e integrato ambiente di discovery dei materiali conservati in

biblioteca, oggetto di particolare attenzione da parte di utenti (spesso studiosi e ricercatori) e bibliotecari: le collezioni che conservano incisioni, volantini e manifesti, cartoline; articoli e citazioni da e-journals e riviste scientifiche, solo per nominare alcuni esempi.

Secondo la scala proposta nell'illustrazione 1, un uso appropriato di linked data si attesta in termini di qualità al livello più alto, perché stabilisce che lo schema RDF è in grado di cambiare con frequenza e di essere estremamente flessibile: appoggiandosi a un'infrastruttura, è garantito l'aggiornamento continuo delle informazioni che hanno vita finché l'ontologia che li sostiene è aggiornata. Il web semantico ha la grande possibilità di rivoluzionare il modo in cui gli utenti e i computer interagiscono con l'informazione in internet: a una impostazione verticalistica dell'informazione, si impone una struttura orizzontale dei dati, sempre vagliati da un contesto di fatto sempre più open. Il web semantico ha due caratteristiche peculiari: ha l'abilità di portare informazione (dati e contenuto) e riproporla rispetto al suo intento originario; ha l'abilità di espandere la ricerca e di mappare i vocabolari. Inoltre, i mainframe di thesauri e vocabolari riducono i tempi di aggiornamento e *maintenance* di cataloghi e archivi e propongono un modello informativo sostenibile e flessibile a cui partecipano utenti, studiosi, bibliotecari, *publishers* e *providers*. Questa impostazione, che prevede una differente organizzazione del lavoro (termine che ritorna con frequenza nei testi proposti da Ed Jones e Michele Seikel) perché sensibile alla divisione delle attività per competenze, trasforma il catalogo e l'archivio (silos): predilige, infatti, le interconnessioni e l'interoperabilità degli archivi, riducendo la presenza di dati non strutturati a favore della loro granularità e accessibilità, evitando la folksonomy propria della

rete per aprire lo spazio controllato della tassonomia dove è privilegiato l'uso consapevole di mashup e tag cloud. L'ontology, il linguaggio per la rappresentazione della conoscenza, è l'oggetto di studio delle ricerche proposte in *Linked data for cultural heritage* e stabilisce le norme per la trasformazione e l'uso di thesauri e vocabolari: ESTC21, per esempio, estende il dataset rappresentato nel catalogo catturando informazione da item che si trovano fuori dalla tradizionale bibliografia, quali la citazione; Elsevier Merged Medical Taxonomy propone la mappatura di concetti dalle principali tassonomie mediche e favorisce l'arricchimento dei contenuti dell'archivio; nel 2004 AGROVOC ha testato il passaggio a Web Ontology Language (OWL) e nel 2009 è stato reso consultabile come linked data in RDF/SKOS-XL; Linked Jazz si configura come un caso di studio rappresentativo sviluppato in uno scenario reale, quello degli artisti jazz; i principi della tecnologia linked data, ampliano la visibilità, l'interpretazione e l'uso delle informazioni. Il dataset, interlinkato con risorse esterne, rivela l'esistenza di una comunità profondamente interconnessa; allinea i termini del vocabolario, assicurando che siano riutilizzabili dove applicabili e che nuovi termini siano integrati con quelli esistenti.

Il modello RFD rappresenta un approccio agnostico all'universo dei dati e olistico alla conoscenza e non chiede aderenza a una particolare prospettiva ontologica: è una struttura agnostica caratterizzata da una adeguata conformità a un principio di ordine flessibile e minimalista. Linked data ha la possibilità di espandere l'identità di ciò che può essere inteso come eredità culturale, riducendo il gap tra collezioni online e outline, ossia non ancora riconosciute e accettate da vocabolari e thesauri. La flessibilità di un archivio supportato dallo schema RDF permette

al catalogo di catturare informazioni su intestazioni che sono estranee allo scopo della tradizionale bibliografia: particolare attenzione è rivolta nel libro all'opera di adattamento artistico-letterario, dove il lavoro di traduzione (*translation*) e di riduzione richiedono il recupero di dati necessari (*translators*) a comprenderne il lavoro creativo (*creative work*). Uno studio, questo, contiguo al bibliotecario e al ricercatore perché diventa sempre più importante la storia e la provenienza dell'intestazione (voce) che, se disponibile, entra a far parte del catalogo. Questo è reso possibile dal fatto che linked data usa vocabolari controllati come il bibliotecario applica le norme dell'authority control, per permettere di comprendere i benefici che offre su altri tipi di modelli di dati. Inoltre, questa somiglianza tra RDF e infrastruttura bibliotecaria fornisce un *mainframe* che integra le pratiche bibliotecarie con i modelli linked data. Proprietà come equivalenza e inferenza, criteri che sottendono allo schema RDF, superano il concetto di archivio come deposito e anticipano il riuso e il remixing dei dati recuperati dalla richiesta degli utenti. Ricerche, queste, che vanno oltre la biblioteca e che possono fornire informazioni contestuali, migliori ricerche e future esplorazioni. Queste ricerche sono generate da vocabolari controllati perché 1) i vocabolari controllati sono meccanismi per mezzo dei quali le risorse sono linkate tra loro; 2) le informazioni tassonomiche a cui si accede per HTTP Uris (generate dal modello RDF) forniscono altre informazioni. Il documento, nella sua manifestazione che ne fa una traduzione, un adattamento, è un costrutto sociale che è prodotto di una comunità che ha differenti interessi, livelli di esperienza e richieste di verità scientifica e che coopera attraverso un network, una rete, un brand istituzionale. Il vocabolario, con le sue voci di authority, si configura come prodotto sociale,

supportato da buone pratiche e da principi di uniformità che accrescono su scala globale il valore della navigazione e del browsing.

Linked data for cultural heritage propone e immagina i prossimi sviluppi di tool e risorse linked data nell'ambito dell'eredità culturale. Lo fa partendo da quanto si è già progettato e ponendo delle criticità su quanto si deve ancora raggiungere per migliorare vocabolari, thesauri, authority file. Se la pubblica amministrazione ha già avuto beneficio da dati openness e linking, l'originalità di questa ricerca è sostenuta dal fatto che questi progetti nascono nell'ambito dell'eredità culturale la cui prima finalità è quella di conservare e preservare il sapere sempre più minacciato da guerra e catastrofi.

It is essential to monitor areas at risk, to advocate for and raise awareness about conflict and disaster prevention. With an increase in cultural heritage being abused for political propaganda and destroyed to serve certain agendas, the protection of cultural heritage has never been more important.²

Proprio in questo ambito il fabbisogno informativo è più alto e sempre di più la comunità scientifica è prosumer, perché maggiore è la necessità della riproducibilità dei dati scientifici, maggiore il vaglio e l'approccio critico all'informazione delle voci di authority, maggiore la richiesta della granularità e dell'interoperabilità dell'informazione, maggiore il bisogno di accessibilità e il riutilizzo.

I principi di linked data, inoltre, identificano il primato degli oggetti del mondo reale (*artifacts*) e interpretano il web come un corpus de facto per quegli oggetti che richiedono uno sforzo collaborativo e un range di expertise per arrivare a stabilire l'identità dell'oggetto stesso.

L'idea che i testi raccolti propongono è quella di una divisione del lavoro

dove l'input consisterebbe nello stimolare a distinguere fra termini che sarebbero formalmente proposti per il vocabolario, che dovrebbe essere a sua volta costantemente aggiornato. Questa osservazione stabilisce quale differenza esista tra il linguaggio comune e quello degli esperti ed è utile in due casi: 1) il lavoro scientifico è supportato da una comunità di ricercatori; 2) è necessario il mantenimento a lungo termine e altre funzioni di biblioteca che siano in grado di rispondere alle richieste degli utenti (come BIBFRAME team della Library of Congress per esempio).

Il volume, tuttavia, non si risparmia nel sottolineare quanto deve ancora essere compreso per un uso completo di tool e risorse di linked data: quanto i *publishers* e gli autori hanno intenzione di investire per rendere i dati accessibili (via API e SPARQL) in linked data e sviluppare la tecnologia RDF? Quanto vogliono di fatto essere autentici prosumers, in modo tale che le voci di authority prevalgano su quelle descrittive, facendo maggiore attenzione al contenuto del documento e non alla sua descrizione, come già prevedono FRBR e RDA? Per ora ancora poche sono le richieste su come i dati sono rappresentati e su come possono essere disponibili e rappresentabili. Ciò richiede la presenza di partner che sarebbero in grado di aggiungere responsabilità nella pubblicazione, *governance* e *maintenance* di liste, vocabolari e altre risorse linked data. Quali passi devono essere fatti per creare dei *linkages* in modo tale da essere implementati e riutilizzabili? Ciò include un insieme di tratti della catalogazione, quali la conoscenza di standard correnti in linked data e nel web semantico, oltre all'abilità di comprendere i casi d'uso e fare valutazioni critiche su risorse che sono appropriate e sostenibili.

Rilevanti, in questo senso, sono due studi: il primo sulle implicazioni in ambito medico scientifico di linked data (*Linked Data Implications for Au-*

thority Control and Vocabularies): l'autore spiega quali siano i criteri di selezione per i vocabolari e i modelli di dati. Il secondo sulle distinzioni che gli autori propongono tra Work dello schema RDF e Manifestation del modello FRBR e RDA in ambito artistico letterario (*A division of labor*): l'esempio è tratto dall'adattamento della fiaba *Nussknacker and Mausekönig* di E.T.A. Hoffmann sia nella forma della traduzione inglese sia in quella del balletto di P. Tchaikovsky, sia in quella di prodotto audiovisivo. Gli autori, volendo dimostrare quanto sia importante comprendere le implicazioni nell'uso delle proprietà e dei risultati delle inferenze per aggiungere un uso ottimale dei cataloghi, intendono comunque stimolare i *vendors* a riconoscere sottoclassi nel modello RDF che si possano sovrapporre alle norme FRBR e RDA, per ottenere un migliore accrescimento informativo. Forse un'utopia, ma forse un'utopia possibile e inevitabile per lo scambio informativo e il fabbisogno di crescita della conoscenza, oltre che al suo riutilizzo.

Questo scenario prevede dunque un cambio culturale, perché questo modello rende possibile a differenti comunità di collaborare senza il bisogno di coordinamento, poiché la condivisione di convenzioni di comune rappresentazione non richiede un consenso a priori. Linked open data porta in biblioteca, nei musei e negli archivi un nuovo modo di pensare l'oggetto dell'informazione e i suoi collegamenti su entità e relazioni che hanno prodotto il documento piuttosto che sul documento stesso. La possibilità di influenzare dati testuali non strutturati, in aggiunta ai metadati bibliografici, può aumentare la ricchezza della tecnologia Linked open data, ma richiede un processo di sviluppo e un insieme strutturato di strumenti non di prossima attuazione.

Oltre all'introduzione, il libro si compone di sei contributi nei quali

rintracciare le criticità sovraespresse:

- Hilary Thorsen (Stanford University) e M. Cristina Pattuelli (Pratt Institute) esaminano l'uso di linked data in alcuni significativi progetti dell'ambito della cultura digitale, compresi European e Digital Public Library of America, sino a illustrare un più dettagliato uso di Linked Jazz, un progetto di ricerca mirato all'uso della tecnologia linked data al fine di stabilire significative connessioni fra documenti e dati della biografia personale e professionale di artisti jazz e descrivere lo sviluppo di tool e metodologie linked data;
- Carl Stahmer (University of California, Davis) ripropone i passaggi della migrazione di English Short Title Catalog da MARC a linked data e le possibilità che si dipanano da questa migrazione, specialmente in termini di coinvolgimento di una comunità più ampia di studiosi che approvino e accrescano i metadati ESTC21;
- Allison Jai O'Deill (University of Florida) porge uno sguardo retrospettivo e pone questioni sul futuro dei thesauri bibliotecari, dei metadati, della ricchezza e adeguatezza dell'informazione, studiando come i vocabolari integrino la pratica bibliotecaria con linked data ed esplorando le pratiche esistenti che sono ricondotte a linked data, così come l'area di espansione di migliori pratiche nello sviluppo di linked data;
- Iker Huerga (Signifikance) e Michael Lauruhn (Elsevier Labs) esaminano linked data e authority control dalla prospettiva STM, descrivendo il ruolo dell'authority control, degli identificatori e vocabolari, includendo l'uso di Web Technology Language per accrescere il significato formale e l'uso di SPARQL per creare mappature fra i vocabolari;
- Carol Jean Godby (OCLC) descri-

ve le sperimentazioni di OCLC's e Schema.org come base per un modello di descrizione di risorse bibliotecarie espresse in forma di linked data. La Godby conclude dicendo che "OCLC's experiments have shown that Schema.org can be used to define a model... which can be expressed in a published standard which institutional backing and potential for widespread adoption";

- Sally McCallum (Library of Congress) illustra lo sviluppo del modello di dati di Bibliographic Framework Initiative (BIBFRAME), il successore di linked data per il modello di dati rappresentato dal formato MARC21, descrivendo le differenze fondamentali tra MARC e BIBFRAME. BIBFRAME è particolarmente adatto al formato di scambio per dati bibliografici creati usando Resource Description and Access.

Ogni capitolo è corredato di una adeguata bibliografia per l'approfondimento del tema, mentre la postfazione è dedicata alla biografia degli autori.

Un libro a cui bisogna prestare attenzione per le domande che gli autori si pongono e che testimonia l'esigenza di coinvolgere le istituzioni pubbliche e private nella tutela delle risorse e dei contenuti conservati e fruibili nell'ambito dell'eredità culturale, in termini di openness linking e produzione di conoscenza. Five stars!

ANTONELLA COSTANZO

costanzo.antonella@tiscali.it

NOTE

¹ Tim Berners-Lee, *Is Your Linked Data 5 Stars?*, in *Linked Data*, www.w3.org/DesignIssues/Linkeddata.html.

² <https://www.ifla.org/cultural-heritage>.

DOI: 10.3302/0392-8586-201801-060-1