

Creazione e ricerca dei dati nell'era digitale

MAURO GUERRINI

Università degli studi di Firenze
mauro.guerrini@unifi.it

TIZIANA POSSEMATO

Direttrice @Cult
tiziana.possemato@acult.it

L'implementazione di RDA e linked data: BIBFRAME e Olisuite/WeCat

1. I dati delle biblioteche nel web semantico

Il web è il luogo in cui si manifestano e si svolgono molti aspetti della vita professionale e privata delle persone; esso ha rivoluzionato il modo di stabilire le relazioni e di fruire le informazioni disponibili e può rivoluzionare il modo di scoprire, accedere, integrare e utilizzare i dati per le sue caratteristiche di ubicità (virtuale) e per la sua natura distribuita e scalabile (modulare). L'assunto è alla base della riflessione sulle modalità e sulle strategie di diffusione dell'informazione nella nostra epoca. RDA, Resource Description and Access, parte da questo assunto e si propone come il nuovo standard che fornisce istruzioni e linee guida per la metadattazione (la versione contemporanea della catalogazione) delle risorse nel web e per facilitare la loro scoperta. Suoi riferimenti teorici sono il modello relazionale FRBR, Functional Requirements for Bibliographic Records, e ICP, International Cataloguing Principles.¹ RDA si rivolge a biblioteche, archivi, musei e ad altri istituti culturali che conservano la memoria registrata e che intendano affrontare il fondamentale passaggio dal catalogo e dall'inventario inteso come insieme di record (web tradizionale) a una nuova struttura informativa intesa come insieme di dati (web semantico) assemblati tra loro e visualizzabili secondo innumerevoli modalità.

La tradizione di pubblicare i dati del catalogo in formato elettronico e nel web è radicata nella storia delle biblioteche come in poche altre comunità. Il formato MARC nasce a metà degli anni Sessanta per lo scambio di record catalografici tra istituzioni diverse; la diffusione del formato in ambito specificatamente biblioteconomico è stata enorme: a settembre 2015, 347.400.683 è il numero di record in formato MARC dichiarati solo

da OCLC. Il nome esteso del formato MARC, MACHINE Readable Cataloguing, indica la sua caratteristica di essere letto e interpretato dalle macchine. Il record in MARC riproduce, tuttavia, quasi fedelmente la tipologia e l'organizzazione delle informazioni del catalogo a schede, da cui eredita altre caratteristiche: la natura confinata entro il dominio di nascita (quello delle biblioteche), la struttura monolitica e testuale delle informazioni in esso contenute (ciascun dato è comprensibile solo all'interno dell'intero record), l'incapacità di esprimere il ricco complesso di relazioni logicamente esistenti tra le entità ed entità esterne. Queste caratteristiche rendono il record bibliografico e d'autorità in formato MARC poco o per niente conforme a quanto richiesto dagli attuali paradigmi del web, affinché i dati del catalogo siano *del web* e non più solo *nel web*, e che i record bibliografici e d'autorità siano fruibili da parte di agenti elettronici. Esiste, infatti, una distanza da colmare tra la realtà dei dati prodotti dagli istituti culturali e le tecniche per la realizzazione del web semantico; occorre che le biblioteche compiano un adeguamento tecnologico e soprattutto di mentalità.

Lo scenario di spazio globale rappresentato dal web, in cui le informazioni sono lette, manipolate e fruite da agenti umani ed elettronici, genera un nuovo atteggiamento in chi produce fonti informative: se in passato l'attenzione era sulle esigenze locali o riferite a una comunità specialistica, adesso ogni decisione dovrebbe essere assunta considerando il contesto globale di condivisione. Ciò che viene prodotto dovrebbe essere usato da chiunque in qualsiasi comunità e in qualsiasi contesto: l'orizzonte si sposta dallo specifico dominio al mondo intero.

La necessità del cambiamento è stata formulata dalle

biblioteche in varie forme e occasioni. La più significativa nella direzione del web semantico è *On the record*, il report della Library of Congress Working Group on the Future of Bibliographic Control,² reso pubblico il 9 gennaio 2008, in cui si richiama alla necessità di:

- trasformare la descrizione testuale in set di dati utilizzabili per processi ed elaborazioni automatiche da parte di macchine;
- rendere gli elementi di dati univocamente identificabili all'interno del contesto informativo del web;
- assicurare la compatibilità dei dati con le tecnologie e gli standard del web;
- usare un linguaggio trasversale e interoperabile nella realtà del web.

La dichiarazione di orientamento verso il web semantico è chiara e ufficializzata, confermando quanto anticipato con la pubblicazione di schemi dati e ontologie in RDF, Resource Description Framework,³ già dal 2005.

2. Dal metadato al dato significativo

Da dove parte l'analisi sul record bibliografico e perché esso, così strutturato e storicamente formalizzato, si è rivelato inadeguato nel contesto del web semantico? La storia dei cataloghi delle biblioteche e degli archivi mostra un impiego antico e diffuso dei metadati, intesi come *informazioni vicarie* della risorsa. Nell'accezione tradizionale del termine un metadato è un dato su un altro dato: nel record descrittivo di un libro o di un film, per esempio, i metadati sono il titolo, il nome dell'autore o del regista, l'anno di pubblicazione ecc. Le caratteristiche principali dei metadati nell'accezione tradizionale sono:

- la natura artefatta, costruita sulla risorsa;
- la finalità descrittiva;
- l'utilizzabilità da parte di macchine – il metadato deve essere strutturato in modo che sia processabile e, dunque, utilizzabile da un computer.

I campi del record utilizzati per esprimere il concetto di metadato rispondono in modo completo alle prime due caratteristiche: sono costruzioni aggiunte all'oggetto, che è la risorsa, e hanno la finalità di descriverla, cioè di fornirne i dati descrittivi. L'utilizzabilità da parte di macchine, nell'accezione tradizionale di web, rimanda alla sola caratteristica di essere *letti* dalle macchine. Perché si parli di web semantico questa terza caratteristica deve essere estesa nel significato, abbracciando l'accezione

di *comprensibilità* da parte delle macchine. Il concetto di metadato, nel contesto del web semantico, estende, pertanto, il suo significato e diventa sinonimo di “dato processabile e interpretabile da una macchina”: le strutture sintattiche del metadato creano così quello strato del web che rende espliciti ad agenti elettronici i significati delle risorse che descrivono e delle relazioni con altre risorse.

3. Record bibliografico vs RDF

La registrazione catalografica, intesa come un insieme di metadati descrittivi della risorsa, è sempre stata al centro dell'attenzione e delle attività delle biblioteche. Il passaggio dal catalogo cartaceo al catalogo elettronico, che ha avuto la massima espressione con la nascita del formato MARC a metà anni Sessanta, con il relativo accrescimento dell'interesse verso gli standard e le formattazioni dell'informazione (per esempio, con l'ISBD, dalla fine del medesimo decennio), ha amplificato la centralità del record, che è diventato il catalizzatore di una serie di servizi differenti offerti a utenti di vario tipo, dai programmi di catalogazione partecipata ai processi di prestito interbibliotecario, alla nascita delle reti cooperative. La centralità del record ha, tuttavia, comportato una serie di inconvenienti:

- la produzione di versioni varianti di record standard, per rispondere a esigenze locali di singole biblioteche, non sempre giustificabili;
- la redazione di record molto dettagliati, di alta qualità ma con elevato costo di produzione;
- la creazione di *silos chiusi* e per lo più non usabili al di fuori del mondo delle biblioteche.

Nel web semantico il record MARC è considerato una fonte d'informazione preziosa per l'enorme quantità di *dati* che contiene e per l'elevato grado di *semantica* in esso presente: il set di metadati raccolto per descrivere la manifestazione *As you like it* di Shakespeare, è ricchissimo e ogni tag rimanda a uno specifico *significato* del campo (esempio riportato a pagina seguente).

La semantica espressa nel record è comprensibile solo all'occhio umano o da macchine che riconoscono la struttura MARC. I *singoli elementi informativi* (vedi sotto) perdono di significato se isolati dal record:

- xv, 240 p.
- Includes bibliographical references (p. 240).
- Cambridge (UK).

Esempio di record MARC

LC Control No.:	2009025024
LCCN Permalink:	http://lccn.loc.gov/2009025024
000	02384cam a2200445 a 450
001	15789102
005	20100204124659.0
008	090624s2009 enka b 000 0 eng
020	__ a 9780521519748 (hbk.)
020	__ a 0521519748 (hbk.)
020	__ a 9780521732505 (pbk.)
020	__ a 0521732506 (pbk.)
035	__ a (OCoLC)ocn416716652
040	__ a DLC c DLC d YDXCP d BWKUK d BWK d FDA d UKM d BTCTA d CEF d CDX d DLC
050	00 a PR2803.A2 b H35 2009
082	00 a 822.3/3 2 22
100	1_ a Shakespeare, William, d 1564-1616.
245	10 a As you like it / c edited by Michael Hattaway.
250	__ a Updated ed.
260	__ a Cambridge, UK ; a New York : b Cambridge University Press, c 2009.
300	__ a xv, 240 p. : b ill. ; c 24 cm.
490	1_ a New Cambridge Shakespeare
504	__ a Includes bibliographical references (p. 240).
505	0_ a Introduction. Journeys ; Plays within the play ; Theatrical genres ; Pastoral ; Counter-pastoral ; The condition of the country ; Politics ; 'Between you and the women the play must please' ; Gender ; Nuptials ; Sources ; Date and occasion ; Stage history ; Recent critical and stage interpretations -- Note on the text -- List of characters -- The play -- Textual analysis -- Appendixes : 1. An early court performance? ; 2. Extracts from Shakespeare's principal source, Lodge's <i>Rosalind</i> ; 3. The songs.
650	_0 a Fathers and daughters v Drama.
650	_0 a Exiles v Drama.
600	10 a Shakespeare, William, d 1564-1616. t As you like it.
655	_7 a Pastoraldrama. 2 gsafd
655	_7 a Comedies. 2 gsafd
700	1_ a Hattaway, Michael.

- New York.
- Cambridge University Press, 2009.
- edited by Michael Hattaway.
- Updated ed.
- Fathers and daughters - Drama.

Trasferito nel web tradizionale, gran parte del *significato* del record in formato MARC viene perduto; molte informazioni sono trattate dai motori di ricerca come *blocchi di testo* e, dunque, non comprese singolarmente e inequivocabilmente. L'intero record d'altra parte, con

la sua complessità e ricchezza, diventa ri-usabile solo da chi trasferisce sul proprio sistema una copia del medesimo record per descrivere la medesima risorsa; nel caso, per esempio, di una nuova edizione della stessa opera, la normativa catalografica richiede la creazione di un nuovo record, che potrebbe differire dal precedente solo per il campo edizione e data. Da ciò possiamo immaginare la moltiplicazione smisurata dei record all'interno di ciascun catalogo. Inoltre i record pubblicati in formato MARC sono stati usati pressoché esclusivamente nell'ambito delle biblioteche; la comunità degli editori, per esempio, utilizza pochissimo questo formato per le procedure di scambio dati con le biblioteche o i fornitori di servizi affini.

Nel contesto del web semantico, gli *elementi singoli* possono riacquistare il significato che l'intero record MARC con i tag conferiva loro, trasformando i dati in *linked open data* e cioè in asserzioni RDF con un soggetto, un predicato e un oggetto. Le asserzioni RDF consentono alle macchine di *capirne il significato*, facendo compiere il salto dal web tradizionale al web semantico. Nella figura 1 è possibile vedere alcuni elementi originali di una registrazione MARC, tradotta in asserzioni RDF: l'*asserzione* (o tripla) nel modello RDF è una breve frase formata da un soggetto (nell'esempio la risorsa denominata "As you like it") un predicato ("has subject") e un oggetto ("Fathers and daughters - Drama"). La risorsa "As you like it" ha diversi attributi (autore, editori, data pubblicazione ecc.) espressi tramite triple di soggetti, predicati e oggetti. Per esempio, l'attributo editore può essere formulato con la tripla: "As you like it" (soggetto), "a cura di" (predicato), "Michael Hattaway" (oggetto). Tale dichiarazione è disponibile per il riutilizzo nel web semantico ed è particolarmente utile quando l'istruzione incorpora URI. Ogni tripla così prodotta (per esempio, "As you like it a cura di Michael Hattaway") è strutturata in un linguaggio che la macchina può capire. Ogni elemento della tripla, che può essere identificato da un URI (o, in caso di tripla "letterale", da un soggetto e un predicato identificati da un URI, e l'oggetto costituito da un testo letterale) può essere "dereferenziato" dalla macchina. Ciò significa che la macchina può leggere una definizione formulata nell'ontologia per ogni entità e per ogni predicato impiegati nella tripla e comprenderne il significato.

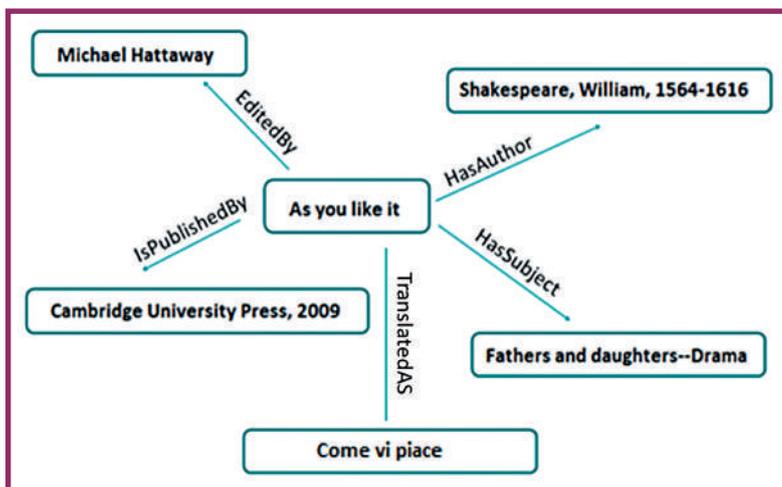


Figura 1 – Elementi e asserzioni RDF per una descrizione bibliografica nel web

Ciascun elemento dell'asserzione (soggetto - predicato - oggetto) così prodotta può essere registrato con un URI. Il trasferimento nel web di queste asserzioni comprensibili alle macchine trasforma i dati contenuti negli originari record MARC in *parti integranti del web semantico*.

4. Nuovi modelli catalografici e linked data

L'adozione delle logiche e delle tecnologie del web semantico può avere un impatto differente nei vari contesti di adozione: pubblicare open data o linked open data in un contesto industriale, per esempio, può significare registrare i propri dati su tabelle excel o convertire dei database in RDF, senza probabilmente cambiare i processi produttivi del dato stesso. Nel caso delle biblioteche il passaggio sta generando un ripensamento profondo del modello del catalogo finora conosciuto, producendo la rifondazione dei principi, degli standard e delle norme. La comunità scientifica biblioteconomica ha predisposto nuovi riferimenti teorici: gli ICP, emanati dall'IFLA nel 2009, parlano di dati e non più di record, come parla di dati FRAD, *Functional Requirement for Authority Data*, rispetto a FRBR che parlava di record. Roy Tennant, nel 2002, in un articolo pubblicato su "Library journal" dal titolo significativo *MARC must die*,⁴ sancisce l'inizio di un nuovo corso evolutivo che si conclude con la dichiarazione ufficiale della morte di quel formato: la pagina wiki del MARC è stata significativamente (e autoironicamente) trasformata da *MARC must die!* in *MARC is dead*.⁵

RDA, concepito inizialmente come un'evoluzione di AACR2, *Anglo-American Cataloguing Rules 2nd edition*, apre

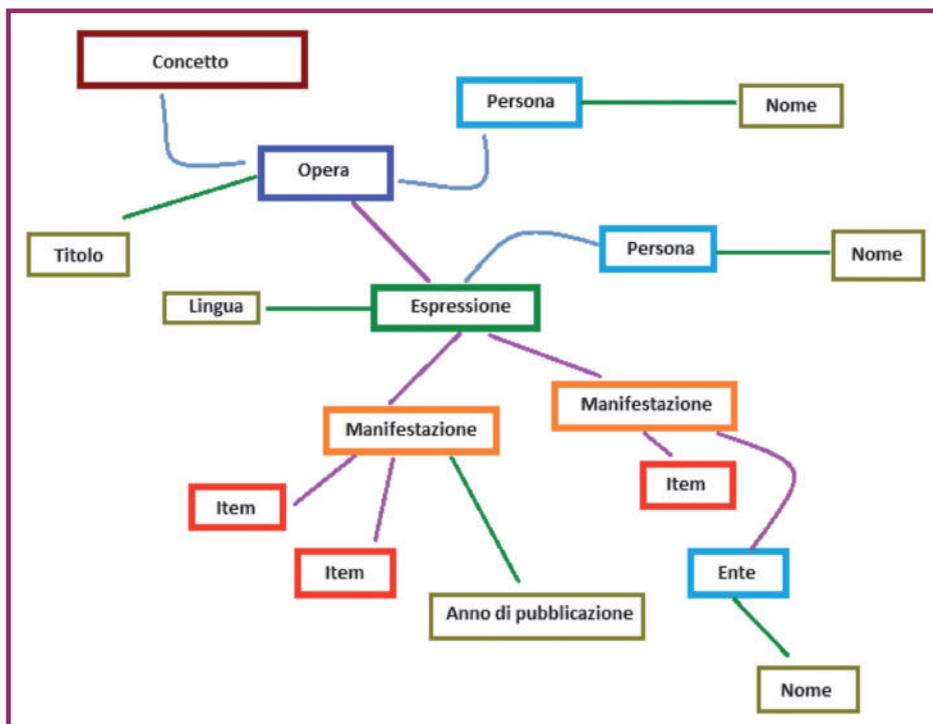


Figura 2 – Modello di FRBR proposto in RDA come schema per l'evoluzione delle norme in termini di linked data

nuove prospettive presentandosi come standard internazionale per la metadatanazione delle risorse nell'era digitale. Le linee guida (e non più un codice di catalogazione) si muovono all'interno della filosofia del web semantico e dei linked data. RDA incoraggiano, per esempio, la pubblicazione di ontologie e vocabolari controllati in RDF, così come dedicano molto spazio alla costruzione di dati di qualità (record d'autorità) che possano essere riutilizzati in contesti diversi e per molteplici scopi. Il modello di FRBR, che si esprime in *entità, relazioni e attributi* (gli *elementi* di RDA) è proposto nelle linee guida come uno schema dati perfetto per l'evoluzione delle norme e degli standard in termini di linked data.

Lo schema in figura 2 mostra le entità, con alcune relazioni e attributi possibili. L'immagine esprime il modello a grafo delle risorse nel web semantico, con una particolare attenzione alle relazioni esplicite tra risorse.⁶ La tappa più significativa e di maggiore rilievo per la comunità catalografica e per il futuro del controllo bibliografico universale è stata la riflessione partita dalla sperimentazione di RDA dal 2010, con l'inadeguatezza, avvertita sempre più consapevolmente, del formato piatto del record MARC21 a esprimere le *relazioni* tra entità bibliografiche considerate fondamentali da FRBR e RDA. A partire da questa conclusione, la Library of

Congress ha annunciato nel maggio 2011 la *Bibliographic Framework Transition Initiative*, dichiarando, nell'ottobre successivo: "The new bibliographic framework project will be focused on the Web environment, Linked Data principles and mechanisms, and the Resource Description Framework (RDF) as a basic data model".⁷

La transizione dal web tradizionale al web semantico con l'adozione ufficiale della tecnologia a esso collegata è sancita. Una testimonianza interessante della transizione è la nascita del W3C Library Linked Data Incubator Group, il cui scopo è sostenere e valorizzare l'interoperabilità globale dei dati

delle biblioteche nel web, favorendo il confronto tra esperti del web semantico.⁸ Con il report *Bibliographic Framework as a Web of Data: Linked Data Model and Supporting Services* (BIBFRAME),⁹ pubblicato dalla Library of Congress il 21 novembre 2012, la prospettiva è tracciata.¹⁰

5. BIBFRAME

BIBFRAME è il risultato di un percorso di riflessione sulle funzioni del catalogo, sull'attualità del MARC e delle norme catalografiche e sulle nuove tipologie di risorse presenti nell'universo bibliografico; esso rappresenta il primo passo nella definizione di una *roadmap* che dovrebbe proporre un nuovo ambiente bibliografico, che prevede l'interconnessione di entità generate da fonti differenti e l'enunciazione di strategie e strumenti capaci di supportare l'evoluzione in atto nel circuito della comunicazione globale. Dovrebbe, in particolare, portare all'enunciazione di un nuovo *ecosistema bibliografico*. Le nuove istruzioni per la descrizione e l'accesso alle risorse dovrebbero caratterizzarsi per:

- un alto livello di analisi e *identificazione* del dato;
- un'enfatizzazione delle *relazioni*;
- un impiego diffuso dei *vocabolari controllati*;
- un accurato dettaglio descrittivo delle risorse;

- un'ampia flessibilità nella gestione delle *voci controllate*.

Il modello esamina l'evoluzione degli standard e delle norme catalografiche e, dunque, compie una riflessione sul ruolo, la diffusione e le funzioni storiche del MARC 21 e sul futuro del controllo bibliografico universale. L'elaborazione teorica si basa sull'esperienza maturata durante la fase di test di RDA. La possibile conversione dei dati bibliografici in linked data viene interpretata in termini di evoluzione piuttosto che di rivoluzione. Le riflessioni nate da questa sperimentazione di RDA sono approdate a considerazioni di opportunità, se non di necessità, di utilizzo della rete come

modello per esprimere e connettere le informazioni, secondo il paradigma del web semantico, ovvero tramite la struttura di URI e attraverso la decentralizzazione e atomizzazione del dato. I risultati sperati del processo evolutivo includono la possibilità di giungere a una semplificazione e ottimizzazione dei processi descrittivi delle risorse che permettano, da una parte, la condivisione della descrizione al di fuori della comunità bibliotecaria, dall'altra la possibilità che il lavoro di catalogazione sia distribuito a livello internazionale tra soggetti diversi, così da estendere la copertura del controllo bibliografico e, quindi, soddisfare al meglio i bisogni degli utenti. BIBFRAME offre un modello dati estremamente semplice, ispirato a FRBR (ma con differenze significative); propone di:

- distinguere tra contenuto concettuale e manifestazione fisica, cioè tra opera e istanza dell'opera;
- identificare un'entità in maniera non ambigua;

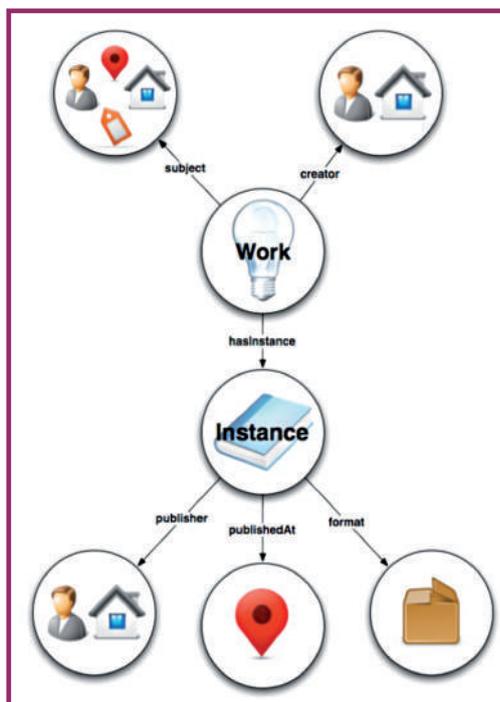


Figura 3 – BIBFRAME: struttura generale del modello

magini di copertina, recensioni, TOC ecc.

L'immagine riprodotta in figura 3 è estratta da BIBFRAME e rappresenta la struttura generale del modello con l'opera al centro del grafo, collegata all'istanza in cui si concretizza e con le relazioni a esse associate.

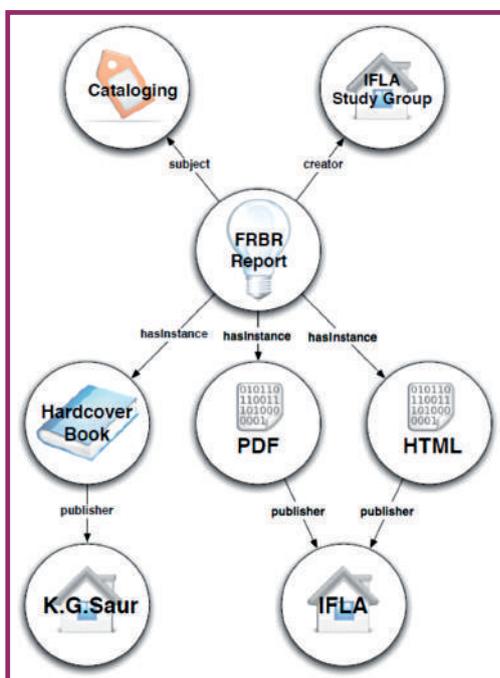


Figura 4 – BIBFRAME: l'esempio dell'opera FRBR Report

- evidenziare le relazioni tra le entità mostrando la natura della correlazione.

Il modello è strutturato in 4 classi principali:

- *opera*: l'essenza concettuale della risorsa catalogata, il cui concetto è assimilabile all'*opera* e all'*espressione* di FRBR;
- *istanza*: una materializzazione individuale dell'opera;
- *authority*: l'identificazione delle entità che hanno una relazione con l'opera o con l'istanza, per esempio, persone, famiglie, enti, concetti, luoghi;
- *annotazioni*: tutto ciò che decora o arricchisce le risorse del modello BIBFRAME con informazioni aggiuntive, per esempio, dati di esemplare, im-

Al centro c'è l'opera con link al creatore e alla voce di soggetto, ovvero al descrittore che esprime il tema di base di cui tratta; l'opera è collegata all'istanza, con link all'editore, al luogo e al formato, ovvero alla modalità di fruizione: edizione cartacea, edizione online, e-book ecc.

Il caso rappresentato in figura 4 presenta al centro l'opera dal titolo *FRBR Report*, con "IFLA Study Group" nel ruolo di ente creatore e "Cataloging" come descrittore di soggetto. *FRBR Report* ha tre istanze, ovvero tre materializzazioni:

- 1) la versione cartacea con copertina cartonata;
- 2) il pdf;
- 3) la versione online in html.

L'edizione cartacea cartonata ha come editore K.G. Saur, mentre il pdf e l'html hanno come editore la stessa IFLA.

BIBFRAME indica alla comunità bibliotecaria e ai produttori di software un percorso di cambiamento, con al centro le esigenze informative dell'utente, il riutilizzo dei dati in contesti differenti (interoperabilità), la necessità che le macchine possano essere parte del processo di creazione e di distribuzione dell'informazione. Il progetto prevede di mettere insieme i dati dell'attuale "risorsa MARC" in una nuova architettura coerente, che consenta la catalogazione cooperativa a un livello descrittivo più granulare. La traduzione dei record bibliografici in linked data è, pertanto, intesa come base per iniziare un confronto e una discussione all'interno della comunità dell'informazione. Il punto centrale è la convinzione dell'utilizzo del web come luogo per esprimere e connettere informazioni. BIBFRAME, dunque, si propone come modello di riferimento, secondo il quale i cataloghi delle biblioteche riusciranno a entrare realmente nel web; un web che sta evolvendosi da una rete di documenti a una rete di dati collegati da relazioni qualificate.

Le tecnologie del web semantico sono già ampiamente diffuse nel mondo dei beni culturali per la pubblicazione dei dati.

Non avviene lo stesso per le modalità e gli strumenti utilizzati per la catalogazione della risorsa, per i quali non abbiamo né realizzazioni, né indicazioni su ciò che questi nuovi sistemi di creazione del dato dovranno garantire. BIBFRAME definisce, tuttavia, alcune linee guida generali che gli ILS (Integrated Library Systems) di nuova generazione dovrebbero seguire, tra cui:

- la prospettiva che essi si orientino su un'architettura "a risorsa", dove cioè l'oggetto catalogato sia il singolo elemento o la risorsa da trattare, non l'intero record;
- la possibilità che ciascuna risorsa così individuata o creata possa essere collegata ad altre risorse, sul modello entità-relazione di FRBR;
- la possibilità che nell'associazione di attributi a ciascuna risorsa il catalogatore possa attingere da informazioni nel web, esposte secondo i requisiti dei linked data, senza creare ogni volta oggetti nuovi;
- l'impiego di liste controllate di termini e, quindi, l'accesso in fase di costruzione dell'entità ai vocabolari e alle liste di termini pubblicate per esempio sull'Open Metadata Registry o su Vocab.org;
- la possibilità di definire in configurazione del siste-

ma la logica di creazione degli URI e di permettere, dunque, che a ogni nuova risorsa creata sia attribuibile automaticamente un URI.

Tutto ciò esprime il quadro generale entro il quale i nuovi ILS vanno ripensati e scritti.

6. Integrated Library System (ILS) di prossima generazione

Il passaggio dal record al dato genera una trasformazione profonda negli strumenti di mediazione tra universo bibliografico e utente che nessun attore che operi in questo ambito può ignorare. La nuova generazione di Integrated Library System tiene conto delle suggestioni dei linked data, proponendo un nuovo modello di creazione, condivisione e rappresentazione dei dati. Lo sviluppo degli strumenti di elaborazione dei dati sta focalizzandosi su alcuni aspetti fondamentali, in particolare:

- maggiore granularità nella strutturazione dell'informazione (elementi descrittivi);
- utilizzo di vocabolari controllati, di ontologie e di authority file, per condividere il più possibile un linguaggio comune e allargare la condivisione alle macchine;
- identificazione chiara degli oggetti e delle entità, tramite codici identificativi univoci e attributi qualificanti;
- creazione, quanto più possibile ampia, di un reticolo di collegamenti tra entità differenti, per fare in modo che ogni oggetto sia connesso ad altri in diverse forme di relazione;
- potenziamento dei processi di cattura e, dunque, di condivisione, dei dati così da coinvolgere attori diversi, tradizionalmente esclusi da questi processi, come per esempio, gli editori, che più di altri sono in contatto con gli autori, cioè con coloro che producono informazione (fonti primarie).

Questi principi sono stati assunti come punti di partenza per la formulazione di RDA. La sfida più urgente per i produttori di nuovi sistemi di catalogazione è perciò assorbire integralmente i principi dei linked data, semplificandone la realizzazione.

I software di nuova generazione non possono non tenere conto delle linee guida RDA per la loro rilevanza mondiale e per essere già in uso dal 31 marzo 2013, proponendo prodotti che favoriscano il passaggio dall'epoca caratterizzata dalla strutturazione dei dati in MARC, durata oltre cinquant'anni, a una nuova era, ancora in cor-

Element Label	Text	RDA Rule
Manifestation		
RIMMF Identifier °	atcult0000013	no rule
Composite Key °	The name of the rose. Harcourt Brace. 1994. Volume	
Title		2.3
Title Proper	The name of the rose	2.3.2
Statement of Responsibility		2.4
Statement of Responsibility Relating to Title Proper	Umberto Eco	2.4.2
Statement of Responsibility Relating to Title Proper	translated from the Italian by William Weaver	2.4.2
Edition Statement		2.5
Designation of Edition	1st Harvest ed	2.5.2
Publication Statement		2.8
Place of Publication	San Diego	2.8.2
Publisher's Name	Harcourt Brace	2.8.4
Date of Publication	1994	2.8.6
Series Statement		2.12
Title Proper of Series	A Harvest book	2.12.2
Title Proper of Series	Harvest in translation	2.12.2
Mode of Issuance	single unit	2.13
Identifier for the Manifestation	LCCN 94013818	2.15
Identifier for the Manifestation	ISBN 0156001314	2.15
Identifier for the Manifestation	http://lcn.loc.gov/94013818	2.15
Note On Manifestation	Including the author's postscript.	2.17
Note On Manifestation	A Helen and Kurt Wolff book	2.17
Note On Manifestation	Includes English translation of: Il postillo a Il nome della rosa	2.17
Media Type	unmediated	3.2
Carrier Type	volume	3.3
Extent statement †	536 p	3.4
Number of units †	536	3.4.1.3
Type of Unit (Text) =	pages	3.4.5
Dimensions	21 cm	3.5
Uniform Resource Locator	http://www.loc.gov/catdir/description/har041/94013818.html	4.6
Uniform Resource Locator	http://www.loc.gov/catdir/enhancements/fy0733/94013818-b.html	4.6
Expression Manifested	Eco, Umberto. Nome della rosa. English <atcult0000014>	17.1+

Figura 5 – RIMMF: maschera d'immissione dati di un'entità di tipo *Manifestation*

so di definizione, ma che, con BIBFRAME, indica la direzione dei linked open data. Il primo sforzo che i nuovi software stanno realizzando è adeguare il linguaggio, con l'introduzione di una terminologia, proposta da ICP nel 2009, che esprime l'evoluzione verso nuovi modelli di metadato: etichette quali *punto d'accesso autorizzato* in sostituzione di *titolo uniforme*; *identificativo dell'opera*, *identificativo dell'espressione* ecc., testimoniano un cambiamento profondo nella strutturazione del modello di catalogazione.¹¹

6.1 RIMMF (RDA in Many Metadata Formats)

Un primo esempio di sistema orientato a RDA è RIMMF (RDA in Many Metadata Formats), sviluppato dalla società The MARC of Quality (TMQ). È disponibile da gennaio 2015 in versione beta e scaricabile, per sistema Windows, con licenza Creative Commons (di uso gratuito, non per uso commerciale). RIMMF è uno strumento illustrativo e didattico, di visualizza-

zione dei dati, nato per aiutare i catalogatori a pensare in RDA e, dunque, in FRBR anziché in AACR e MARC. RIMMF non è, quindi, un ILS, né un modulo di catalogazione, perché i dati creati non sono ancora utilizzabili come prodotto, sebbene siano esportabili in vari formati (RDF, XML, MARC). Esso può essere considerato un prototipo di come potrebbe essere un'interfaccia di catalogazione che adotti RDA. Il suo valore consiste nell'obiettivo di semplificare il cambiamento di approccio alla catalogazione, che dal processo di produzione del record diventa il processo di identificazione e metadato delle entità FRBR (Opera, Espressione, Manifestazione, Item; Persona, Famiglia, Ente; Concetto, Oggetto, Evento, Luogo).

La figura 5 mostra una maschera di immissione dati in RIMMF, per la registrazione dei dati descrittivi di una manifestazione. L'interfaccia consente a un catalogatore di registrare dati per l'opera, l'espressione, la manifestazione, e l'item ("WEMI"), utilizzando template con-

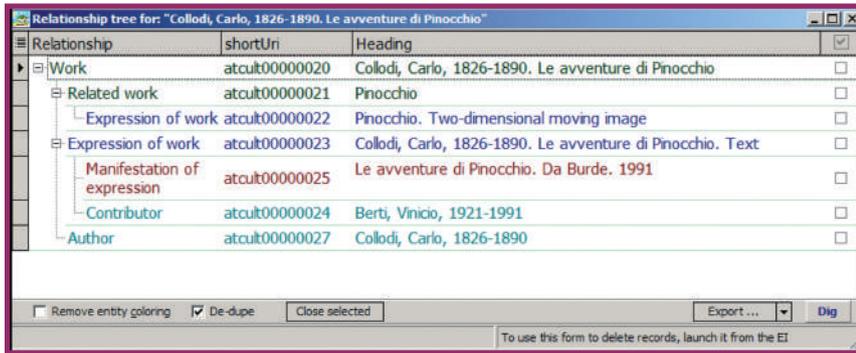


Figura 6 – RIMMF: visualizzazione gerarchica delle relazioni tra differenti entità FRBR

tenenti ciascuno gli attributi previsti dal modello FRBR per ciascuna di queste entità: appare subito chiara l'attenzione posta su FRBR.

Gli elementi riportati per ciascuna entità sono formulati come attributo della risorsa e ripresi direttamente dal set di elementi RDA, disponibile anche come *element set* RDF sull'Open Metadata Registry (<<http://rdvocab.info/>>). Ciascun attributo è associato alla relativa istruzione di RDA Toolkit, al quale si accede previa autenticazione: si semplifica, così, l'utilizzo di uno strumento di lavoro fondamentale per le nuove generazioni di catalogatori. Le funzioni disponibili in RIMMF sono:

- creazione di nuovi modelli di catalogazione;
- creazione di nuove registrazioni catalografiche;
- import di dati da database esterni;
- creazione di relazioni tra entità appartenenti al proprio database;
- visualizzazione della struttura gerarchica (ad albero) creata dalle relazioni tra le varie risorse.

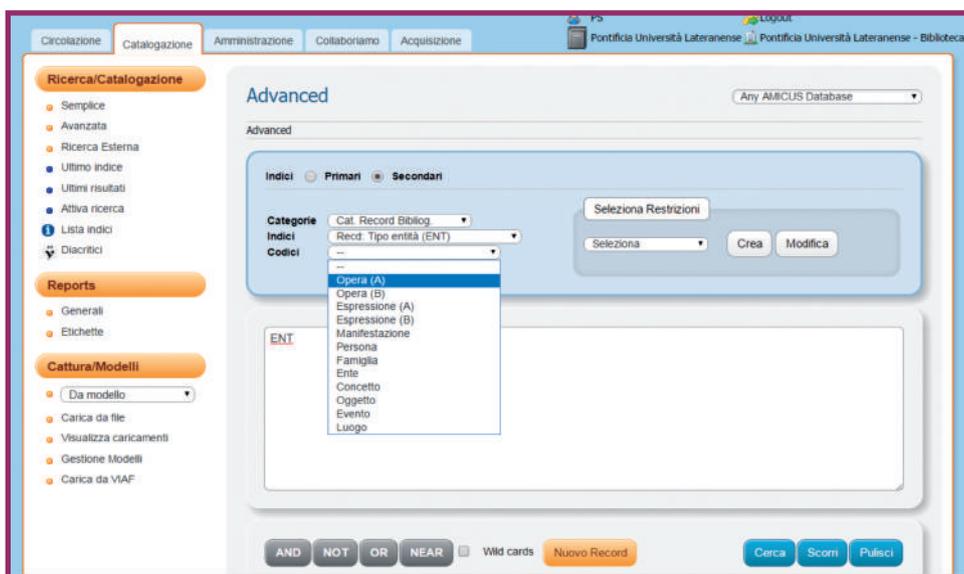


Figura 7 – OliSuite/WeCat: lista degli indici di ricerca per entità FRBR

Particolarmente interessante è la possibilità di visualizzare o importare in RIMMF dati provenienti da banche dati esterne: database bibliografici e OPAC (per esempio, quello della Library of Congress), authority file (per esempio, VIAF), enciclopedie e repertori online (per esempio, Wikipedia). Il meccanismo ha un duplice scopo:

- facilitare le operazioni di condivisione dei dati, per evitare di creare ex novo quanto esistente su altri database;
- consentire, già in fase di strutturazione del dato, di anticipare quella che, secondo il paradigma dei linked data, è una delle attività più importanti nella strutturazione dei dati: la creazione di collegamenti esterni.

La figura 6 mostra la visualizzazione in forma gerarchica delle entità descritte: offre un quadro chiaro e schematizzato della complessa rete di relazioni che può essere costruita tra risorse differenti.

RIMMF3, disponibile da gennaio 2015, offre una serie di nuove funzioni, tra cui:

- un set di elementi basati sull'RDA Registry che facilita la registrazione delle centinaia di nuove relazioni previste dalla linea guida;
- la creazione del namespace "rimmf.com" per gestire gli elementi aggiunti da RIMMF per supportare RDA;
- un'avanzata mappatura MARC con differenti procedure di creazione di dati MARC in RDA;

- le procedure di esportazione di record relativi a entità RDA in RDF (e dunque in linked open data).

6.2 OliSuite/WeCat

In Italia, @Cult, una software house e agenzia bibliografica che lavora per Casalini Libri (e Casalini per la Library of Congress), sta operando per facilitare il passaggio dalla dimensione legata al record alla nuova dimensione legata al dato. Come produttrice e distribu-

trice di Amicus, poi evoluto in OliSuite, @Cult ha assimilato gli aggiornamenti 9-20 di MARC21 per gestire record in RDA, in particolare per quanto riguarda le registrazioni di contenuto, media e supporto (tipi e caratteristiche), gli attributi dei nomi e delle risorse, e tutti i tipi di relazione. Il modulo WeCat, parte dell'ILS OliSuite, è un software che, pur utilizzando ancora la registrazione in MARC21, consente di strutturare i dati secondo RDA, in una modalità, dunque, orientata ai linked open data. Sono diverse le funzionalità che esprimono questo orientamento. Il software propone (già nella definizione degli indici di ricerca per la costruzione di query) codici specifici per le entità FRBR: la catalogazione parte con la selezione di modelli che propongono al catalogatore gli attributi previsti da FRBR per ciascuna entità.

La costruzione di un reticolo FRBR avviene collegando oggetti diversi, ancora identificabili in record ma strutturabili in modalità atomica. Nelle figure 8, 9, 10, 11 sono mostrate le entità opera, espressione, manifestazione e persona, con le loro relazioni, derivanti dallo schema di figura 1.

Le relazioni espresse in figura 1, sono trattate in WeCat tramite oggetti (record d'autorità o bibliografici) relazionati tra loro. L'insieme di questi oggetti, mostrato in figura 13, esprime il reticolo illustrato in figura 1 (vedere pagina seguente).

OliSuite/WeCat ha potenziato le funzioni per il controllo d'autorità, per affrontare la necessità, comune a RDA e ai principi dei linked open data, di descrivere l'entità che si sta

analizzando con un elevato numero di attributi, creando così un oggetto perfettamente identificabile e riutilizzabile in molteplici ambiti. La possibilità di collegare le registrazioni d'autorità (per attributi di un'opera o di un'espressione) con registrazioni bibliografiche (per attributi di una manifestazione) è un elemento di novità, introdotto per consentire la creazione di un reticolo di collegamenti che costituisce il cuore di RDA e dei linked open data. Le relazioni tra entità previste da OliSuite/WeCat rispecchiano le differenti tipologie di relazione previste dalle appendici di RDA:

- Appendice I - relazioni tra una risorsa e persone, famiglie ed enti associate alla risorsa (vedi la figura 14);

Accesso autorizzato	Shakespeare, William, 1564-1616. As you like it
----------------------------	---

Figura 8 – Opera

Accesso autorizzato	As You Like It. Italiano
Visto da	nax@ Come vi piace
Fonte dei dati trovata	L'opera era basata sul romanzo Rosalynde di Thomas Lodge. La data della prima rappresentazione è incerta, anche se è possibile che si sia svolta nel 1603 a Wilton House. Come vi piace segue la vicenda dell'eroina Rosalinda mentre fugge dalle persecuzioni della corte di suo zio e, nella Foresta di Arden, si innamora di Orlando.

Figura 9 – Espressione in lingua italiana

Titolo proprio	As You Like It / William Shakespeare.
Editore	Cambridge : Cambridge University Press, 2009.
Descr. Fisica	250 pagine ; 25 cm
Tipo contenuto	testo rdacontent
Tipo formato	supporti senza mediazione rdamedia
Tipo supporto	volume rdacarrier

Figura 10 – Manifestazione pubblicata dalla Cambridge University Press nel 2009

Titolo proprio	Come vi piace / William Shakespeare.
Editore	Milano : Garzanti, 2012.
Descr. Fisica	208 pagine ; 24 cm
Tipo contenuto	testo rdacontent
Tipo formato	supporti senza mediazione rdamedia
Tipo supporto	volume rdacarrier

Figura 11 – Manifestazione pubblicata da Garzanti nel 2012

Accesso autorizzato	Shakespeare, William, 1564-1616. VIAF ID: http://viaf.org/viaf/96994048
Visto da	nax@ Shakespear, Wilhelm 1546-1616
Fonte dei dati trovata	Wikipedia: William Shakespeare è considerato il più importante scrittore in lingua inglese e generalmente ritenuto il più eminente drammaturgo della cultura occidentale. Spesso considerato il poeta più rappresentativo del popolo inglese e soprannominato il "Bardo dell'Avon" (o semplicemente "Il Bardo") oppure il "Cigno dell'Avon", delle sue opere ci sono pervenuti, incluse alcune collaborazioni, 37 testi teatrali, 154 sonetti e una serie di altri poemi. Le sue opere teatrali sono state tradotte in tutte le maggiori lingue del mondo e sono state messe in scena più spesso di qualsiasi altra opera; inoltre è lo scrittore maggiormente citato nella storia della letteratura inglese e molte sue espressioni linguistiche sono entrate nell'inglese quotidiano.
Dati biografici o storici	William Shakespeare (Stratford-upon-Avon, 23 aprile 1564 – Stratford-upon-Avon, 23 aprile 1616) è stato un drammaturgo e poeta inglese.

Figura 12 – L'entità Persona *Shakespeare, William* espressa con i suoi attributi identificativi essenziali



Figura 13 – Il reticolo di entità e relazioni rappresentato in figura 1, ricostruito tramite record atomici in WeCat

- Appendice J - relazioni tra opere, espressioni, manifestazioni e item (vedi la figura 15);
- Appendice K - relazioni tra persone, famiglie ed enti (vedi la figura 16).

Le liste di termini utilizzate per esprimere entità e relazioni (per esempio, l'entità opera e la sua relazione con l'entità espressione, oppure l'entità manifestazione e la sua relazione con l'editore) sono espresse con classi e proprietà proposte da RDA Registry, che pubblica in modalità linked data gli elementi e i designatori di relazione approvati dal Joint Steering Committee on Development of RDA (JSC).¹²

Il software condivide la finalità di utilizzare un linguaggio formulato in set di dati e in ontologie comprensibili alle macchine. Ciascuna entità, precisamente identificata, diventa dunque nodo di un reticolo, più o meno esteso, che viene mostrato grafica-

mente per consentire al catalogatore di mantenere il controllo della rete che si sviluppa a partire da un pun-

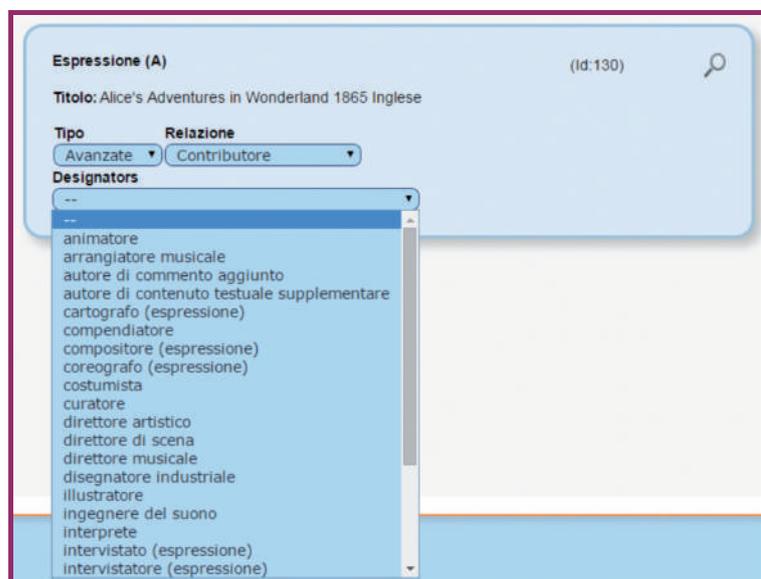


Figura 14 – OliSuite/WeCat: creazione di relazioni tra un'espressione e una persona, nel ruolo di contributore

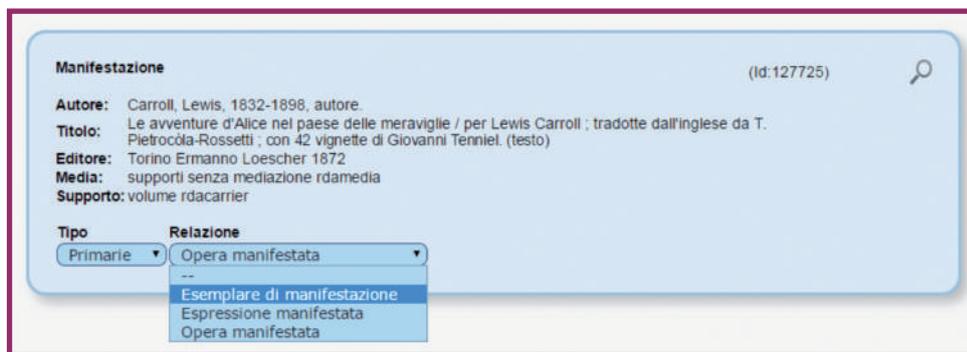


Figura 15 – OliSuite/WeCat: creazione di relazioni tra una manifestazione e un'entità del Gruppo 1 di FRBR

to di osservazione. Si veda nella figura 17 l'esempio di relazioni tra un'opera (*Alice's Adventures in Wonderland*), le sue espressioni (le edizioni del 1872 in italiano e del 1865 in inglese), il suo creatore (Lewis Carroll), le opere derivate (il film di Tim Burton del 2010).

Nell'ottica di identificare in modo univoco le risorse, comune a RDA e ai linked open data, è interessante la funzione di ricerca e cattura, a partire da un'entità, degli identificativi utilizzati per la medesima entità su altri dataset e siti esterni, per predisporre il meccanismo di *interlinking*, che consente ai dati di acquisire la quinta stella prevista dal sistema di valutazione di Tim Berners-Lee.¹³ Ulteriore elemento di novità di WeCat è la funzione

che collega la registrazione a un *framework* di produzione e pubblicazione dei dati secondo il paradigma dei linked open data, frutto di un progetto europeo: ALIADA, Automatic publication under Linked Data Paradigm of Library Data.¹⁴

Il collegamento tra un software di catalogazione

e un convertitore di dati in linked open data ha la finalità di semplificare le funzioni di aggiornamento di un dataset in RDF già pubblicato nel web, consentendo a chi crea i dati di garantire un aggiornamento in tempo reale dei dati pubblicati.

OliSuite/WeCat ha una dimensione internazionale e si rivolge alle biblioteche di tutto il mondo; il piano di sviluppo è orientato in due direzioni:

1. rendere progressivamente il sistema, indipendentemente dal formato MARC, una modalità di catalogazione "object-oriented" che identifichi e descriva gli *elementi* e non i record, in una modalità sempre più vicina alla filosofia del web semantico;

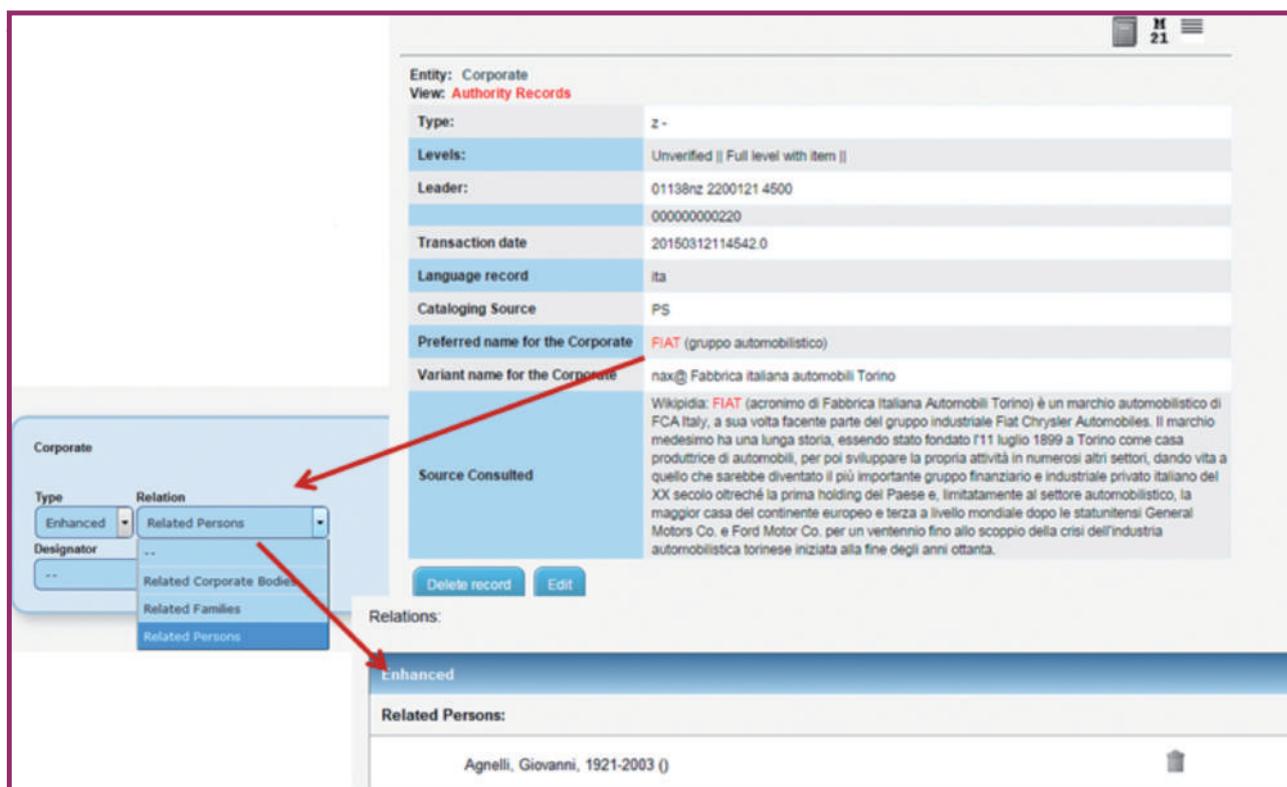


Figura 16 – OliSuite/WeCat: creazione di una relazione tra un ente (FIAT – gruppo automobilistico) e una persona (*Giovanni Agnelli*)

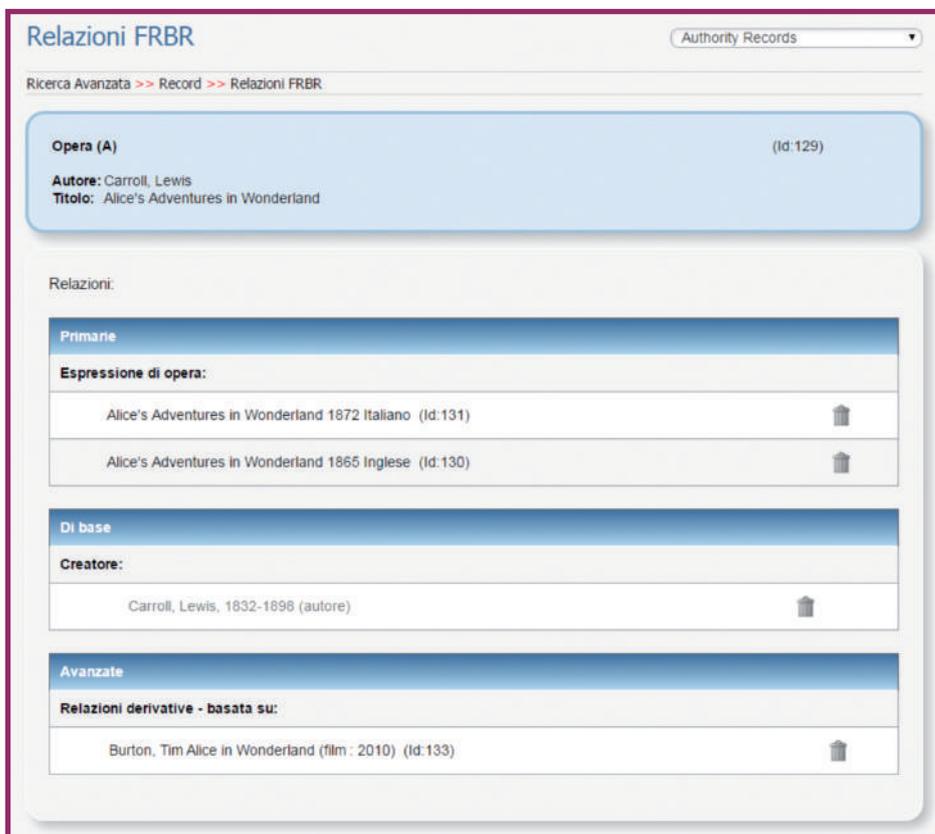


Figura 17 – OliSuite/WeCat: l'opera *Alice's Adventures in Wonderland* con le sue relazioni

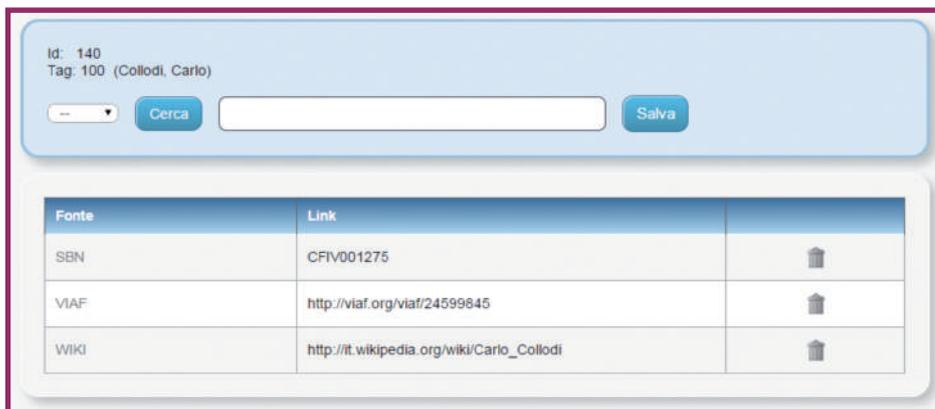


Figura 18 – OliSuite/WeCat: ricerca e cattura degli identificativi dell'entità persona *Collodi, Carlo* su tre fonti di dati differenti: SBN, VIAF, Wikipedia

2. produrre la versione open source del software, in modo da estendere l'interesse per il sistema a una più ampia comunità di esperti, interessati e sostenitori.

7. Conclusioni

La diffusione di RDA, il consolidamento di BIBFRAME, del web semantico e dei linked open data obbligano i gestori delle informazioni a riflettere profondamente sul futuro degli ILS e sulla pratica della catalogazione. La consapevolezza che siamo di fronte a una nuova generazione di sistemi è ormai matura e sempre più dif-

fusa, nonostante siano da chiarire alcune modalità per la transizione dalla vecchia alla nuova dimensione della gestione delle informazioni: come salvare milioni di registrazioni esistenti (strutturate, per esempio, in MARC), garantendo il mantenimento di servizi altrettanto efficaci costruiti su questi dati?

È certo il tempo di investire in questo sviluppo, partendo proprio dai concetti chiave espressi da RDA: *identificare* e *descrivere*, *collegare*, *rappresentare*:

- *identificare* e *descrivere*: i sistemi tradizionali strutturano le informazioni in MARC. Tuttavia, è essenziale realizzare il cambiamento linguistico (segnale di un cambiamento concettuale) prodotto dal modello FRBR e l'uso di identificatori univoci, come gli URI, nella costruzione dei dati. L'applicazione di RDA, anche in un contesto MARC tradizionale, quindi, permette di organizzare i dati in modo tale che una risorsa possa essere facilmente collegata ad altri dati e questi dati essere poi rappresentati nelle molteplici modalità esistenti;

- *collegare*: l'azione del collegare esprime il concetto chiave dei linked open data. La costruzione di collegamenti tra dati av-

viene principalmente tramite l'uso di vocabolari controllati e ontologie, strutturati in RDF: utilizzare termini assunti da vocabolari costruiti per il web nella descrizione di una risorsa (per registrare, per esempio, il tipo di contenuto, il formato, lo strumento utilizzato in un'esecuzione musicale, il supporto) aumenta la possibilità che questi dati siano "compresi" dalle macchine, potenziando così la capacità di condividere le medesime informazioni. Rendere disponibili in un ILS meccanismi di utilizzo di strumenti quali dizionari, liste di parole e ontologie disponibili sul web, consente di esprimere, in una modalità comprensibile alle macchine,

la relazione tra il tipo di risorsa (per esempio, “libro”) e il relativo attributo (per esempio, “titolo”) in termini di rapporto tra soggetto – predicato – oggetto, creando un primo, fondamentale nucleo (la tripla) di una rete più ampia di dati bibliografici.

– *rappresentare*: la pubblicazione dei dati ai fini dell'utilizzo, una funzione tradizionalmente attribuita agli OPAC e formulata in pagine html, si evolve in una varietà di opzioni in relazione alle aree di utilizzo e agli utenti a cui il servizio è rivolto. Il riferimento generale è ancora il modello FRBR. Ogni altra modalità di rappresentazione è possibile e auspicabile, in relazione a:

- *contenuto*: tutte le informazioni tradizionalmente espresse in un OPAC possono essere fuse e arricchite con informazioni provenienti da portali tematici, enciclopedie online, siti specializzati ecc.;
- *contesti disciplinari*: tutte le discipline sono adesso disponibili, non più limitate ai silos tradizionali di biblioteche, archivi, musei, editori ecc.;
- *utilizzo*: tutti gli utilizzi sono disponibili, compreso l'uso da siti scientifici o siti con funzioni commerciali integrate.

Per i manager dell'informazione, la sfida è armonizzare la tradizione con l'innovazione, il vecchio con il nuovo, in modo illuminato, creando così un nuovo sistema di gestione delle informazioni che sia avanzato ed economicamente sostenibile. La rappresentazione delle informazioni diventa, in questo modo, una modalità varia e non univoca, che segue il principio di convenienza dell'utente. L'applicazione di RDA è il primo passo per raggiungere questo cambiamento. BIBFRAME Editor indica per gli sviluppatori di software un modo di pensare in RDA e in linked open data. OliSuite/WeCat fornisce un'implementazione di RDA che integra vocabolari e ontologie già presenti nel web strutturando le informazioni in linked open data.

NOTE

¹ IFLA, *IFLA Cataloguing Principles: Statement of International Cataloguing Principles (ICP) and its glossary*, a cura di Barbara Tillett e Ana Lupe Cristán, München, Saur, 2009, <<http://www.ifla.org/publications/ifla-series-on-bibliographic-control-37>>; traduzione italiana: *Dichiarazione di Principi internazionali di catalogazione*, curata dall'ICCU, <www.ifla.org/files/assets/cataloguing/icp/icp_2009-it.pdf>; per un commento vedi MAURO GUERRINI - GIULIANO GENETASIO, *I Principi internazionali di catalogazione (ICP)*, Milano, Editrice Bibliografica, 2012.

² <www.loc.gov/bibliographic-future/news/lcwg-ontherecord-jan-08-final.pdf>.

³ RDF, Resource Description Framework, è lo strumento base proposto da W3C per la codifica, lo scambio e il riutilizzo di metadata strutturati; consente l'interoperabilità tra applicazioni che condividono le informazioni sul web. RDF Data Model si basa su tre principi chiave: 1) qualunque cosa può essere identificata da un URI (Universal Resource Identifier); 2) “the least power”: utilizzare il linguaggio meno espressivo per definire qualunque cosa; 3) qualunque cosa può dire qualunque cosa su qualunque cosa.

⁴ <<http://lj.libraryjournal.com/2002/10/ljarchives/marc-must-die/>>.

⁵ <http://marc-must-die.info/index.php/Main_Page>.

⁶ Parte dello schema è stato utilizzato da Barbara B. Tillett nel seminario su RDA organizzato dal Master biennale in Catalogazione dell'Università di Firenze, con Casalini Libri, ICCU (Linked Heritage) e AIB Toscana, Firenze, 18-20 marzo 2013.

⁷ <<http://www.loc.gov/marc/transition/news/framework-103111.html>>.

⁸ W3C INCUBATOR GROUP, *Library Linked Data Incubator Group Final Report*, October 25, 2011, <[http://www.w3.org/2005/Incubator/llid/XGR-llid-20111025](http://www.w3.org/2005/Incubator/llid/XGR-llid-20111025/http://www.w3.org/2005/Incubator/llid/XGR-llid-20111025)>.

⁹ <www.loc.gov/marc/transition/pdf/marclid-report-11-21-2012.pdf>.

¹⁰ MAURO GUERRINI, *BIBFRAME. Per un nuovo ruolo delle biblioteche nel contesto del web*, “Digitalia. La rivista del digitale nei beni culturali”, 1 (2014), giugno, p. 125-126, <<http://digitalia.sbn.it/article/view/1061/691>>. In allegato la traduzione di BIBFRAME, di Iolanda Cristalli e di Roberto Morellato, disponibile all'indirizzo <<http://digitalia.iccu.sbn.it>>.

¹¹ I cambiamenti di linguaggio nelle interfacce dei software di nuova generazione non sono ancora pienamente supportati dagli aggiornamenti del MARC21: esso, per esempio, conserva la terminologia *Titolo uniforme* per i tag 130 e 240.

¹² <<http://www.rdaregistry.info/>>.

¹³ Per il sistema a 5 stelle definito da Tim Berners-Lee per valutare la qualità dei dati pubblicati nel web, si veda <<http://www.w3.org/DesignIssues/LinkedData.html>>.

¹⁴ Il progetto, denominato ALIADA (Automatic publication under Linked Data Paradigm of Library Data), è stato realizzato da un gruppo di enti italiani, spagnoli e ungheresi: @Cult s.r.l. (I), ScanBit (S), Tecnalia (S), ARTIUM - Centro Museo Vasco de Arte Contemporáneo (S), Szépművészeti Múzeum (U) ed è consultabile al sito <<http://www.aliada-project.eu/>>. ALIADA intende creare un *framework* di produzione e pubblicazione dei dati con il paradigma dei linked open data. La versione 1.0 è disponibile su GitHub, per la comunità dell'open source.

DOI: 10.3302/0392-8586-201508-040-1

ABSTRACT

The reflection provoked by RDA produced the awareness that the flat format of MARC 21 records is inadequate in expressing the *relationships* between bibliographic entities that the FRBR model and RDA standard consider fundamental. RIMMF and BIBFRAME indicate to software developers a way to think in the RDA. In Italy, @Cult, a software house and bibliographic agency working for Casalini Libri, has taken on the charge of following and facilitating the transition: OliSuite/WeCat provides an implementation of RDA that integrates vocabularies and ontologies already present in the Web by structuring the information in linked open data.