

Tiziana Possemato

# Entity modeling: la terza generazione della catalogazione

Premessa di  
Philip E. Schreur

Prefazioni di  
Carlo Bianchini e Maurizio Vivarelli

Introduzione di  
Mauro Guerrini

FIRENZE UNIVERSITY PRESS  
2024

Entity modeling: la terza generazione della catalogazione / Tiziana Possemato ; premessa di Philip E. Schreur, prefazioni di Carlo Bianchini e Maurizio Vivarelli, introduzione di Mauro Guerrini. – Firenze : Firenze University Press, 2024.  
(Biblioteche & bibliotecari / Libraries & librarians ; 11)

<https://books.fupress.com/isbn/9791221503937>

ISSN 2612-7709 (print)  
ISSN 2704-5889 (online)  
ISBN 979-12-215-0392-0 (Print)  
ISBN 979-12-215-0393-7 (PDF)  
ISBN 979-12-215-0394-4 (XML)  
DOI 10.36253/979-12-215-0393-7

Graphic design: Alberto Pizarro Fernández, Lettera Meccanica SRLs  
Front cover image: © Sathyan Rizzo, *Silicon Mirror*. Direzione creativa, Attilio Cusani

#### *Peer Review Policy*

Peer-review is the cornerstone of the scientific evaluation of a book. All FUP's publications undergo a peer-review process by external experts under the responsibility of the Editorial Board and the Scientific Boards of each series (DOI 10.36253/fup\_best\_practice.3).

#### *Referee List*

In order to strengthen the network of researchers supporting FUP's evaluation process, and to recognise the valuable contribution of referees, a Referee List is published and constantly updated on FUP's website (DOI 10.36253/fup\_referee\_list).

#### *Firenze University Press Editorial Board*

M. Garzaniti (Editor-in-Chief), M.E. Alberti, F. Vittorio Arrigoni, E. Castellani, F. Ciampi, D. D'Andrea, A. Dolfi, R. Ferrise, A. Lambertini, R. Lanfredini, D. Lippi, G. Mari, A. Mariani, P.M. Mariano, S. Marinai, R. Minuti, P. Nanni, A. Orlandi, I. Palchetti, A. Perulli, G. Pratesi, S. Scaramuzzi, I. Stolzi.

#### *FUP Best Practice in Scholarly Publishing* (DOI 10.36253/fup\_best\_practice)

📖 The online digital edition is published in Open Access on [www.fupress.com](http://www.fupress.com).

Content license: except where otherwise noted, the present work is released under Creative Commons Attribution 4.0 International license (CC BY 4.0: <http://creativecommons.org/licenses/by/4.0/legalcode>). This license allows you to share any part of the work by any means and format, modify it for any purpose, including commercial, as long as appropriate credit is given to the author, any changes made to the work are indicated and a URL link is provided to the license.

Metadata license: all the metadata are released under the Public Domain Dedication license (CC0 1.0 Universal: <https://creativecommons.org/publicdomain/zero/1.0/legalcode>).

© 2024 Author(s)

Published by Firenze University Press  
Firenze University Press  
Università degli Studi di Firenze  
via Cittadella, 7, 50144 Firenze, Italy  
[www.fupress.com](http://www.fupress.com)

*This book is printed on acid-free paper  
Printed in Italy*

## Sommario

|   |    |
|---|----|
| Premessa<br><i>Philip E. Schreur</i>  | 11 |
| Prefazione<br><i>Carlo Bianchini</i>  | 13 |
| Prefazione<br><i>Maurizio Vivarelli</i>   | 17 |
| Introduzione<br>L'entity modeling, ovvero la next generation della catalogazione<br><i>Mauro Guerrini</i> | 23 |
| Il contesto   | 29 |
| 1. Il tema  | 30 |
| 2. Il contesto storico  | 32 |
| 3. Lo stato dell'arte   | 33 |
| 4. Fonti della ricerca e articolazione dell'opera   | 34 |
| Capitolo 1  |    |
| Contengo moltitudini  | 35 |
| 1. Introduzione   | 35 |
| 2. Quesiti da risolvere   | 36 |
| 3. Entità e identità  | 40 |
| 4. Le identità diacroniche e sincroniche nell'universo bibliografico                                      | 44 |
| 5. Conclusioni sui concetti di entità e identità  | 75 |

|   |     |
|---|-----|
| Capitolo 2  |     |
| Il mondo è una mia rappresentazione   | 77  |
| 1. Introduzione   | 77  |
| 2. <i>On the Record</i> : la dichiarazione di un nuovo corso per adeguarsi al linguaggio del web        | 78  |
| 3. Cos'è un real world object   | 85  |
| 4. L'entity resolution e l'entity modeling  | 97  |
| 5. Conclusioni sul passaggio dal record all'entity modeling   | 147 |
| Capitolo 3  |     |
| Panta rei, o della fluidità delle cose dell'universo  | 149 |
| 1. Introduzione   | 149 |
| 2. L'entità come prisma   | 150 |
| 3. Cogliere le cose nel cambiamento   | 155 |
| 4. Esempi di entità emergente in ambito catalografico   | 163 |
| 5. Caducità delle proprietà intrinseche, persistenza delle proprietà relazionali                        | 184 |
| 6. Conclusioni sul cambiamento e sulla persistenza  | 193 |
| Capitolo 4  |     |
| O Romeo, Romeo, wherefore art thou Romeo?   | 197 |
| 1. Introduzione   | 197 |
| 2. L'Agente nella tradizione catalografica  | 198 |
| 3. Il trattamento dei papi nella tradizione catalografica   | 235 |
| 4. Identificare entità e identità attraverso URI  | 262 |
| 5. Conclusioni sull'entity modeling e sul significato in termini di cambiamento di mentalità collettiva | 274 |
| Riflessioni finali  | 277 |
| Appendice I   |     |
| Proprietà delle entità svde:Opus, svde:Work e bf:Hub  | 281 |
| Appendice II  |     |
| L'ontologia Share-VDE: un'estensione di BIBFRAME per la discovery dei linked data                       | 285 |
| Appendice III   |     |
| L'ontologia di Share-VDE come estensione di BIBFRAME  | 293 |
| Ringraziamenti  | 303 |
| Riferimenti bibliografici   | 305 |
| Elenco delle figure   | 315 |
| Elenco degli acronimi   | 319 |
| Indice analitico  | 323 |

INTRODUZIONE

## L'entity modeling, ovvero la next generation della catalogazione

Il terreno di coltura della ricerca che ha prodotto questo libro è stata la redazione della tesi di dottorato di Tiziana Possemato che ho avuto l'onore di seguire: la definizione di un linguaggio catalografico tipico dell'era digitale. Decisiva è stata l'esperienza Share-VDE, parte della più ampia iniziativa Share Family, di cui l'autrice è una delle principali protagoniste: un progetto che coinvolge le maggiori biblioteche americane e nordeuropee teso a definire un nuovo modello bibliografico sia sul piano concettuale sia sul piano delle azioni concrete relative alla descrizione delle risorse e alla pubblicazione e all'uso dei loro metadati, possibilmente come linked open data (LOD). Un osservatorio internazionale quanto mai stimolante e collaborativo sul piano teorico e pratico.

Il paradigma della catalogazione in era digitale coinvolge i diversi domini del web, a partire da quelli più vicini al mondo bibliotecario, quali gli archivi e i musei. Non casualmente, *Entity modeling: la terza generazione della catalogazione* reca un titolo emblematico che richiama a un passaggio evolutivo. Uno sviluppo che necessita prima di tutto della consapevolezza di vivere trasformazioni radicali e, quindi, di acquisire il linguaggio tipico del web semantico per favorire una comunicazione tra universo bibliografico e lettore sempre più trasversale.

L'indagine sull'*entity modeling*, tema di base del lavoro dell'autrice, introduce, infatti, una riflessione che si pone come superamento sia delle metodologie che hanno contraddistinto la *Great Tradition* catalografica menzionata da Michael Gorman in *Our enduring values, revised* edito dall'ALA nel 2015 – tradizione che, secondo l'autore inizia da Antonio Panizzi e si conclude con Seymour Lubetzky –, sia il passaggio, avvenuto a cavallo tra XX e XXI secolo, dalla catalogazione alla metadattazione (o *metadata creation*, come viene chiamata in inglese). *L'entity modeling*, in altre parole, va oltre la metadattazione e rappresenta la terza genera-

FUP Best Practice in Scholarly Publishing (DOI 10.36253/fup\_best\_practice)

Tiziana Possemato, *Entity modeling: la terza generazione della catalogazione*, © 2024 Author(s), CC BY 4.0, published by Firenze University Press, ISBN 979-12-215-0393-7, DOI 10.36253/979-12-215-0393-7

zione della catalogazione. È incontrovertibile che la dimensione catalografica si sia oggi spostata interamente nel web e che occorra verificare cosa ancora regge dell'impianto tradizionale e cosa vada concepito di nuovo. Il risultato di questo processo innovativo si chiamerà ancora *catalogo*, ovvero un elenco ordinato di registrazioni bibliografiche compilate in base a convenzioni o a standard?

*Catalogare* ha sempre significato creare metadati, ovvero registrare dati strutturati in accordo a una modalità definita, cioè a un formato funzionale a offrire una rappresentazione degli aspetti distintivi e caratterizzanti della risorsa descritta, di qualsiasi tipo di risorsa – una persona, un'opera, un oggetto d'arte –, ovvero dei metadati funzionali alla sua identificazione e al suo reperimento.

Cos'è, dunque, cambiato?

Tiziana cerca di rispondere alla domanda.

L'affermarsi del digitale ha prodotto la proliferazione delle famiglie dei metadati provocando un'estensione notevole dei confini della tradizione catalografica. In un periodo di profonda trasformazione bibliografica, ritiene insufficiente l'analisi finora compiuta degli elementi che caratterizzano le descrizioni e pone al centro della sua originale riflessione due concetti: *l'entità*, intesa come ogni *real world object*, e *l'identità*. Il lettore troverà la risposta articolata leggendo il volume. S'imbatterà in concetti e termini innovativi in una lettura del testo non facile, come non facile è la lettura di qualsiasi libro che rifletta criticamente su comportamenti considerati acquisiti e che invece necessitano di verifica circa la loro corrispondenza ai cambiamenti avvenuti dell'odierno contesto culturale, editoriale e tecnologico. Il libro, pertanto, introduce uno scenario inedito nella lunga tradizione teorica e pratica della catalogazione, sensibile alle opportunità offerte dai modelli concettuali elaborati sia in ambito biblioteconomico, in particolare emanati da IFLA, sia in ambito informatico in quanto tecnologia contemporanea. L'autrice dimostra la validità del nuovo scenario, *in progress*, maturato e sperimentato in ambito internazionale nella sopra citata iniziativa Share Family.

*Entity modeling* si presenta, pertanto, come il terzo passaggio generazionale nella gestione dei dati bibliografici. Si tratta di un'evoluzione che impone un cambiamento delle modalità di trattamento dei dati e delle modalità della loro formalizzazione; implica un linguaggio che usi i parametri semantici e sintattici propri del contesto digitale. La prospettiva proposta dall'autrice crea un circolo virtuoso in cui la dimensione teorica e applicativa si arricchiscono a vicenda per generare una nuova struttura funzionale alla mediazione tra universo bibliografico e lettore. Ne deriva il superamento del catalogo concepito per l'era convenzionale, con la proposizione, adesso, di *qualcosa* che è ancora in corso di definizione ma che viene percepito come un salto qualitativo significativo rispetto al passato. Nel nuovo paradigma i dati sono definiti singolarmente e sono il più possibile interoperabili. Nella metadatazione, infatti (diversamente dalla catalogazione tradizionale), l'orizzonte descrittivo della risorsa include una serie di dati necessari alla sua identificazione, al suo reperimento e alla sua fruizione, in un contesto informativo sempre più ricco, ampio ed eterogeneo. Ulteriore prerogativa importante della metadatazione rispetto alla catalogazione tradizionale è l'apertura che lascia alla struttura della presentazione dei dati e

alle modalità d'accesso alle descrizioni: esse possono essere definite a posteriori, senza le costruzioni predefinite previste da ISBD. Questo standard, che tanto senso e valore ha avuto internazionalmente nella strutturazione dei metadati al fine della loro condivisione in basi dati diverse, con la sequenza predeterminata di aree, elementi e punteggiatura convenzionale, perde di significato in un'era in cui ciascun singolo metadato diviene autosignificante, ovvero acquista significato e viene creato per essere comprensibile alla macchina. Con la creazione dei metadati, pertanto, non si ha più una sequenza lineare dei metadati descrittivi e la ricerca si estende dai soli *punti d'accesso* – (prima di ICP del 2009 chiamati *intestazioni*) – a un'enorme mole di metadati riferibili alla risorsa ricercata e derivanti o derivabili da varie fonti. Due peculiarità significative. Si può, dunque, dire che la metadattazione è la catalogazione nell'era digitale in cui tutto è connesso. La tradizione catalografica viene confermata nella sua essenza proprio nel momento in cui aggiorna e rafforza le sue tecniche per permettere che il catalogo non sia solo *nel web*, ma che diventi *del web*, ovvero impieghi il linguaggio della contemporaneità, come esso ha sempre fatto nella sua lunga storia. Il focus è ora sulla caratterizzazione delle entità e delle loro relazioni da cui deriva la creazione di metadati descrittivi e d'accesso alle risorse bibliografiche. Il catalogo diventa così (ma lo è sempre stato) lo strumento abilitante ai servizi perché una biblioteca che non dà servizi è un deposito e non una biblioteca.

La proposta dell'*entity modeling* come nuova generazione della catalogazione va nella direzione di trovare un modo comune a tutti i domini della conoscenza per identificare e descrivere le risorse, prescindendo dalle regole di standardizzazione troppo vincolanti e rigide, se legate a uno specifico contesto. L'idea deriva, e ha quale precursore e base, il modello Entity-Relationship (ER), ideato in ambito informatico nella seconda metà degli anni Settanta del secolo scorso, che si è affermato come metodologia di analisi e modellazione dei dati per la progettazione di banche dati e relative applicazioni, recepito in ambito catalografico nel 1998 con FRBR. Un modello la cui importanza ha plasmato gli standard descrittivi e, almeno idealmente, la redazione dei cataloghi.

I modelli concettuali e gli standard catalografici oggi stanno concentrando sempre più la loro attenzione sulle entità e sulle relazioni che intercorrono tra esse piuttosto che su una semplice caratterizzazione delle entità tramite gli attributi. Ciò appare evidente nel passaggio da FRBR a IFLA LRM, ma anche dalle trasformazioni subite dallo standard RDA in seguito all'RDA Toolkit Restructure and Redesign Project e all'emergente standard BIBFRAME con le sue crescenti implementazioni, a cominciare dalla Library of Congress. Ciò porta a una modifica della struttura dei cataloghi, oggi organizzati come semplici database relazionali, per essere database semantici capaci di condividere i metadati con l'esterno, favorendo così l'interoperabilità tra domini e uscendo dalla logica dei cataloghi-silos, dei cataloghi chiusi in sé. Alcune biblioteche hanno già avviato sperimentazioni in questa direzione, come la Bibliothèque nationale de France con data.bnf.fr (dal 2012!), la quale ha rimodellato il proprio catalogo sulla base di FRBR, la famosa *effeerrebierizzazione* del catalogo. Data.bnf.fr è uno strumento di ricerca nettamente potenziato rispetto a un catalogo tradizionale,

capace d'interrogare contemporaneamente più basi dati e offrire innumerevoli percorsi di ricerca. Acquisizione oggi consolidata grazie ai discovery tool. Ciò che più conta è la qualità dei dati, cioè la dimensione bibliografica e di ricerca del catalogo, non solo in quanto strumento di 'scoperta', bensì soprattutto quale dispositivo di un sapere incorporato coerente con gli sviluppi della ricerca storico-bibliografica ed editoriale sempre presente da Antonio Panizzi (ma fin da Audiffredi) in poi, col culmine raggiunto da Ákos Domanovszky.

Il menzionato cambio di paradigma catalografico passa tramite le elaborazioni teoriche che il libro espone dettagliatamente e in modo convinto per offrire le basi a una varietà di applicazioni che permettono d'identificare le entità beneficiando di tecniche di arricchimento dei dati senza perdere le peculiarità di specifiche rappresentazioni o identità. Le conseguenze contribuiscono a costruire le fondamenta di un nuovo ecosistema bibliografico associato alla terza generazione della catalogazione. I benefici si ripercuotono sugli utenti (studiosi e studenti) e sulle macchine che fanno uso dei dati, a cui può essere adesso messo a disposizione uno strumento di ricerca estremamente funzionale per l'accesso alle informazioni, in modo trasparente rispetto alla complessità della strutturazione dei metadati e alla loro interconnessione tra basi dati. I benefici si ripercuotono anche sul mantenimento dei metadati – la cui quantità continuerà a crescere rapidamente – tramite nuove forme di collaborazione internazionale.

Una ristrutturazione del genere è davvero utile per l'utente o, più esattamente, per tutte le tipologie di utente? I nuovi strumenti offerti dal mercato sono facili e intuitivi da utilizzare o la ricerca dei dati risulta più complicata?

In questa fase di transizione bibliografica occorre fermarsi a riflettere sulle conseguenze che ciò potrebbe comportare: le priorità oggi sembrano essere la massima condivisione e visibilità sul web, la possibilità di navigare tra le risorse e di scoprirne di nuove, la possibilità di riutilizzare i dati per ottimizzare i processi catalografici e renderli più veloci e sostenibili.

Gli utenti chiedono tutto ciò? Siamo consapevoli di cosa significhi strutturare uno strumento di mediazione tra raccolta e lettore secondo modalità che consentano questo grado d'integrazione nel web?

Per fare un esempio, pensiamo di descrivere un libro facendo riferimento alla seconda edizione dello standard RDA. La descrizione tende all'atomizzazione e molti metadati che prima erano considerati attributi di una risorsa sono divenuti relazioni tra entità, ciascuna delle quali viene ulteriormente caratterizzata.

Il primo elemento che registriamo per un libro è il titolo, cioè un tipo di *nomen*, ossia un'entità a cui è associato un dataset che è associato a un insieme di attributi o metadati che ne specificano proprietà quali la lingua, il tipo di scrittura, la *stringa del nomen* ecc.; molti altri metadati vengono trattati nello stesso modo come il luogo di pubblicazione, il nome dell'editore o la data di pubblicazione, ciascuno dei quali viene adeguatamente caratterizzato. Ciò è fondamentale quando si opera in un ambiente fondato sul riuso e sullo scambio di metadati, in cui ogni dato per essere utilizzato in maniera corretta dev'essere adeguatamente accompagnato da una serie di metadati che ne specifichino il contesto semantico. Tuttavia, informazioni così dettagliate potrebbero non interessare l'uten-



te di una biblioteca che spesso pone al catalogo domande precise relativamente alla presenza di una determinata risorsa.

Quando parliamo di *entity modeling*, di tecnologie del web, di strumenti di arricchimento da fonti diverse, parliamo di una serie di strategie che nascono proprio per servire fasce di utenza diverse, dalle macchine agli umani. L'identificazione delle entità e, quindi, lo spostamento di attenzione all'oggetto in sé (sia esso un autore, un libro, un luogo geografico, un concetto o qualsiasi altra cosa che sia parte del nostro orizzonte) propone un modello di creazione dell'informazione in cui ciascuno possa mantenere il proprio 'punto di vista' nella fase di descrizione e nella fase di fruizione. L'*entity modeling*, il cui paradigma dei linked open data è solo un 'facilitatore tecnologico' per la sua realizzazione, propone di guardare il mondo e le sue cose ciascuno dalla posizione più comoda (rispetto, dunque al contesto da cui quella cosa si guarda) e utile al proprio scopo (ciascuno, cioè, secondo il proprio bisogno informativo).

Nel processo evolutivo della catalogazione un quesito da sempre aperto e ancora irrisolto è come costruire linguaggi che travalichino i singoli domini conoscitivi, per facilitare lo scambio, il riuso e la comprensione delle informazioni tra domini diversi. Per anni, nel giusto ossequio di regole di catalogazione che hanno standardizzato il linguaggio di comunicazione tra raccolta e lettore, sono stati costruiti patrimoni informativi preziosi, ma chiusi in silos e, per certe loro parti, comprensibili solo ai bibliotecari e agli esperti. Prevale adesso la volontà di permettere il dialogo tra domini differenti, che si persegue con attività di mappatura tra i diversi linguaggi o che sarà perseguito in futuro con un cambiamento radicale all'informazione, in cui, appunto, non sarà più la mappatura da formato a formato a garantire la comunicazione, ma la strutturazione di un linguaggio trasversale (quello proposto dall'*entity modeling*) in cui ciascun dominio possa offrire una propria visione di quanto collaborativamente descritto.

Il rifiuto di riconoscere e accogliere gli standard del web condivisi a livello internazionale mummifica il proprio dominio, il proprio linguaggio disciplinare e si pone fuori dal contesto contemporaneo.

Il libro di Tiziana è un segno di speranza che la catalogazione, la metadattazione o come si chiamerà in futuro, rimanga al centro degli interessi scientifici così che il nuovo strumento di mediazione tra la porzione di universo bibliografico posseduta dalle biblioteche e l'utente sia sempre più funzionale agli scopi e alle necessità del lettore di oggi e del prossimo futuro.

Mauro Guerrini  
*Università di Firenze*