



Il passaggio ai metadati di nuova generazione

Karen Smith-Yoshimura



Il passaggio ai metadati di nuova generazione

Karen Smith-Yoshimura

Senior Program Officer

Tradotto da Smartling, Inc., Maria Cristina Mencato e Sabrina Masoli.



© 2020, 2021 OCLC, Inc.
Pubblicato con Licenza Creative Commons Attribution 4.0 International License.
<http://creativecommons.org/licenses/by/4.0/>



Pubblicazione originale settembre 2020
traduzione italiana ottobre 2021

OCLC Research
Dublin, Ohio 43017 USA
www.oclc.org

ISBN: 978-1-55653-242-9
DOI: 10.2533/k23b-hd86
OCLC Control Number: 1275391975

ORCID iDs
Karen Smith-Yoshimura <https://orcid.org/0000-0002-8757-2962>

Pubblicato originariamente come 2020 Transitioning to the Next Generation of Metadata. Dublin, OH:
OCLC Research. <https://doi.org/10.2533/rqgd-b343>.

Siete pregiati di scrivere a:

OCLC Research
oclcresearch@oclc.org

Citazione suggerita:

Smith-Yoshimura, Karen. (2020) 2021. *Il passaggio ai metadati di nuova generazione*. Tradotto da Smartling, Inc., Maria Cristina Mencato e Sabrina Masoli. Dublin, OH: OCLC Research.
<https://doi.org/10.2533/k23b-hd86>.

INDICE

Sintesi	vi
Introduzione	1
Il passaggio ai linked data e agli identificatori	4
Lo sviluppo dell'uso degli identificatori persistenti	5
Il passaggio dal "controllo di autorità" alla "gestione delle identità"	10
Afrontare la necessità di vocabolari multipli, equità, diversità e inclusione ..	12
Le sfide dei linked data	16
Descrizione delle collezioni "Inside-Out" e "Facilitated".....	17
Collezioni archivistiche	18
Archiviazione di siti web	18
Collezioni audio e video.....	20
Collezioni di immagini	21
Dati di ricerca	24
L'evoluzione: "Metadati come servizio".....	27
Metrica.....	27
Consulenze.....	28
Nuove applicazioni.....	28
Bibliometria	29
Indicizzazione semantica.....	30
Future competenze richieste al personale	31
Il cambiamento culturale	31
Opportunità di formazione	32
Nuovi strumenti e competenze	33
Autoformazione.....	34
Pensare al ricambio del personale	35
Impatto	36

Ringraziamenti.....	37
Appendice.....	38
Note.....	39

IMMAGINI

FIGURA 1	"Modifiche dei flussi di lavoro relativi alla descrizione delle risorse" di OCLC Research.....	4
FIGURA 2	Alcuni dei 300 nomi di autore in forma abbreviata contenuti in nell'articolo, di cinque pagine pubblicato in: <i>Physical Review Letters</i>	7
FIGURA 3	Esempi di alcuni identificatori DOI e ARK	9
FIGURA 4	Un identificatore Wikidata collegato ad altri identificatori ed etichette di lingue diverse	10
FIGURA 5	Estratto dei risultati del sondaggio EDI condotto dal Research Library Partnership nel 2017	14
FIGURA 6	Risposte al sondaggio [2019] sulle sfide relative alla gestione delle raccolte A/V	21
FIGURA 7	L'OCLC ResearchWorks IIIF Explorer recupera immagini relative a "Paris Maps" nelle collezioni contenute nel CONTENTdm	24
FIGURA 8	Distribuzione di 465 codici di lingue indigene nell'Australian National Bibliographic Database dell'Austlang National Codeathon.....	29
FIGURA 9	UK Hachette's "River of Authors" realizzato con i metadati forniti dal catalogo della British Library	30

SINTESI

L'OCLC Research Library Partners Metadata Managers Focus Group, fondato nel 1993, è uno dei gruppi più longevi all'interno dell'OCLC Research Library Partnership (RLP), una rete internazionale di biblioteche di ricerca. Il Focus Group gestisce un forum dedicato agli amministratori responsabili della gestione dei metadati creato per condividere informazioni su argomenti di interesse comune e per identificare questioni relative alla gestione dei metadati. Le questioni sollevate dal Focus Group sono seguite da OCLC Research per supportare il lavoro del RLP e sono fonte di ispirazione per la realizzazione dei prodotti e dei servizi di OCLC.

Questo report dal titolo "*Il passaggio ai metadati di nuova generazione*", riassume 6 anni (2015-2020) di discussioni dell'OCLC Research Library Partners Metadata Managers Focus Group ed indica quali sono le previsioni per i "metadati di nuova generazione". La convinzione che ora, e in futuro, i metadati siano alla base delle prossime innovazioni, indipendentemente dal loro formato, ha permeato tutte le discussioni del Focus Group.

In poche parole, i metadati stanno cambiando. La gestione di metadati specifici basata su accurate stringhe di testo contenute nei record bibliografici, comprensibili esclusivamente dai gestionali di biblioteca, è vicina all'obsolescenza, sia a livello concettuale che tecnico. Le innovazioni introdotte in biblioteconomia stanno creando l'esigenza di nuove pratiche di gestione dei metadati che permettano ai bibliotecari sia di fornire metadati per tutte le tipologie di risorse, sia di collaborare a progetti istituzionali e multi-istituzionali. Questo report descrive come i metadati stiano cambiando e prende in considerazione l'impatto che questo tipo di transizione potrebbe avere sui servizi bibliotecari ponendo domande quali:

- **Perché i metadati stanno cambiando?**
- **Come sta cambiando il loro processo di creazione?**
- **Come cambiano i metadati stessi?**
- **Che tipo di impatto avranno questi cambiamenti sulle future competenze del personale? Come possono le biblioteche prepararsi a questi cambiamenti?**

Il futuro dei linked data è legato al futuro dei metadati: i metadati che biblioteche, archivi e altre istituzioni culturali hanno creato, e continueranno a creare, costituiscono la base dei linked data, che rappresenteranno "dichiarazioni" associate a quei link. L'impatto sarà totale: riguarderà il modo in cui bibliotecari ed archivisti descriveranno le "inside-out" e "facilitated collections", comporterà l'offerta di "metadati come servizio" e determinerà le competenze richieste al personale.

Il passaggio ai metadati di nuova generazione è un processo in continua evoluzione intrecciato all'evoluzione degli standard, delle infrastrutture e degli strumenti. I membri del Focus Group sono giunti ad una visione condivisa delle sfide da affrontare, hanno studiato possibili approcci e hanno confrontato le loro idee con i membri comunità con le quali interagiscono.

INTRODUZIONE

L'OCLC Research Library Partners Metadata Managers Focus Group (di seguito indicato come il Focus Group),¹ fondato nel 1993, è uno dei gruppi più longevi all'interno dell'OCLC Research Library Partnership (RLP),² una rete internazionale di biblioteche di ricerca. Il Focus Group gestisce un forum dedicato agli amministratori responsabili della gestione dei metadati per condividere informazioni su argomenti di interesse comune e per identificare questioni relative alla gestione dei metadati. Le questioni sollevate dal Focus Group sono seguite da OCLC Research per supportare il lavoro del RLP e sono fonte di ispirazione per la realizzazione dei prodotti e dei servizi di OCLC.

La convinzione che ora, e in futuro, i metadati siano alla base del successo di una ricerca, permea tutte le discussioni del Focus Group. I metadati forniscono l'infrastruttura necessaria a tutte le biblioteche che possiedono "sistemi di ricerca" per soddisfare le richieste di informazioni e risorse provenienti dalla propria comunità. I metadati sono fondamentali per realizzare il passaggio. Metadati di qualità prodotti oggi potranno essere facilmente riutilizzati in futuro in ambiente linked data.³ Come riportato dalla *British Library's Foundations for the Future*: "La nostra visione è che, entro il 2023, le collezioni di metadati delle biblioteche saranno riunite in un'unica infrastruttura sostenibile e basata su standard in grado di offrire le migliori condizioni di accesso, collaborazione e riutilizzo".⁴

La gestione di specifici metadati basati su accurate stringhe di testo contenute nei record bibliografici comprensibili esclusivamente da gestionali di biblioteca, è vicina all'obsolescenza, sia a livello concettuale che tecnico.

La gestione di specifici metadati basati su accurate stringhe di testo contenute nei record bibliografici comprensibili esclusivamente da gestionali di biblioteca, è vicina all'obsolescenza, sia a livello concettuale che tecnico. Le innovazioni introdotte in biblioteconomia stanno creando l'esigenza di nuove pratiche di gestione che permettano ai bibliotecari di fornire metadati per tutte le tipologie di risorse e di collaborare a progetti istituzionali o con altri enti. "La tradizionale creazione, gestione e disseminazione dei metadati", suggerisce la *British Library's Collection Management Strategy*, "non è più appropriata in questa era di rapidi cambiamenti digitali, aumento delle aspettative dell'utenza e diminuzione delle risorse".⁵

I membri del Focus Group sono impazienti di applicare la potenza dei metadati sui record che ne sono ancora privi per poter sperimentare interazioni ed usi sia da parte dell'utenza che delle macchine. I metadati creati secondo regole o standard obsoleti devono essere trasformati in nuove strutture.

Perché i metadati stanno cambiando?

I metadati tradizionali sono stati, e continuano ad essere, creati in conformità a regole usate e conosciute principalmente dai bibliotecari. Sono creati in base alla struttura del record, sono costosi da produrre e hanno limiti di dimensioni dovuti a cause storiche. Inoltre, sono utilizzati solo in particolari ambiti, non vengono impiegati per la descrizione degli articoli pubblicati all'interno di riviste accademiche o altri elaborati accademici. I sistemi usati per la loro creazione si sono spesso dimostrati non adeguati alla gestione delle correzioni e dei miglioramenti; ciò ha provocato una ricerca della perfezione durante la loro creazione che, a sua volta, ha generato una incredibile lentezza nello svolgimento del lavoro. Riassumendo: i metadati potrebbero essere migliorati, non ne sono stati creati abbastanza, quelli che sono stati creati non sono ampiamente utilizzati al di fuori delle biblioteche.

Come sta cambiando il loro processo di creazione?

I metadati non sono più generati esclusivamente dal personale delle biblioteche. Al giorno d'oggi, editori, autori, e altre soggetti sono ugualmente coinvolti nella creazione di metadati. La loro creazione è stata anche sollecitata all'interno delle attività accademiche da quegli editori che generano record di metadati prima ancora che la pubblicazione venga sottoposta al tradizionale processo di catalogazione. Attualmente i metadati possono essere migliorati o corretti dalle macchine o su base volontaria.

Come cambiano i metadati stessi?

Il formato MARC (Machine-readable cataloging) è stato creato per replicare i metadati tradizionalmente ricercabili nei cataloghi cartacei a schede. Ora stiamo passando dai record MARC all'assemblaggio di componenti codificati, condivisibili e collegabili, facendo particolare attenzione al modo in cui saranno citati; inoltre, la tendenza è quella di eliminare anacronistiche abbreviazioni non comprensibili alle macchine. Invece di fare affidamento solamente sugli strumenti a disposizione dei bibliotecari (liste di soggetti o liste di codici), le aggregazioni in via di sviluppo potranno accogliere singoli vocabolari creati per specifici domini, aumentando così la possibilità di utilizzo dei metadati.

Riassumendo: i metadati potrebbero essere migliorati, non ce ne sono abbastanza e quelli creati non sono ampiamente utilizzati al di fuori delle biblioteche.

La composizione del Focus Group è variata nel tempo e attualmente è composta dai rappresentati di 63 RLP Partners provenienti da 11 nazioni distribuite in quattro continenti.⁶ Il gruppo include il presidente attuale del Program for Cooperative Cataloging (PCC)⁷ e quello

uscente; i due presidenti collaborano e creano proficue sinergie. Gli argomenti da trattare all'interno del gruppo possono essere proposti da ogni membro e sono selezionati da un Planning Group costituito da otto persone (vedi appendice). Il Planning Group sceglie un tema e, in un apposito documento, spiega il motivo per cui quel determinato argomento deve essere preso in considerazione e propone una serie di questioni per approfondire il tema. Il documento e le questioni vengono poi condivise con ogni membro del Focus Group, che ha a disposizione dalle tre alle cinque settimane per restituire le proprie considerazioni. I riscontri avuti dal Focus Group sono oggetto di discussione durante gli incontri organizzati con l'American Library Association⁸ e vengono poi sviluppati in successivi incontri virtuali.

In qualità di moderatore del Focus Group, ho riassunto e sintetizzato le discussioni in una serie di post pubblicati sul blog di OCLC Research *Hanging Together*.⁹ Negli ultimi sei anni sono stati pubblicati sul blog circa 40 post su argomenti concernenti i metadati. Il Metadata Managers Focus Group è uno dei gruppi attivi all'interno dell'OCLC Research Library Partnership; quest'ultimo si occupa delle ampie opportunità di sviluppo professionale del personale delle biblioteche. I membri del Focus Group considerano l'affiliazione al Research Library Partnership come l'opportunità tramite la quale diventare gli "attori del cambiamento" della futura gestione dei metadati.¹⁰ Inoltre, i riscontri avuti alle questioni poste dal Focus Group hanno generato discussioni all'interno delle istituzioni di appartenenza aiutando gli specialisti di metadati a capire come determinate problematiche fossero comuni.

Le discussioni del Focus Group hanno sollevato la grande varietà di questioni che sono alla base della creazione di questo documento. Il passaggio ad una nuova generazione di metadati è un processo in continua evoluzione e intrecciato all'evoluzione degli standard, delle infrastrutture e degli strumenti. I membri del Focus Group sono giunti ad una visione condivisa delle sfide da affrontare, hanno studiato possibili approcci per affrontarle e si sono confrontati con le comunità con le quali interagiscono.

I membri del Focus Group hanno una vasta esperienza nella gestione dei metadati. Il loro vivo interesse per l'implementazione dei linked data ha dato vita all'OCLC Research's International Linked Data Surveys for Implementers.¹¹ Una parte dei membri del Focus Group ha partecipato a vari progetti sui linked data, incluso il progetto pilota OCLC Research Project Passage and CONTENTdm Linked Data, l'OCLC's Shared Entity Management Infrastructure, il Library of Congress' Bibliographic Framework Initiative (BIBFRAME), il progetto Mellon-funded Linked Data for Production (LD4P), l'iniziativa Share-VDE e il progetto IMLS planning grant Shareable Local Name Authorities che ha posto alcuni problemi relativi agli identificatori in ambiente linked data.¹² Inoltre, i membri del Focus Group contribuiscono in merito agli aspetti che riguardano il lavoro con i linked data, ai gruppi di lavoro PCC: il gruppo su Linked Data Best Practices, il gruppo Identity Management, il gruppo URIs in MARC e il PCC Linked Data Advisory Committee.¹³ Questa contaminazione ha spinto il Focus Group a prendere in considerazione le questioni relative alle entità rappresentate nelle risorse istituzionali.

Questo report riassume 6 anni (2015/2020) di discussioni dell'OCLC Research Library Partners Metadata Managers Focus Group ed indica quali sono le previsioni per i "metadati di nuova generazione". Il documento è organizzato nelle seguenti sezioni, ognuna delle quali rappresenta una tendenza emersa durante le discussioni del Focus Group:

- Il passaggio ai linked data e agli identificatori: ampliamento dell'uso di identificatori persistenti come parte del passaggio dal «controllo di autorità» alla «gestione dell'identità»
- Descrizione delle collezioni "inside-out" e "facilitated": problematiche relative alla creazione e alla gestione di metadati per risorse in vari formati, create o curate dalle istituzioni e condivise all'interno di consorzi
- Evoluzione dei "metadati come servizio": maggior coinvolgimento nella creazione dei metadati al di là del catalogo tradizionale
- Riflessioni sulle future professionalità del personale: il panorama in continua evoluzione richiede nuove competenze, necessarie sia ai professionisti che si stanno formando che ai catalogatori esperti

Il documento si conclude con alcune osservazioni sull'impatto che la nuova generazione di metadati potrà avere sull'intera comunità bibliotecaria.

Il passaggio ai linked data e agli identificatori

I linked data offrono la possibilità di impiegare *dati strutturati* facendo particolare attenzione al contesto nel quale dovranno essere utilizzati. Si basano su identificatori neutri che descrivono oggetti. Particolare attenzione è dedicata a ciò che sostituirà le "stringhe" attualmente utilizzate nei record di autorità e nei cataloghi. Questi identificatori possono essere collegati a dati, vocabolari, termini in altre lingue, discipline e domini, inclusi quelli non bibliotecari. L'applicazione dei linked data può beneficiare del lavoro svolto da altri liberando gli specialisti dall'onere di dover descrivere ciò che è stato già descritto e permettendo loro di focalizzare l'attenzione sulle modalità di accesso a specifiche risorse delle proprie collezioni. La prospettiva è quella di un'esperienza di ricerca più ricca e di una maggior visibilità del proprio patrimonio tramite relazioni non realizzabili tramite gli attuali sistemi. I linked data offrono, inoltre, la possibilità di estendere la ricerca oltre il proprio dominio dato che sono in grado di avvalersi di informazioni provenienti da risorse esterne.¹⁴

Changing Resource Description Workflows

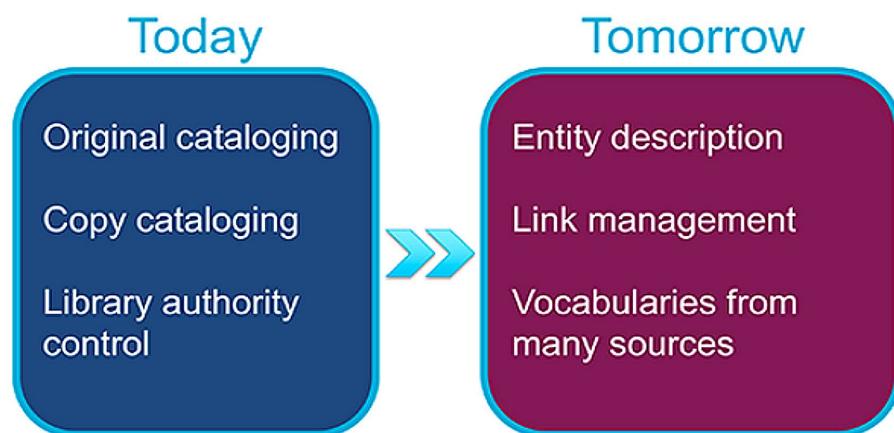


FIGURA 1. "Modifiche dei flussi di lavoro relativi alla descrizione delle risorse" di OCLC Research¹⁵

L'augurio è che i linked data permettano alle biblioteche di offrire nuovi e validi servizi attualmente non realizzabili, che terze parti siano in grado di sfruttare al meglio le descrizioni delle risorse e che i dati, sfruttando la collaborazione nella loro creazione, possano arricchirsi. La transizione ad un ambiente linked data comporta alcune modifiche nei flussi di lavoro dedicati alla descrizione, così come riportato nella figura 1.

Il passaggio dai metadati ai linked data dipende dalla disponibilità di strumenti, dall'accesso a fonti specifiche per il loro riutilizzo, da buone pratiche sugli identificatori e sulle descrizioni ad essi associate ("istruzioni") e dalle risorse da poter impiegare.

LO SVILUPPO DELL'USO DEGLI IDENTIFICATORI PERSISTENTI

Il Focus Group ha focalizzato la propria attenzione sulle "future sfide" che la catalogazione dovrà affrontare. Le problematiche riguardano la possibilità di testare il potere dei metadati sui record tradizionali per poter creare, in futuro, interazioni e nuovi usi. Gli identificatori persistenti sono stati individuati come fondamentali per la transizione dagli attuali metadati alle future applicazioni.¹⁶ Gli identificatori, formati da stringhe alfanumeriche in linguaggio naturale, servono da scorciatoia per aggregare tutti quegli elementi che servono a descrivere in modo univoco un oggetto o una risorsa, e possono essere risolti sulla rete tramite protocolli specifici per la ricerca, l'identificazione e l'uso di quell'oggetto o risorsa. Il numero della previdenza sociale o una sigla identificativa del lavoratore sono esempi di identificatori al di fuori del mondo delle biblioteche. Per quanto riguarda le biblioteche e le università, i membri del Focus Group hanno indicato ORCID (Open Researcher and Contributor ID)¹⁷ come il "collante" in grado di contenere i quattro settori della produzione accademica: editoria, repository, cataloghi e ricercatori—ma ORCID è dedicato solamente ai *ricercatori viventi*. ORCID è sempre più utilizzato nelle riviste delle discipline STEM (scienze, tecnologia, ingegneria, matematica) per identificare autori e curatori¹⁸ ed è integrato con i sistemi di gestione di ricerca delle informazioni (RIM) delle istituzioni. ISNI (International Standard Name Identifier)¹⁹ identifica in modo univoco persone (viventi o meno) ed organizzazioni coinvolte in attività creative di interesse per biblioteche, editori, banche dati e organizzazioni per la gestione dei diritti.

Gli identificatori persistenti sono considerati fondamentali per realizzare la transizione dagli attuali metadati alle future applicazioni.

Gli identificatori persistenti sono usati da Google o HathiTrust per l'integrazione dei servizi.²⁰ Altre istituzioni inseriscono nei metadati, o negli URI (Uniform Resource Identifiers), coordinate geospaziali che puntano ad API (Application Programming Interface) in grado di connettersi con GeoNames,²¹ consentendo quindi la visualizzazione delle mappe. Gli istituti di ricerca, sfruttando i sistemi di gestione delle informazioni di ricerca, stanno utilizzando gli identificatori ORCID per ottimizzare la raccolta dei record di ricerca istituzionali, così come documentato nell'OCLC Research report del 2017 *Convenience and Compliance: Case Studies on Persistent Identifiers in European Research Information Management*.²²

Gli editori giocano un ruolo chiave nel flusso della diffusione dei metadati ma non sempre i dati da loro pubblicati rispondono alle esigenze delle biblioteche. Ad esempio, i dati che gli editori pubblicano per le monografie non includono un identificatore. La British Library sta sviluppando un progetto in collaborazione con cinque editori inglesi per fare aggiungere gli identificatori

ISNI²³ ai loro metadati; questa collaborazione dimostra che editori e biblioteche possono lavorare per un fine comune. La possibilità di aggiungere identificatori tramite algoritmi, o caricarli in batch, è uno dei punti sulla lista dei desideri del Focus Group.

Nessun identificatore di persona è in grado di soddisfare tutte le necessità. I nomi dei ricercatori, viventi o meno, sono stati rappresentati solamente in parte nei file di autorità. Una notevole quantità di nomi è presente solamente nelle stringhe di testo dei record bibliografici. I record di autorità sono creati esclusivamente dalle istituzioni che fanno parte del PCC's Name Authority Cooperative Program (NACO)²⁴ o da istituzioni coinvolte in progetti nazionali. Anche in questo caso, i record di autorità vengono creati solamente per determinate intestazioni. Il file di autorità per autori della LC/NACO contengono solamente il 30% dei nomi riportati nei punti di accesso dei record bibliografici di WorldCat (9 milioni LC/NACO rispetto ai 30 milioni dichiarati nel WorldCat Identities project del 2012).²⁵ Nel 2020, questa percentuale è scesa al 18%: 11 milioni di record di autorità LC/NACO in confronto ai 62 milioni di identità contenute in WorldCat. Queste statistiche dimostrano che i nomi contenuti nei record bibliografici sono di più di quelli contenuti nei record di autorità.

I file di autorità si concentrano sulla "forma preferita" del nome, che può variare a seconda della lingua, della disciplina, del contesto o del periodo temporale. Gli studiosi hanno espresso le loro perplessità riguardo al concetto di "forma preferita" poiché il nome può essere citato in modo diverso a seconda del contesto.²⁶ Quando il nome ha più forme, i ricercatori hanno bisogno di conoscere la forma preferita di ogni singolo nome a seconda delle regole di citazione utilizzate nel loro campo. Un identificatore in grado di collegare le diverse forme di un nome, ognuna associata alla provenienza e al contesto di utilizzo, potrebbe risolvere questo problema.

I nomi dei ricercatori sono solo un esempio delle numerose esigenze non ancora soddisfatte dagli attuali sistemi. Le istituzioni hanno creato propri "identificatori locali" per soddisfare questo specifico bisogno. I casi di utilizzo di identificatori locali prevedono la registrazione di tutti i componenti del campus, comprese quelle entità normalmente poco rappresentate nei file di autorità nazionali. Un esempio sono gli autori di tesi in formato elettronico, gli artisti, gli eventi, i luoghi, i nomi degli edifici, le entità contenute nei progetti di digitalizzazione delle biblioteche e nei repository istituzionali, il multilinguismo della comunità; e la registrazione dei titoli delle collezioni d'archivio a supporto delle attività di amministrazione interna.²⁷

La necessità dei membri del Focus Group di disambiguare i nomi a seconda della disciplina e del formato ha portato alla creazione del gruppo di lavoro Registering Researchers in Authority Files di OCLC Research.²⁸ Il bisogno di registrare chiaramente l'affiliazione di un ricercatore ad una istituzione per testimoniarne la produzione, per promuovere la collaborazione tra enti e per incrementare le iscrizioni e i finanziamenti, ha condotto alla creazione di un altro gruppo: Addressing the Challenges with Organizational Identifiers and ISNI,²⁹ che ha presentato un nuovo modello di dati adattabili ad ogni tipo di istituzione. In seguito è stato creato il Research Organization Registry (ROR) al fine di sviluppare identificativi aperti, sostenibili, univoci ed utilizzabili da qualsiasi istituzione.³⁰

La disambiguazione dei nomi è la parte più impegnativa di controllo di autorità e continuerà ad essere un prerequisito per l'assegnazione di identificatori univoci. Data la presenza di diversi sistemi di identificazione dei nomi, le biblioteche hanno bisogno di uno strumento in grado di riconciliare i nomi. Le revisioni dei file di autorità basati su algoritmi che individuano la corrispondenza delle stringhe presentano dei limiti: i risultati avrebbero in ogni caso bisogno di una revisione da parte degli esperti. Sperimentare personalmente se due identificatori rappresentano la stessa persona potrebbe aiutare a capire l'entità della questione.



FIGURA 2. Alcuni dei [300 nomi di autore in forma abbreviata](#) contenuti nell'articolo di cinque pagine pubblicato in: *Physical Review Letters*

Le forme abbreviate degli autori che appaiono negli articoli delle riviste, rendono spesso difficile—e a volte addirittura impossibile—l'abbinamento alla forma corretta o all'identificatore, sempre che questo esista. L'abbinamento di ORCID agli autori degli articoli rende più semplice la differenziazione di autori che compaiono con la stessa abbreviazione. I sistemi Research Information Management (RIM) gestiscono automaticamente le identità dei ricercatori in modo da associare la forma corretta del loro nome agli articoli che hanno pubblicato. Gli articoli vengono visualizzati nei loro profili. (si veda, ad esempio: Experts@ Minnesota o University of Illinois at Urbana-Champaign's Experts).³³

Per gestire la propria identità i ricercatori devono creare un proprio identificatore ORCID e aggiornarlo costantemente. Le istituzioni incoraggiano l'inclusione di un ORCID nel proprio profilo. I ricercatori, dal loro canto, dimostrano un grande interesse verso l'impiego di ORCID per soddisfare le esigenze di finanziatori come la National Science Foundation o il National Institutes of Health negli Stati Uniti.³⁴ I sistemi di gestione delle ricerche raccolgono i metadati dagli abstract e da database come Scopus, Web of Science o PubMed. Ognuno di questi database ha il proprio identificatore persona e contribuisce in questo modo alla disambiguazione; questi sistemi potrebbero essere anche collegati all'ORCID di un autore. I linked data potrebbero collegare informazioni provenienti da diversi contesti, inclusi i sistemi di gestione delle ricerche: questo richiederebbe un accurato allineamento dei diversi identificatori.

Come primo passo verso un ambiente linked data, e a testimonianza del tentativo di miglioramento dei metadati, alcuni membri del Focus Group hanno provato ad impostare un lavoro di riconciliazione: ricercano termini corrispondenti provenienti da fonti di linked data e aggiungono gli URL nei record di metadati.³⁵ Potenziare la qualità dei dati significa, a breve termine, migliorare l'esperienza degli utenti; in futuro aiuterà la transizione ai linked data. La

Riuscire a disambiguare i curatori o gli autori di articoli pubblicati nelle riviste è un lavoro particolarmente difficile poiché spesso si usano forme abbreviate o possono esserci dozzine, o addirittura centinaia, di collaboratori. Ad esempio, in un articolo di cinque pagine pubblicato in *Physical Review Letters*—[Precision Measurement of the Top Quark Mass in Lepton + Jets Final State](#)—compaiono circa 300 nomi di autori in forma abbreviata contenuti nell'articolo di cinque pagine pubblicato (figura 2).³¹ Questo esempio è la riprova di come siano diverse le abitudini in ogni singola disciplina. All'opposto, altre risorse alle quali partecipano molte persone, come ad esempio i film o le registrazioni orchestrali, sono solitamente rappresentate da un relativo numero di metadati.³² Questi diversi tipi di trattamento rappresentano una sfida per la creazione di metadati univoci, uniformi, comprensibili e riutilizzabili.

maggior parte delle riconciliazioni è effettuata sui nomi di persona, sui soggetti e sui nomi dei luoghi. Le fonti utilizzate per queste riconciliazioni includono l'OCLC's Virtual International Authority File (VIAF), il Library of Congress's linked data service (id.loc.gov), ISNI, la Getty's Union List of Artists Names (ULAN), Art and Architecture Thesaurus (AAT), il Thesaurus of Geographic Names (TGN) e OCLC's Faceted Application of Subject Terminology (FAST), oltre a vari file di autorità di livello nazionale. La selezione delle fonti dipende dall'affidabilità dell'ente produttore, dall'oggetto e dalla ricchezza delle informazioni. Questo tipo di lavoro sulla riconciliazione dei metadati è molto dispendioso e non può essere eseguito meccanicamente.

Alcuni membri del Focus Group hanno cercato di generare degli identificatori (URI persistenti presi da fonti di linked data) da usare al posto delle stringhe di testo. Le istituzioni presso le quali è stato svolto questo esperimento hanno concluso che è più efficace creare da subito un URI in un record di autorità anziché procedere in seguito alla sua riconciliazione. L'Università del Michigan ha sviluppato il programma LCNAF Named Entity Reconciliation³⁶ utilizzando Google's Open Refine che, tramite l'API VIAF, ricerca la corrispondenza con i file VIAF, ricerca le fonti dei record nella Library of Congress ed è in grado di estrarre le intestazioni autorizzate. Tutto questo si traduce in una dataset che associa l'intestazione approvata nel LC Name Authority File heading all'intestazione originale e ad un URI del LCNAF linked data service. Questo servizio potrebbe essere sostituito introducendo un identificatore in VIAF: si otterrebbero risultati corretti anche se basati sulla corrispondenza di stringhe.

È stata identificata un lunga lista di fonti, non appartenenti al mondo delle biblioteche, che potrebbero migliorare gli attuali dati d'autorità o alle quali si potrebbe far riferimento in determinati contesti. Wikidata e Wikipedia sono le prime. Altre risorse potrebbero essere: AllMusic, i siti degli autori e dei fan, Discogs, EAC-CPF (Encoded Archival Context for Corporate Bodies, Persons, and Families), EAD (Encoded Archival Description), family trees, GeoNames, GoodReads, IMDb (Internet Movie Database), Internet Archive, Library Thing, LinkedIn, MusicBrainz, ONIX (ONline Information eXchange), Open Library, ORCID, e Scopus ID. Nel documento del PCC's Task Group on URLs in MARC *Formulating and Obtaining URLs: A Guide to Commonly Used Vocabularies and Reference Sources*,³⁷ viene proposta una esaustiva guida su come importare i dati da queste risorse. Wikidata è considerata un'importante risorsa, rispetto agli attuali sistemi di gestione delle biblioteche, per l'estensione della gamma linguistica e per la fornitura di dati in diverse lingue.³⁸

Gli identificatori di "opere" rappresentano una sfida particolare poiché non esiste una definizione ufficiale di ciò che si intende con "opera distintiva".³⁹ Gli identificatori delle opere generati dalle singole istituzioni non possono essere condivisi o riutilizzabili. I membri del Focus Group hanno espresso preoccupazione per il fatto che le diverse interpretazioni di cosa sia un'"opera" potrebbero ostacolare la possibilità di riutilizzo di dati già creati e la possibilità di rivolgersi ad un repository affidabile e centrale, come OCLC, per pubblicare identificatori persistenti di opere da poter condividere con tutta la comunità.

Esempi di identificatori DOI e ARK

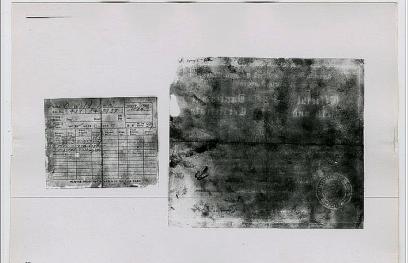
Registration Agency	Examples	[Photographs of Identification Cards]
DataCite	<p>DOI names for accessing registered research datasets:</p> <p>Sets & Subsets:</p> <ul style="list-style-type: none"> Irino, T; Tada, R (2009): Chemical and mineral compositions of sediments from ODP Site 127-797. <i>Geological Institute, University of Tokyo</i>. [doi:10.1594/PANGAEA.726855] <p>Earthquake Event, Authored by Automated System:</p> <ul style="list-style-type: none"> Geofon operator (2009): GEOFON event gfv2009kciu (NW Balkan Region) <i>GeoForschungsZentrum Potsdam (GFZ)</i>. [doi:10.1594/GFZ.GEOFON.gfv2009kciu] 	
Multilingual European DOI Registration Agency mEDRA	<p>Sampling of resources identified using mEDRA DOI names:</p> <ul style="list-style-type: none"> Journal Article: Prodi, Romano. "L'industria dopo l'euro", <i>L'Industria-Rivista di economia e politica industriale</i> 4, 559-566 (2002); [doi:10.1430/8105] Monograph: Attanasio, Piero. <i>"The use of DOI system in eContent value chain: The case of Casalini Digital Division and mEDRA"</i>, White Paper (PDF). [doi:10.1392/BC1.0] 	
Japan Link Center (JaLC)	<p>DOI names for Japanese Journal articles:</p> <ul style="list-style-type: none"> Journal Article: 竹本 賢太郎, 川東 正美, 久保 信行, 左近 多喜男, 大学におけるWebメールとターミナルサービスの研究, 標準化研究 Vol.7(2009), No.1 p.11-20 [doi:10.11467/issn2003.7.1_11] Journal Article: 川崎 英, 植物における免疫誘導と病原微生物の感染戦略, ライフサイエンス 領域融合レビュー, 2, e008 (2013), [doi:10.7875/leading.author.2.e008] 	<p>• Archival Resource Key: ark:/67531/metaph346793</p> <p>Relationships</p> <ul style="list-style-type: none"> [Identification Cards, Photograph #1], DSMA_91-001-photo-344-001, ark:/67531/metaph346554 [Identification Cards, Photograph #2], DSMA_91-001-photo-344-002, ark:/67531/metaph346846 [Identification Cards, Photograph #3], DSMA_91-001-photo-344-003, ark:/67531/metaph346928

FIGURA 3. Esempi di alcuni identificatori DOI (sinistra) e ARK (destra)⁴⁰

Gli identificatori devono essere immutabili nel tempo e indipendenti dalla localizzazione, attuale o futura, dell'oggetto digitale. Ad esempio, identificatori per dataset come risorse digitali o collezioni di repository istituzionali includono ID generati dai sistemi, identificatori locali, PURL, DOI (Digital Object Identifiers), URI, URN, e ARK (Archival Resource Keys). Alcuni esempi di identificatori DOI e ARK sono riportati nella figura 3. La risorsa può avere più copie o le versioni possono cambiare nel tempo. I repository istituzionali, utilizzati come spazi per la condivisione, possono condurre a più pubblicazioni partendo dallo stesso set di dati, la questione si complica quando i ricercatori fanno parte di più istituzioni e depositano le loro pubblicazioni in repository diversi. Inoltre, le biblioteche (così come i finanziatori e i tentativi di valutazione nazionale) vorrebbero essere in grado di collegare ad una pubblicazione tutti i suoi apparati (anteprime, dati supplementari, immagini). Si pone quindi il problema di differenti DOI che puntano allo stesso oggetto. Alcune biblioteche utilizzano DataCite o Crossref per creare e pubblicare identificatori unici e permanenti e ridurre al minimo la possibilità di trovare link che puntano a risorse non più disponibili.⁴¹ Idealmente, le biblioteche dovrebbero contribuire a creare un hub per i metadati che descrivono i set di dati dei loro ricercatori, indipendentemente da dove sono archiviati i set di dati.

Occorre cambiare la modalità di gestione delle identità: dalla fornitura di punti di accesso tramite la descrizione delle risorse, si deve passare alla descrizione delle **entità** che appaiono nella risorsa (opere, persone, enti, luoghi, eventi) stabilendone relazioni e collegamenti.

IL PASSAGGIO DAL "CONTROLLO DI AUTORITÀ" ALLA "GESTIONE DELLE IDENTITÀ"

Il focus del lavoro di gestione delle autorità sta passando dalla costruzione delle stringhe di testo alla gestione dell'identità—differenziare le entità, creare identificatori e stabilire relazioni tra le entità.⁴² Il lavoro intellettuale richiesto per differenziare i nomi è lo stesso. I membri del Focus Group credono che la tendenza vada verso questa direzione: gestione delle identità e progressivo allontanamento dalla "gestione delle stringhe di testo" come base per il controllo delle intestazioni dei record bibliografici.⁴³ La gestione delle identità richiede una nuova prospettiva: dalla fornitura di punti di accesso tramite la descrizione delle risorse, si deve passare alla descrizione delle entità che appaiono nella risorsa (opere, persone, enti, luoghi, eventi) stabilendo relazioni e collegamenti tra esse.

Il passaggio dal "controllo di autorità" ai "punti di accesso autorizzati" nei sistemi per la gestione delle identità richiede la separazione degli identificatori dalle etichette a loro associate. Un identificatore univoco potrebbe essere associato ad un insieme di attributi, il che permetterebbe agli utenti di distinguere un'entità da un'altra.⁴⁴ Idealmente le biblioteche potrebbero avvalersi di identificatori e attributi generati da fonti diverse. Wikidata, ad esempio, aggredisce una varietà di identificatori ed etichette per lingue diverse, così come mostrato nella figura 4.

Un identificatore Wikidata collegato ad altri identificativi e ad altre etichette in lingue diverse

Identificatore Wikidata Q19526



Angelou reciting her poem "On the P Morning" at US President Bill Clint inauguration, January 20, 1993

Born	Marguerite Annie Johns April 4, 1928 St. Louis, Missouri, U.S.
Died	May 28, 2014 (aged 86) Winston-Salem, North C U.S.
Occupation	Writer - poet - civil rights
Period	1951–2014
Subject	Memoir - poetry
Notable works	<i>I Know Why the Caged Sings</i> "On the Pulse of Morning"
Spouses	Tosh Angelos (m. 1951; div. 1954) Paul du Feu (m. 1974; div. 1983)
Children	1
Website	www.mayaangelou.com

Other Identifiers

- BnF ID
- Freebase ID
- GND ID
- IMDb ID
- ISNI ID
- Library of Congress ID
- Music Brains Artist ID
- National Thesaurus for AuthorNames ID
- Viaf ID

Other Labels

ar	مaya أنجيلا	ja	マヤ・アンジェロウ
azb	مايا آنجلو	ko	마야 앤젤로
bg	Мая Анджею	ml	മൊയ് അംജേലൂ
el	Μάγια Αγγέλου	ne	মায়া এন্জেলো
fa	مايا آنجلو	si	මායියා අභ්‍යන්තරේල්
he	מָאיָ אַנְגֵלֹו	zh	马娅·安杰卢
hy	Մայա Էնջելու		



FIGURA 4. Un [identificatore Wikidata collegato ad altri identificatori ed etichette](#) in lingue diverse

Fornire informazioni puntuali è più importante che fornire un'etichetta univoca. Le etichette possono essere diverse a seconda delle comunità—come i vari tipi di ortografia di nomi e termini, lingue diverse, diversi sistemi di scrittura e diverse discipline—senza che una forma sia preferibile ad un'altra. La preferenza per un tipo di etichetta diventa una scelta locale anziché una scelta globale.

Un ostacolo fondamentale al passaggio dalle stringhe di testo alla gestione delle identità è la mancanza di tecnologie e infrastrutture per supportarlo. Sono necessari nuovi strumenti che permettano di indicizzare e visualizzare le informazioni sulle entità descritte, e di inserire i link alle fonti dell'identificatore. Dato che identificatori diversi possono puntare alla stessa entità, sarebbero opportuno dotarsi anche di strumenti in grado di supportare questo tipo di riconciliazione. Alcuni sistemi indicizzano solamente i punti di accesso, il che rappresenta un problema quando si ha a che fare con nomi in lingue differenti. È possibile riconfigurare i sistemi di biblioteca in modo che riescano a gestire gli *identificatori* come punti di corrispondenza, collocazione e come chiavi per tutte le etichette associate visualizzate e indicizzate?⁴⁵

Alcuni membri del Focus Group stanno sperimentando Wikidata come opzione per assegnare identificatori ai nomi non presenti nei file di autorità, opzione che amplierebbe il potenziale bacino di contributori.⁴⁶ Molte biblioteche si rivolgono a Wikidata o a Wikibase—la piattaforma alla base di Wikidata—per risolvere alcune annose questioni riscontrate dal personale tecnico e dagli archivisti.⁴⁷ Wikidata/Wikibase sono percepite come una possibile alternativa al tradizionale controllo di autorità e offrono, inoltre, il vantaggio di supportare sia il multilinguismo, sia il collegamento dei silos che descrivono le risorse di un'istituzione. Le sperimentazioni dei membri del Focus Group con Wikidata, e i progetti portati avanti da OCLC che impiegano Wikibase, indicano quest'ultima come una struttura in grado di realizzare un'implementazione dei linked data. Questa infrastruttura permetterebbe al Focus Group, e all'intera comunità bibliotecaria ed archivistica, di focalizzare la propria attenzione sulla creazione delle entità, sulle loro relazioni reciproche e sul modo migliore per aumentarne la rilevabilità da parte degli utenti finali.

La gestione delle identità potrebbe anche colmare le variazioni dei nomi contenute negli articoli delle riviste, nei profili accademici e nei cataloghi delle biblioteche a prescindere dai domini di appartenenza. Questo è un requisito fondamentale per concretizzare le potenzialità dei linked data.

Wikidata è stata creata sfruttando i dati di Wikipedia. La rappresentazione dei libri in Wikidata è focalizzata sulle opere e sui loro autori. L'attenzione rivolta alle "opere" potrebbe essere considerata una versione alternativa all'intestazione autore/titolo dei file di autorità. I libri più "famosi" hanno maggiori possibilità di essere presenti in Wikidata. Recentemente, lo sforzo di supportare le citazioni negli articoli di Wikipedia - ovvero WikiCite⁴⁸ - dimostra la necessità di registrare e supportare gli identificatori che compongono tali citazioni, comprese le informazioni su un'edizione o un documento specifico.

Uno degli aspetti più pratici—e potenti—della gestione delle identità è la riduzione dei copia/incolla nei flussi di lavoro quando un identificatore viene individuato in una posizione esterna. Gli identificatori potrebbero creare un ponte tra ambienti MARC, non MARC e fonti non appartenenti alle biblioteche. I bibliotecari non dovrebbero essere esperti in tutti i settori.⁴⁹ Molte risorse curate o gestite dalle biblioteche non sono sottoposte a controllo di autorità, come le collezioni digitali, quelle archivistiche, i repository istituzionali e i dati di ricerca. Gli identificatori potrebbero fornire un collegamento a queste risorse. La gestione delle identità potrebbe anche rappresentare un collegamento tra le varie forme dei nomi presenti negli articoli delle riviste, nei profili accademici e nei cataloghi delle biblioteche a prescindere dai domini di appartenenza. Questo è un requisito fondamentale per concretizzare le potenzialità dei linked data.

AFFRONTARE LA NECESSITÀ DI VOCABOLARI MULTIPLI, EQUITÀ, DIVERSITÀ E INCLUSIONE

I concetti o le intestazioni di soggetto sono particolarmente insidiosi poiché i termini possono cambiare a seconda del periodo storico o della disciplina. In alcuni casi i termini possono essere considerati peggiorativi, dannosi o addirittura razzisti da parte di alcune comunità. Affrontare i problemi relativi al linguaggio è importante poiché le biblioteche cercano di sviluppare relazioni e tentano di essere un punto di riferimento per le comunità emarginate. Le questioni relative all'equità, alla diversità e all'inclusione sono complesse e il vocabolario usato nelle intestazioni di soggetto ne rappresenta solo un aspetto; gli identificatori linguistici neutri potrebbero costituire un possibile approccio alla soluzione di questo problema.

La questione relativa alla possibilità di supportare intestazioni di soggetto "alternative" è ascesa alla ribalta quando non è stato possibile implementare la decisione della Library of Congress di cambiare l'intestazione di soggetto da "Illegal aliens" a "Undocumented immigrants". Un membro del Focus Group ha commentato così l'accaduto: "Essere tenuti in ostaggio da un sistema nazionale lento a cambiare rispetto all'evoluzione semantica è dannoso per le biblioteche, dato che siamo orgogliosi della nostra accoglienza e inclusività". Gli utenti ritengono le biblioteche responsabili di ciò che appare nei loro cataloghi. Sebbene il LCSH sia lo strumento di lavoro della *Library of Congress*, viene utilizzato in tutto il mondo, a volte non prendendo in considerazione il contesto nel quale è stato creato.⁵⁰

**Affrontare i problemi relativi al linguaggio
è importante poiché le biblioteche
cercano di sviluppare relazioni e tentano
di essere un punto di riferimento per le
comunità emarginate.**

La Faceted Application of Subject Terminology (FAST)⁵¹ è percepita come uno strumento in grado di coinvolgere la comunità nella soluzione dei problemi riscontrati durante la creazione delle intestazioni di soggetto "alternative" per LCSH. Una parte dei membri del Focus Group ha applicato FAST ad alcuni record sprovvisti di soggetto. FAST è stato sviluppato da OCLC per ottenere uno strumento che fosse una via di mezzo tra parole chiave non controllate da una parte e studio e costruzione di stringhe precoordinate difficili da apprendere e applicare dall'altra.⁵² Le intestazioni FAST forniscono una facile transizione all'ambiente linked data, dato che ogni intestazione prevede un identificatore univoco. Poiché le intestazioni FAST sono generate in base alle intestazioni di soggetto precoordinate della Library of Congress, è possibile che includano alcuni termini considerati inappropriati o irrispettosi.

Il FAST Policy and Outreach Committee⁵³ recentemente istituito rappresenta gli utenti FAST e valuta il coinvolgimento della comunità, i contributi apportati e le procedure utilizzate; propone inoltre i miglioramenti da attuare. Nella sua vision si legge:

FAST sarà un vocabolario di soggetti derivato da LCSH; verrà sviluppato dalla comunità, sarà completamente supportato e ampiamente adottato; comprenderà strumenti e servizi in grado di soddisfare comunità e contesti diversi.⁵⁴

Esistono vocabolari che si sovrappongono, e a volte confliggono, nei dati esistenti delle biblioteche.⁵⁵ Ad esempio, alcuni membri del Focus Group in Nuova Zelanda aggiungono termini tratti dal Māori Subject Headings thesaurus (Ngā Upoko Tukutuku) a record che già sono provvisti di un'intestazione di soggetto della LC; membri australiani del Focus Group aggiungono termini autorizzati dai thesauri dell'Australian Institute of Aboriginal and Torres Strait Islander Studies (AIATSIS).⁵⁶ Il rischio è quello di creare equivalenze non soddisfacenti tra le varie lingue. Concetti contenuti nei vocabolari non sempre possono essere mappati in modo inequivocabile con i concetti inglesi. Il MACS (Multilingual Access to Subjects)⁵⁷ costruisce relazioni tra tre diversi vocabolari: il Library of Congress Subject Headings, il tedesco GND integrated authority file, e il francese RAMEAU (Répertoire d'autorité-matière encyclopédique et alphabétique unifié). È stato un processo molto dispendioso ma non sembra essere ampiamente implementato.⁵⁸

Una percentuale sempre maggiore di dati ricercabili sui discovery deriva da fonti esterne alle biblioteche. I metadati che descrivono dati di ricerca e materiali contenuti nei repository istituzionali sono tradizionalmente trattati in modo diverso—e distinto. In che modo le istituzioni possono provvedere alla normalizzazione e all'accesso alle entità descritte non provocando una "collisione" tra nomi e termini ambigui (o termini che hanno significati diversi a seconda della fonte)? Graphite⁵⁹ - il programma di gestione delle ontologie sviluppato da Synaptica Knowledge Solution - è uno strumento in grado di creare e gestire varie tipologie di vocabolari controllati e sembra essere promettente per la soluzione di questi problemi.

I membri del Focus Group hanno citato diversi esempi di vocabolari o dataset obsoleti o non più rispondenti alle attuali esigenze. Modelli vecchi o non adatti all'implementazione di vocabolari consolidati hanno spinto alla valutazione di nuovi modelli. La necessità di una preparazione specifica, finalizzata all'applicazione dei modelli attuali, ha contribuito allo sviluppo di nuove alternative.⁶⁰

I linked data possono fornire alle comunità il mezzo per scegliere un'etichetta per la descrizione di un concetto, o di un'entità, al posto del termine preferito suggerito da un vocabolario consolidato. Si potrebbe far riferimento alla descrizione di un concetto o di un'entità non rappresentata—o rappresentata in modo non soddisfacente—in un vocabolario consolidato o in una fonte di linked data. Se questo tipo di modifiche, o aggiunte, sono rese possibili in ambiente linked data, altri potrebbero essere d'accordo (o non esserlo) con il punto di vista espresso, collegandosi o meno alla nuova risorsa. Questo "modello distribuito" di gestione dei termini e delle entità crea delle aspettative nei confronti della stabilità, dell'interoperabilità e del mantenimento dei metadati. In che modo un "modello distribuito" potrebbe evitare la duplicazione del lavoro sulla stessa entità o sullo stesso concetto? In che modo un "modello distribuito" potrebbe essere in grado di verificare l'affidabilità di un collaboratore o determinare chi è autorizzato a contribuire?

Istituti educativi e biblioteche hanno intrapreso iniziative in ambito EDI che sono state sostenute dai dipartimenti dedicati ai metadati.

I problemi relativi alla stabilità e alla durata sono stati messi in evidenza in seguito alla creazione di numerosi vocabolari sviluppati per specifici progetti che, una volta esauriti i finanziamenti, non vengono più utilizzati. Come ha notato uno dei membri del Focus Group: "Nulla è più triste di un vocabolario creato da qualcuno e non più implementato". Questi esempi forniscono un importante spunto di riflessione per i bibliotecari che vorrebbero fare affidamento su file di autorità internazionali piuttosto che su soluzioni locali; inoltre, esemplificano e mostrano il valore del lavoro svolto dalla Library of Congress che si assume l'intero costo di creazione e mantenimento di LCSH.

Il report di OCLC Research che presenta i risultati di un sondaggio condotto da Research Library Partnership on equity, diversity, and inclusion (EDI)⁶¹ pubblicato nel 2017, ha stimolato discussioni relative alla complessità di prevedere equità, diversità e inclusione nei vocabolari controllati.⁶² Istituti educativi e biblioteche hanno intrapreso iniziative in ambito EDI che sono state sostenute dai dipartimenti dedicati ai metadati. L'estratto dell'indagine EDI, riportato nella figura 5, mostra come i metadati presenti nei cataloghi delle biblioteche siano in ritardo rispetto agli obiettivi e ai principi EDI stabiliti dalle istituzioni.

Cosa è cambiato o progetti di cambiare per essere in linea con i principi EDI della tua istituzione?

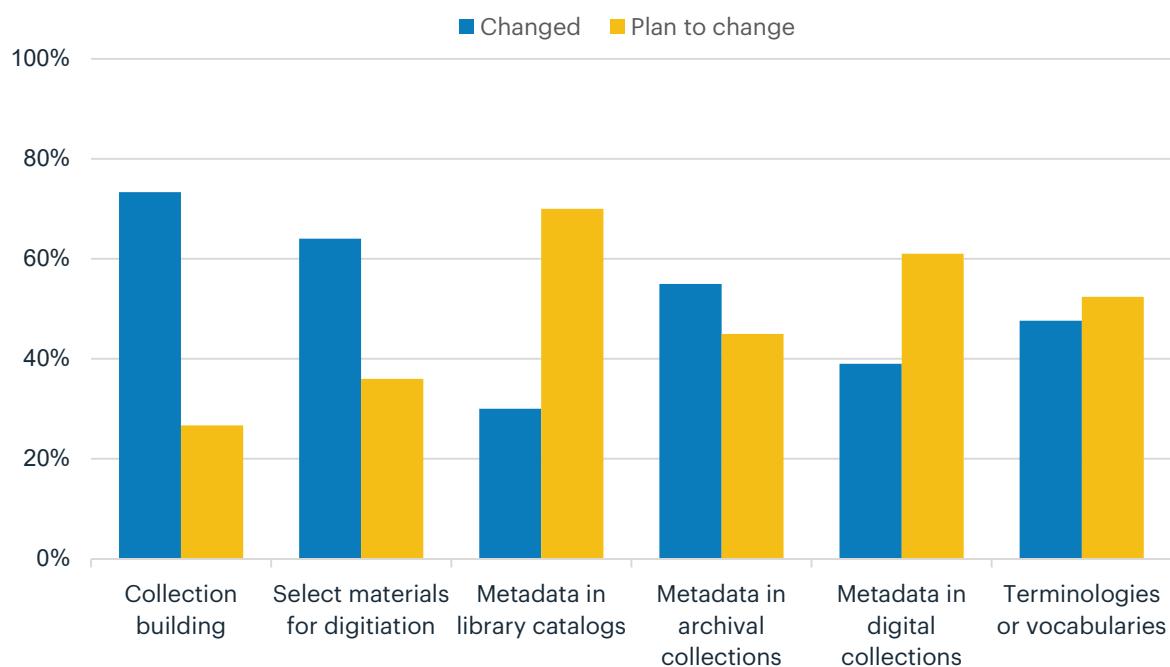


IMMAGINE 5. Estratto dei risultati del sondaggio EDI condotto dal Research Library Partnership nel 2017⁶³

I membri del Focus Group sono impazienti di fornire un accesso ai soggetti più efficace di quello attualmente offerto dai sistemi nazionali, come il LCSH, che risulta essere molto specifico in relazione ai luoghi dell'Europa occidentale e poco accurato in relazione ai luoghi del sud est asiatico o dell'Africa. Il Focus Group è sensibile alla necessità di offrire termini più specifici e tenderebbe a sostituire i termini che riflettono pregiudizi, o considerati offensivi, con termini neutri.

Problematiche che il Focus Group ha identificato per offrire agli utenti una terminologia più "rispettosa": negli accessi per soggetto.

- **Ricerca:** usare termini meno offensivi potrebbe comportare una perdita delle informazioni contenute nei cataloghi compromettendo i risultati della ricerca.
- **Perdita di consenso:** i membri del Focus Group dubitano che ci possa essere un consenso unanime sulle stringhe suggerite. Alcuni termini che possono risultare offensivi per una comunità, potrebbero non esserlo per un'altra. (Ad esempio: "Dissident art" piuttosto che "Non-conformist art").
- **Velocità:** il processo di modifica delle intestazioni di soggetto potrebbe essere molto lento.
- **Capacità:** la modifica delle intestazioni nei record esistenti può richiedere un grandissimo impegno. L'aggiornamento dei punti di accesso avviene generalmente durante la fase di lavorazione. Ad esempio, la Library of Congress ha recentemente modificato l'intestazione "Mentally handicapped" con "People with mental disabilities". L'implementazione di questo tipo di modifiche nel catalogo richiede una serie di azioni automatiche e può comportare alcuni interventi manuali o la decisione di ricollocare alcune risorse.⁶⁴ Alcuni bibliotecari hanno quindi suggerito che potrebbe essere meno impegnativo presentare direttamente nell'interfaccia di ricerca un messaggio di "sensibilità culturale" in cui si avvertono gli utenti che i termini e le annotazioni che si trovano nel catalogo possono riflettere l'atteggiamento del creatore o il periodo in cui l'opera è stata creata e potrebbero risultare non appropriati in determinati contesti.
- **Condivisione:** i vocabolari locali non sono condivisibili da altri sistemi.
- **Manutenzione:** sono stati creati vocabolari locali poiché adatti ad un determinato contesto, ma sono stati reputati troppo onerosi e non sono più stati implementati.
- **Barriere linguistiche:** il linguaggio impiegato nei nostri vocabolari controllati potrebbe risultare non comprensibile alle persone che non conoscono la nostra lingua. L'Ohio State University Libraries ha cercato di risolvere questo problema sviluppando script non latini equivalenti a soggetti in lingua inglese.
- **Classificazione:** gli attuali sistemi di classificazione tendono a proporre una divisione dei gruppi etnici. Piuttosto che includerli come parte di un generico concetto come la storia, l'istruzione o la letteratura, i sistemi di classificazione tendono a raggruppare le etnie in un unico insieme. Man mano che le istituzioni usano depositi esterni il bisogno di collocare i materiali tramite un'unica classificazione è venuto meno; pochi, però, applicano più di una classificazione ai record.

I requisiti di un sistema collaborativo che ospiti più vocabolari e sia in grado di accogliere anche EDI, devono rispondere alla necessità di supportare relazioni semantiche attraverso vocabolari differenti. I centri di sperimentazione hanno bisogno di un ambiente in cui aggregare e riconciliare i termini all'interno dei loro domini. Durante le sperimentazioni è stato notato che un termine usato può essere in conflitto con un altro o avere un significato diverso. La PCC Linked Data Advisory Committee's Linked Data Infrastructure Models: Areas of Focus for PCC Strategies⁶⁵ propone una serie di requisiti e modelli affinché le istituzioni culturali adottino i linked data come strategia per la condivisione dei dati. I modelli devono essere applicabili e modulabili, devono avere la capacità di accogliere la proliferazione di termini e gli argomenti specifici delle discipline umanistiche e devono facilitare il contributo dei ricercatori. Il sistema collaborativo necessita dunque di flessibilità per poter permettere la coesistenza di vocabolari diversi.

La sostituzione delle stringhe con identificatori stabili e persistenti faciliterebbe l'impiego di etichette diverse a seconda del contesto. In questo modo potrebbero essere contenute lingue diverse e differenti ortografie (come ad esempio l'American-english e il British English) così come termini più rispettosi delle comunità marginali. Il 19 ottobre 2017 durante il webinar della serie OCLC Research Works in Progress dal titolo "Decolonizing Descriptions: Finding, Naming and Changing the Relationship between Indigenous People, Libraries, and Archives", ⁶⁶ è stato presentato il progetto lanciato dalla Association for Manitoba Archives e dalle biblioteche della University of Alberta che ha per obiettivo l'esame delle intestazioni di soggetto e degli schemi di classificazione al fine di capire come potrebbero essere più rispettosi ed inclusivi rispetto alle popolazioni indigene.

Espandere i vocabolari per poter includere quelli usati da altre comunità richiede la costruzione di rapporti di fiducia. Un modello di "comunità collaborativa" che vota la creazione di nuovi termini potrebbe essere più inclusivo. Attualmente i processi di verifica all'interno delle biblioteche escludono molte persone. La maggior parte dei metadati è attualmente creata rispetto ai costrutti della cultura occidentale e ai sistemi progettati per essa. La comunicazione della cronologia dei cambiamenti e la provenienza di ogni nuovo termine garantirebbe l'affidabilità della fonte e una maggiore trasparenza delle informazioni. La cronologia delle modifiche e le pagine di discussione che fanno parte di ogni entità descritta in Wikipedia rappresentano un possibile modello da seguire. Prevedere l'obbligo della citazione della provenienza in un vocabolario condiviso comporterebbe la creazione di un ambiente maggiormente equo, diversificato ed inclusivo.

LE SFIDE DEI LINKED DATA

Identifieri e vocabolari sono solamente due delle componenti richieste per la transizione ai linked data. Una parte fondamentale della descrizione delle entità sono le *dichiarazioni* ad esse associate. Come potranno le biblioteche risolvere i conflitti generati da dichiarazioni diverse? ⁶⁷ Potranno verificarsi diversi tipi di incongruenze rispetto ad ora: ad esempio, le diverse date di nascita di una persona. La provenienza di ogni dichiarazione diventa quindi un aspetto cruciale. Anche negli attuali ambienti di lavoro una fonte può essere considerata dai catalogatori più attendibile rispetto ad un'altra. Le biblioteche hanno spesso una lista di "fonti preferite". ⁶⁸ OCLC Research, nel 2015, durante il "Works in Progress Webinar: Looking Inside the Library Knowledge Vault" ⁶⁹ ha valutato come le biblioteche potrebbero avvalersi del "Knowledge Vault" di Google per identificare dichiarazioni più "veritieri" di altre. Alcuni membri del Focus Group pensano che aggregatori come WorldCat, il Virtual International Authority File (VIAF) e Wikidata, fonti riconosciute come attendibili, potrebbero consentire alle biblioteche di visualizzare le loro dichiarazioni. I bibliotecari potrebbero mettere a disposizione le loro competenze per stabilire le relazioni tra dichiarazioni diverse provenienti da fonti diverse.

Tuttavia linked data affidabili dipendono dalla qualità dei metadati. Gli amministratori sono consapevoli delle problematiche relative alla fornitura di una buona descrizione e all'accesso tempestivo alle collezioni. La descrizione contenuta nei metadati deve essere sufficientemente dettagliata per permettere alle biblioteche di gestire le proprie collezioni e assicurarne la ricercabilità per l'utente finale. Molte biblioteche devono scendere ad un compromesso tra velocità e accuratezza, velocità e profondità o brevità rispetto a niente. Questa problematica viene in parte risolta tramite l'impiego di record non accurati messi a disposizione dai fornitori. In questo modo vengono creati record con descrizioni a livello minimo per determinati tipi di risorse e viene limitato il controllo di autorità. ⁷⁰ Pur di non lasciare i materiali inutilizzati, viene spesso utilizzata la catalogazione a livello minimo, spesso a causa del grande volume di materiale da catalogare a fronte di personale insufficiente. Queste descrizioni incomplete si tradurranno in dichiarazioni di linked esigue e data meno accurate.

Linked data affidabili dipendono dalla qualità dei metadati.

La transizione dai sistemi di catalogazione basati sul MARC ad un ambiente linked data basato su entità e dichiarazioni deve affrontare molte sfide poiché sia gli standard che le procedure stanno cambiando. Non è ben chiaro in che modo le biblioteche condivideranno non più record ma dichiarazioni nel contesto dei linked data. I membri del Focus Group hanno individuato due direzioni: la prima prevede un archivio di dati centralizzato e in grado di fornire informazioni sulla provenienza dei dati; la seconda auspica una condivisione di dati tra pari.⁷¹ Dichiarazioni diverse potrebbero essere corrette all'interno di un determinato contesto. Le divergenze tra dichiarazioni potrebbero anche rappresentare punti di vista diversi. La selezione delle dichiarazioni in base alla loro provenienza potrebbe rappresentare una sfida per i nostri principi di equità, diversità ed inclusione.

I membri del Focus Group hanno pensato a come coinvolgere i fornitori di record MARC nel processo di transizione ai linked data. Nel Regno Unito, il progetto di Jisc, "Plan M" (dove M sta per metadata) mira a semplificare la fornitura di metadati a biblioteche, editori, fornitori di dati e infrastrutture.⁷² Uno dei problemi sollevati dalla National Bibliographic Knowledgebase (NBK) del Regno Unito rispetto alla visione decennale del Plan M è stato: "I linked data della NBK devono essere creati e implementati sulla base di esperienze in grado di testimoniare l'impatto che questi possono avere sulla ricerca".⁷³ Il lavoro nell'ambito dei linked data coinvolge persone che non hanno familiarità con l'ambiente bibliotecario; questo comporta, da parte degli specialisti, la disponibilità a far presenti le proprie esigenze usando un linguaggio comprensibile anche ai non addetti ai lavori.

Descrizione delle collezioni "Inside-Out" e "Facilitated"

Lorcan Dempsey, vice presidente e Chief Strategist di OCLC, fa riferimento al mutevole interesse delle biblioteche nei confronti della creazione, dell'implementazione e della ricercabilità delle proprie risorse (ad esempio le "inside-out collection"); atteggiamento che contrasta con quello riservato alle "outside-in collection", per le quali vengono acquistate licenze e materiali da fornitori esterni con il fine di renderli disponibili alla propria utenza locale. Una "facilitated collection"⁷⁴ consente l'accesso ad un'ampia gamma di risorse locali, esterne o condivise rispondenti ai bisogni degli utenti. Le attività svolte dai membri del Focus Group sono sempre più incentrate sulla creazione di metadati in grado di fornire l'accesso alle risorse uniche dell'istituzione, dei consorzi o delle reti nazionali.

Tutte le risorse raccolte, create e implementate dalle biblioteche hanno bisogno dei metadati per essere ricercabili. In ogni caso i membri del Focus Group hanno concentrato la propria attenzione sulle problematiche relative ai seguenti tipi di formato:

- Collezioni archivistiche
- Siti web archiviati
- Raccolte audio e video
- Collezioni di immagini
- Dati di ricerca

Questo tipo di contenuti può essere considerato parte delle collezioni "inside-out" e presenta diverse problematiche. I membri del Focus Group, ad esempio, descrivono come "eccezionali" gli sforzi che devono essere compiuti per recuperare i metadati da sistemi completamente diversi. Inoltre, molte di queste risorse non sono sottoposte ad alcun tipo di controllo di autorità. La riconciliazione dei punti di accesso attraverso l'analisi di thesauri diversi e la mappatura dei metadati, richiede notevole esperienza e competenze tecniche.⁷⁵ Questo tipo di riconciliazione sarebbe necessaria anche nell'ambiente linked data descritto precedentemente.

La sezione che segue sintetizza le discussioni sui diversi tipi di formato.

COLLEZIONI ARCHIVISTICHE

Le collezioni archivistiche sono spesso considerate le punte di diamante delle collezioni poiché in grado di fornire informazioni attraverso i secoli e i luoghi; spesso rappresentano le fonti principali per la creazione di nuove conoscenze. Archivi, biblioteche e studiosi potrebbero beneficiare dell'aumentata visibilità di queste collezioni. In ogni caso gli archivi sono organismi molto complessi e, parlando di metadati, presentano problematiche diverse rispetto a quelle delle biblioteche. Man mano che le istituzioni si rivolgono ad ArchiveSpace o ad altri sistemi di gestione in grado di fornire infrastrutture per la gestione dei metadati strutturati degli archivi, stanno emergendo le problematiche da affrontare.⁷⁶

All'interno delle istituzioni di appartenenza, gli archivi godono di maggiore autonomia rispetto alle biblioteche a causa dell'unicità delle loro collezioni, della utenza specialistica, degli standard usati per la gestione dei metadati e dei loro sistemi. Sebbene alcune istituzioni abbiano integrato i processi di elaborazione degli archivi all'interno dei propri servizi tecnici, la maggior parte mantiene una gestione separata. Gli archivisti tradizionalmente non creano record di autorità e non condividono identificatori per le entità così come succede tra i bibliotecari. Inoltre, essi tendono ad utilizzare la forma completa di un nome basandosi sulle informazioni contenute nelle collezioni, mentre i bibliotecari si basano sulle forme "preferite" individuabili nelle pubblicazioni. A prescindere da questo, si sta verificando un significativo passaggio da un approccio artigianale ad una standardizzazione dei metadati.

Come possono, quindi, archivisti e bibliotecari integrare i rispettivi metadati e le forme dei nomi? Il numero di "autore persona" all'interno delle collezioni archivistiche può essere così ampio che la maggior parte rischia di non essere sottoposto al controllo di autorità e di restare privo di un identificatore. È altresì vero che le informazioni fornite da un archivista relativamente a persone o ad organizzazioni potrebbero arricchire i file di autorità—possibilità sperimentata nel 2017-2018 con il progetto pilota Project Passage,⁷⁷ ed esaminata in modo più approfondito tra il 2019 e il 2020 dall'OCLC Research Library Partners Archives and Special Collections Linked Data Review Group.⁷⁸

L'incremento dell'uso delle risorse elettroniche e digitali riscontrato durante la pandemia di COVID-19 stimolerà le istituzioni a digitalizzare i loro archivi e le collezioni che finora sono state disponibili solamente in formato cartaceo.⁷⁹ Dalla versione digitalizzata di queste risorse sarà possibile creare ulteriori metadati.

ARCHIVIAZIONE DI SITI WEB

Archivi e biblioteche hanno iniziato da alcuni anni ad archiviare le risorse web di interesse accademico o istituzionale al fine di garantire la loro accessibilità e la loro sopravvivenza nel tempo. Alcuni siti web sono effimeri o intenzionalmente temporanei, come quelli creati per uno specifico evento. Le istituzioni vorrebbero archiviare e conservare i contenuti dei loro siti come testimonianza storica. La maggior parte dei contenuti web viene analizzata dai web crawler, ma i metadati generati dalla sola raccolta non assicurano la loro ricercabilità.⁸⁰

Sono stati archiviati alcuni siti web istituzionali e alcune collezioni tematiche che supportano una specifica area di ricerca, come ad esempio la Columbia University's Human Rights, Historic Preservation and Urban Planning e New York City Religions.⁸¹ Anche le biblioteche nazionali archiviano i propri siti web all'interno dei propri domini, ad esempio la Library of Australia's Archivato il websites (dal 1996 ad oggi)⁸² contiene, in collaborazione con altre istituzioni culturali australiane, i siti web istituzionali precedentemente accessibili tramite l'Australian Government Web Archive e i siti appartenenti al dominio .au; questi siti vengono analizzati annualmente su larga scala tramite web crawler. Queste collezioni, suddivise in base al soggetto, forniscono un'esauriva immagine della storia culturale e sociale dell'Australia. Esempi di archivi di siti web archiviati da consorzi sono: the Ivy Plus Libraries Confederation's Collaborative Architecture, Urbanism, and Sustainability Web Archive (CAUSEWAY), Contemporary Composers Web Archive (CCWA) e New York Art Resources Consortium (NYARC), che acquisisce dati dinamici delle versioni web di cataloghi di aste, siti di artisti, di gallerie e di musei disponibili.⁸³

Il Focus Group ha affrontato anche il tema relativo alla creazione e alla gestione dei metadati necessari per migliorare i metadati raccolti automaticamente dai siti web. Ecco alcune delle questioni rilevate:

- **Importanza della tipologia di sito web.** La creazione dei metadati può dipendere dal tipo di sito archiviato (ad es. siti transitori, dati di ricerca, social media, siti aziendali). A volte, quando l'esperienza dell'utente non viene considerata significativa (relativamente al suo aspetto), viene archiviato solamente il contenuto del sito.
- **Gestioni diverse.** Alcune caratteristiche dei siti web non sono contemplate nelle attuali regole descrittive tipo RDA (*Resource Description and Access*) e DACS (*Describing Archives: A Content Standard*). I metadati, in relazione al contesto in cui vengono creati, tendono a seguire le tradizionali descrizioni bibliografiche o le pratiche in uso negli archivi.
- **Quantità dei dati e usabilità.** I requisiti necessari alla creazione di metadati possono essere differenti a seconda della mole del materiale che deve essere archiviato e dell'uso che se ne dovrà fare. Ad esempio: gli umanisti analizzano i contenuti dei siti web con lo scopo di identificare tendenze, mentre altri utenti hanno solamente bisogno di consultare singole pagine. Il livello di granulosità dei metadati (collezioni, URL, documenti) può variare anche in base alle previsioni dei bisogni dell'utenza, alla previsione della mole di lavoro e al personale disponibile.
- **Frequenza degli aggiornamenti.** Molti siti web sono costantemente aggiornati e richiederebbero un costante monitoraggio. Alcuni tipi di modifica possono causare errori di acquisizione.
- **Siti web multi-istituzionali.** Alcuni siti web vengono archiviati da più istituzioni. Ogni istituzione potrebbe catturare lo stesso sito in momenti differenti e con indicazioni diverse per il web crawling. Come possono essere ricercati e utilizzati insieme?

Un sondaggio del 2015 dell'OCLC Research Library Partnership sottolineava la "mancanza di linee guida per metadati descrittivi" come la più grande sfida da affrontare relativamente all'archiviazione dei siti web; questo ha portato alla creazione dell'OCLC Research Library Partnership Web Archiving Metadata Working Group.⁸⁴ Le problematiche sollevate dal Focus Group sono state indagate in modo approfondito da questo gruppo di lavoro che ha rese note le sue raccomandazioni: in un report del 2018 dal titolo *Descriptive Metadata for Web Archiving*.⁸⁵

COLLEZIONI AUDIO E VIDEO

I membri del Focus Group hanno notato che le istituzioni possiedono repository contenenti una grandissima quantità di materiale audio e video, che spesso rappresenta una parte fondamentale della collezione locale.⁸⁶ In ogni caso, come Chela Scott Weber afferma nella pubblicazione *Research and Learning Agenda for Archives, Special, and Distinctive Collections in Research Libraries* "Per decenni i materiali A/V sono stati separati dal materiale manoscritto correlato (spesso messi da parte per essere trattati in un secondo momento) o trattati come singoli oggetti. Ciò ha contribuito a creare una quantità di materiale A/V non quantificato e non descritto".⁸⁷ Gran parte di questo materiale richiede urgentemente di essere preservato, trattato, digitalizzato, descritto e occorre anche dichiararne le condizioni d'uso.

Non è semplice trovare le competenze e i finanziamenti necessari. La gestione delle risorse A/V richiede una conoscenza del contesto d'uso e delle problematiche relative ai metadati; è necessaria una riflessione su una serie di questioni relative ai requisiti per la descrizione e l'accesso. Inoltre, le biblioteche devono gestire i media che vengono prodotti localmente: strumenti per la ricerca o l'apprendimento, streaming di contenuti con licenza commerciale.

Le discussioni del Focus Group si sono focalizzate sulle risorse A/V contenute nelle collezioni archivistiche — collezioni spesso deteriorate, molto consistenti e bisognose di interventi con apparecchiature rare e costose. Per quanto riguarda i contenuti generati a livello locale, le istituzioni generalmente preferiscono che le proprie risorse siano trattate dal proprio personale.

**Metadati che descrivono lo stesso
materiale A/V possono essere diversi a
seconda della biblioteca, dell'archivio e dei
sistemi di gestione delle risorse digitali.**

La sfida è stata quella di quantificare l'investimento necessario per descrivere questa tipologia di materiali unici. Le istituzioni hanno utilizzato strutture gerarchiche per aggregare materiali simili avvalendosi dell'aiuto fornito dall'Encoded Archival Description standard,⁸⁸ che fornisce utili informazioni da applicare a singoli item appartenenti ad una specifica collezione. Accade spesso, però, che un approccio di tipo "aggregativo" comporti la perdita di informazioni dettagliate relative a dati singoli ed importanti per la ricerca, come, ad esempio, la trascrizione del titolo o della data. Si tratta di una notevole mancanza per dati che devono descrivere registrazioni degli anni passati. Metadati che descrivono lo stesso materiale A/V possono essere diversi a seconda della biblioteca, dell'archivio e dei sistemi di gestione delle risorse digitali. Si auspica che migliori strumenti di rilevamento diminuiscano la necessità di ripetere le stesse informazioni nei diversi database; in ogni caso la presentazione delle informazioni all'utenza richiederebbe l'utilizzo di punti di accesso coerenti tra sistemi. Lo stesso accadrebbe in ambiente linked data. La sfida di mantenere collegati item, aiuti per la ricerca e link stabili, resta aperta.

I metadati per materiali A/V devono includere importanti informazioni tecniche, come i dettagli sui processi di acquisizione, digitalizzazione, compressione, anno della digitalizzazione, tecnologia impiegata e compatibilità dei file. Questo tipi di dati sono fondamentali per assicurare un accesso perpetuo a file così grandi e formati di riproduzione incostanti. Alcuni membri del Focus Group hanno applicato, su alcuni materiali A/V, PREMIS (Preservation Metadata: Implementation Strategies),⁸⁹ uno standard internazionale di metadati che incentiva la conservazione degli oggetti digitali e assicura la loro usabilità nel tempo.

Chela Scott Weber, OCLC Senior Program Officer sta continuando a lavorare nell'ambito del Research Library Partnership sui bisogni e le sfide relative alla gestione dei materiali A/V; il suo lavoro è riassunto nei post sul blog *OCLC Research Hanging Together Blog* dal titolo: "Assessing Needs of AV in Special Collections" e "Scale & Risk: Discussing Challenges to Managing A/V Collections in the RLP".⁹⁰ All'interno delle attività della Research Library Partnership, una parte dei membri del Focus Group ha risposto al sondaggio di Weber del 2019 creato per valutare le esigenze relative alla gestione dei materiali A/V nelle collezioni: incorporare le collezioni A/V nei flussi di lavoro delle collezioni archivistiche e digitali sono risultate essere le questioni più interessanti, così come raffigurato nell'immagine 6.

A quale tipo di problematiche relative alla gestione delle collezioni A/V saresti interessato? (n=137)

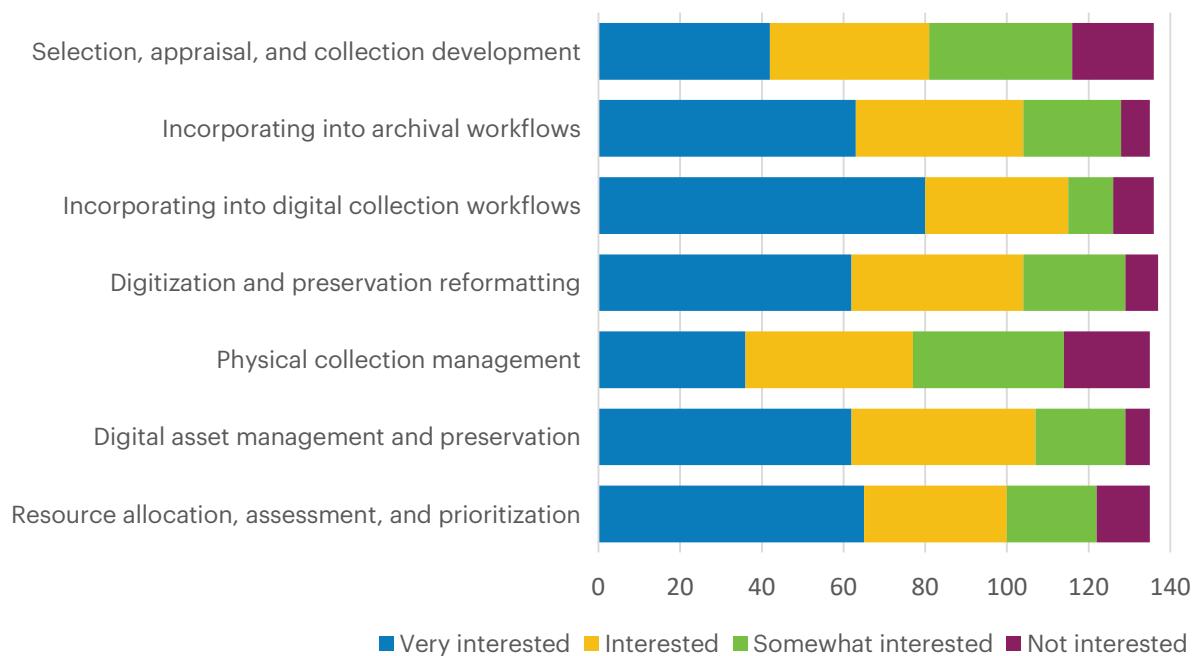


FIGURA 6. Risposte al sondaggio (2019) sulle sfide relative alla gestione delle raccolte A/V

COLLEZIONI DI IMMAGINI

I membri del Focus Group gestiscono un'ampia varietà di collezioni di immagini che presentano diverse problematiche per la gestione dei metadati. In alcuni casi, le raccolte di immagini sviluppate all'esterno delle biblioteche e i relativi modelli di dati, devono essere integrate con altre collezioni o in nuovi ambienti di ricerca. A seconda della natura della collezione, e degli utenti, sorgono questioni relative all'identificazione delle opere, alla rappresentazione delle entità, alla cronologia, alla geografia, alla provenienza, al genere o al soggetto. Le raccolte di immagini offrono anche opportunità di sviluppo di progetti collettivi e di ricerca multidisciplinare.⁹¹

Molte biblioteche descrivono le risorse di immagini digitali a livello di collezione, e i singoli item in modo selettivo. Per quanto possibile i miglioramenti vengono eseguiti in batch. A seconda della qualità dei metadati associati, alcuni professionisti svolgono il controllo di autorità. Alcune biblioteche hanno messo a disposizione delle linee guida sui metadati per contribuire a migliorarne la coerenza.

Tra le problematiche discusse dal Focus Group:

- **Varietà di sistemi e schemi:** le collezioni di immagini create nei diversi reparti delle istituzioni, come ad esempio il dipartimento di arte o quello di antropologia, hanno scopi diversi e seguono schemi e sistemi diversi da quelli usati dalle biblioteche. Spesso i metadati vengono forniti su fogli di calcolo e in formati non strutturati. Spesso i metadati creati dai diversi dipartimenti hanno bisogno di essere modificati manualmente. La situazione si semplifica quando tutte le digitalizzazioni sono gestite da un'organizzazione centrale e la biblioteca è l'unica a creare i metadati. Alcune biblioteche utilizzano i metadati Dublin Core per le collezioni di immagini, altre MODS (Metadata Object Description Schema).⁹² Alcune usano METS (Metadata Encoding and Transmission Standard),⁹³ uno schema implementato dalla Library of Congress progettato per riflettere l'organizzazione gerarchica degli oggetti digitali, i nomi e la posizione dei file che contengono tali oggetti e i metadati ad essi associati. Alcuni suggeriscono che MODS dovrebbe essere utilizzato in congiunzione con MADS (Metadata Authority Description Schema).⁹⁴
- **Duplicazione dei metadati per oggetti diversi:** i metadati relativi ad una serie di disegni scansionati potrebbero essere identici, anche in presenza di lievi differenze tra i vari disegni. La duplicazione dei metadati per oggetti simili è probabilmente dovuta alla mancanza di personale. Le facoltà di fotografia potrebbero darci maggiori informazioni a riguardo.
- **Mancanza di dati relativi all'origine:** spesso si ricevono collezioni di immagini accompagnate da dati relativi alla scansione e nessuna informazione riguardante la loro origine. Ad esempio, ad una istituzione sono stati consegnati OCR recuperati da un ricercatore tramite HathiTrust. Milioni di immagini hanno perso le informazioni relative alla loro provenienza e questo—a rischio di essere screditati—ne limita ogni ulteriore impiego.
- **Mantenere i collegamenti tra metadati ed immagini:** in che modo le biblioteche dovrebbero conservare le immagini e sincronizzarle con i metadati? Se affidassero questo compito ad una piattaforma potrebbero crearsi problemi relativi alla gestione dei diritti? Dove dovrebbero stare le anteprime?
- **Relazioni tra diverse visualizzazioni e versioni dello stesso oggetto:** disporre di versioni dello stesso oggetto in tempi diversi può rivelarsi molto utile, ad esempio per le discipline forensi. La Brown University ha deciso di descrivere lo stesso oggetto in formati differenti e di fornire una descrizione per ogni specifico formato. Questo lavoro è stato sperimentato anche se attualmente non esiste un sistema in grado di visualizzare le relazioni tra immagini, nemmeno se i relativi metadati sono memorizzati in METS.
- **Gestione dei rapporti tra docenti e curatori:** è importante venire incontro alle esigenze delle facoltà. È necessario creare una sinergia tra chi possiede i materiali, gli specialisti di metadati e gli sviluppatori poiché ciascuno contribuisce dal proprio punto di vista. La sfida è quella di soddisfare singoli bisogni, necessità comuni e assicurare una ricerca efficace.
- **Aggregazione delle collezioni digitali:** le istituzioni hanno condiviso i metadati delle loro collezioni digitali con servizi di discovery nazionali e internazionali. All'interno delle singole organizzazioni i bibliotecari hanno creato, e continuano a creare, metadati per risorse digitali o digitalizzate utilizzando tantissimi sistemi — cataloghi, sistemi per la gestione di archivi, sistemi per la conservazione delle risorse digitali, repository istituzionali, sistemi di gestione della ricerca e anche repository esterni a pagamento. Gli obiettivi per la condivisione di questi metadati vanno da servizi di discovery personalizzati basati su ricerche per soggetto, ad aggregatori nazionali ed internazionali tipo Google Scholar, HathiTrust, Digital Public Library of America (DPLA), Internet Archive, Trove, e WorldCat, a sistemi che permettono di realizzare mostre online come Google Arts and Culture, o sequenze di immagini come Flickr o Unsplash. Consultare questo tipo di aggregatori può essere d'aiuto per sviluppare la propria collezione perché i bibliotecari possono vedere i loro contributi nel contesto del contenuto di altri e identificare le lacune che potrebbero voler colmare a livello locale.⁹⁵

Gli aggregatori si servono di linee guida diverse e supportano formati diversi. Gli aggregatori affermano, in modo ragionevole, che non possono supportare molte variazioni nei metadati inviati; ma questo è in conflitto con l'altrettanto sensata affermazione dei contributori che sostengono di non poter supportare le diverse esigenze di un'ampia gamma di aggregatori. Fare correzioni o aggiornare simultaneamente la fonte e l'aggregatore può essere problematico. Alcune correzioni fatte nell'aggregatore potrebbero non essere riportate nella fonte; questo significa che lo stesso errore deve essere corretto più volte. Non sempre è chiaro quali elementi sono stati aggiornati, quando o da chi.

Aggregare e riunire immagini o versioni differenti dello stesso oggetto era l'obiettivo del OCLC Research Europeana Innovation Pilots 2012-2013,⁹⁶ che ha sviluppato un metodo per organizzare gerarchicamente oggetti di interesse culturale in base a "cluster semantici"—quelli che includono termini con un significato simile. Nel 2017, OCLC ha implementato l'International Interoperability Image Framework (IIIF)⁹⁷ Presentation Manifest protocol in CONTENTdm, un sistema per la gestione di contenuti digitali che aggredisce più di 70 milioni di record forniti da oltre 2500 biblioteche sparse nel mondo. Nel 2019 OCLC Research ha sviluppato un prototipo sperimentale: IIIF Explorer, per fare ricerca su tutte le immagini contenute in CONTENTdm che usano IIIF Presentation Manifest protocol⁹⁸ così come mostrato nella figura 7. Aggregare contenuti su sistemi conformi a IIIF potrebbe facilitare la ricerca sulle diverse piattaforme prima citate.

Nel 2020 OCLC Research ha lanciato il progetto CONTENTdm Linked Data⁹⁹, focalizzato sullo sviluppo di metodi e approcci scalabili per produrre rappresentazioni leggibili dalla macchina di entità e relazioni, e rendere visibili le connessioni precedentemente invisibili. I metadati esistenti basati sui record vengono convertiti in linked data tramite la sostituzione delle stringhe di caratteri con identificatori provenienti da file di autorità e vocabolari scelti dalle singole biblioteche; i grafi risultanti di entità e relazioni possono recuperare informazioni provenienti da fonti come GeoNames e Wikidata. Questo progetto sta affrontando molte delle problematiche descritte precedentemente dal Focus Group.

L'OCLC ResearchWorks IIIF Explorer recupera immagini di "Paris Maps" tramite le collezioni del CONTENTdm

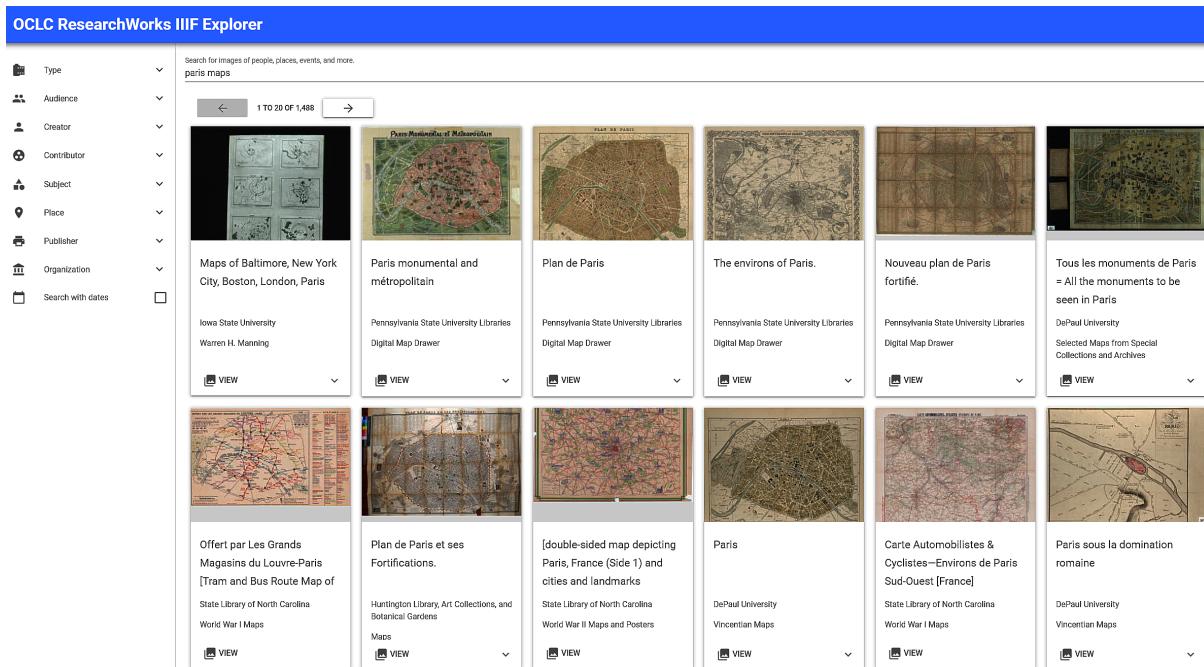


FIGURA 7. L' [OCLC ResearchWorks IIIF Explorer](#) recupera immagini relative a "Paris Maps" nelle collezioni contenute nel CONTENTdm

DATI DI RICERCA

I finanziatori dei progetti si aspettano che i dati relativi alle ricerche da loro supportate vengano archiviati e resi disponibili. Le istituzioni hanno destinato molte risorse alla raccolta e alla cura di questi dati al fine di riutilizzarne queste informazioni nella documentazione accademica. L'articolo di Ixchel Faniel, ricercatore di OCLC, intitolato "Data Management and Curation in 21st Century Archives" (settembre 2015)¹⁰⁰ ha suscitato una discussione tra i membri del Focus Group sui metadati necessari alla gestione dei dati di ricerca.¹⁰¹ Per aumentare la possibilità di condivisione e riusabilità dei dati di ricerca, le comunità dovrebbero condividere i loro progetti e mettere a punto piani per soddisfare le esigenze emergenti.

I metadati sono importanti sia per la ricerca che per il riutilizzo dei dataset. Il report del 2016 di OCLC: *Building Blocks: Laying the Foundation for a Research Data Management Program* riporta:

I dataset sono utili solamente se possono essere compresi. Occorre invitare i ricercatori a fornire informazioni strutturate sui loro dati in modo da poterne individuare provenienza e significato; ciò permetterebbe di condividerli, citarli e usarli in modo corretto. Si dovrebbe quantomeno consigliare ai ricercatori di indicare chiaramente il modo in cui hanno raccolto e utilizzato i dati e per quale scopo. Il posto migliore dove registrare questo tipo di informazioni è un file *readme.txt* che includa informazioni sul progetto e sui metadati (ad es. il nome dei file, il loro formato e il software utilizzato, il titolo, l'autore, la data, il finanziatore, il titolare del copyright, le parole chiave, tipo di dati, tipo di data e lingua).¹⁰²

La conoscenza dei bibliotecari di sistemi tecnici e di condivisione, degli standard, degli identificatori e dei linked data, può essere preziosa per la ricerca.

La serie di quattro webinar *The Realities of Research Data Management*¹⁰³ svoltisi tra il 2017 e il 2018 e presentati da Rebecca Bryant, OCLC Senior Program Officer, ribadiscono l'importanza dei metadati. Le infrastrutture che devono essere create hanno bisogno di avvalersi di alcuni aspetti chiave della professione di bibliotecario. I metadati sono fondamentali in un ambiente complesso come quello della ricerca, a partire dalla pianificazione che i ricercatori fanno prima e durante la creazione dei dati, alla gestione dei dati stessi, alla diffusione delle conoscenze acquisite; fino ad arrivare, attraverso la comprensione dell'impatto, al coinvolgimento e alla conseguente reputazione delle istituzioni di appartenenza.¹⁰⁴

La conoscenza dei bibliotecari di sistemi tecnici e di condivisione, degli standard, degli identificatori e dei linked data, può essere preziosa per la ricerca. Faniel sottolineava questo aspetto nel novembre del 2019 nel post "Let's Cook Up Some Metadata Consistency":

Se vogliamo che i dati e i metadati ad essi associati possano essere in qualche modo riutilizzati, durante la catalogazione devono essere usati termini e definizioni coerenti tra i vari repository... I bibliotecari e gli archivisti possono generare metadati coerenti, che permettano di creare collegamenti tra ricercatori e repository. Questo aumenterebbe la ricercabilità, il riutilizzo e costituirebbe un ritorno degli investimenti fatti nella ricerca dalle istituzioni.¹⁰⁵

Diversi sono i contesti nei quali si opera. Ad esempio, il nostro collega australiano può avvalersi dell'Australia's National Computational Infrastructure per i big data e dell'Australian Data Archive per le scienze sociali.¹⁰⁶ Il Canada ha lanciato Portage, un network nazionale per la "gestione condivisa dei dati di ricerca".¹⁰⁷

Alcune istituzioni hanno sviluppato dei modelli per l'acquisizione dei metadati in formati strutturati. Alcuni membri del Focus Group hanno notato la necessità di mantenere questi modelli il più semplici possibile poiché potrebbe essere difficile convincere i ricercatori a compilarli. Tutti concordano sul fatto che chi crea dati dovrebbe essere il principale fornitore dei relativi metadati; cosa potrebbe motivare la creazione di metadati di qualità? Si rendono dunque necessari nuovi metodi di formazione e sensibilizzazione; questo è oggetto di discussione del Metadata 2020's Research Communications project.¹⁰⁸

I membri del Focus Group pensano che gli elementi necessari al riutilizzo dei dati siano: licenze, fasi del lavoro, strumenti, documentazione sui dati, definizione dei dati, gestione dei dati, numeri identificativi, coordinate geospaziali e temporali (ove pertinenti). Gli schemi di metadati utilizzati includono Dublin Core, MODS (Metadata Object Description Schema) e DDI (Data Documentation Initiative's metadata standard). Nel Regno Unito, il Digital Curation Centre mette a disposizione un catalogo degli standards per i metadati.¹⁰⁹ Il Research Data Alliance's Metadata Standards Directory Working Group ha sviluppato una directory, implementata su base volontaria, di standard di metadati per diverse discipline.¹¹⁰ La disomogeneità riscontrata negli schemi di metadati rappresenta una limitazione delle possibilità di ricerca.

L'importanza degli identificatori, sia per i dati di ricerca che per i creatori dei dati, è ampliamente riconosciuta. DOI, Handle e ARK (Archival Resource Key) sono stati utilizzati per fornire accessi stabili ai dataset. Identificatori sono disponibili per interi gruppi di dati o per singole componenti, e possono essere utilizzati per monitorare i download e, potenzialmente, per misurare l'impatto del loro utilizzo. Sia ORCID che ISNI sono usati per identificare in modo univoco i creatori di dati, e il lavoro è continuato dal Research Organizational Registry fornisce che si occupa delle affermazioni istituzionali fornendo un identificatore per ogni istituzione.

Una delle maggiori criticità individuate dai membri del Focus Group è la necessità di coinvolgere maggiormente gli specialisti dei metadati nelle prime fasi del ciclo di vita della ricerca. I ricercatori devono capire quanto sia importante la creazione di metadati per le loro ricerche. La mancanza di gestione dei metadati all'interno di un'istituzione rende problematica l'integrazione dei flussi di lavoro tra repository e attuali piattaforme di ricerca.

Alcuni membri del Focus Group hanno iniziato ad analizzare i requisiti che dovrebbero avere i metadati relativi ad un intero ciclo di ricerca, e non solo al prodotto finale, ponendosi domande quali: chi sono i collaboratori?¹¹¹ In che modo i progetti si servono di diversi file di dati? Che tipo di strumenti di analisi utilizzano? Qual è la relazione tra i vari tipi di file di dati all'interno di un singolo progetto, tra progetti correlati e altri prodotti accademici come, ad esempio, articoli simili? Si stanno sviluppando servizi di supporto alla ricerca come quelli offerti dall'University of Michigan;¹¹² questi servizi, nati in collaborazione con altre unità del campus, sono stati pensati per assistere i ricercatori durante tutte le fasi della ricerca.

Una delle maggiori criticità individuate dai membri del Focus Group è la necessità di coinvolgere maggiormente gli specialisti dei metadati nelle prime fasi del ciclo di vita della ricerca. I ricercatori hanno bisogno di capire quanto sia importante la creazione di metadati per le loro ricerche. La mancanza di gestione dei metadati all'interno di un'istituzione, rende problematica l'integrazione dei flussi di lavoro tra repository e attuali piattaforme di ricerca.

Alcune biblioteche hanno iniziato a fornire un supporto alla gestione dei dati di ricerca in vari modi. Ad esempio, alcuni specialisti di metadati collaborano con i settori dedicati alle comunicazioni o all'editoria che gestiscono i repository istituzionali. Questi repository istituzionali, a volte forniscono solamente record relativi alla "citazione" o "solamente i metadati" con un link al full text o ad un dataset ospitato in un repository disciplinare. Potrebbero essere organizzati dei "servizi di consultazione dei metadati" per fornire consulenza sulla gestione dei dati che include standard appropriati, vocabolari controllati, strategie efficaci per l'organizzazione dei dati e approcci che possano consentire il loro riutilizzo una volta conclusa la ricerca. La serie di report *The Realities of Research Data Management* di OCLC Research, considera il supporto ai metadati uno dei servizi da garantire e, nei casi di studio,

vengono riportate diverse esperienze.¹¹³ All'University of Illinois a Urbana-Champaign, i consulenti per i metadati offrono il loro aiuto ai ricercatori indipendentemente dal luogo dove i dati sono conservati; la Monash University offre consulenze solamente per i metadati depositati localmente.¹¹⁴

La comunicazione è fondamentale per far comprendere ai ricercatori l'importanza dell'organizzazione dei metadati durante un progetto di ricerca. Alcune università offrono degli "sprint per la ricerca": momenti i cui i ricercatori collaborano con bibliotecari specializzati per organizzare la creazione dei metadati prevedendo la loro analisi e la loro conservazione. Lo "Shared BigData Gateway for Research Libraries", ospitato dall'Indiana University e parzialmente finanziato dall'Institute of Museum and Library Services, sta sviluppando una piattaforma basata sul cloud per condividere dati e competenze, inclusi dataset come i record dell'US Patent and Trademark Office e di Microsoft Academic Graph.¹¹⁵

La gestione dei dati di ricerca, intesa come evoluzione della documentazione prodotta in ambito accademico, richiede specifiche professionalità e competenze che riguardano la progettazione dei dati, la conoscenza approfondita del dominio e lo sviluppo delle ontologie. Le biblioteche si stanno impegnando per poter partecipare ai processi di ricerca condotti dalle proprie facoltà offrendo servizi che aiutino i ricercatori a rendere i dati delle loro ricerche accessibili e sempre consultabili. La realizzazione di buoni metadati da una parte aiuterà altri ricercatori a capire quali possono essere i dati di cui aver bisogno, dall'altra, darà ai ricercatori che li hanno creati, la soddisfazione di sapere che altri hanno beneficiato del loro lavoro.¹¹⁶

L'evoluzione: "Metadati come servizio"

I metadati sono alla base della scoperta di tutte le risorse che fanno parte delle "inside-out" e "facilitated" collection. I membri del Focus Group, al di là del catalogo tradizionale, prevedono un maggior coinvolgimento dei bibliotecari nella creazione dei metadati e nello sviluppo di nuovi servizi in grado di sfruttarne al meglio le potenzialità.

METRICA

Gli obiettivi strategici di una biblioteca spesso includono frasi come "favorire la scoperta e l'uso", "migliorare l'esperienza degli utenti" e "esplorare nuovi modi per continuare a supportare borse di studio", tutte affermazioni che si basano sulla disponibilità di metadati di qualità. Le statistiche d'uso—con che frequenza un articolo è andato in prestito, è stato citato, scaricato o richiesto—potrebbero essere sfruttate per progettare un insieme di servizi e attività. I membri del Focus Group hanno individuato alcuni possibili servizi da realizzare: orientare chi gestisce le collezioni nelle decisioni nell'ambito dei progetti di scarto e individuare i materiali da destinare a storage esterni; valutare gli abbonamenti; confrontare ciò che viene citato dai ricercatori con ciò che non è stato acquistato; migliorare gli algoritmi di rilevanza, personalizzare i risultati di ricerca, offrire servizi di raccomandazione a livello di discovery, e misurare l'impatto che ha l'uso della biblioteca sulla ricerca, sui risultati degli studenti e sull'analisi dell'apprendimento.¹¹⁷ La University of Minnesota ha condotto uno studio per indagare le relazioni esistenti tra uso della biblioteca accademica da parte degli studenti del primo anno, risultati accademici e fidelizzazione.¹¹⁸ I risultati dimostrano che esiste una correlazione tra l'uso della biblioteca e delle sue risorse—sono stati analizzati gli accessi ai database, i prestiti effettuati, gli accessi alle riviste elettroniche, e gli accessi alle aree di lavoro—e voti superiori alla media. Nel Regno Unito, il Jisc Library Impact Data Project ha ottenuto risultati simili conducendo le stesse analisi.¹¹⁹

CONSULENZE

Il valore dei metadati può essere dimostrato tramite la loro integrazione nelle biblioteche e nelle altre strutture del campus. Ad esempio, gli specialisti di metadati possono offrire un "servizio di consulenza per i metadati"—una consulenza nelle fasi preliminari dei progetti di ricerca o degli studi condotti dalle biblioteche.¹²⁰ Da parte dei dipartimenti di informatica umanistica c'è stata una grande richiesta di consulenze relative all'uso degli standard e dei vocabolari controllati. La maggior richiesta di specialisti di metadati si evince anche dalle proposte di lavoro. Un annuncio per un posto di lavoro per bibliotecario a Cornell¹²¹ prevede un 20% del tempo da dedicare alla "diffusione e all'implementazione dei metadati"; nell'annuncio si legge anche: "Intrattiene solide relazioni di lavoro con tutto lo staff e promuove la collaborazione tra il Servizio metadati e la comunità di Cornell". La Georgia Tech sta cercando un bibliotecario esperto di metadati che "svolgerà funzioni di consulenza per progetti e iniziative di maggiore portata curati dalla biblioteca. Lavorerà a stretto contatto con le biblioteche degli altri dipartimenti, con le biblioteche della Emory University, GALILEO, con lo University System of Georgia Libraries".¹²²

NUOVE APPLICAZIONI

L'uso coerente e condiviso dei campi MARC supporta nuove applicazioni. Le biblioteche attualmente utilizzano gli identificativi nei record bibliografici per recuperare indici, abstract, immagini delle copertine e per generare mappe per localizzare risorse appartenenti ad uno specifico numero di classe (ad esempio tramite l'integrazione di StackMap in OCLC).¹²³ I metadati bibliografici vengono utilizzati per popolare i sistemi di gestione delle risorse digitali e i repository istituzionali e, con l'aiuto di strumenti come Tableau o OpenRefine, si possono ottenere approfondite analisi delle collezioni nonché le loro visualizzazioni. I metadati MARC stanno collegando gli studiosi con i dati bibliografici dei loro progetti e possono creare relazioni a risorse correlate tramite applicazioni come Yewno,¹²⁴ vengono inoltre utilizzati per calcolare quanto è stato prodotto, per monitorare l'efficacia delle collaborazioni e per ricostruire la storia di un'organizzazione. La provenienza implicita nei metadati bibliografici di un'istituzione è stata anche utile nella documentazione di casi di furto. L'analisi dei dati del catalogo tramite data mining può essere finalizzata all'arricchimento dei metadati, come ad esempio generare codici linguistici mancanti o identificare i titoli originali delle opere tradotte. I dati MARC sono stati anche alla base della costruzione di mappe di soggetti che hanno permesso di scoprire relazione non esplicite nei metadati di catalogazione.¹²⁵

La provenienza implicita nei metadati bibliografici di un'istituzione è stata anche utile nella documentazione di casi di furto.

Le visualizzazioni rappresentano un altro tipo di impiego dei metadati. Un esempio eclatante è quello della Auslang national codeathon svoltosi nel 2019: una collaborazione tra National Library of Australia, l'Australian Institute of Aboriginal and Torres Strait Islander Studies, Trove, Libraries Australia, e State and Territory libraries—un code-a-thon nazionale in grado di identificare le diverse lingue indigene australiane nei record delle biblioteche.¹²⁶ La figura 8 mostra i risultati: una mappa che individua le 465 lingue indigene presenti nell'Australian National Bibliographic Database: questo è un esempio di come le comunità possono contribuire al miglioramento dei metadati.

Distribuzione di 465 codici di lingue indigene nell'Australian National Bibliographic Database

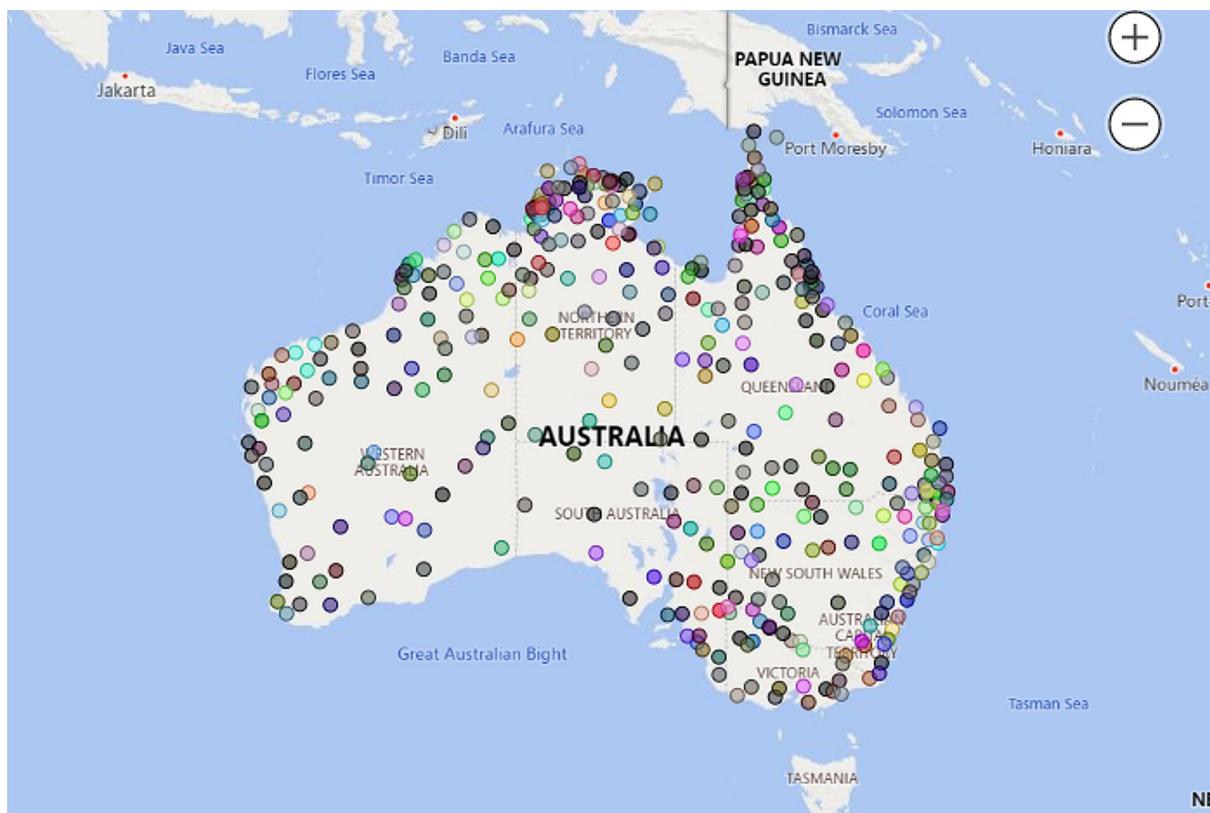


FIGURA 8. Distribuzione dei 465 codici di lingue indigene nell'[Australian National Bibliographic Database](#) dell'Austlang National Codeathon

BIBLIOMETRIA

I metadati vengono anche usati per generare metodi statistici per analizzare libri, articoli e altre pubblicazioni. Usare i metadati nei progetti di ricerca delle digital humanities può essere molto utile. Ad esempio, un ricercatore della Library of Congress ha utilizzato metadati bibliografici per studiare la storia dell'editoria e dei copyright; i ricercatori della UCLA hanno utilizzato i metadati per monitorare la commercializzazione di invenzioni come l'insulina.

Usare i metadati nei progetti di ricerca delle digital humanities può essere molto utile.

Un impiego particolare dei metadati è stato promosso da Hachette UK, il secondo rivenditore di libri più importante del Regno Unito. Hachette UK ha commissionato alla Graphic History Company l'illustrazione della storia delle sue nove case editrici. Ha quindi chiesto alla British Library di fornire i nomi degli autori e i titoli delle opere di queste nove case editrici. Nomi ed

opere coprono un arco temporale di oltre 250 anni. La British Library ha fornito un elenco di oltre 55.000 autori, 5000 dei quali sono stati selezionati per creare il più bell'esempio di utilizzo di metadati: un gigantesco murale che si estende per otto piani e raffigura i nomi degli autori in ordine cronologico. (La figura 9 mostra una parte del murale; per vedere altre immagini, cercate Hachette's River of Authors).¹²⁷

"River of Authors" di Hachette UK, generato con i metadati forniti dal catalogo della British Library



FIGURA 9. UK Hachette's "[River of Authors](#)" realizzato con i metadati forniti dal catalogo della British Library

INDICIZZAZIONE SEMANTICA

Quando i vocabolari controllati e i thesauri saranno convertiti in linked open data e condivisi pubblicamente, verrà meno il loro tradizionale ruolo di aiuto nella consultazione della collezione, ma potranno trovare un nuovo utilizzo all'interno dei sistemi di organizzazione della conoscenza basati sul web (KOS).¹²⁸ Come Marcia Zeng sottolinea in Knowledge Organization Systems (KOS) in the Semantic Web: a multi-dimensional review:

un vocabolario KOS è molto più che una semplice fonte di valori da utilizzare nella descrizione dei metadati: modellando le strutture semantiche, KOS può essere utilizzato come road map semantica e costituire un orientamento per futuri indicizzatori e utenti, siano essi umani o macchine.¹²⁹

Un buon esempio è quello dei Getty Vocabularies che consentono di esplorare le loro rappresentazioni della conoscenza e aiutano gli utenti a generare proprie SPARQL utilizzabili anche in applicazioni esterne. Un altro esempio è quello della Social Networks and Archival Context (SNAC),¹³⁰ che permette di esplorare entità e relazioni indipendentemente dalla loro collezione di appartenenza. In questi casi, lo strumento di ricerca è incentrato sulla persona (o sulla famiglia e sul tema) anziché (solamente) sulla collezione.

Invece di creare un "dominio globale" gli specialisti di metadati possono potenziare il valore dei metadati costruendo ponti tra le biblioteche e il mondo esterno. Wikidata è un esempio di piattaforma che aggredisce entità provenienti da fonti differenti e le collega a dati più specifici in varie lingue. Alcune istituzioni hanno ospitato dei Wikimediani per dar spazio a questo tipo di progetti.

I membri del Focus Group sperano che l'intelligenza artificiale—o il machine learning—sia in grado di minimizzare gli sforzi manuali necessari a collegare nomi e concetti nei dati di ricerca. Forse alcuni algoritmi potranno essere impiegati per abbinare nomi in base a metadati o a fonti, correlate per mettere in relazione argomenti, per disambiguare nomi sulla base di altri metadati disponibili e per analizzare set di dati al fine di individuare eventuali difformità nelle collezioni.¹³¹ Alcuni partner della Research Library partecipano all'Artificial Intelligence for Libraries, Archives & Museums (AI4LAM),¹³² una "comunità internazionale e partecipativa che supporta lo sviluppo dell'intelligenza artificiale applicata alle biblioteche, agli archivi e ai musei".¹³³ Alcune raccomandazioni su come migliorare le descrizioni su larga scala ed incrementare la visibilità sono suggerite nel position paper *Responsible Operations: Data Science, Machine Learning, and AI in Libraries*¹³⁴ di Thomas Padilla, OCLC Research.

Future competenze richieste al personale

Le modifiche previste dalla transizione alla nuova generazione di metadati richiederanno, in futuro, anche nuove competenze. I membri del Focus Group hanno identificato queste nuove competenze come necessarie sia a chi si affaccia al mondo del lavoro, sia ai catalogatori esperti. La ragione è dovuta da una parte al mutevole panorama informatico, dall'altra all'obsolescenza delle conoscenze. I membri del gruppo hanno anche descritto un professionista come colui che "introduce innovazioni" che si trasformano in seguito in lavoro quotidiano. Queste discussioni confermano le affermazioni di Padilla sull'importanza dello sviluppo di nuove competenze, sull'opportunità di investire nelle persone talentuose e sulla necessità di formare il personale direttamente sul campo.¹³⁵

IL CAMBIAMENTO CULTURALE

I membri del Focus Group hanno segnalato una sorta di cautela, da parte delle istituzioni, nell'allocare i membri dello staff anziché nella catalogazione tradizionale (attività come catalogazione originale o derivata, lavoro su authority file) nei nuovi progetti di ricerca e sviluppo, come ad esempio i progetti sui linked data, l'esplorazione di nuovi modelli di dati o di nuove tecnologie come Wikidata, lo studio degli standard emergenti e gli identificatori. Un **cambiamento culturale** è necessario: dall'orgoglio della produzione locale si deve passare alla valutazione delle opportunità per apprendere, esplorare e provare nuovi approcci al lavoro sui metadati. Gli specialisti di metadati devono capire che l'attenta creazione di metadati è molto più importante rispetto alla produttività individuale. Questo cambiamento culturale implica l'autorizzazione e il finanziamento di programmi di formazione che permettano allo staff di apprendere nuovi flussi lavoro. Il cambiamento culturale comporterà anche una nuova visione di questa professionalità: lo specialista non sarà più solamente una "macchina per la produzione".

Gli specialisti di metadati, di fronte alle riduzioni del personale, sono comunque tenuti a mantenere gli stessi livelli di produzione e devono quindi giustificare il tempo impiegato in attività di studio o ricerca finalizzate a rispondere a domande quali: "Cosa possiamo smettere di fare?", "Qual è la cosa più importante che hai imparato e che dovremmo condividere?", "Di cosa hai bisogno per migliorare il tuo lavoro?", "Quale tipo di software ci permetterebbe di lavorare in modo più efficiente?", "Quali nuovi metodi potrebbero migliorare la reperibilità, l'accesso e l'uso delle facilitated collection?". È arrivato il momento di stabilire obiettivi che non siano basati esclusivamente sui numeri.¹³⁶

Un cambiamento culturale è necessario: dall'orgoglio della produzione locale, si deve passare alla valutazione delle opportunità per apprendere, esplorare e provare nuovi approcci al lavoro sui metadati.

Questo cambiamento prevede l'esternalizzazione di una parte del lavoro, la formazione di personale "meno specialistico" in grado di occuparsi di lavori più semplici e la specializzazione di alcuni catalogatori per svolgere lavori particolari. Gli specialisti di metadati dovranno definire quali materiali possono essere trattati dallo staff o dagli studenti, mettendo a loro disposizione template sui quali lavorare. Se si escludono queste attività, ciò che resta richiede l'impiego di specialisti con competenze nel campo delle lingue, nel trattamento dei diversi formati fisici, nella disambiguazione e descrizione delle persone, delle organizzazioni e altre entità.

OPPORTUNITÀ DI FORMAZIONE

Per promuovere questo tipo di cambiamento si deve incoraggiare una riflessione rispetto ai tradizionali metodi di lavoro e si deve stimolare l'interesse nei confronti delle opportunità di formazione. I membri del Focus Group hanno utilizzato diversi approcci:

- Identificazione, all'interno dello staff, della persona che ha maggiori capacità di acquisire nuove competenze. In una istituzione, il membro selezionato per la formazione ha condiviso ciò che aveva imparato migliorando, così, il rendimento dell'intera unità. Questo approccio ha generato una sensibilità verso la "formazione permanente" e tutto lo staff ha presentato le proprie attività durante alcune conferenze nazionali.
- Convocazione di discussioni di gruppo in cui si parla dei problemi legati ai metadati e si trovano soluzioni; questo approccio si dimostra utile ad un continuo miglioramento delle attività. Il personale meno interessato ad acquisire nuove competenze potrà comunque beneficiare della maggior preparazione di chi segue la formazione producendo gli stessi risultati con minor fatica.
- Promozione di momenti di lettura in cui l'intero staff legge un articolo e risponde a tre domande; le risposte stimoleranno gli specialisti dei metadati e solleveranno problemi diversi da quelli affrontati quotidianamente.

- Organizzazione di un "brown bag lunch" settimanale durante il quale visionare video sui nuovi sviluppi dei metadati. In questo modo tutto lo staff potrà "guardare ed imparare" insieme.
- Partecipazione a progetti multistituzionali al fine di stimolare la collaborazione tra pari nella soluzione di problemi e favorire la contaminazione di idee.
- Partecipazione alle conferenze professionali e allo sviluppo degli standard.

La formazione dei catalogatori è stata uno dei nodi principali di molte discussioni all'interno della comunità che si occupa di metadati. Sia i professionisti che le persone che si affacciano al mondo del lavoro, hanno bisogno di nuove competenze per affrontare con successo la transizione al nuovo ambiente di lavoro dei linked data. I catalogatori stanno studiando e sperimentando BIBFRAME pur restando responsabili del tradizionale controllo bibliografico delle loro collezioni. Gli specialisti di metadati utilizzano strumenti per mappare i metadati, per riconciliarli e per migliorarli; identificano e mappano relazioni semantiche tra diverse tassonomie al fine di rendere i thesauri più comprensibili agli utenti finali. La gestione degli aspetti più tecnici dei metadati ha suscitato l'interesse da parte di altri settori. Questo interesse potrebbe aumentare man mano che i metadati diventeranno centrali per la promozione delle attività governative, delle organizzazioni non profit e delle imprese private.¹³⁷

NUOVI STRUMENTI E COMPETENZE

La collaborazione tra specialisti di metadati, personale IT e sistemisti varia a seconda delle istituzioni. Questo tipo di collaborazione è fondamentale per diverse ragioni, tra cui la gestione dei dati che sono fuori dal raggio d'azione delle biblioteche. È stato notato che esistono "differenze culturali" tra le vari professioni: gli sviluppatori tendono ad essere più dinamici e si focalizzano sulla prototipizzazione e sull'iterazione, mentre i bibliotecari si concentrano più sulla documentazione di cui hanno bisogno e sono più "schematici". Quale soluzione sarà la migliore? Insegnare le competenze informatiche agli specialisti dei metadati o insegnare i principi dei metadati al personale IT? La soluzione è reclutare personale con una preparazione tecnica e competente anche in materia di metadati. Fidelizzare personale con competenze IT è difficile—spesso questo tipo di personale è richiesto per lavori ben più pagati nel settore privato. Le esperienze dei membri del Focus Group hanno dimostrato che è più facile per i bibliotecari acquisire competenze sulla programmazione, piuttosto che assumere specialisti IT in grado di assimilare la "mentalità dei servizi tecnici". Idealmente, i membri del Focus Group auspicherebbero che parte dello staff fosse in grado di eseguire azioni su gruppi di dati, o almeno conoscesse gli strumenti disponibili per automatizzare il numero più ampio possibile di attività.

Per tanti anni i membri del Focus Group hanno usato MarcEdit e/o altri strumenti come OpenRefine, script (Python, Ruby o Perl) e macro per la riconciliazione dei metadati e per la loro l'elaborazione in batch.¹³⁸ MarcEdit è lo strumento più noto ed ha una comunità di utenti ampia, globale ed attiva così come indicato nelle sue statistiche di utilizzo nel 2017.¹³⁹ Terry Reese, sviluppatore di MarcEdit, stima che circa un terzo degli utenti lavora in ambiente non-MARC e due terzi degli utenti più attivi sono membri di OCLC. I membri del Focus Group hanno dichiarato di utilizzare MarcEdit per la trasformazione dei dati, per migliorare i record dei fornitori, per creare record MARC da fogli di calcolo, per riconciliare metadati, per de-duplicare record all'interno di un file, per fondere due o più record, per la ricerca Z39.50 e per riconciliare i metadati prima di inviarli ad altri sistemi. La versione MarcEdit 7 lanciata nel 2017 include nuove funzionalità come quella di clustering leggero: un nuovo e potente modo per trovare relazioni tra i dati senza aver bisogno di una specifica formazione. Inoltre questa versione supporta i linked data¹⁴⁰. Reese ha creato una serie di tutorial disponibili sul suo canale YouTube nella playlist MarcEdit.¹⁴¹

I manager vogliono concentrarsi meno su schemi specifici e più, invece, sulle regole dei metadati che possono essere applicati a una gamma di formati e ambienti diversi. Le competenze trasversali includono la capacità di soluzione dei problemi, di collaborazione efficace, la disponibilità—perfino l'entusiasmo—alla sperimentazione e alla comprensione dei bisogni dei ricercatori. Sebbene alcuni specialisti di metadati apprezzino la sperimentazione di nuovi approcci, spesso non hanno il tempo per studiare nuovi strumenti o nuove metodologie poiché costretti a portare avanti il lavoro quotidiano. Le biblioteche dovrebbero promuovere, nelle scuole di biblioteconomia, o tramite l'ALA's New Members Roundtable, i metadati come un'entusiasmante possibilità di carriera per i nuovi professionisti. Sottolineare che i metadati comprendono molto più della catalogazione—identificazione delle entità; standard descrittivi utilizzati nelle varie discipline accademiche; descrizione di dati digitali, archivistici e dati di ricerca che possono interagire con il web semantico—potrebbe aumentare il loro successo. Come ha osservato un membro del Focus Group: "Portiamo ordine nel vuoto".¹⁴²

AUTOFORMAZIONE

Sempre più spesso i metadati vengono creati al di fuori delle biblioteche da accademici o studenti che hanno una formazione minima, il che richiede catalogatori in grado di verificare i record creati. I membri del Focus Group pensano che la capacità di creare di semplici script, saper correggere i dati, essere in grado di gestire le identità per riconciliare dati equivalenti provenienti da archivi diversi, siano alcune delle competenze tecniche necessarie. Le fonti più spesso consultate sono Library Juice Academy, MarcEdit tutorials, LinkedIn Learning (che ha acquisito Lynda.com), Library of Congress Training Webinars, ALCTS Webinars, Code Academy, Software Carpentry, e conferenze come Code4Lib e Mashcat.¹⁴³ Letture raccomandate sono: W3C's Data on the Web Best Practices e Semantic Web for the Working Ontologist.¹⁴⁴ Fondamentale per questo tipo di formazione è la possibilità di applicare rapidamente ciò che si è imparato. Se le competenze acquisite non vengono messe in pratica velocemente si rischia di dimenticare ciò che è stato studiato. Lo staff spesso soffre di aver dedicato del tempo per imparare qualcosa che non può utilizzare nel lavoro quotidiano.

Ultimamente i membri del Focus Group hanno notato che l'autoformazione viene condotta su più fonti anziché limitarsi alle istruzioni fornite dalla Library of Congress. Alcuni approcci citati dai partecipanti:

- Sottolineare la continuità delle regole dei metadati quando si estende l'ambito del lavoro.
- Approfittare del Library Workflow Exchange,¹⁴⁵ un sito progettato per promuovere la condivisione dei flussi di lavoro delle biblioteche e le buone pratiche delle istituzioni.
- Dalla Electronic Resources and Libraries Conference tenutasi nel 2017: " Non aspettate, reiterate" In altre parole: piuttosto che attendere che lo staff acquisisca tutte le competenze necessarie, lasciare la possibilità di svolgere attività iterative e imparare dall'esperienza, in questo modo le persone saranno preparate a svolgere nuovi compiti quando sarà il momento.
- Organizzare piccoli gruppi di specialisti di metadati che seguano corsi di programmazione e, in seguito, continuino ad incontrarsi e a discutere i modi più appropriati di applicare ciò che hanno imparato alle attività quotidiane.
- Incoraggiare la partecipazione ad eventi quali OCLC's DevConnect Webinars,¹⁴⁶ per imparare dalle biblioteche che utilizzano le API come migliorare i servizi della propria biblioteca.
- Creare gruppi di studio o di lettura tra campus diversi o tra uffici diversi.

- Ampliare gli ambiti di lavoro per permettere agli specialisti dei metadati di applicare le loro competenze a nuovi domini o terminologie, come l'utilizzo di Dublin Core per le collezioni digitali. Coinvolgere il personale nei progetti digitali, dalle fasi iniziali fino a quelle di sviluppo, farlo partecipe delle procedure di controllo della qualità e della selezione degli strumenti. Tutto ciò favorirà la nascita di relazioni collaborative.
- Assumere neolaureati in informatica per impieghi a breve termine finalizzati alla creazione, ad esempio, di script.

PENSARE AL RICAMBIO DEL PERSONALE

Il ricambio di una figura professionale all'interno di una unità che si occupa di catalogazione o di metadati, comporta il rischio significativo che potrebbe essere impossibile convincere la dirigenza a mantenere quella posizione e riconfermarla. Questo rischio aumenta particolarmente quando si tratta di operatori che hanno eseguito lavori "tradizionali" come la catalogazione originale in MARC. Le probabilità di mantenere aperta la posizione sono maggiori se si pensa attentamente a come questa potrebbe essere riconfigurata, o riproposta, per soddisfare le esigenze emergenti.¹⁴⁷

La maggior parte dei membri del Focus Group ha dovuto affrontare dei ricambi di personale dovuti a pensionamenti o a cambiamenti di lavoro. Circa la metà ha dovuto rivedere la posizione dei bibliotecari in uscita. Osservare quali sono i profili ricercati da altre istituzioni può essere d'aiuto per creare avvisi di lavoro. Molte posizioni aperte per la catalogazione non richiedono una laurea in MLS, quindi il reclutamento di professionisti si è concentrato sull'adattabilità delle nuove posizioni con le priorità universitarie, sulla motivazione all'apprendimento e sulla capacità di essere propositivi nello sviluppo dei metadati per la promozione della ricerca, per implementare gli open access, le collezioni digitali e i linked data. La progettazione di strategie future e la descrizione di come i metadati possano interagire tra sistemi diversi sono stati alcuni delle questioni discusse durante i recenti colloqui di lavoro. Ai neoassunti vengono richieste competenze relative alla programmazione, per poter applicare tecniche batch ai metadati che compensino la perdita di personale. L'impiego della tecnologia nei servizi bibliotecari aiuta i catalogatori a "fare di più con meno".

I membri del Focus Group auspicano che il nuovo personale sia messo in relazione con la comunità di catalogazione e sia consapevole delle sovrapposizioni con altre organizzazioni culturali come archivi e biblioteche. L'ambiente bibliotecario è in continua evoluzione e i bibliotecari hanno dovuto riflettere sulle proprie priorità per affrontare questi cambiamenti. I gestori di metadati devono ripensare i ruoli degli specialisti al di là del tradizionale lavoro di catalogazione. I candidati con competenze flessibili hanno più probabilità di successo di quelli con un background di catalogazione tradizionale; che potrebbero non adattarsi bene a lavorare in nuovi ambienti. Probabilmente molti ruoli necessitano di essere riconsiderati. Le uniche attività che resteranno sempre uguali saranno la gestione, la scoperta di nuove tendenze, la pianificazione, la partecipazione alla progettazione di nuovi standard internazionali, la creazione e l'implementazione di cambiamenti e le riflessioni sui temi generali.

IMPATTO

I metadati di nuova generazione si focalizzeranno sempre più sulle entità piuttosto che sulla descrizione contenuta nei record. Le attività svolte dai membri del Focus Group relativamente ai linked data, inclusa la partecipazione a progetti come OCLC Research's Project Passage e CONTENTdm Linked Data pilots, hanno permesso a OCLC di ottenere ad hoc il finanziamento di Andrew W. Mellon per la realizzazione del progetto Shared Entity Management Infrastructure¹⁴⁷ della durata di due anni e iniziato nel gennaio 2020. I membri dello Shared Entity Management Infrastructure Advisory Group fanno parte anche del Focus Group. Questo progetto si basa sui lavori condotti da OCLC sui linked data e fornirà *un'infrastruttura* in grado di supportare identificatori persistenti e autorevoli per persone ed opere; sarà basata sulle API e sarà personalizzabile rispetto ai flussi di lavoro delle singole biblioteche. Questa infrastruttura è stata desiderata a lungo dal Focus Group poiché sarà in grado di soddisfare molte delle questioni prese in considerazione precedentemente, in particolare gli identificatori per le opere.

I metadati di nuova generazione si
focalizzeranno sempre più sulle entità
piuttosto che sulla descrizione contenuta
nei record record di un'istituzione.

Gli identificatori disponibili tramite la Shared Entity Management Infrastructure saranno in grado di fornire collegamenti linguisticamente neutri a fonti affidabili. I metadati che biblioteche, archivi e altre istituzioni culturali creano costituiranno la base di queste entità, come "dichiarazioni" associate a quei collegamenti. L'impatto sarà globale: riguarderà il modo in cui bibliotecari e archivisti descriveranno le "inside-out" e "facilitated collections", comporterà l'offerta di "metadati come servizio" e determinerà le competenze future richieste al personale.

RINGRAZIAMENTI

OCLC Research desidera ringraziare tutti i membri del Research Library Partners Metadata Managers Focus Group che hanno condiviso le loro esperienze e riassunto in questo documento le proprie riflessioni. Desideriamo inoltre ringraziare il Metadata Managers Planning Group che, attraverso il proprio lavoro, ha posto una serie di questioni che hanno stimolato le nostre discussioni. Infine desideriamo ringraziare i seguenti membri del Focus Group che hanno esaminato la prima stesura di questo documento; i loro commenti sono stati preziosi per la stesura finale.

- Charlene Chou, New York University
- Suzanne Pilsk, Smithsonian Institution
- Greg Reeve, Brigham Young University
- Alexander Whelan, Columbia University
- Helen K. R. Williams, London School of Economics

Desidero estendere i miei ringraziamenti ai colleghi: Rebecca Bryant, Jody DeRidder, Annette Dortmund, Rachel Frick, Janifer Gatenby, Jean Godby, Shane Huddleston, Andrew Pace, Merrilee Proffitt, Nathan Putnam, Stephan Schindehette, in particolare ringrazio Chela Weber per l'attento lavoro di revisione della prima stesura di questo documento. Ringrazio Erica Melko per l'editing, Jeanette McNicol per la parte grafica e JD Shipengrover per la copertina di questo lavoro.

Una considerazione personale: ho tratto gran beneficio dall'interazione con OCLC Research Partners Metadata Managers Focus Group e sono orgogliosa di aver dato un piccolo contributo ai metadati di nuova generazione.

APPENDICE

OCLC Research Library Partners Metadata Managers Planning Group

2015-2020

I membri del Planning Group hanno selezionato gli argomenti per le discussioni su OCLC Research Library Partners Metadata Managers, hanno descritto l'importanza di un determinato argomento e hanno sviluppato una serie di domande alle quali il Focus Group ha risposto. I membri del Planning Group hanno inoltre esaminato le bozze preliminari che sono state pubblicate sul blog *Hanging Together*.

Gli attuali membri del Planning Group sono riportati in grassetto; le affiliazioni istituzionali sono riportate per il periodo in cui hanno fatto parte del gruppo:

- **Jennifer Baxmeyer, Princeton University**
- Sharon Farnel, University of Alberta
- Steven Folsom, Harvard University and Cornell University
- **Erin Grant, University of Washington**
- Dawn Hale, Johns Hopkins University
- Myung-Ja Han, University of Illinois, Urbana-Champaign
- Kate Harcourt, Columbia University
- Corey Harper, New York University
- **Stephen Hearn, University of Minnesota**
- **Daniel Lovins, Yale University**
- **Roxanne Missingham, Australian National University**
- Chew Chiat Naun, Cornell University and Harvard University
- **Suzanne Pilsk, Smithsonian**
- **John Riemer, University of California, Los Angeles**
- Carlen Ruschoff, University of Maryland
- Philip Schreur, Stanford University
- Jackie Shieh, George Washington University
- **Melanie Wacker, Columbia University**

NOTE

- 1 OCLC Research Library Partnership Metadata Managers Focus Group.
<https://www.oclc.org/research/areas/data-science/metadata-managers.html>.
 2. OCLC Research. "The OCLC Research Library Partnership."
<https://www.oclc.org/research/partnership.html>.
 - 3 Smith-Yoshimura. 2017. "Metadata Advocacy" *Hanging Together: the OCLC Research Blog*, 17 ottobre 2017. <https://hangingtogether.org/?p=6282>.
 - 4 British Library. 2019. *Foundations for the Future: The British Library's Collection Metadata Strategy 2019-2023*. London: British Library. <https://www.bl.uk/bibliographic/pdfs/british-library-collection-metadata-strategy-2019-2023.pdf>.
 - 5 Ibid, 4.
 - 6 Statistiche al 1 giugno 2020.
 - 7 Library of Congress. "Program for Cooperative Cataloging." <https://www.loc.gov/aba/pcc/>.
 - 8 Tranne nel giugno 2020, quando tutte le discussioni si sono svolte virtualmente solo a causa della pandemia di COVID-19.
 - 9 Vedi *Hanging Together: The OCLC Research Blog*, search-category Metadata.
<https://hangingtogether.org/?cat=81>.
 - 10 Benefits from affiliating with the RLP are cited in Smith-Yoshimura. 2018. "What Metadata Managers Expect from and Value about the Research Library Partnership," *Hanging Together: The OCLC Research Blog*, 16 aprile 2018. <https://hangingtogether.org/?p=6683>.
 - 11 Sono disponibili le analisi dei tre International Linked Data Survey for Implementers 2014-2018 e il foglio di calcolo di tutte le risposte ai sondaggi. Vedi OCLC Research. 2020. "Linked Data." International Linked Data Survey. <https://www.oclc.org/research/themes/data-science/linkeddata/linked-data-survey.html>.
 - 12 Godby, Jean, Karen Smith-Yoshimura, Bruce Washburn, Kalan Davis, Karen Detling, Christine Fernsebner Eslao, Steven Folsom, Xiaoli Li, Marc McGee, Karen Miller, Honor Moody, Holly Tomren, and Craig Thomas. 2019. *Creating Library Linked Data with Wikibase: Lessons Learned from Project Passage*. Dublin, OH: OCLC Research.
<https://doi.org/10.2533/faq3-ax08>;
- OCLC Research. 2020. "CONTENTdm Linked Data pilot." <https://www.oclc.org/research/themes/data-science/linkeddata/contentdm-linked-data-pilot.html>;
- OCLC. 2020. "WorldCat®: OCLC and Linked Data." Shared Entity Management Infrastructure. <https://www.oclc.org/en/worldcat/linked-data/shared-entity-management-infrastructure.html>;
- Library of Congress. "BIBFRAME." Bibliographic Framework Initiative.
<https://www.loc.gov/bibframe/>;

Futornick, Michelle. 2019. "LD4P2 Linked Data for Production: Pathway to Implementation." LS4P2 Project Background and Goals. Lyris. Postato 14 gennaio 2019. <https://wiki.lyris.org/display/LD4P2/LD4P2+Project+Background+and+Goals;>

Share-VDE (Share Virtual Discovery Environment). "An Effective Environment for the Use of Linked Data by Libraries." Ultimo accesso 17 settembre 2019. <https://www.share-vde.org/sharevde/clusters?l=en;>

Casalini, Michele, Chiat Naun Chew, Chad Cluff, Michelle Durocher, Steven Folsom, Paul Frank, Janifer Gatenby, Jean Godby, Jason Kovari, Nancy Lorimer, Clifford Lynch, Peter Murray, Jeremy Myntti, Anna Neatrour, Cory Nimer, Suzanne Pilsk, Daniel Pitti, Isabel Quintana, Jing Wang, and Simeon Warner. 2018. *National Strategy for Shareable Local Name Authorities National Forum: White Paper*. Ithaka, New York: Cornell University Library eCommons digital repository. <https://hdl.handle.net/1813/56343>.

- 13 Library of Congress. 2019. PCC (Program for Cooperative Cataloging) Task Group on Linked Data Best Practices. 2019. PCC Task Group on Linked Data Best Practices Final Report: Submitted to PCC Policy Committee 12 settembre 2019. Washington DC: Library of Congress. <https://www.loc.gov/aba/pcc/taskgroup/linked-data-best-practices-final-report.pdf>;
- Library of Congress. 2018. "Charge for PCC Task Group on Identity Management in NACO," 5. American Bar Association, Program for Cooperative Cataloging, revisionato il 22 maggio 2018. <https://www.loc.gov/aba/pcc/taskgroup/PCC-TG-Identity-Management-in-NACO-rev2018-05-22.pdf>;
- Library of Congress. 2020 "PCC Task Group on URIs in MARC." Programs of the PCC. Charge. Ultimo accesso 19 settembre 2020. <https://www.loc.gov/aba/pcc/bibframe/TaskGroups/URI-TaskGroup.html>;
- Library of Congress. 2018. "PCC Linked Data Advisory Committee: Linked Data Advisory Committee Charge." PCC Task Groups 2018. Task Groups. Revisionato il 24 luglio 2018. [Word doc; 28KB]. <https://www.loc.gov/aba/pcc/taskgroup/task-groups.html>.
- 14 Smith-Yoshimura, Karen. 2015. "Shift to Linked Data for Production." OCLC Research Hanging Together Blog, 13 maggio 2015. <https://hangingtogether.org/?p=5195>.
- 15 OCLC Research. 2020. "LInked Data." Linked Data Overview. <https://www.oclc.org/research/areas/data-science/linkeddata/linked-data-overview.html>. [Tutte le figure CC BY 4.0]
- 16 Smith-Yoshimura, Karen. 2019. "'Future Proofing' of Cataloging." OCLC Research Hanging Together Blog, 10 novembre 2019. <https://hangingtogether.org/?p=7526>.
- 17 ORCID: Connecting Research and Researchers. "What is Orcid." Our Vision. Ultimo accesso 19 Settembre 2020. <https://orcid.org/about/what-is-orcid/mission>.
- 18 Si veda ad esempio l'elenco dei firmatari degli editori di riviste che richiedono ID ORCID per gli autori. ORCID. "ORCID Open Letter - Publishers." Ultimo accesso 19 Settembre 2020. <https://orcid.org/content/requiring-orcid-publication-workflows-open-letter>.
- 19 ISNI. "What is ISNI." Ultimo accesso 19 Settembre 2020. <https://isni.org/page/what-is-isni/>.

- 20 HathiTrust è una collaborazione senza scopo di lucro di biblioteche accademiche e di ricerca che conserva oltre 17 milioni di articoli digitalizzati. Vedi: HathiTrust Digital Library. "Welcome to HathiTrust." Ultimo accesso 19 settembre 2020.
<https://www.hathitrust.org/about>.
- 21 GeoNames. "Browse the Names." Ultimo accesso 19 settembre 2020.
<https://www.geonames.org/>.
- 22 Bryant, Rebecca, Annette Dortmund, and Constance Malpas. 2017. *Convenience and Compliance: Case Studies on Persistent Identifiers in European Research Information*. Dublin, OH: OCLC Research. <https://doi.org/10.25333/C32K7M>.
- 23 ISNI detiene attualmente 11,02 milioni di identità: 10,26 milioni di individui (di cui 2,91 milioni di ricercatori) e 933.039 organizzazioni. Statistiche recuperate da ISNI. Vedi ISNI. "Key Statistics." Ultimo accesso 5 maggio 2020. <https://isni.org/>.
- 24 Library of Congress. 2020. "NACO – Name Authority Cooperative Program." Documents and Updates. Programs for Cataloging and Acquisitions (PCC). Ultimo accesso 19 settembre 2020. <http://www.loc.gov/aba/pcc/naco/index.html>.
- 25 Smith-Yoshimura, Karen. 2015. "Getting identifiers Created for Legacy Names." *Hanging Together: The OCLC Research Blog*, 30 ottobre 2015. <https://hangingtogether.org/?p=5463>.
- 26 Smith-Yoshimura, Karen. 2013. "Irreconcilable Differences? Name Authority Control & Humanities Scholarship" *Hanging Together: The OCLC Research Blog*, 27 marzo 2013. <https://hangingtogether.org/?p=2621>.
- 27 Smith-Yoshimura, Karen. 2017. "Use Cases for Local Identifiers." *Hanging Together: The OCLC Research Blog*, 5 maggio 2017. <https://hangingtogether.org/?p=5938>.
- 28 OCLC Research. 2020. "Registering Researchers in Authority Files."
<https://www.oclc.org/research/themes/research-collections/registering-researchers.html>.
- 29 Smith-Yoshimura, Karen, Janifer Gatenby, Grace Agnew, Christopher Brown, Kate Byrne, Matt Carruthers, Peter Fletcher, Stephen Hearn, Xiaoli Li, Marina Muilwijk, Chew Chiat Naun, John Riemer, Roderick Sadler, Jing Wang, Glen Wiley, and Kayla Willey. 2016. *Addressing the Challenges with Organizational Identifiers and ISNI*. Dublin, Ohio: OCLC Research. <https://doi.org/10.25333/C3FC9Q>.
- 30 Research Organization Registry (ROR). "About." <https://ror.org/about>.
- 31 V. M. Abazov, B. Abbott, B. S. Acharya, M. Adams, T. Adams, J. P. Agnew, G. D. Alexeev et al. (2014) 2020. "Precision Measurement of the Top-Quark Mass in Lepton+jets Final States." (Archiviato il 24 febbraio 2020) ArXiv.org: 1501.07912.
<https://arxiv.org/pdf/1405.1756.pdf>.
- 32 Smith-Yoshimura, Karen. 2017. "How Much Metadata Is Practical?" *Hanging Together: The OCLC Research Blog*, 14 novembre 2017. <https://hangingtogether.org/?p=6328>.
- 33 University of Minnesota. 2020. "Experts@Minnesota." Find Profiles.
<https://experts.umn.edu/en/persons/>
- University of Illinois at Urbana-Champaign. 2020. "Illinois Experts." Find U of I Research, View Scholarly Works, and Discover New Collaborators. <https://experts.illinois.edu/>.

- 34 National Institute of Health (NIH). Il National Institute of Allergy and Infectious Diseases (NIAID) il 7 aprile 2020 incarica ORCID per la formazione, le borse di studio, l'istruzione o i premi per l'avanzamento di carriera nell'anno fiscale 2020. Vedi NIH: NIAID. 2019. "ORCID iD: Required for Some, Encouraged for All." *NIAID Funding News*. Ultimo accesso 7 agosto 2019. <https://www.niaid.nih.gov/grants-contracts/orcid-id-required-some-encouraged-all>; Vedi anche Lyrasis. 2020. "SciENcv and ORCID to Streamline NIH and NSF Grant Applications." *LyrasisNow* (blog), 8 aprile 2020. <https://lyrasisnow.org/sciencv-and-orcid-to-streamline-nih-and-nsf-grant-applications/>.
- 35 Smith-Yoshimura, Karen. 2016. "Metadata Reconciliation." *Hanging Together: The OCLC Research Blog*, 28 settembre 2016. <https://hangingtogether.org/?p=5710>.
- 36 Carruthers, Matt. (2014) 2020. *mcarruthers/LCNAF-Named-Entity-Reconciliation*. GitHub Repository. <https://github.com/mcarruthers/LCNAF-Named-Entity-Reconciliation>.
- 37 Deliot, Corine, Steven Folsom, Myung-Ja Han, Nancy Lorimer, Terry Reese, and Adam Schiff. 2019. *Formulating and Obtaining URLs: A Guide to Commonly used Vocabularies and Reference Sources*. Library of Congress PCC Task Group on URLs in MARC. https://www.loc.gov/aba/pcc/bibframe/TaskGroups/formulate_obtain_URI_guide.pdf.
- 38 Smith-Yoshimura, Karen. 2019. "New Ways of Using and Enhancing Cataloging and Authority Records." *Hanging Together: The OCLC Research Blog*, 2 aprile 2019. <https://hangingtogether.org/?p=5710>.
- 39 Smith-Yoshimura, Karen. 2015. "Persistent Identifiers for Local Collections." *Hanging Together: The OCLC Research Blog*, 27 ottobre 2015. <https://hangingtogether.org/?p=5445>.
- 40 Si vedano gli esempi DOI in dettaglio: DOI. 2020. "DOI System Examples." Ultimo accesso 20 settembre 2020. <https://www.doi.org/demos.html>; and Si vedano gli esempi ARK in dettaglio: Department, Dallas (Tex.) Police. 1963. "[Photographs of Identification Cards]." Collection. University of North Texas. *The Portal to Texas History digital repository*. <https://texashistory.unt.edu/ark:/67531/metaph346793/>.
- 41 DataCite. "Assign DOIs." <https://datacite.org/dois.html>;
- Wilkinson, Laura J. 2020. "Constructing your DOIs." Crossref: The Crossref Curriculum. Ultimo aggiornamento 8 aprile 2020. <https://www.crossref.org/education/member-setup/constructing-your-dois/>.
- 42 La "gestione delle identità" qui riflette il suo utilizzo tra gli specialisti dei metadati (vedi per esempio, Library of Congress. 2018. "Charge for PCC Task Group on Identity Management in NACO," 5. American Bar Association, Program for Cooperative Cataloging. Riveduto 22 maggio 2018. <https://www.loc.gov/aba/pcc/taskgroup/PCC-TG-Identity-Management-in-NACO-rev2018-05-22.pdf>.) But the term has other meanings depending on the audience; for example, identity access management, as described in:
- Wikiwand. "Identity Management." https://www.wikiwand.com/en/Identity_management.
- 43 Smith-Yoshimura, Karen. 2018. "The Coverage of Identity Management Work." *Hanging Together: The OCLC Research Blog*, 8 ottobre 2018. <https://hangingtogether.org/?p=6805>.

- 44 Smith-Yoshimura, Karen. 2017. "Beyond the Authorized Access Point? *Hanging Together: The OCLC Research Blog*, 10 ottobre 2017. <https://hangingtogether.org/?p=6271>.
- 45 Smith-Yoshimura, "Coverage of Identity Management." (Vedi nota 43.)
- 46 Si veda il webinar di Andrew Lih e Robert Fernandez. 2018. "Works in Progress Webinar: Introduction to Wikidata for Librarians: Structuring Wikipedia and Beyond." Produced by OCLC Research, 12 giugno 2018. MP4 video presentation, 1:1:51. <https://www.oclc.org/research/events/2018/06-12.html>.
- 47 Smith-Yoshimura, Karen. 2020. "Experimentations with Wikidata/Wikibase, *Hanging Together: The OCLC Research Blog*, 18 giugno 2020. <https://hangingtogether.org/?p=8002>.
- 48 Wikimedia. "WikiCite." Home. <https://meta.wikimedia.org/wiki/WikiCite>.
- 49 Smith-Yoshimura, Karen. 2016. "Impact of Identifiers on Authority Workflows. *Hanging Together: The OCLC Research Blog*, 22 marzo 2016. <https://hangingtogether.org/?p=5603>.
- 50 Smith-Yoshimura, Karen. 2019. "Strategies for Alternate Subject Headings and Maintaining Subject Headings. *Hanging Together: The OCLC Research Blog*, 29 ottobre 2019. <https://hangingtogether.org/?p=7591>.
- 51 OCLC 2020. "FAST (Faceted Application of Subject Terminology)." <https://www.oclc.org/en/fast.html>.
- 52 Smith-Yoshimura, Karen. 2016. "Faceted Vocabularies." *Hanging Together: The OCLC Research Blog*, 31 ottobre 2016. <https://hangingtogether.org/?p=5739>.
- 53 OCLC 2020. "FAST." (Vedi nota 51.)
- 54 OCLC 2020. "FAST (Faceted Application of Subject Terminology)." Heading #3, FAST Policy and Outreach (FPOC) Committee, : <https://www.oclc.org/en/fast.html>.
- 55 Smith-Yoshimura, Karen. 2017. "Vocabulary Control Data in Discovery Environments." *Hanging Together: The OCLC Research Blog*, 5 ottobre 2017. <https://hangingtogether.org/?p=6264>.
- 56 National Library, New Zealand Government. "Ngā Upoko Tukutuku / Māori Subject Headings" <http://mshupoko.natlib.govt.nz/mshupoko/>;
- AIATSIS Pathways: Gateway to the AIATSIS Thesauri. "Pathways." <http://www1.aiatsis.gov.au/>.
- 57 Deutsche Nationalbibliothek. 2019. "MACS - Multilingual Access to Subjects." (Archiviato 13 gennaio 2019.) https://web.archive.org/web/20190113003823/https://www.dnb.de/EN/Wir/Kooperation/MACS/mac_node.html.
- 58 Smith-Yoshimura, Karen. 2019. "Knowledge Organization Systems." *Hanging Together: The OCLC Research Blog*, 17 marzo 2019. <https://hangingtogether.org/?p=7135>.
- 59 Synaptica. "Ontology Management – Graphite." <https://www.synaptica.com/graphite/>.

- 60 Smith-Yoshimura, Karen. 2018. "Are Distributed Models for Vocabulary Maintenance Viable?" *Hanging Together: The OCLC Research Blog*, 12 aprile 2018. <https://hangingtogether.org/?p=6672>.
- 61 OCLC Research. 2020. "Equity, Diversity, and Inclusion in the OCLC Research Library Partnership Survey." Overview. Ultimo accesso 20 settembre 2020. <https://www.oclc.org/research/areas/community-catalysts/rlp-edi.html>.
- 62 Smith-Yoshimura, Karen. 2018. "Creating Metadata for Equity, Diversity, and Inclusion." *Hanging Together: The OCLC Research Blog*, 7 novembre 2018. <https://hangingtogether.org/?p=6833>.
- 63 Smith-Yoshimura. "Distributed Models." (Vedi nota 60.)
- 64 Smith-Yoshimura, Karen. 2019. "Strategies for Alternate Subject Headings and Maintaining Subject Headings." *Hanging Together: The OCLC Research Blog*, 29 ottobre 2019. <https://hangingtogether.org/?p=7591>.
- 65 Baxmeyer, Jennifer, Karen Coyle, Joanna Dyla, MJ Han, Steven Folsom, Phil Schreur, and Tim Thompson. 2017. *Linked Data Infrastructure Models: Areas of Focus for PCC Strategies*. Library of Congress PCC Linked Data Advisory Committee. <https://www.loc.gov/aba/pcc/documents/LinkedDataInfrastructureModels.pdf>.
- 66 Bone, Christine, Sharon Farnel, Sheila Laroque, and Brett Lougheed. 2017. "Works in Progress Webinar: Decolonizing Descriptions: Finding, Naming and Changing the Relationship between Indigenous People, Libraries and Archives." Prodotto dal OCLC Research, 19 ottobre 2017. MP4 video presentation, 54:35:00. <https://www.oclc.org/research/events/2017/10-19.html>.
- 67 Smith-Yoshimura, Karen. 2015. "Shift to Linked Data for Production." *Hanging Together: The OCLC Research Blog*, 13 maggio 2015. <https://hangingtogether.org/?p=5195>.
- 68 Smith-Yoshimura, Karen. 2015. "Working in Shared Files." *Hanging Together: The OCLC Research Blog*, 7 aprile 2015. <https://hangingtogether.org/?p=5091>.
- 69 Bruce Washburn and Jeff Mixter, 2018. "Works in Progress Webinar: Looking Inside the Library Knowledge Vault." Prodotto dal OCLC Research, 12 agosto 2018. MP4 video presentation, 57:45:00. <https://www.oclc.org/research/events/2015/08-12.html>.
- 70 Smith-Yoshimura, Karen. 2019. Systematic Reviews of our Metadata, *Hanging Together: The OCLC Research Blog*, 10 aprile 2019. <https://hangingtogether.org/?p=7117>.
- 71 Smith-Yoshimura, Karen. 2015. "Working in Shared File." *Hanging Together: The OCLC Research Blog*, 7 aprile 2015. <https://hangingtogether.org/?p=5091>.
- 72 Jisc Library Services. n.d. "What Is 'Plan M'?" Ultimo accesso 21 settembre 2020. <https://libraryservices.jiscinvolve.org/wp/2019/12/plan-m/>
- Smith-Yoshimura, Karen. 2020. "Knowledge Management and Metadata." *Hanging Together: The OCLC Research Blog*, 9 aprile 2020. <https://hangingtogether.org/?p=7845>;
- Per saperne di più sull'attuale fase del "Piano M" (maggio-novembre 2020), si veda Grindley, Neil. "Moving Plan M Forwards – We Need Your Help!" *Library Services (PlanM)* (blog), Jisc, 6 maggio 2020. https://libraryservices.jiscinvolve.org/wp/2020/05/planm_nextphase/.

- 73 Grindley, Neil. 2019. "Plan M: Definition, Principles and Direction." Jisc. (Word docx.) <http://libraryservices.jiscinvolve.org/wp/files/2019/12/Plan-M-Definition-and-Direction-1.docx>.
- 74 Dempsey, Lorcan. 2016. "Library Collections in the Life of the User: Two Directions." *LIBER Quarterly* 26(4): 338–359. <http://doi.org/10.18352/lq.10170>.
- 75 Smith-Yoshimura, Karen. 2019. "Presenting Metadata from Different Sources in Discovery Layers." *Hanging Together: The OCLC Research Blog*, 16 aprile 2019. <https://hangingtogether.org/?p=7880>.
- 76 Smith-Yoshimura, Karen. 2017. "Metadata for Archival Collections." *Hanging Together: The OCLC Research Blog*, 30 maggio 2017. <https://hangingtogether.org/?p=5903>.
- 77 Godby, Jean, Karen Smith-Yoshimura, Bruce Washburn, Kalan Knudson Davis, Karen Detling, Christine Fernsebner Eslao, Steven Folsom, Xiaoli Li, Marc McGee, Karen Miller, Honor Moody, Craig Thomas, and Holly Tomren. 2019. *Creating Library Linked Data with Wikibase: Lessons Learned from Project Passage*, 49-51. Dublin, OH: OCLC Research. <https://doi.org/10.25333/faq3-ax08>.
- 78 L'OCLC Research Library Partnership Archives and Special Collections Linked Data Review Group è descritto qui <https://www.oclc.org/research/partnership/working-groups/archives-special-collections-linked-data-review.html>.
- 79 Smith-Yoshimura, Karen. 2020. "Metadata Management in Times of Uncertainty." *Hanging Together: The OCLC Research Blog*, 15 giugno 2020. <https://hangingtogether.org/?p=7998>.
- 80 Smith-Yoshimura, Karen. 2016. "Metadata for Archiviato il Websites." *Hanging Together: The OCLC Research Blog*, 14 marzo 2016. <https://hangingtogether.org/?p=5591>.
- 81 Archive-It. 2008. "Human Rights." Columbia University Libraries Collection. (Archiviato maggio 2008). <https://archive-it.org/collections/1068>;
- Archive-It. 2010. "New York City Places and Spaces." Columbia University Libraries Collection. (Archiviato gennaio 2010). <https://archive-it.org/collections/1757>;
- Archive-It. 2010. "Burke Library New York City Religions." Columbia University Libraries Collection. (Archiviato maggio 2010). <https://archive-it.org/collections/1945>.
- 82 NLA. "Trove." Archiviato il Websites. Sub Collections. Ultimo accesso 20 settembre 2020. <https://trove.nla.gov.au/website>.
- 83 Archive-It. 2014. "Collaborative Architecture, Urbanism, and Sustainability Web Archive (CAUSEWAY)." Ivy Plus Libraries Confederation Collection. (Archiviato giugno 2014.) <https://archive-it.org/collections/4638>;
- Archive-It. 2013. "Contemporary Composers Web Archive (CCWA)." Ivy Plus Libraries Confederation Collection. (Archiviato ottobre 2013.) <https://archive-it.org/collections/4019>
- NYARC: New York Art Resources Consortium. "Web Archiving." <http://www.nyarc.org/content/web-archiving>.

- 84 OCLC Research. 2020. "Web Archiving Metadata Working Group." The Problem, Addressing the Problem, Outputs. <https://www.oclc.org/research/themes/research-collections/wam.html>.
- 85 Dooley, Jackie, and Kate Bowers. 2018. *Descriptive Metadata for Web Archiving: Recommendations of the OCLC Research Library Partnership Web Archiving Metadata Working Group*. Dublin, OH: OCLC Research. <https://doi.org/10.25333/C3005C>.
- 86 Smith-Yoshimura, Karen. 2018. "Metadata for Audio and Videos." *Hanging Together: The OCLC Research Blog*, 29 ottobre 2018. <https://hangingtogether.org/?p=6814>.
- 87 Weber, Chela Scott. 2017. *Research and Learning Agenda for Archives, Special, and Distinctive Collections in Research Libraries*. Dublin, OH: OCLC Research. <https://doi.org/10.25333/C3C34F>.
- 88 Library of Congress. "Standards." Encoded Archival Description (EAD) Official Site. Ultimo accesso 21 settembre, 2020. <https://www.loc.gov/ead/>.
- 89 Library of Congress. "Standards." Preservation Metadata Maintenance Activity (PREMIS). Ultimo accesso 21 settembre, 2020. <https://www.loc.gov/standards/premis/>.
- 90 Weber, Chela Scott. 2019. "Assessing Needs of AV in Special Collections." *Hanging Together: The OCLC Research Blog*, 23 luglio 2019. <https://hangingtogether.org/?p=7405>;
- Weber, Chela Scott. 2019. "Scale & Risk: Discussing Challenges to Managing A/V Collections in the RLP." *Hanging Together: The OCLC Research Blog*, 1 ottobre 2019. <https://hangingtogether.org/?p=7479>.
- 91 Smith-Yoshimura, Karen. 2015. "Managing Metadata for Image Collections." *Hanging Together: The OCLC Research Blog*, 9 aprile 2015. <https://hangingtogether.org/?p=5130>.
- 92 Library of Congress. "Standards." Metadata Object Description Schema (MODS). Ultimo accesso 21 settembre 2020. <http://www.loc.gov/standards/mods/>.
- 93 Ibid.
- 94 Library of Congress. "Standards." Metadata Authority Description Schema (MADS)." Ultimo accesso 21 settembre 2020. <http://www.loc.gov/standards/mads/>.
- 95 Smith-Yoshimura, Karen. 2016. "Sharing Digital Collections Workflows." *Hanging Together: The OCLC Research Blog*, 2 novembre 2016. <https://hangingtogether.org/?p=5744>.
- 96 OCLC Research. 2020. "Europeana Innovation Pilots." Ultimo accesso 20 settembre 2020. <http://www.oclc.org/research/themes/data-science/europeana.html?urlm=168921>.
- 97 IIIF (International Image Interoperability Framework): Enabling Richer Access to the World's Images. "Home." Ultimo accesso 20 settembre 2020. <https://iiif.io/>.
- 98 OCLC Research. 2020. "OCLC ResearchWorks IIIF Explorer." <https://www.oclc.org/research/themes/data-science/iiif/iiifexplorer.html>.
- 99 OCLC Research. 2020. "CONTENTdm Linked Data Pilot." Introduction. <https://www.oclc.org/research/themes/data-science/linkeddata/contentdm-linked-data-pilot.html>.

- 101 Smith-Yoshimura, Karen. 2015. "Data Management and Curation in 21st Century Archives – Part 1." 21 settembre 2015. <http://hangingtogether.org/?p=5375>.
- 101 Smith-Yoshimura, Karen. 2016. "Metadata for Research Data Management." *Hanging Together: The OCLC Research Blog*, 18 aprile 2016. <https://hangingtogether.org/?p=5616>.
- 102 Erway, Ricky, Laurence Horton, Amy Nurnberger, Reid Otsuji, and Amy Rushing. 2015. *Building Blocks: Laying the Foundation for a Research Data Management Program*, 8. Dublin, OH: OCLC Research. <https://doi.org/10.25333/C39P86>.
- 103 Si veda OCLC Research Data Management Planning Guide at <https://www.oclc.org/research/areas/research-collections/rdm/guide.html>.
- 104 Smith-Yoshimura, Karen. 2020. "Knowledge Management and Metadata." *Hanging Together: The OCLC Research Blog*, 9 aprile 2020. <https://hangingtogether.org/?p=7845>.
- 105 Faniel, Ixchel M. 2019. "Let's Cook Up Some Metadata Consistency." Next (blog), OCLC, 21 novembre 2019. <http://www.oclc.org/blog/main/lets-cook-up-some-metadata-consistency/>.
- 106 NCI (National Computational Infrastructure): Australia. "Home." Ultimo accesso 21 Settembre 2020. <http://nci.org.au/>;
- ADA (Australian Data Archive). "Home." Ultimo accesso 21 settembre 2020. <https://www.ada.edu.au/>.
- 107 Portage Network. "Home." Ultimo accesso 21 settembre 2020. <https://portagenetwork.ca/>.
- 108 Metadata 2020 è una "collaborazione che promuove metadati più ricchi, connessi, riutilizzabili e aperti per tutti i risultati della ricerca" (<http://www.metadata2020.org/>). The Metadata 2020 Researcher Communications project is outlined here: <http://www.metadata2020.org/projects/researcher-communications/>.
- 109 Digital Curation Centre. "Disciplinary Metadata." List of Metadata Standards. Ultimo accesso 21 settembre 2020. <http://www.dcc.ac.uk/resources/metadata-standards/list>.
- 110 RDA Metadata Directory. "Metadata Standards Directory Working Group." GitHub Repository. Ultimo accesso 21 settembre 2020. <http://rd-alliance.github.io/metadata-directory/>.
- 111 NISO sta per rendere standard il CRediT (Contributor Roles Taxonomy), che identifica 14 ruoli che descrivono il contributo specifico di ciascun contributore alla produzione accademica. CRediT è stato sviluppato da CASRAI, il Consortia Advancing Standards in Research Administration Information. Si veda CASRAI. "CRediT – Contributor Roles Taxonomy." Ultimo accesso 21 settembre 2020. <https://casrai.org/credit/>.
- 112 University of Michigan Library. 2020. "Data Services." <http://www.lib.umich.edu/research-data-services>.
- 113 OCLC Research. 2020. "The Realities of Research Data Management." Overview. <https://www.oclc.org/research/publications/2017/oclcresearch-research-data-management.html>.

- 114 Bryant, Rebecca, Brian Lavoie, and Constance Malpas. 2017. *Scoping the University RDM Service Bundle*. The Realities of Research Data Management, Part 2, pp. 16, 21. Dublin, OH: OCLC Research. <https://doi.org/10.25333/C3Z039>.
- 115 Indiana University. 2018. "IU will Lead \$2 Million Partnership to Expand Access to Research Data: IU Libraries and IU Network Science Institute Are Leading a Public-Private Partnership to Create the Shared BigData Gateway for Research Libraries" News at UI, (Science and Technology.) Indiana University, 18 Ottobre 2018. <https://news.iu.edu/stories/2018/10/iu/releases/18-shared-bigdata-gateway-for-research-networks.html>;
- Microsoft. 2020. "Microsoft Academic Graph." Established 5 giugno 2015. <https://www.microsoft.com/en-us/research/project/microsoft-academic-graph/>;
- Per maggiori dettagli, si veda la registrazione dell'agosto 2019 di "Democratizing Access to Large Datasets through Shared Infrastructure." See Wittenberg, Jamie, and Valentin Pentchev. "Works in Progress Webinar: Democratizing Access to Large Datasets through Shared Infrastructure." Prodotto dal OCLC Research, 8 agosto 2019. MP4 video presentation, 58:34:00. <https://www.oclc.org/research/events/2019/080819-democratizing-access-large-datasets-shared-infrastructure.html>.
- 116 Il Reproducibility Badging and Definitions di NISO, ora disponibile pubblicamente, può aiutare i ricercatori a estendere il beneficio della loro ricerca ad altri. Si veda "Taxonomy, Definitions, and Recognition Badging Scheme Working Group | NISO Website." n.d. Ultimo accesso 22 settembre 2020.
<https://www.niso.org/standards-committees/reproducibility-badging>.
- 117 Smith-Yoshimura, Karen. 2015. "Services Built on Usage Metrics." *Hanging Together: The OCLC Research Blog*, 30 settembre 2015. <https://hangingtogether.org/?p=5430>.
- 118 Krista M. Soria, Jan Fransen, Shane Nackerud. 2014. "Stacks, Serials, Search Engines, and Students' Success: First-Year Undergraduate Students' Library Use, Academic Achievement, and Retention." *Journal of Academic Librarianship* 40: 84-91.
<https://doi.org/10.1016/j.acalib.2013.12.002>.
- 119 See Jisc. "Library Impact Data Project (LIDP)." Ultimo accesso 21 settembre 2020. <http://www.activitydata.org/LIDP.html>.
- 120 Smith-Yoshimura, Karen. 2019. "Alternatives to Statistics for Measuring Success and Value of Cataloging." *Hanging Together: The OCLC Research Blog*, 15 aprile 2019. <https://hangingtogether.org/?p=7122>.
- 121 DLF (Digital Library Federation). 2015. "Metadata Librarian, Cornell University Library." *DLF (blog)*, 11 giugno 2015. <https://www.diglib.org/metadata-librarian-cornell-university-library/>.
- 122 Salary.com. (2019) 2020. "Metadata Librarian." Posted by Georgia Tech University 13 Novembre 2019. (Archiviato 2 settembre 2020)
<https://web.archive.org/web/20200903061830/https://www.salary.com/job/gt-library/metadata-librarian/e5644ece-c847-4cfb-994f-c4c80fa81e3d>.
- 123 OCLC. 2020. "Locate Items in the Library with StackMap." https://help.oclc.org/Discovery_and_Reference/WorldCat_Discovery/Search_results/Locate_items_in_the_library_with_StackMap.
- 124 Yewno: Transforming Information into Knowledge. 2020. "Home."
<https://www.yewno.com/>.

- 125 Smith-Yoshimura, Karen. 2019. "New Ways of Using and Enhancing Cataloging and Authority Records" *Hanging Together: The OCLC Research Blog*, 2 aprile 2019. <https://hangingtogether.org/?p=7805>.
- 126 National Library of Australia (NLA). "Austlang National Codeathon." Ultimo accesso 21 settembre 2020. <https://www.nla.gov.au/our-collections/processing-and-describing-the-collections/Austlang-national-codeathon> [Map of Australia. 2020 HERE, Bing, Microsoft Corporation];
- NLA. "Trove." Search. Uncover. Australia. Ultimo accesso 21 settembre 2020. <https://trove.nla.gov.au/>.
- 127 The Graphic History Company – Hachette UK. "River of Authors." Ultimo accesso 21 settembre 2020. <http://theghc.co/project.php?project=hachette-uk-a-river-of-authors>.
- 128 Smith-Yoshimura, Karen. 2019. "Knowledge Organization Systems." *Hanging Together: The OCLC Research Blog*, 17 aprile 2019. <https://hangingtogether.org/?p=7135>.
- 129 Zeng, Marcia Lei, and Philipp Maggior. 2019. "Knowledge Organization Systems (KOS) in the Semantic Web: A Multi-dimensional Review." *International Journal on Digital Libraries* 20: 209-230. <https://doi.org/10.1007/s00799-018-0241-2>.
- 130 SNAC (Social Networks and Archival Context). "About SNAC." What is SNAC? <https://portal.snaccooperative.org/about>.
- 131 Smith-Yoshimura, Karen. 2020. "Knowledge Management and Metadata." *Hanging Together: The OCLC Research Blog*, 9 aprile 2020. <https://hangingtogether.org/?p=7845>.
- 132 AI4LAM (Artificial Intelligence for Libraries, Archives & Museums). Updated 18 maggio 2020 <https://sites.google.com/view/ai4lam/home>.
- 133 La missione di AI4LAM è organizzare, condividere ed elevare la conoscenza e l'uso dell'intelligenza artificiale da parte di biblioteche, archivi e musei. È stata fondata nel 2018, ispirata dal successo dell'International Image Interoperability Framework (IIIF) nel coordinare la collaborazione su larga scala sulla tecnologia interoperabile per far progredire i LAM.
- Vedi AI4LAM. "About." Our Mission. <https://sites.google.com/view/ai4lam/about>.
- 134 Padilla, Thomas. 2019. *Responsible Operations: Data Science, Machine Learning, and AI in Libraries*. Dublin, OH: OCLC Research. <https://doi.org/10.25333/xk7z-9g97>.
- 135 Ibid, 17-19.
- 136 Smith-Yoshimura, Karen. 2019. "Alternatives to Statistics for Measuring Success and Value of Cataloging." *Hanging Together: The OCLC Research Blog*, 15 aprile 2019. <https://hangingtogether.org/?p=7122>.
- 137 Smith-Yoshimura, Karen. 2017. "New Skill Sets for Metadata Management." *Hanging Together: The OCLC Research blog*, 17 aprile 2017. <https://hangingtogether.org/?p=5929>.
- 138 Smith-Yoshimura, Karen. 2018. "MarcEdit and Other Tools for Batch Processing and Metadata Reconciliation." *Hanging Together: The OCLC Research Blog*, 26 marzo 2018. <https://hangingtogether.org/?p=6646>.

- 139 Reese, Terry. 2018 "MarcEdit 2017 Usage Information." *Terry's Worklog* (blog), 9 settembre 2020. <http://blog.reeset.net/archives/2572>.
- 140 Reese, Terry. 2020. "Working with Linked Data In MarcEdit." *MarcEdit Development* (blog). Ultimo accesso 21 settembre 2020. <https://marcedit.reeset.net/working-with-linked-data-in-marcedit>.
- 141 Reese, Terry. 2018. "MarcEdit Playlist." 139 YouTube videos. Last updated 26 dicembre 2018. <https://www.youtube.com/playlist?list=PLrHRsJ91nVFScJLS91SWR5awtFfpewMWg>.
- 142 Smith-Yoshimura, Karen. 2017. "New Skill Sets for Metadata Management." *Hanging Together: The OCLC Research blog*, 17 aprile 2017. <https://hangingtogether.org/?p=5929>.
- 143 "XML and RDF-Based Systems Archives." n.d. *Library Juice Academy* (blog). Ultimo accesso 22 settembre 2020. <https://libraryjuiceacademy.com/certificate/xml-and-rdf-based-systems/>;
- Reese, Terry. 2013. "Tutorials." YouTube (selected). *MarcEdit Development* (blog). 14 marzo 2013. <http://marcedit.reeset.net/tutorials>;
- "Lynda: Online Courses, Classes, Training, Tutorials." n.d. Lynda.com - from LinkedIn Learning. Ultimo accesso 22 settembre 2020. <https://www.lynda.com/>;
- "Learn to Code - for Free." n.d. Codecademy. Ultimo accesso 22 settembre 2020. <https://www.codecademy.com/>;
- Software Carpentry. "Teaching Basic Lab Skills for Research Computing." Upcoming Workshops. Ultimo accesso 22 settembre 2020. <https://software-carpentry.org/>.
- 144 "Data on the Web Best Practices." n.d. Ultimo accesso 22 settembre 2020. <https://www.w3.org/TR/dwbp/>;
- Semantic Web for the Working Ontologist. (2008) 2020. <http://workingontologist.org/>.
- 145 Library Workflow Exchange. n.d. "About." Ultimo accesso 21 settembre 2020. <http://www.libraryworkflowexchange.org/about/>.
- 146 OCLC Developer Network. 2020. "DevConnect Webinars. <https://www.oclc.org/developer/events/devconnect-workshops.en.html>.
- 147 Smith-Yoshimura, Karen. 2019. "Stewardship of Professional FTEs In Metadata Work and Turnover." *Hanging Together: The OCLC Research Blog*, 18 ottobre 2019. <https://hangingtogether.org/?p=7580>.
- 148 OCLC. 2020. "WorldCat®: OCLC and Linked Data." Shared Entity Management Infrastructure. <https://www.oclc.org/en/worldcat/linked-data/shared-entity-management-infrastructure.html>.



Per ulteriori informazioni riguardo il nostro lavoro in ambito
digitalizzazioni di collezioni di biblioteche vi invitiamo a visitare:
oclc.org/digitizing



6565 Kilgour Place
Dublin, Ohio 43017-3395

T: 1-800-848-5878

T: +1-614-764-6000

F: +1-614-764-6096

www.oclc.org/research

ISBN: 978-1-55653-242-9
DOI: 10.25333/k23b-hd86
RM-PR-216787-WWIT-A4 2110