**Controllo bibliografico e repository istituzionali: benvenuti nella giungla**

In biblioteconomia parliamo spesso di “ecosistemi”[[1]](#footnote-1).

Volendo mantenere questo parallelismo “naturalistico”, parlando di repository la prima immagine che mi era venuta in mente quando mi era stato proposto l’argomento era stata quella di una giungla: caotica, fitta e impenetrabile, per camminare nella quale è necessario spesso avere un machete.
In realtà vorrei, però, cercare di vedere i repository piuttosto come una foresta pluviale: un ambiente altrettanto complesso, denso di varietà, stratificato, pieno di reti di comunicazione interne ed esterne e di interdipendenze nascoste e visibili. Un ambiente che poggia su un terreno ricco di humus, ma sostanzialmente pulito, nel quale è possibile muoversi e camminare.

E, in effetti, la relativamente breve storia dei repository ci mostra una grande varietà in termini di materiale custodito (pre print, pubblicazioni di ricerca, materiali didattici, articoli e libri, tesi, multidisciplinari o specialistici), di modalità di popolamento e organizzazione (auto archiviazione, recuperi via batch, collezioni interne, inserimento mediato dai bibliotecari), di software utilizzati (Digital Commons, DSpace, Eprints…).

Non dobbiamo dimenticare che i repository vedono la loro nascita per spinta e all’interno delle comunità scientifiche, per volontà dei singoli autori, a partire dalla “sovversiva” proposta nel 1994 di Stevan Harnad, docente di scienze cognitive al Virginia Polytechnic Institute: condividere le proprie ricerche all’interno dell’istituzione tramite l’autoarchiviazione dei contributi in rete, per rendere più efficace la loro diffusione. Nasce da qui l’embrione di un nuovo tipo di archivio aperto: quello istituzionale, promosso e gestito da un ente, che va ad affiancarsi a quelli disciplinari sorti per aggregazione di documenti riguardanti singole aree di ricerca.
E tuttora “the environment and context in which a repository is situated will unequivocally play a part in the decisions that are made and the quality of metadata that is produced” (Moulaison Sandy and Dykas 2016, 105).

Archivi, pertanto, che, in linea generale, affondano le proprie radici in un terreno creato e alimentato dagli autori stessi tramite la pratica che abbiamo imparato a definire “self-archiving”. Non a caso gli elementi su cui poggiano le fondamenta dei repository sono l’utilizzo di set di metadati minimali (come il Dublin Core Metadata Element Set, adottato dalla maggior parte dei repository providers) e gli standard di interoperabilità, come l’OAI-PMH, che sfruttano proprio tale essenzialità.
Tale scelta minimale ricadeva, secondo alcuni, nella volontà di incoraggiare la partecipazione degli autori al riempimento dei repository, portando però a scoraggiare qualunque attività che potesse essere percepita come barriera tra gli studiosi e le proprie istituzioni; questa percezione ha investito anche gli interventi e le misure di controllo di qualità nel processo di creazione dei metadati (Barton, Currier and Hey 2003).

Di conseguenza, diversamente da altri ambienti a noi più familiari, come gli OPAC e i Discovery Tools, i repository istituzionali non sono stati “creati” dai bibliotecari per organizzare e rendere disponibile l’universo bibliografico; non sono strumenti la cui nascita ricade nella nostra sfera, seppure indubbiamente ampio sia stato e tuttora sia il contributo della comunità bibliotecaria (ad esempio, nel caso dello sviluppo del Dublin Core, pur visto alla sua nascita dalla “bibliographic control community as simplistic”[[2]](#footnote-2)).

A rimarcare la differenza tra IR e strumenti bibliografici veri è propri la consapevolezza che “[for IRs] there is no universally accepted practice or standard defining quality metadata; similarly, there is no set of rules for describing institutional repository materials.” (Stein, Applegate and Robbins 2017, 650)

A partire da questa consapevolezza la prima domanda che mi faccio e vi faccio è: come ci dobbiamo porre nei confronti di questi ambienti? Vogliamo mantenerne la naturalezza (con la relativa entropia) o vogliamo trasformare queste foreste in boschi, di dimensioni controllate, ordinati e organizzati secondo la nostra visione del mondo?

Personalmente la mia risposta è di affidarsi ad un comportamento improntato all’”aurea mediocritas” cui ci invitava il poeta latino Orazio: con moderazione e con il rispetto dovuto a un ambiente che ha le sue caratteristiche cercando di intervenire nella maniera più discreta possibile, non per snaturarlo, ma per evitare che chi vi entri si perda e non ne apprezzi tutta la bellezza .

In questo solco ritengo che si dovrebbero muovere anche le scelte sul controllo bibliografico e di autorità, il cui scopo finale dovrebbe essere fornire le informazioni necessarie nella forma migliore per rendere l’esplorazione piacevole, sicura e soddisfacente a qualunque visitatore (umano o macchina che sia).

Seppure nati in seno alle comunità accademiche e fatti propri dalle istituzioni, i repository sono infatti indubbiamente rivolti sempre più ad un’utenza la più ampia possibile, avendo come primo obiettivo - voluto dagli autori stessi - non tanto di organizzare e sistematizzare quanto prodotto da un ente (compito che ben può essere svolto da un catalogo), quanto di garantirne la maggiore visibilità possibile per il maggior tempo possibile.
Accesso e conservazione di lunga durata realizzabili grazie a due elementi strettamente collegati tra loro: metadati di qualità e repository system di qualità; infatti “quality metadata may be underutilized due to weakness in indexing, navigation, and display options” (Moulaison Sandy and Dykas 2016, 103).

Definire cosa si intenda per “metadati di qualità” non è, però, di per sé attività facile: rimane forte la componente soggettiva e locale, si rimarca spesso lo stretto legame tra requisiti funzionali e idoneità allo scopo (Powell, Day and Guy 2004) “by defining both the internal requirements related to the needs of end-users in a local setting and by defining external requirements related to disclose and exposed local metadata relating to external service providers” (Park 2009, 214).

Contemporaneamente, pertanto, bisogna garantire l’interoperabilità, all’interno di un quadro di Search Engine Optimization, che ricomprende una serie molto più ampia di attività tecniche e scelte operative che dovrebbero far parte del piano di sviluppo di un repository.

Come bibliotecari abbiamo il compito di fornire un supporto per la creazione di metadati di qualità, come ci ricorda la NISO (National Information Standards Organization) nel suo *A Framework of Guidance for Building Good Digital Collections[[3]](#footnote-3)*, quando dice che Good Metadata:

1. Conforms to community standards in a way that is appropriate to the materials in the collection, users of the collection, and current and potential future uses of the collection.
2. Supports interoperability.
3. Uses authority control and content standards to describe objects and collocate related objects.
4. Includes a clear statement of the conditions and terms of use for the digital objects.
5. Supports the long-term curation and preservation of objects in collection.
6. Good metadata records are objects themselves and therefore should have the qualities of good objects, including authority, authenticity, archivability, persistence, and unique identification.

Parlando di qualità di metadati, oltre ai requisiti, che sempre ritroviamo in letteratura, di accuratezza, completezza e consistenza (Park 2009) non possiamo, quindi, trascurare un quarto aspetto: l’efficacia in termini di recupero dell’informazione, in particolare in relazione all’indicizzazione da parte dei motori di ricerca.
Uno studio del 2019 (Mering 2019) condotto sull’IR della University of Nebraska-Lincoln ci dice che l’accesso al 57% della collezione è stato via Google Search e solo il 17% direttamente dal repository.
Già nel 2012 Arlitsch e O’Brienci ricordavano che “digital repositories [...] face a common challenge: having their content found by interested users in a crowded sea of information on the Internet” (Arlitsch and O’Brien 2012, 64) e nelle loro ricerche mostravano che tutti i repository analizzati, indipendentemente dalla piattaforma, avevano un low index ratio rendendoli di fatto invisibili a Google Scholar.

Il fine ultimo comunque rimane sempre quello di garantire la più ampia visibilità (Swan and Carr, 2008), che si ottiene anche rendendo il contenuto del repository condivisibile verso l’esterno.
“Shareability implies an adherence to internal but also extra-organizational standards and best practices; when every repository uses recommended file types, metadata schemas, and the same controlled vocabularies, information is more easily searched and retrieved across them” (Moulaison Sandy and Dykas 2016, 102).
E’ importante pertanto che “for metadata to be effective, enforcement of standards of quality must take place at the community level” (Bruce and Hillman 2004, 3) e che sia necessario “establish effective policies for the management of authorities in these types of digital collection through cooperative efforts that will permit the development of corpora of authority entries that will aid the processes of cataloguing, metadata creation and information retrieval” (Barrionuevo Almuzara, Alvite Díez and Rodríguez Bravo 2012, 101).

L’economia, però, ci insegna che l’efficacia, nell’analisi dei costi benefici, è sempre affiancata dall’efficienza, mentre la nostra esperienza quotidiana di bibliotecari ci ricorda continuamente che le attività di controllo bibliografico sono fortemente “time and resource consuming” e richiedono competenze elevate.

Ad esempio, un sondaggio (Moulaison Sandy and Dykas 2016), condotto nel 2015 tra alcuni amministratori di repository statunitensi, indicava tra i maggiori ostacoli alla fornitura di metadati di alta qualità proprio time limitations and staff hours e skills levels of staff, anche per coloro che avevano autovalutato la qualità dei metadati del proprio repository come superiore alla media.

Da qui ne deriva la necessità di fare scelte accorte su dove concentrare gli sforzi e di utilizzare tutti gli strumenti che possano aiutare nella riduzione dei costi.
E’ evidente, infatti, che - di fronte alla crescita continua delle pubblicazioni scientifiche, frutto anche di sempre più pervasive attività di valutazione - all’interno degli IR, soprattutto quando la fonte principale è l’autoarchiviazione, uno dei compiti dei bibliotecari sia di costruire dei percorsi chiari e veloci, che garantiscano contemporaneamente la facilità di utilizzo e la qualità dell’informazione, sia per gli autori che per i lettori, lavorando su quei metadati che impattano sull’accesso.

Un primo livello, minimo, di supporto poggia sull’offrire agli autori o ai redattori note esplicative, istruzioni, drop-down menus o pop-up windows in fase di compilazione della scheda bibliografica: in questo caso non possiamo certamente parlare di controllo bibliografico. É pur sempre, però, una base da cui non possiamo prescindere: fornire informazioni chiare, strumenti semplici e affiancare al supporto tecnico attività di formazione ed educazione, costituiscono comunque un tassello indispensabile per creare una comunità di utenti consapevoli nella creazione di metadati di qualità, così da ridurre ex ante la necessità di successivi interventi di controllo.
Allo stesso livello si pone anche la definizione preliminare di best practices e linee guida rivolte ai gestori del repository e allo staff: “metadata guidelines seem to be fundamental in ensuring a minimum level of consistency in resource description within a collection and across distributed digital repositories” (Park and Tosaka 2010, 711).

Un secondo livello confida sulla qualità dei dati recuperabili da fonti terze tramite identificativi.
Entra in campo un’altra parola molto cara al contesto “ambientalista”: riciclare, intesa come la possibilità di recuperare informazione strutturata e valida da fonti esterne in maniera rapida.

Al momento la maggior parte della letteratura, per il controllo bibliografico e di autorità, sembra mostrare due strumenti principali:

* Identificativi univoci di pubblicazione per il recupero di metadati verificati;
* Identificativi univoci di persona (ORCID, VIAF, ISNI, LoC Authority Name) per garantire il controllo di qualità degli autori.

Entrambe le strade presentano attualmente dei limiti e delle difficoltà, ma sulle potenzialità del loro utilizzo sembra esserci una certa convergenza.

Sempre più, anche nell’ottica di sostenibilità economica, si tende ad offrire nei repository la possibilità di recuperare (in fase di creazione, ma non solo) tramite DOI, Scopus identifier o Web of Science Accession number, ISBN, PMID molte informazioni da banche dati esterne o direttamente dall’editore.
Una scelta di questo genere, comporta da parte nostra, come bibliotecari l’accettazione di delegare a terzi l’organizzazione e la presentazione dei dati, ma non è una novità per noi: lo abbiamo fatto con la catalogazione derivata, lo facciamo oggi in parte con i discovery tool.

Se, da un lato, questo può comportare in linea di principio una certa omogeneità nella presentazione dell’informazione, riducendo il rischio di refusi ed errori derivanti dall’inserimento manuale, dall’altro riduce solo in parte l’eventuale necessità di un controllo bibliografico, prima di tutto perché gli editori e le singole banche dati hanno ognuna le proprie regole di metadatazione e catalogazione, poi perché le attività di harvesting dipendono strettamente dai protocolli di interoperabilità applicati e dalla mappatura tra i differenti set di metadati (Chapman, Reynolds and Shreeves 2009).
Ad esempio: il titolo di un articolo all’interno delle banca dati Pubmed è sempre presentato in inglese, anche se l’articolo è pubblicato in altra lingua. In tale forma viene solitamente importato del campo dc.title del Core Set Dublin Core, laddove, però, dovrebbe essere presente il titolo in lingua originale.

Inoltre, sappiamo bene quanto siano parziali le coperture delle grandi banche dati internazionali, sia da un punto di vista linguistico sia disciplinare, e ancora come non esistano identificativi specifici per tipologie di pubblicazioni quali, ad esempio, il saggio all’interno di un volume (anche se molti editori stanno cominciando a dotare anche i singoli capitoli di libro di un proprio DOI).

Anche in caso di caricamenti batch effettuati da personale dedicato la qualità dei dati non è garantita, anzi in alcuni casi sembra essere addirittura inferiore (Stein, Applegate and Robbins 2017), a meno che non si preveda un’intensa attività di pulizia dei metadati prima della batch ingestion tramite strumenti tipo Open Refine[[4]](#footnote-4).

Problema simile lo riscontriamo soprattutto nella gestione dei nomi degli autori, le cui forme varianti - sia dipendenti dalla volontà dell’autore (utilizzo discontinuo del middle name o di forme abbreviate, del cognome da sposato o da nubile/celibe…) che indipendenti (presentazioni diversi nelle varie fonti, varianti linguistiche, traslitterazioni, etc.) costituiscono da sempre una grossa sfida per il controllo di autorità.

Per cercare di ovviare a tale problema una prima forma di assistenza, non presente in tutti i software utilizzati, prevede di offrire la funzionalità di autocompletamento del nome in fase di inserimento: sebbene tale funzionalità possa contribuire alla riduzione delle forme varianti, tuttavia non rafforza il controllo di autorità.
Ad un livello lievemente superiore, tanto da poter parlare di controllo di autorità a livello locale si pone l’aggancio del nome dell’autore alle anagrafiche (institution identity management system) dell’ente, ma è evidente che tale soluzione non garantisce appieno l’interoperabilità e la shareability e “can be, at best, [only] one part of the repository authority control puzzle” (Downey 2019, 130).
È per raggiungere questo risultato che si lavora all’implementazione degli identificativi univoci dei nomi nei repository, tramite l’aggancio ad authority schemes esterni.
L’aspetto interessante, determinato dall’evoluzione verso un modello LOD, è il passaggio dal concetto di “name authority work” a quello di “identity management”, proprio grazie all’associazione di identificativi registrati: “Identity management won’t work the same way as the traditional authority control because identity management emphasizes the process of associating a registered identifier (or a URI) with a single entity and the differentiation of names or headings is only of secondary importance in identity management” (Zhu 2019, 227).
La coesistenza, però, di numerosi progetti per il controllo di autorità dei nomi con la conseguente produzione di diversi identificativi costituisce un ulteriore elemento di rumore e può portare alla necessità, nuovamente, di fare delle scelte.

Nel 2019, una collega della Duke University, Moira Downey, ha pubblicato un’analisi (Downey 2019) di tre tra le principali international authority sources - Library of Congress Name Authority Files (LCNAF), Virtual International Authority File (VIAF), and Open Researcher and Contributor Identifier (ORCID) - in vista dello sviluppo del controllo di autorità in chiave Linked Data nel proprio repository istituzionale, data la capacità di tali sistemi di fornire author URI’s tramite API.

Secondo l’autrice, LCNAF e soprattutto VIAF, che ha sviluppato un modello cooperativo per l’aggregazione di dati d’autorità da fonti nazionali e regionali con un’intensa attività di clustering, merging and deduplication, costituiscono “a broader step forward in preparing library data for better integration with the broader web” (Ibid., 120), ma poggiano ancora su meccanismi tradizionali di catalogazione partecipata e controllo di autorità che impattano sulla creazione di identificativi, in particolare per autori di articoli su rivista che costituiscono in molti repository accademici la percentuale maggiore, vista l’iperproduzione di letteratura nei settori delle scienze mediche e STEM (Science, Technology, Engineering and Mathematics). Garantiscono, però, affidabilità sulla loro persistenza, grazie alle professionalità coinvolte, con la creazione di “record in structured, machine-actionable format that did not require additional resources or inferences to ascertain” (Ibid., 131).

ORCID, invece, opera come un “self claim researcher registry”, che pare delegare il controllo di autorità, tradizionalmente svolto dalle biblioteche, direttamente ai ricercatori, dando loro anche una certa autonomia di scelta sulla propria online identity nell’universo della comunicazione accademica.
Un atto di fiducia da parte dei bibliotecari o il riconoscimento di non poter più essere i detentori assoluti dell’organizzazione e della presentazione delle informazioni?
Come spesso succede, la realtà sta nel mezzo: gli ORCID URI’s si dimostrano una buona soluzione per l’autoarchiviazione, ma “the undifferentiated nature of the current ORCID database system seems unhelpful for bulk remediation of existing repository content or for large scale batch operations” (Ibid., 131).
In particolare, non abbiamo nessuna certezza sulla persistenza di tale identificativo, dato che l’autore potrebbe decidere di rimuovere il proprio profilo in qualunque momento. Lo stesso problema lo abbiamo, comunque, anche con le anagrafiche locali, per loro natura strettamente collegate alla durata della presenza dell’autore all’interno dell’istituzione.

Personalmente la scelta di offrire all’autore uno strumento, pur perfettibile, per presentare se stesso mi pare una scelta coraggiosa di onestà intellettuale che possiamo supportare rendendo il più agevole possibile il suo utilizzo e la sua implementazione all’interno dei “nostri” repository, continuando a investire su una parallela attività di educazione e di informazione sulle best practices.

Numerose sono ormai le esperienze in tal senso nel mondo.

Ad esempio, in Italia, nel 2015 in occasione dell’esercizio nazionale di valutazione della qualità della ricerca (VQR 2011-2014), venne lanciato il progetto IRIDE (Italian Research IDentifier for Evaluation), con lo scopo di sfruttare tale occasione per dotare tutti coloro che in Italia operano nella ricerca (docenti, ricercatori universitari e degli enti di ricerca, dottorandi e post-doc) di un identificativo persistente ORCID, attivando la procedura di registrazione direttamente dal repository della propria istituzione. La maggior parte dei repository italiani universitari, che sfruttano un software proprietario, permettono da allora tramite un sistema di push and pull di dialogare in maniera bidirezionale con il proprio account ORCID.

Altrettanto interessante l’esperienza (Svantesson and Steletti 2019) dell’European University Institute di Fiesole (uno dei nostri ospiti) di integrazione dei propri database - CADMUS, EUI Central Persons Registry - con ORCID, anche come soluzione per il controllo di autorità sui nomi degli autori, sopperendo in parte alla mancanza di un CRIS (Current Research Information System). Particolarmente interessante la scelta di utilizzare come preferred name la forma presente nel repository, a ricordarci che “the criteria of selecting which of the various IDs to use will depend on the stakeholder. Among the factors to be considered is to select the ID system which attracts the “critical mass” representing one’s peers” (Smith**-**Yoshimura et al. 2014, 9).

Se sulla validità di perseguire queste strade c’è un certa convergenza rimangono ancora distanti le posizioni sull’opportunità di investire tempo e risorse per il controllo bibliografico sulla componente semantica.

In un contesto come quello di un repository istituzionale, in molti casi multidisciplinare, spesso alimentato dagli autori stessi, la profondità, l’ampiezza e la varietà delle discipline comporta che l’uso di termini controllati di soggetti sia possibile solo ad un livello alto, se vogliamo mantenere un’omogeneità all’interno del repository.

Se, invece, rispettando tale eterogeneità lasciamo le comunità autodisciplinarsi, avremo un repository in cui la consistenza dei metadati semantici sarà estremamente varia: dalla loro totale assenza al popolamento tramite thesauri riconosciuti, passando per la complessità determinata dall’utilizzo di sinonimi, omonimi e varianti grammaticali, ortografiche e linguistiche.

Ci troviamo nuovamente di fronte al dilemma entropia/controllo: quanto oltre deve spingersi il nostro intervento come repository manager? Sempre in un’ottica di rispetto Lubas, parlando di repository di theses and dissertation, sosteneva che, l’eventuale intervento di soggettazione da parte dello staff, avrebbe dovuto affiancarsi e non sostituirsi alla scelta delle parole chiave fatte dall’autore, anche se questo avrebbe comportato un inevitabile aumento del rumore (Lubas 2009). Una normalizzazione delle keywords, o la loro mappatura su un preesistente vocabolario controllato avrebbe, infatti, eliminato quella prospettiva unica con cui l’autore si riferisce alla propria opera (Radio 2014).

Un interessante studio del 2018 (White, Chen and Liu, 2018) ha cercato di analizzare la relazione tra la presenza di alcuni metadati nel repository della Law School della Duke University e il numero di download, per comprendere l’efficacia dei metadati stessi.
Tra i risultati più sorprendenti, il fatto che il numero di coautori, la presenza e il numero di keywords (fossero esse a testo libero o derivanti da vocabolari controllati) avessero una relazione sostanzialmente negativa con i downloads e non fossero essenziali agli utenti per arrivare al contenuto; mentre la presenza e lunghezza dell’abstract ne avessero una significativamente positiva.

Questo contraddice in parte la nostra certezza sull’importanza della soggettazione e sulla sua efficacia. Certezza già minata negli anni da studi che, anche all’interno dei cataloghi, invitavano ad abbandonarne l’uso (Calhoun 2006), sebbene altri studi ne confermassero l’efficacia (Gross, Taylor and Joudrey 2015).
Tale divario di visioni si è attualmente ampliato con l’adozione, da parte di molte biblioteche, dei Discovery tools, che - in nome di una presunta volontà di rendere l’esperienza di recupero dell’informazione il più possibile soddisfacente per l’utente - hanno ricreato un immaginifico “Google like world” senza avere, però, al momento la potenza del suo Page rank. Forse il termine “giungla” si attaglia meglio a tali strumenti, ancor più che agli IR.

Un grosso aiuto, per garantire un equilibrio, potrebbe arrivare dalle potenzialità dei Linked Open Data e del Web semantico, che, in particolare per la soggettazione, potrebbero dare un importante contributo all’arricchimento dei contenuti e al controllo bibliografico, tramite a simpler management of multiple languages, better linkability of resources, simpler reuse of authority registries in applications. Inoltre la “semantic search enables a new set of queries that are based on the power of inference engines and are not possible with traditional keyword based search” (Solomou and Koutsomitropoulos 2015, 66).

Pur prevedendo un inevitabile sforzo iniziale di adeguamento del repository, la scelta tra le differenti tecnologie dipende dalle complessità e dal rigore richiesti dallo specifico ambiente (Zhu 2019). Come già visto all’inizio per i metadati di qualità, anche per la scelta degli strumenti propri del web semantico ritorna il legame con l’idoneità allo scopo e con le peculiarità dell’ambiente cui gli strumenti si applicano, permettendo una possibile scalarità nelle scelte.

Anche in quest’ambito vi sono numerose esperienze interessanti: alla Central University of Gujarat, in India, nel 2017 hanno sviluppato un prototipo (Khumar 2018) in cui hanno collegato ad un repository basato su Dspace la knowledge base Dbpedia, scelta come linked dataset per l’ampia copertura disciplinare, i meccanismi automatizzati di aggiornamento e il supporto di informazioni multilingue. Altrettanto interessante, e non è la sola[[5]](#footnote-5), la ricerca condotta da studiosi greci su “a transformation engine which can be used to convert an existing institutional repository installation into a Linked Open Data repository: the data that exist in a DSpace repository can be semantically annotated to serve as a Semantic Web (meta)data repository” (Konstantinou, Spanos, Houssos and Mitrou 2013, 834).

Come avrete potuto capire da questa carrellata assolutamente non esaustiva, e come preannunciato in apertura, avviandomi alla conclusione, ho più domande che risposte, più dubbi che certezze.
Le strade che si stanno provando a percorrere per costruire repository di qualità, in cui il controllo bibliografico sia efficace e funzionale, sono tante: alcune saranno dei vicoli ciechi, altre si avviano a diventare percorsi ben segnalati per potersi godere al meglio la foresta pluviale che sono i nostri repository.

Non resta che continuare a esplorare.

**Ringraziamenti:**
Vorrei ringraziare il collega, Dott. Paolo Baldi, per gli interessanti e utili spunti di riflessione e il Dott. Emiliano Wass per l’indispensabile aiuto nella traduzione del testo.

BIBLIOGRAFIA

Almuzara Barrionuevo, Leticia, Maria Luisa Díez Alvite, and Blanca Rodríguez Bravo. 2012. "A Study Of Authority Control in Spanish University Repositories." *Knowledge Organization* 39 (2): 95-103. <https://doi.org/10.5771/0943-7444-2012-2-95-1>

Arlitsch, Kenning, and Patrick S. O'Brien. 2012. "Invisible institutional repositories: Addressing the low indexing ratios of IRs in Google Scholar." *Library Hi Tech* 30 (1): 60-81. <https://doi.org/10.1108/07378831211213210>

Barton, Jane, Sarah Currier, and Jessie M. N. Hey. 2003. "Building Quality Assurance into Metadata Creation: An Analysis based on the Learning Objects and e-Prints Communities of Practice." *International Conference on Dublin Core and Metadata Applications; DC-2003--Seattle Proceedings*. <https://dcpapers.dublincore.org/pubs/article/view/732>

Bruce, Thomas R., and Diane Hillmann. 2004. "The Continuum of Metadata Quality: Defining, Expressing, Exploiting." In *Metadata in Practice*, eds. Diane I. Hillmann and Elaine L. Westbrooks (Chicago: ALA Editions)

Chapman, John W., David Reynolds, and Sarah A. Shreeves. 2009. "Repository Metadata: Approaches and Challenges." *Cataloging & Classification Quarterly* 47 (3-4): 309-325. <https://doi.org/10.1080/01639370902735020>

Downey, Moira. 2019. "Assessing Author Identifiers: Preparing for a Linked Data Approach to Name Authority Control in an Institutional Repository Context." *Journal of Library Metadata* 19 (1-2): 117-136. <https://doi.org/10.1080/19386389.2019.1590936>

Gross, Tina, Arlene G. Taylor, and Daniel N. Joudrey. 2015. "Still a Lot to Lose: The Role of Controlled Vocabulary in Keyword Searching." *Cataloging & Classification Quarterly* 53 (1): 1-39. <https://doi.org/10.1080/01639374.2014.917447>

Karen, Calhoun. 2006. The Changing Nature of the Catalog and its Integration with Other Discovery Tools: Final report. Library of Congress. <http://www.loc.gov/catdir/calhoun-report-final.pdf>

Konstantinou, Nikolaos, Dimitrios-Emmanuel Spanos, Nikos Houssos, and Nikolaos Mitrou. 2014. "Exposing scholarly information as Linked Open Data: RDFizing DSpace contents." *The Electronic Library* 32 (6): 834-851. <https://doi.org/10.1108/EL-12-2012-0156>

Kumar, Vinit. 2018. "A Model for Content Enrichment of Institutional Repositories Using Linked Data." *Journal of Web Librarianship* 12 (1): 46-62. <https://doi.org/10.1080/19322909.2017.1392271>

Lubas, Rebecca L. 2009. "Defining Best Practices in Electronic Thesis and Dissertation Metadata." *Journal of Library Metadata* 9 (3-4): 252-263. <https://doi.org/10.1080/19386380903405165>

Mering, Margaret. 2019. "Transforming the Quality of Metadata in Institutional Repositories." *The Serials Librarian* 76 (1-4): 79-82. <https://doi.org/10.1080/0361526X.2019.1540270>

Moulaison Sandy, Heather, and Felicity Dykas. 2016. "High-Quality Metadata and Repository Staffing: Perceptions of United States–Based OpenDOAR Participants." *Cataloging & Classification Quarterly* 54 (2): 101-116. <https://doi.org/10.1080/01639374.2015.1116480>

Niso Framework Working Group. 2007. A Framework of Guidance for Building Good Digital Collections. 3rd edition. National Information Standards Organization (NISO). [http://www.niso.org/publications/rp/framework3.pdf](https://translate.google.com/translate?hl=it&prev=_t&sl=it&tl=en&u=http://www.niso.org/publications/rp/framework3.pdf)

Park, Jung-Ran. 2009. "Metadata Quality in Digital Repositories: A Survey of the Current State of the Art." *Cataloging & Classification Quarterly* 47 (3-4): 213-228. <https://doi.org/10.1080/01639370902737240>

Park, Jung-Ran, and Yuji Tosaka. 2010. "Metadata Quality Control in Digital Repositories and Collections: Criteria, Semantics, and Mechanisms." *Cataloging & Classification Quarterly* 48 (8): 696-715. <https://doi.org/10.1080/01639374.2010.508711>

Powell, Andy, Michael Day, and Marieke Guy. 2004. "Improving the Quality of Metadata in Eprint Archives." *Ariadne* (38). <http://www.ariadne.ac.uk/issue/38/guy/>

Radio, Erik. 2014. "Information Continuity: A Temporal Approach to Assessing Metadata and Organizational Quality in an Institutional Repository." In *Metadata and Semantics Research*, edited by Sissi Closs, Rudi Studer, Emmanouel Garoufallou and Miguel-Angel Sicilia, 226-237. Cham: Springer International Publishing. [https://doi.org/10.1007/978-3-319-13674-5\_22](https://translate.google.com/translate?hl=it&prev=_t&sl=it&tl=en&u=https://doi.org/10.1007/978-3-319-13674-5_22)

Smith-Yoshimura, Karen, Micah Altman, Michael Conlon, Ana Lupe Cristǹ, Laura Dawson, Joanne Dunham, Thom Hickey, Daniel Hook, Wolfram Horstmann, Andrew MacEwan, Philip Schreur, Laura Smart, Melanie Wacker, Saskia Woutersen, and Oclc Research. 2014. *Registering Researchers in Authority Files*. [http://www.oclc.org/content/dam/research/publications/library/2014/oclcresearch-registering-researchers-2014.pdf](https://translate.google.com/translate?hl=it&prev=_t&sl=it&tl=en&u=https://www.oclc.org/content/dam/research/publications/library/2014/oclcresearch-registering-researchers-2014.pdf)

Solomou, Georgia, and Dimitrios Koutsomitropoulos. 2015. "Towards an evaluation of semantic searching in digital repositories: a DSpace case-study." *Program* 49 (1): 63-90. <https://doi.org/10.1108/PROG-07-2013-0037>

Stein, Ayla, Kelly J. Applegate, and Seth Robbins. 2017. "Achieving and Maintaining Metadata Quality: Toward a Sustainable Workflow for the IDEALS Institutional Repository." *Cataloging & Classification Quarterly* 55 (7-8): 644-666. <https://doi.org/10.1080/01639374.2017.1358786>

Svantesson, Lotta, and Monica Steletti. 2019. "DSpace ORCID integration: name authority control solution at the European University Institute." Presented at the The 14th International Conference on Open Repositories (OR2019), Hamburg, Germany <https://doi.org/10.5281/ZENODO.3553926>

Swan, Alma, and Leslie Carr. 2008. "Institutions, Their Repositories and the Web." *Serials Review* 34 (1): 31-35. <https://doi.org/10.1016/j.serrev.2007.12.006>

Thomas, R. Bruce, and Hillmann Diane. 2004. "The Continuum of Metadata Quality: Defining, Expressing, Exploiting." In *Metadata in Practice*. Chicago: ALA editions.

White, H. C., S. Chen, and G. Liu. 2018. "Relationships between metadata application and downloads in an institutional repository of an American law school." *LIBRES* 28 (1): 13-24. <https://www.libres-ejournal.info/2608/>

Zhu, Lihong. 2019. "The Future of Authority Control: Issues and Trends in the Linked Data Environment." *Journal of Library Metadata* 19 (3-4): 215-238. <https://doi.org/10.1080/19386389.2019.1688368>

1. In questo convegno almeno 5 o 6 relatori hanno inserito tale parola nei titoli dei propri interventi. [↑](#footnote-ref-1)
2. In his keynote address to the Library of Congress Bicentennial Conference on Bibliographic Control for the New Millennium, Michael Gorman (2001, xxv) referred to metadata as “a fancy name for an inferior form of cataloguing,” and as “unstandardized, uncontrolled, ersatz cataloguing.” Citato in nota 6 da Howarth 2005, Metadata and Bibliographic control: soul mates o two solitudes, Cataloging and classification Quarterly, 40 (3-4), 37-56. [↑](#footnote-ref-2)
3. NISO Framework Working Group, *A Framework of Guidance for Building Good Digital Collections*, 3rd edition (Baltimore, MD: National Information Standards Organization (NISO), 2007), 61–62.<http://www.niso.org/publications/rp/framework3.pdf> [↑](#footnote-ref-3)
4. Problema segnalato anche per la gestione delle “name entries” da Salo Dorothea, 2009. “Name authority control in Institutional repositories”, *Cataloging and classification quarterly*, 47 (3-4), 249-261. doi: 10.1080/01639370902737232 [↑](#footnote-ref-4)
5. Vedi anche H. Fari, S. Khan and M. Y. Javed, "Publishing institutional repositories metadata on the semantic web," *Eighth International Conference on Digital Information Management (ICDIM 2013)*, Islamabad, 2013, pp. 79-84, doi: 10.1109/ICDIM.2013.6694008. e Robert J. Hilliker, Melanie Wacker & Amy L. Nurnberger (2013) Improving Discovery of and Access to Digital Repository Contents Using Semantic Web Standards: Columbia University's Academic Commons, Journal of Library Metadata, 13:2-3, 80-94, DOI: [10.1080/19386389.2013.826036](https://doi.org/10.1080/19386389.2013.826036) [↑](#footnote-ref-5)