

## Bibliographic control and institutional repositories: welcome to the jungle

Tessa Piazzini<sup>(a)</sup>

a) Università degli Studi di Firenze, <http://orcid.org/0000-0002-8876-371X>

**Contact:** Tessa Piazzini, [tessa.piazzini@unifi.it](mailto:tessa.piazzini@unifi.it)

**Received:** 3 April 2021; **Accepted:** 23 May 2021; **First Published:** 15 January 2022

### ABSTRACT

In 1994 cognitive scientist Stevan Harnad made what he defined a “subversive proposal” to his colleagues: «immediately start self-archiving their papers on the Internet». Since then, institutional repositories have been chaotically developing, alongside disciplinary repositories. In the early XXI Century the public debate was centered on their purposes and therefore on what they were supposed to contain; librarians joined the discussion and contributed to it by implementing descriptive standards such as Dublin Core and interoperability protocols (OAI-PMH). The themes under discussion were closely related to bibliographic and authority control, given that the quality of metadata has a profound impact on the quality of the services offered to users. Presently, we are still trying to answer some of those old questions: what (or whom) are IRs for? Is bibliographic control so necessary within an environment that has never failed in self-archiving? Can we consider IRs a bibliographic tool? We also need to deal with a wider vision: in a scenario that saw the transition from OPACs (created, managed and controlled by librarians) to current discovery tools (with their information redundancy and the related problems on data correctness and quality control) can librarians still be authoritative and act effectively?

### KEYWORDS

Authority control; Institutional repository; Bibliographic control; Metadata.

## 1. Introduction

In library sciences we often talk about “ecosystems”<sup>1</sup>. Within this naturalistic metaphor, I used to think of repositories as a jungle: chaotic, dense and impenetrable. Now, however, I look at repositories rather as a rain forest: an equally complex environment full of variety, multi-layered, characterized by a lot of internal and external communication networks and also hidden and visible interdependencies. An environment that rests on a clean, rich soil, on which it is possible to move and walk.

The relatively short history of repositories shows us a great variety in terms of stored material (pre-print, research publications, teaching materials, articles and books, theses, multidisciplinary or specialized), in terms of population and organization (self-archiving, batch retrieval, internal collections, librarian mediated insertion) and also in terms of software (Digital Commons, DSpace, Eprints, Fedora...).

We must not forget that the repositories were born because of the initiative of the scientific community, in particular by the will of single authors; it all started with the “subversive” proposal by Stevan Harnad, professor of cognitive sciences at the Virginia Polytechnic Institute, in 1994: sharing their own research within the institution through the self-archiving of online contributions, in order to make their dissemination more effective. Hence the embryo of a new type of open archive: the institutional archive, promoted and managed by an institution, which goes alongside the disciplinary ones created by aggregation of documents concerning single research areas. We should also remember that “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” (Moulaïson Sandy and Dykas 2016, 105).

Such repositories, therefore, are fed by the authors themselves through the practice of “self-archiving”. It is no coincidence that the elements on which the foundations of repositories rest are the use of minimal metadata sets (such as the Dublin Core Metadata Element Set, adopted by most repository providers) and interoperability standards, such as the OAI-PMH, which go hand in hand with such minimality. According to some scholars, this choice was due to the desire of encouraging the participation of authors in filling the repositories, however it did discourage any activity that could be perceived as a barrier between scholars and their own institutions; this perception has also affected any intervention and quality control measure in the process of metadata creation (Barton, Currier and Hey 2003). Consequently, unlike other more familiar environments, such as OPACs and Discovery Tools, institutional repositories were not created by librarians to organize and make available the bibliographic universe; they are not tools whose birth falls within our sphere, although the contribution of many librarians has undoubtedly been, and still is, vast (for example, in the case of the development of the Dublin Core, despite the “bibliographic control community” seeing it at its birth as “simplistic”<sup>2</sup>).

The awareness that “[for IRs] there is no universally accepted practice or standard defining quality

---

<sup>1</sup> In this issue many speakers have included this word in the titles of their articles.

<sup>2</sup> 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.” Cited in footnote 6 by Howarth 2005, *Metadata and Bibliographic control: soul mates o two solitudes*, *Cataloging and classification Quarterly*, 40 (3-4), 37-56.

metadata; similarly, there is no set of rules for describing institutional repository materials” (Stein, Applegate and Robbins 2017, 650) marks even more the difference between IRs and actual bibliographic tools.

## 2. Repositories as rain forests or woodlands?

This awareness leads to a first question: what should we do about these environments? Should we maintain their relatively unorganized nature (with the relative entropy) or should we transform these rain forests into controlled, orderly and organized woodlands, according to our vision of the (bibliographic) world?

Maybe we should be looking for the golden middle way, or the “*aurea mediocritas*”, to quote the Latin poet Horace, who reminded us “*est modus in rebus*”: we should tread carefully and with the respect that is due to an environment with its own characteristics; we should try to intervene most discretely, in order not to deform it, but simply to provide an orientation to its users, so that they not get lost and can appreciate all its beauty.

Therefore, perhaps bibliographic and authority control should also be careful, aiming at providing the users (both humans and machines) with the necessary information in the best possible form, for them to enjoy a pleasant, safe and satisfying journey.

Although born within the academic communities and made their own by the institutions, the repositories, in fact, cater to the widest possible audience, and their main goal is – desired by the authors themselves – not so much to organize and systematize what is produced by an institution (a task that can well be carried out by a catalog), but to ensure the widest visibility for the longest time.

Long-term access and storage can be achieved through two closely linked elements: good metadata and a good repository system; in fact “quality metadata may be underutilized due to weakness in indexing, navigation, and display options”(Moulaison Sandy and Dykas 2016, 103).

Defining what is meant by “quality metadata” is not, however, an easy task in itself: the subjective and local elements remain strong, and the close link between functional requirements and suitability for purpose is often emphasized (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 expose local metadata relating to external service providers” (Park 2009, 214). At the same time, therefore, it is necessary to guarantee interoperability within a Search Engine Optimization framework, and that implies a much wider series of technical activities and operational choices that should be part of the development plan for a repository.

As librarians we have the task of providing support for the creation of quality metadata, as the NISO (National Information Standards Organization) reminds us in its *A Framework of Guidance for Building Good Digital Collections* (NISO 2007, 61-62), when it says that 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.

When speaking of quality of metadata, we cannot, therefore, neglect a fourth aspect besides those of in accuracy, completeness and consistency, which we always find in the literature (Park 2009): their effectiveness in terms of information retrieval, in particular in relation to indexing by search engines. A 2019 study (Mering 2019) conducted upon the University of Nebraska-Lincoln IR tells us that 57% of the collection was accessed via Google Search and only 17% was accessed directly from within the repository. Already in 2012 Arlitsch and O'Brien reminded us that "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). Their research also showed that all the analyzed repositories, regardless of the platform, had a low index ratio making them virtually invisible to Google Scholar.

The ultimate goal, however, remains that of ensuring the widest visibility (Swan and Carr, 2008), which is also achieved by making the repository content externally shareable. "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).

It is therefore important that "for metadata to be effective, enforcement of standards of quality must take place at the community level" (Bruce and Hillman 2004, 3) and that it is necessary to "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 cataloging, metadata creation and information retrieval" (Barrionuevo Almuzara, Alvite Díez and Rodríguez Bravo 2012, 101).

Economics, however, teaches us that effectiveness, in cost-benefit analysis, is always accompanied by efficiency, while our daily experience as librarians continually reminds us that bibliographic control activities are very time-consuming and resource-consuming and also require high professional skills.

For example, a survey (Moulaison Sandy and Dykas 2016), conducted in 2015 among some US repository administrators, indicated time limitations and staff hours and skills levels of staff among the greatest obstacles to the provision of high quality metadata, even for those who had self-assessed the quality of their own metadata repository as above average.

### **3. Different levels of supports for authors and staff in bibliographic control**

Hence the need to make prudent choices on how to distribute efforts and how to use all the tools that can help reduce costs.

Now, in the face of the continuous growth of scientific publications – a result also of increasingly pervasive evaluation activities – especially when the main source is self-archiving, one of the tasks of librarians within IRs is to build clear and fast paths, which simultaneously guarantee ease of

use and quality of information both for authors and readers, by working on those metadata that have an impact on access.

The first basic form of support consists in offering authors or editors explanatory notes, instructions, drop-down menus or pop-up windows when filling in the bibliographic form online: in this case we certainly cannot speak of bibliographic control; nevertheless, it is a first step which we cannot ignore. Providing clear information, simple tools and technical support goes along with training and education activities, and they all constitute an indispensable step towards creating a community of users who have awareness of the creation of quality metadata, as to reduce *ex ante* the need for subsequent controls.

At the same, initial, layer of support we find the preliminary definition of best practices and guidelines destined to repository managers and 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).

A second level consists of recovering quality data from third-party sources through identifiers. Here is where another naturalistic (or rather, environmentalist) metaphor comes into play: recycling, intended as the possibility of quickly recovering structured and valid information from external sources.

Currently, when it comes to bibliographic and authority control, scholars seem to agree on two main tools:

- Unique publication identifiers for retrieving verified metadata;
- Unique personal identifiers (ORCID, VIAF, ISNI, LoC Authority Name) to ensure the quality control of the authors.

Both tools currently have some limitations and difficulties, but there seems to be a certain agreement among scholars on their effectiveness.

With an eye to economic sustainability, there is a growing tendency for repositories to offer the possibility to retrieve a lot of information from external databases or directly from the publisher through DOI, Scopus identifier or Web of Science Accession number, ISBN, PMID.

A choice of this kind implies on our part, as librarians, the acceptance of delegating the organization and presentation of data to third parties, but this is not new to us: we have done it with derived cataloging, we do it today in part with discovery tools.

If, on the one hand, this can in principle lead to a certain homogeneity in the presentation of information, reducing the risk of typos and errors deriving from manual entry, on the other hand it only partially reduces the need for bibliographic control. This is first of all because publishers and individual databases each have their own metadata and cataloging rules, and secondly because harvesting activities strictly depend on the interoperability protocols applied and on the mapping between the different sets of metadata (Chapman, Reynolds and Shreeves 2009).

For example: the title of an article in the Pubmed database is always presented in English, even if the article is published in another language. In this form the dc.title field of the Core Set Dublin Core is usually imported, where, instead, there should be the title in the original language.

Furthermore, we know well how partial in their coverage large international databases are when it comes to languages or subject matters, and how there are no specific identifiers for certain kinds of publications such as, for example, the essay within a volume – although many publishers are beginning to equip individual book chapters with their own DOI.



Even in the case of batch uploads carried out by dedicated staff, the quality of the data is not guaranteed, indeed in some cases it seems to be even lower (Stein, Applegate and Robbins 2017), unless an intense metadata cleanup is planned prior to batch ingestion using tools like Open Refine<sup>3</sup>.

#### 4. Authority control as part of bibliographic control

We find a similar problem especially in the management of authors' names, whose ever-changing forms have always been a great challenge for the authority control. It could be due to the will of authors (discontinuous use of the middle name or of abbreviated forms, change of surname after marriage...) or for other reasons (different presentations in various sources, linguistic variants, transliterations, etc.).

In order to overcome this problem, as already noted, a first form of assistance is the auto-completion function that – although not present in every relevant software – can kick in during the insertion phase: although this functionality can contribute to the reduction of variant forms, this does not strengthen the authority control.

At a slightly higher level – so much so that we can speak of authority control at a local level – we find the automatic linking of the author's name to the institution identity management system; nevertheless it is evident that this solution does not fully guarantee the interoperability and shareability and “can be, at best, [only] one part of the repository authority control puzzle” (Downey 2019, 130).

In order to achieve this result, unique identifiers of the names are being implemented within the repositories, by linking to external authority schemes.

The interesting aspect, determined by the evolution towards an Linked Open Data model, is the transition from the concept of “name authority work” to that of “identity management”, thanks to the association of registered identifiers: “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). The coexistence, however, of numerous projects for the name authority control with the consequent production of different identifiers constitutes an additional element of noise and can lead to the necessity, once again, to make choices.

In 2019 Moira Downey, a colleague from Duke University, published an analysis (Downey 2019) of three among the major international authority sources – Library of Congress Name Authority Files (LCNAF), Virtual International Authority File (VIAF), and Open Researcher and Contributor Identifier (ORCID) – looking to develop a Linked Data authority control within their institutional repository, given the ability of the mentioned systems to provide author URI's via API. According to the author, LCNAF and above all VIAF, which has developed a cooperative model for the aggregation of authority data from national and regional sources with an intense activity of clustering, merging and deduplication, constitute “a broader step forward in preparing library data

---

<sup>3</sup> Problem also reported for the management of name entries by Salo Dorothea, 2009. “Name authority control in Institutional repositories”, *Cataloging and classification quarterly*, 47 (3-4), 249-261. doi: 10.1080/01639370902737232

for better integration with the broader web” (Ibid., 120), but still rely on traditional mechanisms of participatory cataloging and authority control that have an impact on the creation of identifiers, in particular for authors of articles in journals, who happen to represent the category with the biggest presence within many academic repositories, given the hyperproduction of literature in the fields of medical sciences and STEM (Science, Technology, Engineering and Mathematics). Nevertheless, they guarantee reliability on their persistence, thanks to the professionals involved, with the creation of “record in structured, machine-actionable format that did not require additional resources or inferences to ascertain” (Ibid., 131).

ORCID, on the other hand, operates as a “self claim researcher registry”, which seems to delegate the authority control traditionally carried out by libraries directly to researchers, also giving them a certain autonomy of choice on their online identity in the universe of academic communication. A leap of faith by librarians or the recognition that we can no longer be the absolute rulers of the organization and presentation of information?

As often happens, the reality lies somewhere in between: ORCID URI’s prove to be a good solution for self-archiving, but “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 particular, we have no certainty about the persistence of this identifier, given that any author could decide to remove their profile at any time. We also have the same problem with local registries, which by nature are closely linked to the duration of the author’s presence within the institution.

I believe that offering authors a tool, even if not a perfect one, to present themselves is a courageous and intellectually honest choice that we can support by making its use and implementation as easy as possible within “our” repositories, and continuing to invest in a parallel education and information activity on best practices.

There are now numerous experiences in this sense in the world.

For example, in Italy, in 2015, during the national research quality assessment exercise (VQR 2011-2014), the IRIDE project (Italian Research IDentifier for Evaluation) was launched, aiming to equip the Italian academic community (professors, university researchers and research institutions, doctoral students and post-docs) with a persistent ORCID identifier, by activating the registration procedure directly within the repository of their institution. Most of the Italian university repositories, which use proprietary software, have since then allowed, through a push and pull system, a bidirectional communication with their ORCID account.

Equally interesting is the experience (Svantesson and Steletti 2019) of the European University Institute of Fiesole (one of the hosts of 2021 Florence Conference on Bibliographic Control in the Digital Ecosystem) for the integration of its databases – CADMUS, EUI Central Persons Registry – with ORCID, which is also a solution for the authority control over the names of authors that partially compensates for the absence of a CRIS (Current Research Information System). The choice of using the form in the repository as the preferred name is particularly interesting, reminding us that “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).

## 5. The challenge of the semantic control

If there is a certain agreement on pursuing these paths, that is not the case regarding the opportunity to invest time and resources for bibliographic control on the semantic component.

In the context of an institutional repository, in many cases a multidisciplinary one, often fed by the authors themselves, the depth, breadth and variety of disciplines means that the use of subject-controlled terms is possible only at a high level, if we are to maintain homogeneity within the repository itself.

If, on the other hand, we want to respect the heterogeneity and we let the communities self-discipline, we will end up with a repository in which the consistency of semantic metadata will be extremely varied: from their total absence to populating via recognized thesauri, and in between the complexity determined by the use of synonyms, homonyms and grammar, spelling and linguistic variants.

We are again faced with the entropy vs. control dilemma: how far must our intervention as repository managers go? Lubas, speaking of PhD theses and dissertation repositories, argued that any subject indexing intervention by staff should have complemented and not replaced the choice of keywords made by the authors, even if this would have led to an inevitable increase in noise (Lubas 2009). A normalization of the keywords, or their mapping on a pre-existing controlled vocabulary would have, in fact, eliminated the unique perspective with which authors refer to their work (Radio 2014).

An interesting study from 2018 (White, Chen and Liu, 2018) tried to analyze the relationship between the presence of some metadata in the Duke University Law School repository and the number of downloads, to understand the effectiveness of the metadata itself.

The results were surprising: the number of co-authors and the presence and the number of keywords (whether they were free text or derived from controlled vocabularies) had a substantially negative correlation with downloads and were not essential for users to reach the content. On the contrary, the presence and length of the abstract had a significantly positive impact.

This partly contradicts our certainty about the importance of subject indexing and its effectiveness. A certainty already undermined over the years by studies which invited us to abandon their use (Calhoun 2006), even within the catalogs. At the same time, other studies confirmed its efficacy (Gross, Taylor and Joudrey 2015). This polarization has now widened with the adoption of Discovery tools by many libraries, which reproduce a sort of “Google-like” environment – in the name of an alleged desire to make the information retrieval experience as satisfying as possible for the user – without, however, the power of Google’s Page Rank. Perhaps the term “jungle” is better suited to define such tools, more than IRs.

In order to ensure a balance, a great help could come from Linked Open Data and the Semantic Web, which, with regard to subject indexing, could make an important contribution to the enrichment of contents and bibliographic control, through a simpler management of multiple languages, better linkability of resources and simpler reuse of authority registries in applications. Furthermore, the “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).

While repositories go through an inevitable initial effort of adaptation, the choice between differ-



ent technologies depends on the complexity and rigor required by the specific environment (Zhu 2019). As already seen at the beginning for quality metadata, also when choosing semantic web tools the relationship between suitability for the purpose and the peculiarities of the environment to which the tools are applied must be addressed, allowing a possible scalarity of choice.

There are also many interesting experiences in this area: in 2017, at the Central University of Gujarat, India, a prototype (Khumar 2018) was developed, in which they linked the Dbpedia knowledge base to a Dspace-based repository, chosen as a linked dataset for its broad disciplinary coverage, for its automated updating mechanisms and its multilingual information support. Equally interesting is the research conducted by Greek scholars on “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). And it’s not the only research of this kind<sup>4</sup>.

## 6. Conclusion

At the end of this absolutely non-exhaustive overview, the conclusion is that there are more questions than answers, more doubts than certainties.

However, it is clear that librarians will not be able to fail in their task of helping to build reliable, rich and “clean” repositories, while exploiting the potential offered by third parties for the creation of quality metadata and bibliographic control.

There are many roads that are being tried to build quality repositories, in which bibliographic control is effective and functional: some will be dead ends, others will become well-marked paths through which librarians and users will be able enjoy the rich rainforest that institutional repositories represent.

All we need to do is to keep on exploring.

## Acknowledgements

I would like to thank my colleague, Paolo Baldi, for the interesting and useful food for thought and Emiliano Wass for his fundamental help for the translation.

---

<sup>4</sup> See also H. Fari, S. Khan and MY Javed, “Publishing institutional repositories metadata on the semantic web,” *Eighth International Conference on Digital Information Management (ICDIM 2013)*, Islamabad, 2013, 79-84, DOI: <https://dx.doi.org/10.1109/ICDIM.2013.6694008> and Robert J. Hilliker, Melanie Wacker and 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: <https://doi.org/10.1080/19386389.2013.826036>

## References

- 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*. Accessed 22 November 2021. <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. Westbrook (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. Accessed 22 November 2021. <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>
- 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). Accessed 22 November 2021. <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://doi.org/10.1007/978-3-319-13674-5_22)
- Smith-Yoshimura, Karen, Micah Altman, Michael Conlon, Ana Lupe Cristán, 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*. Accessed 22 November 2021. <http://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>