



The Magazine of Digital Library Research

[HOME](#) | [ABOUT D-LIB](#) | [CURRENT ISSUE](#) | [ARCHIVE](#) | [INDEXES](#) | [CALENDAR](#) | [AUTHOR GUIDELINES](#) | [SUBSCRIBE](#) | [CONTACT D-LIB](#)

## D-Lib Magazine

March/April 2017

Volume 23, Number 3/4

[Table of Contents](#)

---

## Broken-World Vocabularies

Daniel Lovins

New York University (NYU) Division of Libraries

daniel.lovins [at] nyu.edu

Diane Hillmann

Metadata Management Associates LLC

metadata.maven [at] gmail.com

---

<https://doi.org/10.1045/march2017-lovins>

---

## Abstract

There is a growing interest in vocabularies as an important part of the infrastructure of library metadata on the Semantic Web. This article proposes that the framework of "maintenance, breakdown and repair", transposed from the field of Science and Technology Studies, can help illuminate and address vulnerabilities in this emerging infrastructure. In particular, Steven Jackson's concept of "broken world thinking" can shed light on the role of "maintainers" in sustainable innovation and infrastructure. By viewing vocabularies through the lens of broken world thinking, it becomes easier to see the gaps – and to see those who see the gaps – and build maintenance functions directly into tools, workflows, and services. It is hoped that this article will expand the conversation around bibliographic best practices in the context of the Web.

Keywords: Semantic Web, Vocabularies, Science and Technology Studies

---

## 1 Introduction

The Semantic Web is creating new opportunities for distributed knowledge management and even "universal bibliographic control" (Coyle, [2010a](#), [2010b](#); Dunsire, Hillmann & Phipps, [2012](#)). Many projects are underway to convert library data into forms that can be consumed and published by Semantic Web applications. At the same time, the Web is also a distributed and volatile "web of dependencies," and services based on Web infrastructure face constant risk of breakdown. This can seem a bit counter-intuitive, as the naive experience of cyberspace is one of an alternative reality, seemingly free from material constraints like maintenance, repair, and property rights (cf. Fidler, [2016](#)). But the reality of the Web involves server farms, electricity grids, fiber-optic cables, routers and many other components, all subject to laws of governments and entropy. The software and data that inform these systems are similarly embodied and constrained.

A research network has developed among science and technology scholars around "maintenance and repair studies," an orientation that raises awareness of community norms and processes that underlie successful infrastructure and sustainable innovation. One of its progenitors, Jérôme Denis ([2016](#)) describes how it emerged from research David Pontille and he were conducting on the Paris subway wayfinding system, which turned up "a much less stabilized world than we thought: from day to day, the supervision and repair operations ensure the sturdiness, the sustainability and the efficiency of the signs network intended to guide the travels."

Stephen Jackson, Lee Vinsel, Andrew Russell, Nathan Ensmenger, David Edgerton, and others have explored these topics in recent publications and conferences in the field of [Science and Technology Studies \(STS\)](#) and allied disciplines.

This article picks up on a particular strand of maintenance and repair studies, namely Jackson's framework of "broken world thinking", i.e., where infrastructure is viewed not as a fundamentally sound system with occasional lapses, but rather as a fragile network of dependencies in various states of disrepair, kept from further collapse only by diligent attention and intervention of persons one might call "maintainers."

In discussing maintenance, one needs to keep in mind not just the care of hardware and software systems, but also the community norms and commitments that prevent those systems from breaking, or enable them to be repaired on a regular basis. Russell and Vinsel (2016) suggest in their essay "Hail the Maintainers," that conversations about innovation and infrastructure are at their best when they "move away from narrow technical matters to engage deeper moral implications." Noting that "contemporary discourse treats innovation as a positive value in itself, when it is not," they remind us that "a focus on maintenance provides opportunities to ask questions about what we really want out of our technologies. What do we really care about? What kind of society do we want to live in? Will this help get us there?"

In the case of bibliographic vocabularies, as with other socially-embedded technologies, the idea is that by making maintenance work more visible, one can begin to shift attention to under-resourced tasks that will help vocabularies reach their full potential on the Web. The Web, in turn, stands to benefit as well, in the form of universal access to stable, scalable knowledge management data and tools. We suggest that a lack of appreciation for broken world constraints leads to unbalanced priorities, where investments flow largely to novelty projects and prototypes, and insufficiently to persons and protocols that enable maintenance and repair, and thus *sustainable* innovation. This is not to suggest that innovative projects are unimportant, only that the community needs to pay close attention to collateral activities if those projects are to succeed and flourish. Moreover, some of the most consequential innovations involve new workflows for seemingly mundane tasks. Consider the profound impact of [Git](#), for example, on distributed version control or [Slack](#) on team messaging and file management.

Given this complex and volatile environment, forward-thinking managers of bibliographic vocabularies are considering the role of best practices for publishing, maintaining and sharing their data on the Web. Members of a Vocabulary Special Session at the meeting of the 2011 Dublin Core Metadata Initiative (DCMI) identified specific issues and opportunities for such shared practices, and more recently the National Information Standards Organization (NISO) has commissioned a Bibliographic Roadmap Initiative to advance these efforts.

---

## 2 The Semantic Web and Linked Bibliographic Vocabularies

Libraries have been developing and using bibliographic vocabularies for many decades. The revolution in information technology that began in the 20th century, and continues unabated into the 21st, has affected bibliographic vocabularies in three fundamental ways. First, the [explosion of knowledge-creation](#) after World War II outstripped the capacity of libraries to collect and describe resources of interest to scholars. Second, concurrent with that explosion of new (often digital) resources, libraries and other institutions were increasingly subject to financial pressure by demands for new user services. Third, the World Wide Web created a networked publishing, discovery and retrieval environment (Berners-Lee, 1989) that risked leaving libraries behind if they failed, as Eric Miller (DCMI Announce, 2014) puts it, to "speak the language of the Web".

Bibliographic data fall into three general categories: instance data (such as MARC bibliographic records), which contain structured and/or unstructured information representing members of a particular class or collection; structural vocabularies (such as Resource Description & Access (RDA)), which standardize how information is organized, recorded, exchanged, interpreted, etc; and value vocabularies (such as Library of Congress Subject headings (LCSH) and the Getty Art and Architecture Thesaurus (AAT)), which provide controlled terms for use with a structural vocabulary. While advances in Web technology and artificial intelligence have in some ways lessened the need for structured vocabularies (cf. Shirky, 2005; Weinberger, 2007; Doctorow, 2001), these vocabularies continue to play a key role in discovery and interoperability, providing contextual cues, relationships, and consistent facet labeling, for example, that make it possible to find the "needle in a haystack" among millions of objects in research collections.

Other challenges to the traditions of sharing library data come into play. Even now, most libraries focus their data creation and sharing activities on published materials available in many libraries, each of which maintains a local catalog (with few exceptions made available in digital form). Those libraries maintaining collections of unique resources, whether rare or archival in scope, have only recently articulated the value of the 'special' collections to broader audiences and shifted resources towards their description and availability on the web. Data outside the traditional library sharing environment, whether crowdsourced or created by machines (or those not familiar with library rules of description) are treated with suspicion and rarely integrated into bespoke data silos. The result has been that most activity based on shifting to more Semantic Web practices focuses almost entirely on exposing data on the Web, rather than figuring out how to use data provided by others. This reality has tended to limit creation and distribution of vocabularies not part of the traditional practices of library description.

That said, since the [Resource Description Framework](#) (RDF) was introduced by the W3C in 1999, followed by the RDF-linked DCMI Abstract Model in [2005](#) (revised in [2007](#)), there has been growing interest among libraries to model their data in alignment with the Semantic Web. Considerable progress has been made, as evidenced in RDF datasets published by national libraries, research institutes, OCLC (using bibliographic extensions to [Schema.org](#)), and other organizations. As Harper & Tillett ([2007](#)) have pointed out, by modeling vocabularies in RDF, OWL, SKOS (etc.), library metadata will become more interoperable with Dublin Core and other RDF datasets, and thus more useful for applications based on machine inferencing. Baker, Vandenbussche & Vatan ([2013](#)) have described the critical role that vocabulary management plays in scholarly communication: "As the givers of meaning to datasets, vocabularies are of vital importance to the scholarly record and cultural memory."

---

### 3 RDA, FRBR, and BIBFRAME

Arguably the most advanced and well-supported vocabularies for multi-language bibliographic description are those associated with RDA. The importance of RDA is reflected in its diverse and growing community, led by a [steering committee](#) that is determined to bring in participants from outside the traditional Anglo-American nexus. Various translations have been completed or are in progress, signaling RDA's determination to become an international standard. RDA's data model is the Functional Requirements of Bibliographic Records (FRBR) as published by the International Federation of Library Associations (IFLA Study Group, [2008](#)), and more recently the IFLA Library Reference Model, described by Riva & Žumer ([2015](#)), and nearly complete. While RDA began as an instruction manual (a "content standard") published as the RDA Toolkit and intended for construction of bibliographic descriptions, the structural and value vocabularies managed through the [RDA Registry](#) are the basis for the current shift to the Web.

Similarly important is the Bibliographic Framework Initiative (BIBFRAME), developed by the Library of Congress in collaboration with [Zepheira](#), and designed to succeed MARC21 as a metadata encoding and exchange standard. The Library of Congress describes [BIBFRAME](#) as, "the foundation for the future of bibliographic description that happens on the web and in the networked world." BIBFRAME-based services are in development at OCLC, [Casalini Libri](#), Mellon-funded projects, Zepheira's [Library.Link](#), and in other contexts. It is also true that by simplifying the modeling of resources, BIBFRAME harks back to a flatter "record" model, familiar in MARC, and as such undermines interoperability with other library-domain vocabularies. There are also questions about its extensibility to non-book formats and non-English-language resources. For example, a BIBFRAME AV Modeling Study (Van Malssen, [2014](#)) highlighted issues with the concept of a creative "work", which fails to accommodate documentary artifacts like birdsong recordings.

In the meantime, the obsolescent MARC21 standard continues to be the indispensable data format for library data management and exchange, despite Roy Tennant's ([2002](#)) compelling argument on why "MARC must die". What allows MARC to endure, even while ostensibly better options exist is that MARC is still the basis for the Library Management Systems widely in use. Library vendors have been slow to look at new formats for data and integrate them into their offerings, largely because they represent a significant investment in an environment shrinking rapidly by vendor mergers and acquisitions.

The important concerns for all of these standards revolve around underinvestment in critical infrastructure change, as well as high-level risk-aversion within the commercial vendor community asked to respond to those changes. It has been suggested that a preoccupation with "Linked Open Data" and "Semantic Web" places an unfortunate emphasis on conceptual purity, rather than on inevitable compromises needed to make systems work (cf. Rochkind, [2015](#)). Or perhaps there is wishful thinking that a future system infused with artificial intelligence will be able to bypass metadata experts and maintainers. In any case it seems clear that the remaining commercial vendors are waiting for assurances about winners and losers that may never come, and the community may need to find willing partners outside of the current group.

---

### 4 The Need for Best Practices

Technical guidelines established by the W3C provide a foundation for creating and maintaining Web-based vocabularies. Of special importance are "Data on the Web Best Practices ([2016](#))," and "Best Practice Recipes for Publishing RDF vocabularies" ([2008](#)). The Dublin Core Metadata Initiative, NISO and other organizations have similarly contributed guidelines and recommended practices for their respective communities of practice. The NISO Bibliographic Roadmap Initiative (report forthcoming) is reviewing vocabulary maintenance and sustainability practices, including version control, documenting terms-of-use (e.g., licenses) namespaces, ownership and governance, and preservation.

The W3C ([2016](#)) recommends reuse of pre-existing vocabularies rather than creation of new ones. As Heath & Bizer ([2011](#)) explain, "Reuse of existing terms is highly desirable as it maximizes the probability that data can be consumed by applications that may be tuned to well-known vocabularies, without requiring further pre-processing of the data or modification of the application."

Note, however, that the principle of vocabulary reuse is not universally valid or held. Bill de hÓra [sic] ([2007](#)) has expressed the trade-off as follows:

*There are two schools of thought on vocabulary design. The first says you should always reuse terms from existing vocabularies if you have them. The second says you should always create your own terms when given the chance. The problem with the first is you are beholden to someone else's sensibilities should they change the meaning of terms from under you (if you think the meaning of terms are fixed, there are safer games for you to play than vocabulary design). The problem with the second is term proliferation, which leads to a requirement for data integration between systems (if you think defining the meaning of terms is not coveted, there are again safer games for you to play than vocabulary design). What's good about the first approach is macroscopic – there are less terms on the whole. What's good about the second approach is microscopic – terms have local stability and coherency. Both of these approaches are wrong insofar as neither represents a complete solution. They also transcend technology issues, such as arguments over RDF versus XML. And at differing rates, they will produce a need to integrate vocabularies.*

Vocabulary reuse and integration are familiar principles to librarians, as knowledge organization systems (KOSs) depend on a shared (if not necessarily universal) understanding of key concepts and definitions. Moreover, by sharing common vocabularies, libraries benefit from economies of scale and spreading the financial cost of common infrastructure. In order to encourage reuse or extension on the Web, however, vocabularies need to be discoverable and stable. Otherwise, developers may falsely, but understandably, assume that they do not exist.

What does it mean for a vocabulary to be discoverable on the Web? At the very least, some sort of description of the vocabulary as a whole must be indexed by search engines. However, there is a deeper level of discovery, one that involves key characteristics of a vocabulary: how many classes, how many properties, how widely used and by whom? Is a license required? Is there a service-level agreement, so that the vocabulary will not suddenly change its namespace structure or disappear from the Web? And if terms are changed or deprecated, how will downstream applications be notified? The Linked Open Vocabularies (LOV) project provides precisely these kinds of metadata and has subsequently become an effective discovery portal for structural vocabularies. This deeper level of documentation and discovery is essential for enabling reuse, and thus contributes to the health of the bibliographic ecosystem.

---

## 5 Broken Vocabulary Thinking and Articulation Work

In his book chapter entitled "Rethinking Repair" (2014), Stephen Jackson proposes a kind of "broken world thinking [that] asserts that breakdown, dissolution and change, rather than innovation, development, or design as conventionally practiced and thought about are the key themes and problems facing new media and technology scholarship today." He does this through a conceptual re-framing, including the idea of "articulation work" (p. 223) or "the art of fitting, the myriad (often invisible) activities that enable and sustain even the most seemingly natural or automatic forms of order in the world [...] When articulation fails, systems seize up, and our sociotechnical worlds become stiff, arthritic, unworkable." Articulation work is ever-present in bibliographic services. Traditional catalogers saw their role as 'fitting' new materials in their existing catalogs; currently, metadata librarians and vocabulary managers practice the art of 'fitting' when adding new terms, updating schema crosswalks, adjusting normalization rules, creating [semantic maps](#), and otherwise enabling KOSs and discovery services.

In a sense, bibliographic vocabularies have always been "broken". Anyone who has participated in library standards committees knows how much effort is required to keep MARC, RDA, LCSH, etc. in stable condition. This is partly from internal inconsistencies born of compromise, and partly because the world around descriptive vocabularies is itself constantly breaking. When Czechoslovakia broke apart in 1993, for example, to be succeeded by the new nation states of the Czech Republic and Slovakia, subject headings and descriptive notation had to change as well. This sort of thing (i.e., shifting geopolitical realities) happens all the time, if not always as dramatically. Moreover, the library world has never managed to integrate title-level descriptors (e.g., LCSH) with article-level index terms (e.g., [ERIC descriptors](#)), a kind of long-standing, if largely forgotten, brokenness from the users' perspective. Indeed, the sheer number of vocabularies used in bibliographic systems, and the difficulty in harmonizing them, has long imposed a burden on library users. Riley and Becker's (2010) celebrated info-graphic "Seeing Standards", illustrates the great diversity of vocabularies and other cultural heritage metadata that must be (but often are not) reconciled for effective resource discovery.

Software applications are at constant risk of breakdown because of their many dependencies (external code libraries, scripts, services, databases, etc.) that frequently change. An important mitigating factor has been the practice of "semantic versioning", which flags changes in dependencies likely to prove incompatible with local software. The basic idea is to number software releases as X.Y.Z, where X is a major change, Y a minor change, and Z a patch, so that, for example, moving from versions 1.1.0 to 1.1.1 – an incremental change at the patch level – is unlikely to break functionality in the application that depends on it. The practice of semantic versioning has spread beyond software design, and benefits any technical infrastructure that is internally-complex and/or part of a larger, constantly-changing ecosystem. Given that bibliographic vocabularies used outside traditional centralized distribution are, by nature, highly interdependent, they stand to benefit greatly from semantic versioning.

Dunsire, Hillmann & Phipps (2014) note that structural vocabularies such as ISBD and value vocabularies such as the UniMARC code list are increasingly published in Web-friendly formats, but the practices around their maintenance and support have lagged. That is to say, an older pre-Web workflow remains in place, including centralized, human-centered quality assurance, publishing, and versioning. In order to be sustainable on the Web, these workflows need updating to be as distributed and machine-actionable as possible. For this reason, and following the lead of the software developer community, maintainers of the [RDA Registry](#) have implemented semantic versioning where changes at the vocabulary or property level can be flagged and acted upon by machine or human agents (Versions and Releases, 2014).

One of the challenges in managing this kind of distributed content is moving from a traditional "filter, then publish" to "publish, then filter". The publish-then-filter model allows any authorized person to add new information, after which it will be discussed, edited, enhanced. In contrast, a contributor to certain Library of Congress/PCC value vocabularies, for example, must undertake [extensive training](#) before being allowed to submit proposals for new terms, and then wait for the term to be approved and forwarded to the various distribution nodes. This process affords a high measure of quality assurance, but does not scale well to a fast-moving global information space such as the Web. In 2016, the Program for Cooperative Cataloging (PCC) established a "[Task Group on Identity Management in NACO](#)," in part to address this challenge, for example to study ways to contribute terms "at all states of completeness, so that the intellectual effort expended in baseline work is shared as a foundation on which other institutions can build."

---

## 6 Further Thoughts on Supporting Maintenance: Tools, Community, and Funding

Dean Kolkman (2016) describes a process of "embedding" in efforts to implement computer modeling tools in government agencies. It turned out that the technical challenge was less formidable than the organizational one, i.e., "In order for computer models to be effective in informing the policy making process, they have to be understood, trusted, and relevant to experts and non-experts alike." This is familiar from "[Diffusion of Innovations](#)" theory, that in order to endure, innovations must be perceived as "consistent with the values, past experiences, and needs of potential adopters," available for experimentation, and easy to use. Vocabulary innovation is facing a similar set of challenges, as will be shown below.

Those who maintain bibliographic vocabularies may be expert at metadata analysis but not necessarily the intricacies of data processing and encoding, much less systems design. There are still relatively few tools that provide a simple interface, without forcing users to wrangle with command line utilities and XML or JSON serializations. As Kolkman suggests, there may be a deficit in education, outreach, and/or community-building, which inhibits the adoption of the most promising vocabulary management tools. Technologists are not always accustomed to dealing with maintainers, but they need to work with those who fully understand the data, and not insist that maintainers understand the full technology stack.

In the case of bibliographic vocabularies, there is sometimes a tendency toward wishful thinking, that the mere presence of Linked Open Data will solve the challenges of sustainability, scalability, and quality control; that if one only had more RDF instance data and more controlled vocabularies, a stable ecosystem would emerge of its own accord. Since that emergence has not yet happened, we suggest that significant attention must be paid to the realities of how to engage user communities to develop methods to embed these vocabularies into existing and future community policies and service commitments.

GitHub provides an interesting model for community building, as it layers a user-friendly interface over the Git distributed version control system, lowering the barrier to participation. Tools like Git support back-ups, documentation, log files, branching and forking, all of which integrate maintenance functions in the daily work of project members. Moreover, issue tracking functionality provides a central place to communicate about specific features or bugs, while also relaying updates to email accounts as desired. GitHub was designed for complex software projects, and includes a significant learning curve and technical overhead, but is increasingly used for documentation and other writing tasks.

Wikipedia, with its open access editing, open data mining, discussion pages, version control, and encyclopedic scope is another model for bibliographic community building. Users link to it and rely on it as if it were permanent infrastructure. However, there is an army of volunteers who keep constant watch and prevent breakages. The results have the \*feel\* of a self-healing organism, where vandalism or propaganda is quickly found, discussed through "talk pages," and corrected. One does not need to endorse Wikipedia as a reference tool in order to learn from it, to understand the dynamics of a relatively decentralized editorial structure, with gradations of privilege and responsibility based on reputation. Continuous versioning, quality assurance, the ability to lock down controversial topics, persistent URIs and open dissemination of data, are major ingredients to its success (Zittrain, [n.d.](#)). Libraries and Wikipedia have much to offer one another. The former are looking at version control, crowd-sourcing, open data, i.e., areas where Wikipedia excels. The latter, in turn, is being enriched with library-curated identity management data (Klein, [2012](#)), and through initiatives like the Remixing Archival Metadata Project (RAMP) (Thompson *et al.*, [2013](#)) and One Librarian One Reference (1Lib1Ref) (Stinson, [2017](#)).

Funding for bibliographic infrastructure and maintenance is a related challenge. In general, money tends to flow to high-profile projects and individuals. Nathan Ensmenger (2016) has pointed out that software debugging may be called "development" or "maintenance," depending on whether it occurs before or after software release. The former gets most of the attention and

resources, but the latter is just as important to sustained success. Declaring a program complete and releasing it into the wild, without a proper maintenance strategy, is a recipe for failure. And yet it happens in software as in vocabularies. Discussion of infrastructure and maintenance must go beyond technology to consider the wider social and cultural context, and ongoing commitments that ensure sustainability. Dean Kolkman (2016) writes about "decay as a status quo." When one recognizes decay as the normal state of systems, maintenance tasks are seen in a new light, as essential, not peripheral, to what makes a thing useful.

In the case of libraries, even if funding is targeted specifically for bibliographic infrastructure, there is a risk that a vocabulary or tool will become "orphaned" once the project is complete. Hillmann (2016) describes her experience working on the National Science Digital Library Registry (now Open Metadata Registry (OMR)), where her group struggled to maintain its operation after National Science Foundation (NSF) funding ended in 2007. She noted, "As far as I know, the OMR is one of the only free general-purpose vocabulary development and maintenance tools that has survived past its initial funding."

## 7 Conclusion

Value and structural vocabularies are increasingly important parts of library services, and play a crucial role on Web as part of the "backbone of trust" (Hanneman & Kett, 2010, p. 2). But the narrative of innovation needs to change, so that activities mostly invisible today are made visible in the future. Standards and best practices being developed by W3C, NISO, DCMI, and others, are raising the profile of sustainable vocabulary efforts. Such activities include versioning, licensing, quality assurance, documentation – roles usually hidden while systems are running smoothly, or even forgotten until point of failure. Maintenance and repair have long been integral to how libraries manage their vocabularies, but practices need to adjust for Web-scale implementation.

The literature on "breakdown, maintenance and repair" and Jackson's notion of "broken world thinking" provide a framework to shift the terms of discourse, and potentially strengthen the way we maintain our network of bibliographic vocabularies. What does broken world thinking mean for vocabularies? Recognition of the fragility of current systems and preparation for inevitable breakdowns; building maintenance functions directly into tools, workflows, and budgets, and including documentation, preservation, and terms of use from the moment projects are conceived.

GitHub and Wikipedia illustrate what it looks like to have maintenance tasks "baked-into" regular operations. Much of this happens without conscious effort, thanks to automated versioning and roving "bots". But what lies below the surface, in Wikipedia for instance, is a kind of social compact: that most participants are acting in good faith, that they are building something of value for the general public, that the Web itself is a public good. There are opportunities to monetize services that make use of the collective knowledge-base, but trying to monetize the knowledge-base itself is self-defeating as potential contributors hit paywalls or become demoralized by advertising. Moreover, the publish-then-filter model avoids the bottleneck of editorial approval, allowing any visitor to find new content, spot errors or omissions, and correct them in real time (if authorized), and become new members of the community.

Diffusion of Innovations theory helps us understand why good ideas often fail to survive. In particular, one must bridge the gap between new technologies and the values, needs, and past experiences of potential adopters. Tools like the RDA Registry show how vocabulary and metadata tools can be embedded in a community of practice for the long term. There are hints of growing collaboration between [LC/PCC](#) and other identity management communities like [ORCID](#), [ISNI](#), and [SNAC](#), which point the way toward lower barriers to participation and wider network of maintainers for those systems. But the continuing tendency to ignore maintenance when conceiving new projects or writing new grant applications promises to keep us in a broken world for some time to come. We can go beyond (though not leave behind) the never-ending cycle of breakdown, maintenance, and repair, but it requires that we take very seriously the need to invest in the community of maintainers and tools that keep bibliographic vocabularies available and sustainable over time.

## References

- [1] Baker, T., Vandenbussche, P.-Y., & Vatan, B. (2013). Requirements for vocabulary preservation and governance. *Library Hi Tech*, 31(4), 657-668. <https://doi.org/10.1108/LHT-03-2013-0027>
- [2] Berners-Lee, T. (1989, March). [Information Management: A Proposal](#).
- [3] Berrueta, D., & Phipps, J. (2008, August 28). [Best Practice Recipes for Publishing RDF Vocabularies](#).
- [4] Coyle, K. (2010a). 1. Library Data in a Modern Context. Chapter 1 of Understanding the Semantic Web: Bibliographic Data and Metadata. *Library Technology Reports*, 46(1), 5-13.
- [5] Coyle, K. (2010b). 2. Changing the Nature of Library Data. Chapter 2 of Understanding the Semantic Web: Bibliographic Data and Metadata. *Library Technology Reports*, 46(1), 14-29.

- [6] [Data on the Web Best Practices: W3C Candidate Recommendation](#) (30 August 2016).
- [7] DCMI Announce. (2014). [DCMI Webinar: Eric Miller on libraries and speaking the language of the Web from DCMI Announce on 2014-12-30](#) (public-sweo-ig@w3.org from December 2014).
- [8] de hÓra, B. (2007, April 8). [Vocabulary Design and Integration](#).
- [9] Denis, J. (2016). [Investigating maintenance and repair – Carnet de recherche](#).
- [10] Doctorow, C. (2001, August 1). [Metacrap](#).
- [11] Dunsire, G., Hillmann, D., & Phipps, J. (2012). Reconsidering Universal Bibliographic Control in Light of the Semantic Web. *Journal of Library Metadata*, 12(2/3), 164-176.
- [12] Ensmenger, N. (2016). [When Good Software Goes Bad: The Surprising Durability of an Ephemeral Technology](#). Presented at the The Maintainers, Stevens Institute of Technology.
- [13] [ERIC – Education Resources Information Center](#). (n.d.).
- [14] Fidler, B. (n.d.). [The Dependence of Cyberspace](#). Presented at the Maintainers.
- [15] Hannemann, J., & Kett, J. (2010). [Linked Data for Libraries](#). In *Linked Data for Libraries*. Gothenburg, Sweden.
- [16] Harper, C. A., & Tillett, B. B. (2007). Library of congress controlled vocabularies and their application to the Semantic Web. *Cataloging and Classification Quarterly*, 43(3-4), 47-68. [https://doi.org/10.1300/J104v43n03\\_04](https://doi.org/10.1300/J104v43n03_04)
- [17] Heath, T., & Bizer, C. (2011). Linked Data: Evolving the Web into a Global Data Space. *Synthesis Lectures on the Semantic Web: Theory and Technology*, 1(1), 1-136. <https://doi.org/10.2200/S00334ED1V01Y201102WBE001>
- [18] Hillmann, D. (2016, May 5). [The Conundrum of Research Funding](#).
- [19] IFLA Study Group. (2009). [Functional Requirements for Bibliographic Records](#).
- [20] Jackson, S. J. (2014). [Rethinking Repair](#). In *Media Technologies: Essays on Communication, Materiality and Society* (pp. 222-260). Cambridge, Mass: MIT Press.
- [21] Klein, M. (2012, August 3). [Authority control integration proposal/RFC](#).
- [22] Kolkman, D. (2016, June 3). [Maintenance in progress?](#) [blog post].
- [23] Library of Congress. (n.d.). [BIBFRAME Frequently Asked Questions](#).
- [24] [Linked Open Vocabularies \(LOV\)](#). (n.d.).
- [25] [Maintainers Conference: 2016 Program](#). (n.d.).
- [26] [Maintainers Blog](#). (n.d.).
- [27] Powell, A., Nilsson, M., Naeve, A., & Johnston, P. (2005, March 7). [DCMI Abstract Model](#).
- [28] Powell, A., Nilsson, M., Naeve, A., Johnston, P., & Baker, T. (2007, June 4). [DCMI Abstract Model](#).
- [29] Program for Cooperative Cataloging. (2016, March 31). [Charge for PCC Task Group on Identity Management in NACO](#).
- [30] [RAMP Editor](#). (n.d.).
- [31] [RDA Registry](#). (n.d.).
- [32] [RDA Toolkit](#). (n.d.).
- [33] Riley, J., & Becker, D. (2010). [Seeing Standards: A Visualization of the Metadata Universe](#).
- [34] Riva, P., & Žumer, M. (2015). [Introducing the FRBR Library Reference Model](#).
- [35] Rochkind, J. (2015 Nov. 23). [Linked Data Caution](#). [blog post].

- [36] Russell, A., & Vinsel, L. (n.d.). [Innovation is overvalued. Maintenance often matters more.](#) Aeon.
- [37] Shirky, C. (n.d.). [Ontology is Overrated – Categories, Links, and Tags.](#)
- [38] Stinson, A. (2017, January 15). [Librarians offer the gift of a footnote to celebrate Wikipedia's birthday.](#)
- [39] Tennant, R. (2002, October 15). [MARC Must Die.](#)
- [40] Thompson, T. A., Little, J., González, D., Darby, A., & Carruthers, M. (2013). [From Finding Aids to Wiki Pages: Remixing Archival Metadata with RAMP.](#) *The Code4Lib Journal*, (22).
- [41] Van Malssen, K. (2014). [BIBFRAME AV Modeling Study: Defining a Flexible Model for Description of Audiovisual Resources.](#) AVPreserve.
- [42] Versions and Releases. (2014, December 29). [RDA Registry.](#)
- [43] Vinsel, L. (2016). [The Stories We Tell, or Mary Poppins, Maintainer.](#) In *The Maintainers*. Stevens Institute of Technology.
- [44] W3C. (1999, February 22). [Resource Description Framework \(RDF\) Model and Syntax Specification.](#)
- [45] Weinberger, D. (2007). *Everything Is Miscellaneous: The Power of the New Digital Disorder* (First Edition edition). New York: Times Books.
- [46] Zittrain, J. (2015). [Why Wikipedia Works Really Well in Practice, Just Not in Theory – Video.](#)
- 

## About the Authors

**Daniel Lovins** is Head of Knowledge Access Design & Development at New York University (NYU) Division of Libraries. Before coming to NYU (in October 2011) he served as Yale University's Metadata & Emerging Technologies librarian, before which he was Yale's Hebraica Catalog Librarian. His research interests include resource discovery systems, multilingual indexing, open-source software development, agile project management, and linked library data. Among his current professional activities, he is co-chair of the NISO Bibliographic Roadmap working group on vocabulary use and reuse. In 2016 he published a book chapter entitled "From User Stories to Working Code: A Case Study from NYU's Digital Collections Discovery Initiative", in Ken Varnum (ed.) *Exploring Discovery: The Front Door to Your Library's Licensed and Digitized Content* (ALA Editions).

---

**Diane Hillmann** is currently a principal in the consulting firm Metadata Management Associates LLC. From 1977 to 2008 she was associated with Cornell University Library, as a law cataloger, technical services manager, and manager of authorities and maintenance processes for the Cornell Library's MARC database. She also participated in the Cornell portion of the [National Science Digital Library](#) Core Infrastructure as Director of Library Services and Operations between 2000-2005. Ms. Hillmann was a liaison to and member of MARBI from the late 1980's to 2000, specializing in the Holdings and Authorities formats, which led to her early participation in the Dublin Core Metadata Initiative. She is currently a member of the DCMI Advisory Board, was Vocabulary Maintenance Officer for DCMI from 2011-2013, and was co-Program Chair for the DC-2010 and DC-2011 conferences in Pittsburgh and The Hague. Diane edited (with Elaine Westbrook) *Metadata in Practice*, published by ALA Editions (2004) and she publishes frequently on digital library and metadata issues, particularly in the [Metadata Matters](#) blog.

---

(On March 22, 2017, several minor typographical errors appearing in this article were corrected at the request of the authors.)

Copyright © 2017 Daniel Lovins and Diane Hillmann