Metadata & Discovery Group

Book review: Ethics in Linked Data

Reviewed by: **Elizabeth Cooper**
University of Bristol

Received: 5 August 2025 | Published: 22 September 2025

Watson, B. M., Provo, Alexandra and Burlingame, Kathleen (eds) (2023) *Ethics in Linked Data*. Sacramento, CA: Library Juice Press. ISBN 978-1-63400-133-5

CONTACT Elizabeth Cooper liz.cooper@bristol.ac.uk University of Bristol

Ethics in Linked Data, edited by B.M. Watson, Alexandra Provo, and Kathleen Burlingame, and conceived by the LD4 Ethics in Linked Data Affinity Group, is an incredibly engaging, thought-provoking and ambitious work. Contributing authors present a wide range of ethical linked data project case studies from across libraries, museums and archives in North America.

There is a consensus among contributing authors that an ethical framework should be the foundation and guiding principle of any linked data project, from its technological underpinnings through to its conceptual models, vocabularies, workflows, and overall practice. This book positions linked data as a vehicle for bringing the work of marginalised communities and creators to a wider, worldwide audience, through increasing the visibility, discoverability, and connectivity of creators and works. Use cases to achieve this aim include converting datasets to linked open data; repurposing traditional metadata into linked data ontologies; creating linked open data from scratch to render a collection of works more visible; and enhancing existing authority files with linked data entities. There is an emphasis on collaborating with the communities represented, particularly with living creators: gaining consent to create or re-use legacy metadata; returning agency and sovereignty to communities through collaborative linked data workflows. Other common threads are working slowly and carefully to create accurate metadata that adheres to ethical policies and procedures; and working at local level with critical reflection and feedback mechanisms in place at the start. Overall, this is a book about creating a community of practice with ethical consideration at its forefront.

Ethics in Linked Data is divided into three main sections: the first applies ethical theories to linked data ontologies and schemas; the second challenges the legacy of hegemonic power structures inherent in the knowledge organisation systems that inform linked data frameworks and workflows; and the third examines the outdated naming and representation of identity in traditional authority files and controlled vocabularies that get perpetuated in linked data entities.

In the first section, notable chapters look at the development of linked data technologies over the decades by wealthy organisations in an inequitable way, emphasising that linked data technology must lower barriers to entry and place end-users at the centre of its aims if it's to have relevant application in cash-strapped GLAM institutions. James Kalwara and Erik Radio challenge the very structure of RDF triples, arguing that this syntactical structure is based on Western language syntax (subject/object/predicate), an expression of the dominant languages of the Global North. They argue that RDF syntax excludes participants from nations whose language and ways of knowing/organising data are radically different. The authors' Marxist critique of RDF triple stores made for interesting reading but did not really add much weight to their argument. Practical recommendations to alter the already-established RDF structure, other than examining linked data software with a critical eye, were not forthcoming.

The second section looks at historic dominant power structures and institutions, including the Library of Congress, and the continuing prevalence of the white, male, Christian, cisgender worldview in our knowledge organisation systems. Chapters document reparative work undertaken through linked open data projects to redress harm done, particularly through colonialism.

Many indigenous cultures were erased through the collection (and stealing) of cultural artefacts by institutions linked to the Library of Congress in the eighteenth and nineteenth centuries. Systematic oppression eradicated indigenous culture as it sought to replicate national heritage collections like the anthropological collections of museums in Europe. This practice imposed inaccurate or false descriptions and narratives on Native American culture, artefacts, and history. Authors posit that the harmful practice of collecting artefacts continues in large-scale metadata harvesting via APIs and completionist practices, perpetuating harm. Outdated descriptions and inappropriate naming also affect findability. The conflation of indigenous identity and nationhood with land in Wikidata means that tribes not associated with a specific geographical area, such as the Metis community, have remained invisible. Authors highlight that these injustices in linked data need to be addressed at a local level with community collaboration and careful metadata practices.

Another aspect of identity misrepresentation occurs through the problematic assignment of gender to person entities in Wikidata, particularly for people identifying as non-binary or trans. Daniele Metilli and Chiara Paolini have undertaken the first qualitative and quantitate study of gender modelling in Wikidata since its inception in 2012. Their findings highlight that the Wikidata knowledgebase is far from inclusive of non-binary identities. This issue is sometimes amplified in data augmentation by bots and by the 'anyone can edit anything at any time' (AAA) ethos of Wikidata (with its overreliance on assumed goodwill of Wikidata editors). There is still much work to be done to ensure identities are described accurately and fairly in open knowledgebases like Wikidata.

The third section expands on the second section focusing on case studies looking at creating equitable entries in authority files, linking traditional authority files to more flexible diachronous linked data entities, such as those in Wikidata. Traditional authority files centre on single author descriptors (such as dates), presented as facts, as a means of disambiguation. In contrast, Wikidata items and properties lend themselves to connecting a multitude of descriptors to a single data entry point to reflect multiple identities. In a chapter entitled “The Oklahoma Native Artists Project : Oral History to Linked Open Data”, Megan Machen et al detail a project to increase the discoverability of work of living Native American artists in Oklahoma through recording interviews and encouraging self-representation, with linked open data as the end product. The emphasis is placed on getting consent to showcase work and artists, working collaboratively with the artists to create accurate metadata. In addition, linked open data becomes a vehicle for moving away from the dominance of metadata describing printed works, to metadata for works created in different media.

Some of the early chapters examine linked data through theories such as pragmatic sociology, reflective/diachronous language theory and Marxism. Whilst these provide an interesting angle to ethical consideration, they seem based on the academic background of the writers and are not always useful paradigms. While this book highlights injustice and asks challenging questions about inclusivity and equity in linked data practices, authors are not always able to suggest clear solutions. It should be noted that this book is primarily based on linked data developments in North America - where the richer institutions have had funding for linked data projects and infrastructure - itself an inequality in the world of linked data (which authors acknowledge).

Overall, *Ethics in Linked Data* provides much food for thought. It should be considered key reading for any institution embarking on a small-scale linked data project, with stand-alone chapters and an appendix containing an extremely extensive ethics toolkit. Whilst this book assumes a basic understanding of linked data, its chapters would be accessible for the uninitiated. Key concepts and theories are clearly presented and provide critical reflection on what could have been approached differently. The focus is on open knowledgebases such as Wikidata and locally created ontological frameworks, with little mention of the library-focused BIBFRAME ontology that tends to be maintained by large library management system vendors. Case studies are detailed, with descriptions of decision-making and practical suggestions that could be adapted to any local linked data project. Wikidata lends itself to small projects to showcase local collections. Ethical considerations presented would also have application in other areas within GLAM institutions.

Ethics in Linked Data is a tool for learning more about ethics, for challenging biases and cultural assumptions, as well as extending knowledge about the history of oppressive colonial power structures and how those continue to have an impact on our knowledge organisation and vocabularies today. What becomes clear throughout this book is that ethical decision-making in linked data often needs to be made at local level

and with care; metadata quality needs to be maintained in a way that is sustainable; and a collaborative approach between linked data practitioners and creators of works implemented where possible. Above all linked data practice needs to be an iterative process as meaning and vocabularies are always historically situated and constantly evolve over time.