# Migrating special collections to linked open data environments

Dilnavozxon Iqboljon-qizi Soliyeva
Scientific advisor: Rashid Turgunbayev
Kokand State University

***Abstract:*** *The migration of special collections to linked open data environments represents a paradigm shift that extends far beyond technical metadata conversion, touching upon the epistemological foundations, professional practices, and ethical commitments of research libraries. This article examines the intellectual, practical, and organisational dimensions of transitioning rare books, manuscripts, archives, and other unique cultural heritage materials from traditional siloed catalogues to the interconnected semantic web of linked open data. It argues that the case for such migration rests upon three interrelated pillars: the enhancement of scholarly utility through large-scale aggregation and computational analysis, the preservation and expression of archival complexity through rich relational modelling, and the advancement of accessibility and social justice through open, reusable, and community-enrichable data. The article critically engages with the significant obstacles that libraries confront in this undertaking, including the heterogeneity of legacy metadata, the demands of long-term sustainability and persistent identification, and the profound human challenges of professional culture, cross-training, and interdisciplinary collaboration. Drawing upon exemplars from major national libraries and smaller institutional projects, the discussion advocates for a phased, iterative, and community-engaged approach that prioritises careful data modelling, user-centred evaluation, and incremental scaling. Ultimately, the article contends that linked open data migration is not merely a technical upgrade but an act of intellectual hospitality that transforms special collections from guarded repositories into participatory nodes within a global knowledge network, ensuring their continued relevance and vitality in an era of computational, networked, and democratised scholarship.*
***Keywords:*** *linked open data, special collections, semantic web, cultural heritage metadata, digital preservation, archival description*

The migration of special collections to linked open data environments represents one of the most intellectually demanding and professionally transformative undertakings currently facing research libraries. It is not, despite initial appearances, a mere technical exercise in format conversion or metadata upgrading. Rather, it is a fundamental reimagining of what a special collection is, what it does, and whom it serves. For generations, special collections have been understood as treasured enclaves within larger library systems, physical or digital spaces where rare books, manuscripts, archives, maps, photographs, and ephemera are preserved with exceptional care and consulted under controlled conditions. Their value has been traditionally measured in terms of uniqueness, material authenticity, provenance, and the scholarly mystique that attaches to direct contact with primary sources. Yet the digital age has unsettled every one of these assumptions, not by diminishing the importance of rarity or authenticity, but by exposing the deep contradictions between the curated isolation of special collections and the networked, participatory, and computationally driven modes of knowledge production that now dominate academic research. Linked open data, or LOD, offers a way out of this contradiction, but only if librarians are willing to embrace a level of semantic rigor, infrastructural investment, and collaborative openness that challenges the very identity of special collections librarianship.

At its core, linked open data is a set of principles and technologies for publishing structured data on the web in such a way that it can be interlinked and queried across disparate systems. The foundational tenets, often summarized as the four rules of linked data, require that things be identified by HTTP URIs, that those URIs be dereferenceable to provide useful information, that the information returned include links to other URIs, and that the data be made openly available without restrictive licensing. When applied to special collections, these principles transform the catalogue record, the finding aid, the digitised image, and the authority file from isolated silos into interconnected nodes within a global web of cultural heritage knowledge. A medieval manuscript in a university library in Oxford ceases to be merely a local asset described in a local system; it becomes a resource that can be semantically linked to the same text in a Parisian archive, to the modern edition in a digital repository, to the biographical data of its scribe, to the iconographic database of its illuminations, and to the geographical gazetteer of its place of origin. This is not simply enhanced discoverability; it is a new epistemology of the archival object, one in which meaning emerges from the network of relationships rather than from the isolated description.

The intellectual case for migrating special collections to LOD is compelling, yet it rests on several distinct arguments that must be carefully distinguished. The first is the argument from scholarly utility. Contemporary humanities scholarship, particularly in fields such as digital history, computational literary studies, and network analysis, increasingly depends on the ability to aggregate, compare, and analyse cultural heritage data at scale. A researcher studying the transmission of classical texts across medieval Europe does not want to search fifty separate catalogues using fifty different vocabularies; they want a single query that returns all known manuscript witnesses, their dates, their locations, their textual variants, and their digital surrogates, ideally in a machine-readable format that can be fed directly into their analytical toolkit. Linked open data makes this possible by providing a common semantic framework, typically expressed in RDF, the Resource Description Framework, and by reusing well-established ontologies such as the CIDOC Conceptual Reference Model for cultural heritage, the Functional Requirements for Bibliographic Records, or the Europeana Data Model. When special collections are published as LOD, they cease to be obstacles to large-scale research and become instead its primary fuel.

The second argument is the argument from archival integrity and contextual richness. Traditional metadata, whether encoded in MARC, EAD, or Dublin Core, tends to flatten the multidimensional complexity of archival objects into linear, text-based descriptions. A single manuscript might be described in terms of its title, author, date, extent, and subject headings, but these fields cannot easily capture the dense web of agents, events, places, and works that constitute its true historical significance. Linked open data, by contrast, excels at representing complex relationships. It can express that a particular codex was commissioned by one patron, copied by a scribe in a specific scriptorium, annotated by a later reader, owned by a sixteenth-century humanist, bequeathed to a college library, rebound in the nineteenth century, and digitised in the twenty-first, with each of these entities and acts given its own URI and linked to external authority files. This does not replace the narrative richness of a scholarly catalogue; it complements it by providing a formal, machine-actionable graph that can support visualisation, querying, and inference. Moreover, because LOD encourages the reuse of external vocabularies and identifiers, such as VIAF for persons, Getty TGN for places, or Wikidata for virtually anything, the collection becomes embedded in a wider ecosystem of shared knowledge, reducing local cataloguing burdens and increasing semantic interoperability.

The third argument, and perhaps the most ethically significant, is the argument from accessibility and social justice. Special collections have long been criticised for their exclusivity. Their physical

security requirements, their reading room protocols, their often intimidating aura of scholarly privilege, and their historical accumulation through colonial, aristocratic, or otherwise unequal channels have made them symbols of epistemic gatekeeping. Digitisation has alleviated some of these barriers, but digital surrogates are frequently locked behind proprietary platforms, restricted by copyright overreach, or described in language that presupposes advanced subject knowledge. Linked open data, when implemented with genuine commitment to openness, reverses this logic. It makes the underlying descriptive data freely available for anyone to download, reuse, and build upon. It enables multilingual access through alignment with multilingual thesauri. It allows community-driven enrichment, so that descendant communities, citizen historians, and independent researchers can contribute their own knowledge, corrections, and interpretations directly to the semantic graph. In this sense, LOD migration is not merely a technical upgrade but a reparative practice, a way of acknowledging that special collections are not private treasures but public goods, and that their full value can only be realised when they are woven into the open web of shared human memory.

Nevertheless, the path from traditional collection management to linked open data publishing is strewn with practical, organisational, and conceptual obstacles that no library should underestimate. The first and most obvious challenge is the sheer heterogeneity of special collections metadata. A typical research library might hold manuscripts described in hand-written finding aids, early printed books catalogued in legacy MARC records, born-digital archives with minimal provenance documentation, and digitised images with technical metadata in METS, all of which reside in separate databases managed by different units. Converting this polyglot array into a coherent RDF graph requires not only technical proficiency in semantic web technologies but also deep curatorial judgment about what entities to define, what relationships to assert, and what ontologies to employ. There is no single correct way to model a collection, and the choices made at the outset have profound consequences for how the data will be queried, linked, and reused. Libraries that rush into LOD migration without thorough data modelling, stakeholder consultation, and user needs analysis often end up with graphs that are technically valid but intellectually thin, offering many triples but little insight.

The second challenge concerns sustainability. Linked open data is not a one-time export but a living resource that must be maintained, updated, and versioned over time. As new acquisitions are added, as scholarly research refines attributions and dates, as digital surrogates are improved, and as external authority files evolve, the LOD graph must keep pace. This demands ongoing staffing, computational infrastructure, and governance policies that many libraries, particularly those with shrinking budgets, struggle to provide. Moreover, the very openness that makes LOD so attractive also creates dependencies on external resources that may change or disappear. If a library links its manuscript scribes to a name authority that is later deprecated or reorganised, those links may break or become misleading. If a geospatial gazetteer changes its coordinates, the spatial queries built upon them may yield erroneous results. Responsible LOD publishing therefore requires not only technical skill but also a robust approach to persistent identification, provenance tracking, and long-term preservation of the semantic data itself, a requirement that adds another layer of complexity to already overburdened digital preservation programmes.

The third challenge, and one that is frequently overlooked in technical discussions, is the human factor. Migrating special collections to linked open data is not primarily a problem of software or syntax; it is a problem of professional culture, training, and collaboration. Many special collections librarians have spent their careers mastering the intricacies of descriptive bibliography, palaeography, codicology, and archival theory, but have had little exposure to semantic web technologies, SPARQL

querying, or ontology engineering. Conversely, many digital librarians and data scientists possess the technical skills but lack the subject expertise to make informed modelling decisions. Bridging this gap requires sustained investment in cross-training, interdisciplinary teamwork, and, perhaps most importantly, a shared vocabulary that allows curators and coders to communicate effectively. This is not achieved through a single workshop or a one-year pilot project; it demands a fundamental shift in how library schools educate future professionals and how research libraries structure their organisational charts. It also demands patience, because the benefits of LOD migration are often diffuse and long-term, while the costs are immediate and tangible, making it difficult to sustain institutional commitment through leadership changes and budget cycles.

Despite these challenges, a growing number of libraries have demonstrated that LOD migration is not only feasible but genuinely transformative when approached with realism and humility. The British Library, for example, has published its collection metadata as LOD for years, linking its holdings to external datasets such as the Virtual International Authority File and the Library of Congress Subject Headings, while also experimenting with more ambitious projects that connect manuscripts to historical place data and biographical resources. The Bibliothèque nationale de France has undertaken similarly extensive work with its data.bnf.fr platform, which aggregates catalogue data, digitised content, and external references into a rich semantic portal. At a smaller scale, many university libraries have developed targeted LOD projects for specific collections, such as papyrological archives, incunabula, or personal papers of notable literary figures, often in collaboration with digital humanities centres and research consortia. These projects, while modest in scope, provide invaluable lessons in data modelling, user engagement, and workflow redesign that can inform larger institutional strategies.

For a library contemplating its own migration, a phased and iterative approach is almost always superior to a grand, all-at-once conversion. The first phase should be diagnostic: an audit of existing metadata assets, an assessment of user needs and research use cases, and a selection of one or two representative collections to serve as testbeds. The second phase should be experimental: the creation of a small but well-modelled LOD graph for the test collections, using established ontologies wherever possible and documenting all modelling decisions thoroughly. This phase should also include active engagement with potential users, including scholars, students, and external developers, to evaluate whether the LOD representation actually supports the kinds of queries and analyses they wish to perform. The third phase should be iterative: refining the data model based on feedback, scaling up to additional collections, and developing sustainable workflows for ongoing maintenance. Throughout all phases, libraries should participate actively in the broader LOD community, sharing their models, their vocabularies, and their lessons learned, because the success of linked open data depends not on any single institution but on the collective intelligence of the network.

The future of special collections in linked open data environments is not a fixed destination but an ongoing horizon. As artificial intelligence and machine learning tools become more sophisticated, the semantic graphs produced by libraries will serve as training data, as validation sets, and as inference substrates for computational analysis. As the web itself evolves toward more structured and agent-mediated forms, the ability of special collections to participate in that semantic web will determine their visibility and relevance for generations of researchers yet unborn. But this future will not arrive automatically. It requires deliberate choices, sustained investment, and a willingness to question long-held professional orthodoxies about the nature of cataloguing, the boundaries of the collection, and the relationship between libraries and their publics. The migration to linked open data is, in the final analysis, an act of intellectual hospitality, a way of saying that the rare and the precious do not become

less so when they are shared, but rather more so, because their meaning expands with every new connection, every new query, and every new interpreter. For the special collections librarian, there is no more urgent or more honourable task than to facilitate that expansion, to transform the archive from a fortress into a crossroads, and to ensure that the treasures of the past are not merely preserved but actively, openly, and perpetually brought into dialogue with the questions of the present. That is the true promise of linked open data, and it is well worth the cost of the journey.

## References

1. Yo‘Ldasheva, S. (2023). AXBOROT ASRIDA ZAMONAVIY KUTUBXONALARNING O ‘RNI. Oriental Art and Culture, 4(6), 268-274.
2. Yuldasheva, S. (2025). BIBLIOGRAFIYA FANNI O ‘QITISH METODIKASINING MAQSAD VA VAZIFALARI. World of Philology, 4(4), 37-47.
3. Yuldasheva, S. (2023). The Experience of Learning to Read from Foreign Countries. European Journal of Innovation in Nonformal Education, 3(10), 42-46.
4. Manzirova, M., & Tuychiyeva, S. (2023). KUTUBXONACHI KUTUBXONA-AXBOROT JARAYONINING YETAKCHI ISHTIROKCHISI SIFATIDA. Oriental Art and Culture, 4(1), 62-65.
5. To‘Ychiyeva, S. (2023). O ‘ZBEKISTONDA ELEKTRON KUTUBXONALAR VA XALQARO HAMKORLIKNING RIVOJLANISH SAMARALARI. Oriental Art and Culture, 4(6), 38-43.
6. Tuychieva, S. (2023). The Impact OF the National Process OF Ascension on the Socio-Spiritual Development OF Society. European Journal of Innovation in Nonformal Education, 3(12), 1-3.
7. To‘Ychiyeva, S. N. M. (2022). OILADA BOLANI KITOBGA QIZIQTIRISHNING YO ‘LIARI VA VOSITALARI. Oriental Art and Culture, 3(2), 367-372.
8. Туйчиева, Д. (2022). О ЛИТЕРАТУРНОЙ ЭТИКЕ АМИРА ТЕМУРА. Oriental Art and Culture, 3(4), 810-816.