Foundations for the future, the British Library’s Collection Metadata Strategy was published in April 2019. *Foundations for the future* sets out the Library’s metadata strategy to 2023.

“Collection metadata is an umbrella term for the structured information required to efficiently manage, access and preserve the collection. It is a key organisational asset, representing centuries of resource investment. Efficient exploitation and stewardship of collection metadata underpins delivery of The British Library’s core purposes, strategies and portfolios.”

“Our vision is that by 2023 the Library’s collection metadata assets will be unified on a single, sustainable, standards-based infrastructure offering improved options for access, collaboration and open reuse”

Strategic activities will be guided by the following priorities:

- Enhance the Library’s ability to exploit its collection metadata assets in order to deliver its strategic priorities, portfolios and programmes
- Ensure that the value of the Library’s Collection Metadata investment is maintained through effective stewardship
- Enable open access to collection metadata in order to improve resource discovery and promote wider community reuse

The period covered will see the Library facing several key metadata management challenges.

- Unifying Collection Metadata Infrastructure
- Maintaining and Developing the Value of Collection Metadata Assets
- Delivering Efficient and Sustainable Collection Metadata Processes
- Collection Metadata for Preservation, Rights and Management Information
- Spreading and Maintaining Best Practice

*Foundations for the future* combines continuity of purpose with our previous strategy, *Unlocking the value*, with a shift of focus onto unification of the Library’s metadata infrastructure.

Unification of the metadata infrastructure comprehends metadata standards; metadata creation systems, discovery and access. The opportunity arises from the need to replace the Digital Library System (DLS); the Integrated Archives and Manuscripts System (IAMS); the Sound and Moving Image System (SAMI) and the Aleph ILS over the next 4-5 years. New systems will be phased in by the Library Systems Transformation programme (LST).


By maintaining and developing the value of our metadata we ensure it becomes (and remains) fit for contemporary purposes.

The volume and complexity of demands for metadata exceed the capacity of traditional cataloguing processes. We will continue to remove barriers to efficiency and to develop workflows that combine machine processing with manual exception handling.

Traditional assumptions about metadata requirements for preservation and access are not valid for digital content and we have to respond to much more complex requirements.

The success of the Strategy will rely on effective communication to stakeholders and adoption and maintenance of best practices by Library staff.

**Strategic Priorities 2019-23**

**Enhance the Library’s ability to exploit its collection metadata assets in order to deliver its strategic priorities, portfolios and programmes by:**

- Managing migration to a new unified infrastructure for collection metadata management to reduce operational costs and complexity
- Developing and maintaining innovative collection metadata creation and enhancement processes to support efficient exploitation of third party data sources e.g. automated entity extraction from full text, crowdsourcing and bulk record enhancement.
- Designing and supporting accurate and flexible rights management metadata for both content and descriptions to enable the Library to take advantage of new licensing options and increase reuse of the collections
- Providing colleagues with self-service tools, expert assistance and best practice guidance to improve consistency and efficiency in the exploitation of collection metadata and integrate with established training initiatives
- Ensuring requirements for new management information tools utilising collection metadata are included in LST procurement specifications to improve accurate assessment of collection strengths and weaknesses

**Ensure that the value of the Library’s Collection Metadata investment is maintained through effective stewardship by**

- Collaborating with international, cross sectoral partners to ensure appropriate collection metadata standards are developed and implemented to support operational efficiency and preserve long-term value, e.g. creating opportunities to consume high-quality metadata from external sources
- Ensuring collection metadata accurately and persistently incorporates all necessary information to deliver the Library’s portfolios and wider strategic objectives
- Developing processes to prioritise collection metadata enhancements that support delivery of improved services to internal and external users
- Managing long-term risk to the integrity of collection metadata against short-term efficiencies
- Undertaking collection metadata maintenance, enhancement and QA processes to support increased operational efficiencies and smooth transition to new systems
- Ensuring licensing of third party collection metadata is negotiated with a long-term view
Enable open access to collection metadata in order to improve resource discovery and promote wider community reuse by:

- Ensuring suitable collection metadata is available to enable and support new shared service initiatives, open research and commercial collaborations
- Increasing opportunities for interaction with our collection metadata to show the richness of the collection and its relevance to all
- Ensuring our collection metadata is accurately represented and accessible via all relevant national or global discovery channels and platforms
- Exploring opportunities for community engagement in metadata enhancement, e.g. via development and promotion of crowdsourcing challenges
- Taking steps to expose any collection metadata assets currently unavailable for resource discovery, partnership or collection management
- Ensuring comprehensive, accurate and timely rights metadata is available to support the sharing or purchase of any digital content
- Making sure our collection metadata is openly available in full compliance with current public sector best practice and related legislation

Annual Implementation Plan

Delivery of the strategy is tracked through a series of annual plans.

Fig. 1 Road map Collection Metadata Strategy 2019-23
Priorities for 2019-20 include:

- Metadata assessment, migration and configuration activities for the Digital Asset Management and Preservation System (DAMPS) and other LST Project strands
- A single target metadata model for use on next generation systems will be developed in consultation with stakeholders.
- Activities supporting publisher circulation of good quality metadata via trade supply chains will be investigated and piloted (e.g. use of the ISNI)
- Collection Metadata for Library holdings will be available via the new JISC National Bibliographic Knowledgebase

A graphical representation of the implementation plan has been published on the Library’s website.³

**Unifying Collection Metadata Infrastructure**

The collection metadata infrastructure includes systems, metadata and business processes.

> **Fig. 2 the metadata silos**

Separate workflows have developed into silos that cut across generic functions. The resulting infrastructure is increasingly difficult to maintain. The complexity and number of metadata transformations required are barriers to efficiency. These silos extend into the discovery layer, providing a fragmented view of the Library’s collection.

Unification will deliver a flexible infrastructure in which metadata is transformed on import to the library and on publication to services (the green lines indicate where transitions are required). No systems are illustrated on the diagram because the number and scope of replacement systems required is dependent on the outcome of future procurement exercises. The Library is currently working with Libnova to implement the Digital Assets Management and Preservation System (DAMPS). A proof of concept to migrate SAMI data to Aleph is planned for autumn, 2019. Requirements gathering to initiate the replacement of IAMS is also scheduled for 2019. It is anticipated that Aleph will be the last of the strategic metadata repositories to be replaced.

Collection Metadata Standards
Unification is expected to simplify and reduce the Library’s standards portfolio. It has been explicitly recognised in the strategy that there is a dependency in some cases on system replacement. For example, we can’t stop using MARC 21 until we replace Aleph.

But what *should* we do about MARC? In addition to MARC 21, we also use an internal variant, SAMIMARC, for our sound and vision resources. Our options will be constrained by the systems available in the marketplace, but the LST is our opportunity to replace MARC as an input format. We will continue to supply MARC data, even if we no longer use it internally. But, how will we replace MARC 21 and SAMIMARC? BIBFRAME is a linked data alternative to MARC, but the BIBFRAME model does not implement the IFLA Library Reference Model (LRM),\(^4\) nor does it implement RDA.\(^5\) In general, it would make more sense to implement a more expressive schema for import and export of less expressive schemata.

---


5. RDA: Resource Description and Access [https://www.rdatoolkit.org](https://www.rdatoolkit.org)
There are instances of competing standards. For example FAST (Faceted Application of Subject Terminology)\(^6\) uses the same vocabulary as LCSH, but with a much simpler syntax. FAST is already in use for some workflows because it is easier and cheaper to implement. The Library has consulted\(^7\) with the community on the feasibility of FAST as a replacement for LCSH. As a consequence of that consultation we have collaborated with OCLC and others to form the FAST Policy and Outreach Committee (FPOC).\(^8\) FPOC will promote FAST and work with OCLC to develop and deliver an efficient service.

In addition to MARC and LCSH, we are reviewing standards in use for cataloguing unpublished resources, including archives and manuscripts. In general, we follow ISAD(G)\(^9\) but specialised workflows are using either TEI\(^10\) or an internal format to catalogue manuscripts. Maintaining three different input methods may provide for richer descriptions but represents a significant overhead. The solution once again, may be to provide a richer input format for all unpublished resources and transform data to Encoded Archival Description (EAD)\(^11\) or TEI, as required, on export.

The Library remains committed to RDA. We are contributing to the implementation of the 3R\(^12\) project, through EURIG,\(^13\) and work is underway to develop supporting documentation and training materials. RDA 3R provides flexible options for how data elements and their provenance are recorded; concepts such as manifestation statement, aggregations and representative expression will prove valuable for integrating legacy data and metadata from other cultural heritage traditions. As an implementation of the Library Reference Model, 3R is aligned with the museum sector’s Conceptual Reference Model (CRM)\(^14\) and RDF (Resource Description Framework).\(^15\) The LRM provides a rich set of entities many of which, such as Agent, Place, Timespan, can be reused independently of RDA.

**Target Metadata Model (TMM)**

The TMM is a framework model that will help us focus on the commonalities between our different metadata standards and models rather than obsess about their differences.

---


10. Text Encoding Initiative (TEI) [https://tei-c.org/](https://tei-c.org/)

11. Encoded Archival Description [https://www.loc.gov/ead](https://www.loc.gov/ead)

12. RDA FAQ 3R Project [http://www.rda-rsc.org/node/551](http://www.rda-rsc.org/node/551)

13. European RDA Interest Group (EURIG) [http://www.rda-rsc.org/europe](http://www.rda-rsc.org/europe)


15. W3C Resource Description Framework [https://www.w3.org/2001/sw/wiki/RDF](https://www.w3.org/2001/sw/wiki/RDF)
The TMM developed from comparison of the Library’s legacy metadata models with the IFLA Library Reference Model (LRM). The analysis identified 3 levels of metadata common to the different models.

- **Level 1**: metadata that puts an orderable unit into context (e.g. authority data)
- **Level 2**: metadata that describes an orderable unit (e.g. bibliographic description)
- **Level 3**: metadata that describes a dependent unit (e.g. analytic)

The TMM has to support comprehensive, coherent discovery of the collection; enable dissemination of collection metadata in the schemata required by stakeholders; and deliver efficiencies, e.g. by reducing the number of systems, workflows, standards and transformation to be maintained.

Figure 4 illustrates a RIMMF3\(^{16}\) “r-tree”\(^{17}\) for Pickwick Papers by Charles Dickens. The “r tree” presents metadata for sound recordings, published books and unpublished papers taken from different silos in a unified RDA structure. Experimentation using RIMMF illustrated that while RDA offers potential for convergence around Levels 2 & 3, RDA is designed to contextualise them in relation to their content. This is not surprising but it is not clear that we could support a collection based view, as required by ISAD(G).

Initial work on the Target Metadata Model has mainly focused on reconciliation of RDA and ISAD(G). The TMM should be expressive enough to enable collection items to be shown in either context. The TMM won’t be RDA or ISAD(G) or TEI or SAMIMARC, although cataloguers in specific workflow will be working with some of those standards.

---


17. An r-tree is a tool for visualising relationships between RDA entities
Figure 5 Illustrates core entities and relationships in the SAMI model; the Library Reference Model, and ISAD (G). Although entities and attributes from different models are known by different names and may be recorded in different ways, they generally identify the same things, e.g. Persons, Items, Works, and Places.

Both LRM and SAMI models make a distinction between content and carrier and there is a relatively straightforward correlation between them. ISAD(G) makes provision for describing the carrier, but it is concerned with the context of items (if described) in archives or collections rather than their intellectual content.
The SAMI data model is very similar to FRBR and can be merged into LRM. The Sound and Moving Image Collection also contains archival textual and graphic materials which may be more efficiently handled using archival standards.

The LRM and ISAD(G) models offer alternative contexts for collection items. An LRM view brings together collection items with the same or related content.
An ISAD(G) view shows the place of a collection item in relation to the rest of the collection and it exposes collections that have not been fully analysed.

**Fig. 9 ISAD(G) view of unpublished resources related to Pickwick Papers**

**Conclusion**

*Foundations for the future* continues the Library’s collection metadata strategy, begun in 2015. In conjunction with the Library Systems Transformation programme and other strategic initiatives, the plan of work is intended to lay the foundations for the new collection metadata infrastructure necessary to realise the British Library’s strategic objectives.