

Alan Danskin

Former Collection Metadata Standards Manager, British Library

## Introduction

My first CIG conference was at Retford, about 30 years ago. I am very honoured to have been invited to give the keynote today.

“[Collection metadata] is a key organisational asset, representing centuries of resource investment”



Figure 1. British Library Collection Metadata Strategy documents

If this presentation has a theme, it is that collection metadata is a key organisational asset for any library. This has been the core message of the British Library's Collection Metadata Strategies from 2015 to the present day. Obviously, metadata is essential for discovery and management of collections and, as was recently acknowledged by George Osborne, uncatalogued collections are not secure.

The title of the presentation refers to my intention that, after a couple of years working as a cataloguer at the British Library, I would leave and return to Scotland to work in a university library. Thirty-six years later I have retired from the British Library. So, what went wrong?

When I joined the British Library in 1987. The British Library itself was still a relatively new institution. It was just 14 years old. The St Pancras site was still shrouded in hoardings and the collections were distributed over at least 10 sites spread across London and at Boston Spa in West Yorkshire.

The Library then was very different from the library today. The management style was very hierarchical. Information was communicated very much on a need-to-know basis. It was generally exclusive. The reference services located at the British Museum were regarded as a library of last resort. It was difficult to get a reader's ticket. Some services, such as Science Reference Information Service (SRIS), were open access. There were conflicts between the free to access public services and the commercial imperatives of the British National Bibliography and Document Supply Centre. The physical separation of staff and collections across the Library's estate meant that old loyalties died hard. My manager, Pat Oddy (second from the left in the photograph below), was inclined to compare the different directorates to mediaeval fiefs ruled by competing robber barons.

The world of cataloguing was also very different. It surprised me that cataloguers didn't do direct data entry. Pat Oddy had a desktop computer for word processing, but cataloguers had fiche reader for routine

consultation of BNB and Precis subject packages. Catalogue records were written by hand on UKMARC input sheets and sent to a keying bureau in Birmingham, before being dispatched on tape to the IBM mainframe in Harlow that the Library rented from Rank Hovis McDougall. There were no mobile phones, no World Wide Web, but we could look up the catalogues for post 1975 collections and the British National Bibliography on BLAISE (the British Library Automated Information Service) – which you can see displayed on the monitor in the photograph below.



**Figure 2.** *The British Library Catalogue Conversion Project team (1990). The Project Manager, Pat Oddy is second from left.* ©

I was assigned to the project to convert the British Library Catalogue of Printed Books to UKMARC. This photo was taken in the Novello Room at Novello House on Sheraton Street in Soho, which was where the Bibliographic Services Division was based. As you can imagine, Soho was an interesting and lively place in which to start working in London – lunchtime drinking was not compulsory, but it was considered a normal thing to do.

The past may be another country and we did drink differently there. But, as William Faulkner put it, “The past isn’t dead. It isn’t even past.” While much has changed since 1987, the legacy of the past lives on and a large part of my career has been spent trying to address the deficiencies of the metadata the Library inherited with its disparate collections and services. So, I want to start where I began, with the project to convert the British Museum catalogue of printed books into machine readable format – MARC.

British Library catalogue of printed books. K.G. Saur

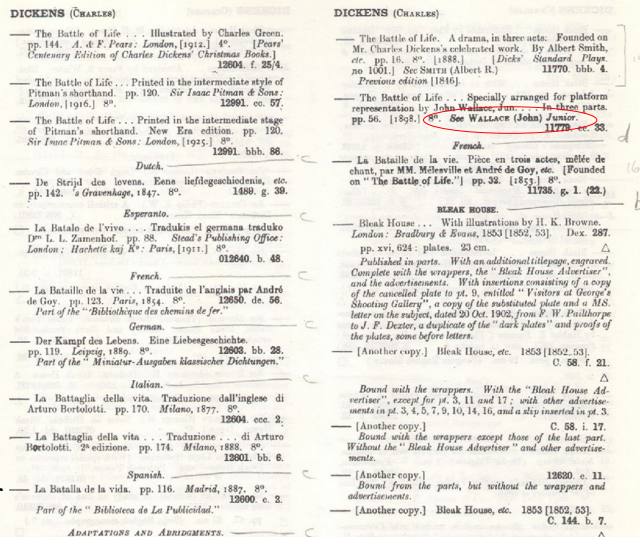


Figure 3. Detail of a page from British Library Catalogue of Printed Books. This is from the conversion team's copy and contains our mark-up and annotations

### British Library Catalogue Conversion Project

In the late 1980s the Library was planning for the move to St Pancras. However, there was no room at St Pancras for GK (the General Catalogue) whose guard books took pride of place in the centre of the round reading room at Bloomsbury. The decision had been taken to convert the catalogue into UKMARC format so that it could be consulted through computerised catalogue.

The source for the work was the 360 volume *British Library Catalogue of Printed Books* and its six supplements, published by K.G. Saur. This catalogue was the final printed manifestation of the catalogue of the British Museum Library.

The catalogue makes wonderfully effective use of the print technology of its time to render the complexity of publications intelligible to its users. Entries are arranged by alphabetical sequence of authors and titles. Different editions of the same work are gathered and collocated with translations and other related works. But the structure is implicit in the arrangement on the page. The structure was exploited to render the catalogue into UK MARC and later MARC 21.

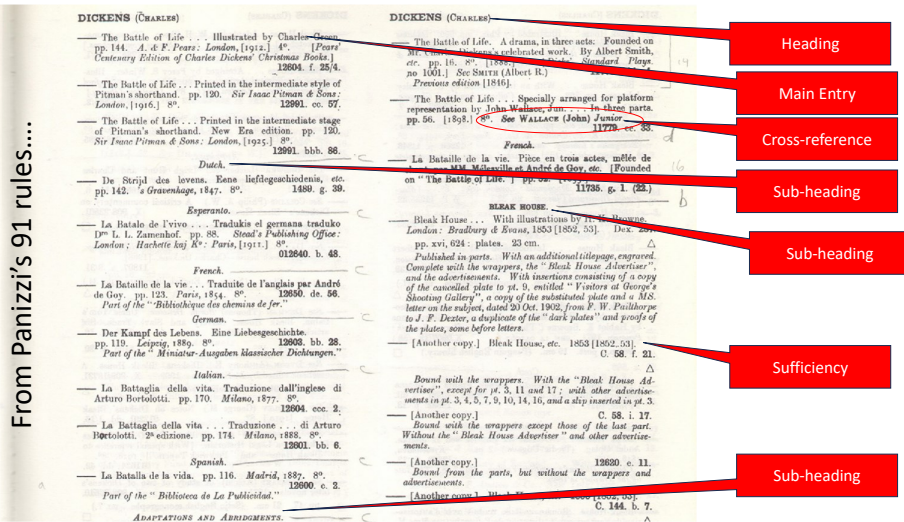


Figure 4. Illustrating the structure of BLC.

The catalogue rules used here are those of the British Museum, developed from Panizzi's 91 rules. By adding some labels, we can see the structure more clearly. At the top of each column is the heading. Followed by a series of main entries for publications in English of the *Battle of life* by Charles Dickens. Subheadings for other languages follow. Then another subheading for adaptations and abridgements, which we would think of as added entries pointing to the main entry under the name of the person who adapted it. Then another subheading introduces a different work, *Bleak house*.

from Panizzi's 91 rules...

DICKENS (CHARLES)

— The Battle of Life . . . Illustrated by Charles Green . . . pp. 144. A. & F. Pears: London, [1912.] 4°. [Pears' Centenary Edition of Charles Dickens' Christmas Books.] 12604. f. 25/4.

— The Battle of Life . . . Printed in the intermediate style of Pitman's shorthand. pp. 120. Sir Isaac Pitman & Sons: London, [1916.] 8°. 12601. cc. 57.

— The Battle of Life . . . Printed in the intermediate stage of Pitman's shorthand. New Era edition. pp. 120. Sir Isaac Pitman & Sons: London, [1925.] 8°. 12991. bbb. 86.

Dutch.

— De Strijd des levens. Eene liefdesgeschiedenis, etc. pp. 142. 's Gravenhage, 1847. 8°. 1488. g. 39.

Esperanto.

— La Batalo de l'vivo . . . Tradukis el germana traduko D<sup>no</sup> L. L. Zamenhof. pp. 88. Stead's Publishing Office: London; Hachette kaj K<sup>o</sup>: Paris, [1917.] 8°. 012640. b. 48.

French.

— La Bataille de la vie. . . Traduite de l'anglais par André de Goy. pp. 123. Paris, 1854. 8°. 12650. de. 56. Part of the "Bibliothèque des chemins de fer."

German.

— Der Kampf des Lebens. Eine Liebesgeschichte. pp. 119. Leipzig, 1889. 8°. 12603. bb. 28. Part of the "Miniatur-Ausgaben klassischer Dichtungen."

Italian.

— La Battaglia della vita. Traduzione dall'inglese di Arturo Bortolotti. pp. 170. Milano, 1877. 8°. 12604. ecc. 2.

— La Battaglia della vita . . . Traduzione . . . di Arturo Bortolotti. 2<sup>a</sup> edizione. pp. 174. Milano, 1888. 8°. 12601. bb. 6.

Spanish.

— La Batalla de la vida. pp. 116. Madrid, 1887. 8°. 12690. c. 2. Part of the "Biblioteca de La Publicidad."

DICKENS (CHARLES)

— The Battle of Life. A drama, in three acts: Founded on Mr. Charles Dickens's celebrated work. By Albert Smith, etc. pp. 16. 8°. [1885.] no. 1001. See SMITH (Albert R.) Previous edition [1846].

— The Battle of Life . . . Specially arranged for platform representation by John Wallace, Jun. . . in three parts. pp. 56. [1898.] See WALLACE (John) Junior 11779. cc. 33.

French.

— La Bataille de la vie. Pièce en trois actes, mêlée de vaudeville. MM. Méliès et André de Goy, etc. [Founded on "The Battle of Life."] pp. 52. [1897.] 11735. g. 1. (22.)

BLEAK HOUSE.

— Bleak House . . . With illustrations by H. M. Howne. London: Bradbury & Evans, 1853 [1852, 53]. Dex. 20. pp. xvi, 624: plates. 23 cm. Published in parts. With an additional titlepage, engraved. Complete with the wrappers, the "Bleak House Advertiser", and the advertisements. With insertions consisting of a copy of the cancelled plate to pt. 9, entitled "Visitors at George's Shooting Gallery", a copy of the substituted plate and a MS. letter on the subject, dated 20 Oct. 1902, from P. W. Paillhorpe to J. F. Dexter, a duplicate of the "dark plates" and proofs of the plates, some before letters.

— [Another copy.] Bleak House; etc. 1853 [1852, 53]. C. 58. f. 21.

Bound with the wrappers. With the "Bleak House Advertiser", except for pt. 3, 11 and 17; with other advertisements in pt. 3, 4, 5, 7, 9, 10, 14, 16, and a slip inserted in pt. 3.

— [Another copy.] C. 58. i. 17. Bound with the wrappers except those of the last part. Without the "Bleak House Advertiser" and other advertisements.

— [Another copy.] 12690. c. 11. Bound from the parts, but without the wrappers and advertisements.

— [Another copy.] C. 144. b. 7.

Figure 5. Illustrating assignment of MARC encoding.

BLC MARC (a simplified form of UK MARC) was used to encode the data so that the BLC file could be made available on BLAISE in the BL OPAC. I will return to this later, but in 2004 the BLC was one of about twenty UKMARC files migrated to the Library's new Library Management System, Aleph. As part of that process, the BLC MARC data had to be converted to MARC 21, as Aleph did not support UKMARC.

The conversion to MARC was obviously an essential step, but the flat structure of MARC traded the collocation and sufficiency of the printed catalogue for efficient data exchange and keyword searching. By the mid to late 1990s the impacts of digitisation and the World Wide Web had begun to highlight limitations of both MARC and the Anglo-American Cataloguing Rules, leading to the development of *Functional Requirements for Bibliographic Records* (FRBR), *RDA: resource description and access*, and eventually the *IFLA Library Reference Model* (LRM).

...to RDA's 4,000+ options

DICKENS (CHARLES)	DICKENS (CHARLES)
— The Battle of Life . . . Illustrated by Charles Green pp. 144. A. & F. French, London, [1911.] 8°. [Pears Anniversary Edition of Charles Dickens' Christmas Books.] 12804. f. 25/4.	— The Battle of Life. A drama, in three acts. Founded on Mr. Chivers' original novel. By Albert Smith, etc. pp. 16. 8°. [1888.] 11770. bbb. 4. no 1001. See SMITH (Albert R.) Previous edition [1816].
— The Battle of Life . . . Printed in the intermediate style of Pitman's shorthand. pp. 120. Sir Isaac Pitman & Sons, London, [1916.] 8°. 12801. cc. 57.	— The Battle of Life . . . Specially arranged for platform representation by John Wallace, Jun. . . . In three parts. pp. 56. [1898.] 8°. See WALLACE (John) Junior. 11779. cc. 33.
— The Battle of Life . . . Printed in the intermediate stage of Pitman's shorthand. New Era edition. pp. 120. Sir Isaac Pitman & Sons, London, [1925.] 8°. 12801. bbb. 86.	
	French.
— De Strijdl des levens. <i>Wann Heftgeschiedenis, etc.</i> pp. 143. La Gravenhage, 1847. 8°. 1400. a. 20.	— La Bataille de la vie. Pièce en trois actes, mêlée de vaudeville. Par M. de Goy, etc. [Founded on "The Battle of Life."] pp. 56. [1898.] 11735. g. 1. (22.)
	Dutch expression of The Battle of Life
	Attribute: language of expression
	Related Work (adaptation of Work)
	Work by Charles Dickens
	Person (author of The battle of life)
	Related Works

Figure 6. Illustrating alignment with RDA.

If we review the BLC example above, it is easy to see how its structure can also be expressed in RDA. Charles Dickens is a Person; *The battle of life* is a Work. "Charles Dickens is the author of *The battle of life*" expresses a relationship between a Person and a Work. The subtitles help to identify multiple language expressions of *The battle of life* and each description includes attributes of the Manifestation and Item.

Other subheadings also provide evidence for relationships between the original work and derivative works, such as adaptations. FRBR, LRM and RDA, represent a return to a more user centred approach by making a clear separation of content from carrier and by the relative weight given to relationships placing the resource described in context within the collection and the wider universe of human discourse.

Enhancement of Legacy Metadata



Figure 7. The "Title room" British Museum pre-1914.

These gentlemen were the custodians of the Library's metadata in the Edwardian era. On the shelves around them are millions of title slips, which are the British Museum's original catalogue records. The title slips have since been moved to new accommodation at Boston Spa and are still consulted regularly to answer queries from staff and the public.

"I. Titles to be written on slips, uniform in size."

*Rules for the compilation of the catalogue*

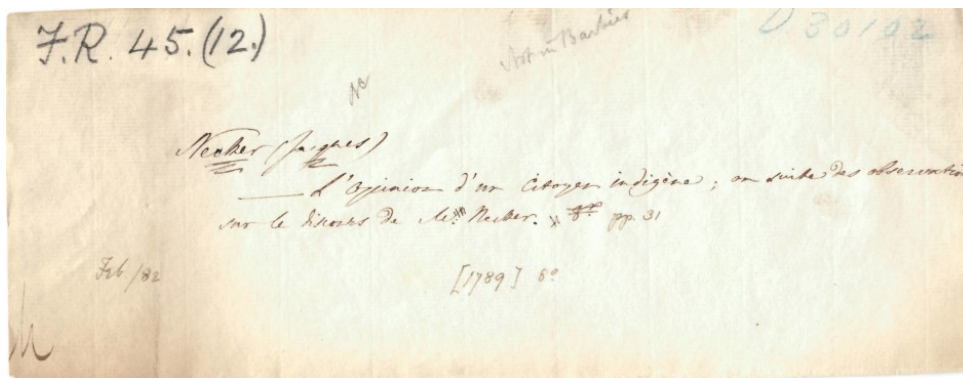


Figure 9. Title slip, attributed to Sir Anthony Panizzi

FMT	BK
LDR	nam a22001573 4500
001	002614806
005	20201009183539.0
008	890802 1789               fre
040	a Uk  c Uk
1001	a Necker, Jacques,  d 1732-1804.
24010	a Appendix
24510	a L'Opinion d'un citoyen indigène; ou suite des observations, sur le discours de M. Necker.
264 1	c [1789]
300	a 31 pages ;  c (8°)
336	a text  2 rdacontent
337	a unmediated  2 rdamedia
338	a volume  2 rdacarrier
340	m 8vo  2 rdabf
85241	a British Library  b HMNTS  j F.R.45.(12.)
SYS	002614806

Figure 9. Current MARC record

This is the current MARC 21 record, copied from Explore the British Library. As you can see, there is quite a lot more data here than on the original slip. So, what has changed? When did it change? And why has it changed?

FMT	BK
LDR	nam a22001573 4500
001	002614806
005	20201009183539.0
008	890802 1789               fre
040	la Uk lc Uk
1001	la Necker, Jacques, ld 1732-1804.
24010	la Appendix
24510	la L'Opinion d'un citoyen indigène; ou suite des observations, sur le discours de M. Necker.
264 1	lc [1789]
300	la 31 pages ; lc (8°)
336	la text  2 rdacontent
337	la unmediated  2 rdamedia
338	la volume  2 rdacarrier
340	lm 8vo  2 rdabf
883	\$aLanguage codes assigned by Bayesian analysis of words present in record\$dd20200826\$qUk \$c0.9919
88241	la British Library  b HMNTS  j F.R.45.(12.)
SYS	002614806

KEY
Original
UKMARC/MARC 21
Enhancements

Transcribed from the original and still enabling discovery 150 years later

Figure 10. Illustrating legacy transcribed data in green.

The original data is still there, highlighted in green. Panizzi's instruction to transcribe from the book in, "as few words [as necessary] and those only of the author" meant that the data was accurate and remains effective more than a century after it was created. That seems like a decent return on investment.

FMT	BK
LDR	nam a22001573 4500
001	002614806
005	20201009183539.0
008	890802 1789               fre
040	la Uk lc Uk
1001	la Necker, Jacques, ld 1732-1804.
24010	la Appendix
24510	la L'Opinion d'un citoyen indigène; ou suite des observations, sur le discours de M. Necker.
264 1	lc [1789]
300	la 31 pages ; lc (8°)
336	la text  2 rdacontent
337	la unmediated  2 rdamedia
338	la volume  2 rdacarrier
340	lm 8vo  2 rdabf
883	\$aLanguage codes assigned by Bayesian analysis of words present in record\$dd20200826\$qUk \$c0.9919
88241	la British Library  b HMNTS  j F.R.45.(12.)
SYS	002614806

KEY
Original
UKMARC/MARC 21
Enhancements

Converted to UKMARC 1988-1991  
Converted to MARC 21, 2004  
Enables exchange of data, keyword searching

Figure 11. Illustrating conversion to MARC.

In converting the record to UK MARC and later to MARC 21, the encoding was added, as well as various data elements required for data exchange, such as system number, leader, latest transaction, and source of metadata. BLC MARC proved quite difficult to transform to MARC 21 because of the ambiguity of some of the simplified encoding. For example, all main entries were encoded 100 and the whole publication statement was bundled into 260 \$a (anticipating RDA's manifestation statement). The simplification was probably an essential component in the success of the BLC Conversion project, but it is another example of how decisions taken for perfectly good reasons can have consequences down the line.

Alignment with North American Standards was a major component of the library's cataloguing strategy, devised by to improve efficiency Pat Oddy in the early 1990s. Replacing local authority files with NACO and

LCSH enabled reuse of Library of Congress and OCLC records with minimal intervention. MARC 21 was also essential for compatibility with new library management systems, which did not have native support for UKMARC.

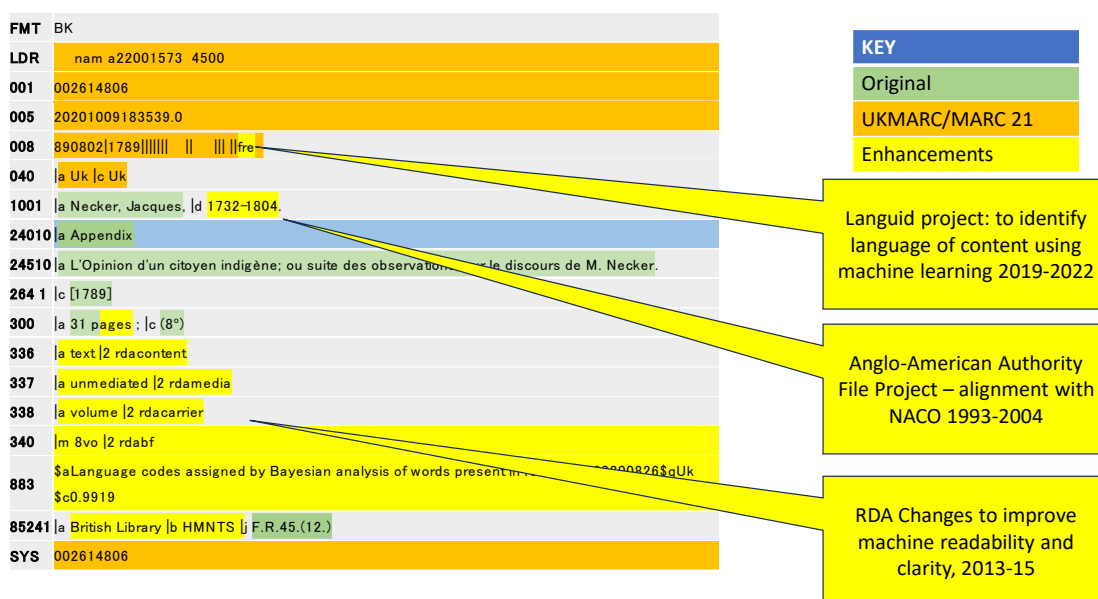


Figure 12. Illustrating enhancements over time.

Beginning in 1993, the Anglo-American Authority File was a collaboration with Library of Congress by which the Library was able to join NACO. Following the implementation of Aleph, Richard Moore ran several projects to align internal name headings with NACO. This greatly improved collocation of legacy records with more recent records created under AACR2. The focus for these projects was to convert the most common headings, so emphasis was on best-selling or canonical authors, sacred works, and corporate bodies. Richard gave his team some leeway regarding choice of headings that they worked on to maintain enthusiasm and quality. During this period, Andrew MacEwan led the implementation of LCSH to replace the unlamented CompASS subject vocabulary.

More recently, Thurstan Young undertook a substantial project in 2014/15 to assign RDA content, media, and carrier types to legacy records. Other RDA influenced changes can be seen in the expansion of abbreviations in the 300 field to make the catalogue more accessible to non-cataloguers, we also took advantage of the expansion of MARC 21 to copy book format from the extent to a specific subfield in 340.

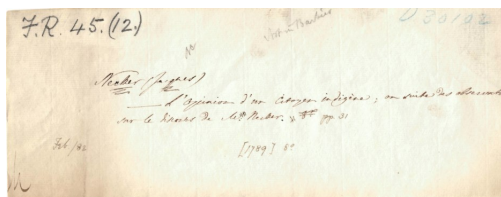
Another major enhancement has been the assignment of language codes in 008/35-37 using statistical methodologies and machine learning, which I want to describe in a bit more detail. As well as the language code, note that the 883 field contains provenance information for the project, including the confidence level for each assignment recorded in \$c.



## Languid Project 2018-2022

This book is in French

- Implicit to humans
- Opaque to computers
- Estimated 4.5 million records lacked MARC language codes
- Mostly pre-1975 collections



Title: L'Opinion d'un citoyen indigène; ou suite des observations, sur le discours de M. Necker.  
Author: Jacques Necker, 1732-1804.  
Publication Details: [1789]  
Language: French  
Uniform Title: Appendix  
Identifier: System number: 002614806  
Physical Description: 31 pages ; (8°)  
Shelfmark(s): General Reference Collection FR.45.

Figure 13. Languid project

The Languid project addressed the longstanding problem of identifying the language of items in the collection. With the exceptions of canonical authors, official publications and sacred works, the language of the expression was not recorded in legacy records because it was implicit to users and there was no way to search a linear name-title catalogue by language. I estimated from our annual audit of the catalogue metadata that around 4.5 million records lacked this information.

The project began at the end of 2018 and was the work of Victoria Morris – who will be speaking later today on another project. Victoria used a machine learning approach to identify the probability of words in the title belonging to a specific language. This was possible because we already had millions of records which contained accurate language codes. Using this corpus, the machine could be trained to recognise languages and assign a confidence level. Setting a high confidence level was essential because, with such a large dataset, assignment errors would be very hard to detect. Victoria was also able to engage the aid of Library's community of language experts to resolve ambiguous results. This proved to be a very popular activity during lock down, when curatorial staff and cataloguers were unable to access the collections and it meant that the confidence level for assignments was well above 99%.

## Languid Project 2018-2022

Dependencies

- Transcription of title
- Corpus of accurate metadata
- Expert professionals
- Capacity to try things out

Benefits:

- Improved knowledge of collections
- Better discoverability
- Inclusivity

**“XVIII The title of the book next to be written, and that expressed in as few words, and those only of the author, as may be necessary to exhibit to the reader all that the author meant to convey in the titular description of his work; the original orthography to be preserved.” A Panizzi Rules of the compilation of the catalogue. 1841**

Figure 14. Languid project overview

This approach depended absolutely on the Anglo-American cataloguing tradition of accurate transcription from the title page, which can be traced back to Panizzi's codification of cataloguing rules in 1841. The availability of good quality metadata from which to create a language corpus was also vital. Nothing could have been achieved without the enthusiasm, technical competence, and linguistic expertise of the staff involved.

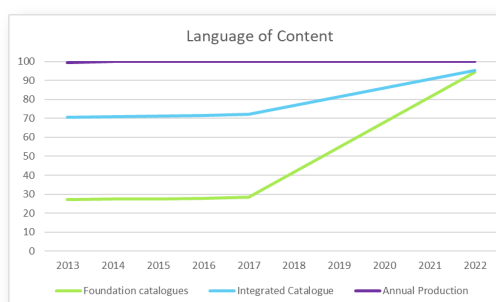
The benefits are substantial. Curators covering the African and Australian collections have commented that it has enabled them to get a much clearer picture of strengths and weaknesses of the collections for which they are responsible. Discoverability is improved, searches refined by language will achieve a much better recall and precision. Better understanding of the language content of the collection also enables us to serve different language communities more effectively. I hope that the more granular identification of minority and endangered languages will allow us to reach out to audiences who may have felt excluded in the past.

## Languid Project 2018-2022

### Outcomes

Language gap closed:

- 5 million records enhanced
- 424 languages identified
- High confidence levels
- Identification more granular than MARC 21/ISO 638-1



**Figure 15. Closing the language gap.**

This graph from the annual audit of the catalogue illustrates the quantitative outcome from the project. The blue line along the top shows that 100% of new MARC records produced each year contain a valid language code. The green line at the bottom represents “foundation collections”, including the British Museum catalogues, only about a quarter of which contained language codes prior to the Languid project. As shown by the pale blue line in the middle the gap between language coverage for current material and the catalogue as a whole was 30% in 2013, when we started monitoring this. The gap between current practice the foundation collections was more than 70%. As you can see, by end of 2022 the gap had closed to less than 10%.

Languid resulted in the assignment of million language codes to more than 5 million records, representing all of the 400 + languages identified by the MARC 21 language codes. We hoped that the methodology would be applicable to other catalogues and collections, including our sound recordings and our archives and manuscripts. Unfortunately, in those catalogues title information is much less likely to have been transcribed.

There is always more that could be done. It would be nice to make use of the more granular identification of languages by assigning codes from ISO 639-3 to supplement the MARC 21 codes for language groups and I hope it will be possible to take this forward through future work on equity and inclusion.

We have made a modest start towards addressing outdated or inappropriate demonyms by updating LCSH headings containing the term blacks. There is much more to do, and work is under way on evaluation of terminological and provenance issues in the Caribbean and South Asian collections.

The introduction of FAST to replace LCSH was motivated in part by the need for a more agile approach to subject indexing. Subject coverage across the catalogue is far from comprehensive, with the legacy data being particularly weak. This is unfortunately not easily addressed by artificial intelligence. During lockdown we were able to map FAST headings to the Watts *Elastic Classification* used by the British Museum and assign subject and genre terms to thousands of records. There are obviously limitations to how far contemporary terms can be mapped to 19<sup>th</sup> century classifications, but we discovered that by far the biggest problem, was that what was actually shelved at specific locations did not always correspond with the classification. We also benefit from the enrichment of our 19<sup>th</sup> century books data base by UCL students on the *Information through the ages* course, who have contributed thousands of subject, genre and geographical terms since 2019. I hope that much more can be done by matching and merging our legacy records with the richer records in the *English Short Title Catalogue*.

Further projects combining the *Georeferencer* tool with the wisdom of a select crowd of enthusiasts have enabled us to add coordinates and geographic headings to cartographic materials to improve visualisation of the collection in future.

In 2022, we took the decision to supersede LCSH with FAST as our preferred subject vocabulary. This was not an easy decision. One of the benefits of FAST is that each term has a unique identifier, which can be expressed as a URI (Uniform Resource Identifier). Encoding information as URI's makes it machine actionable and provides new ways of linking and reusing metadata. In addition to FAST, we have also begun to embed ISNI's in our bibliographic records, as a result of a successful collaboration between the Library, major publishers and our Cataloguing in Publication agent, BDS.

## Linked Data and the Share Family

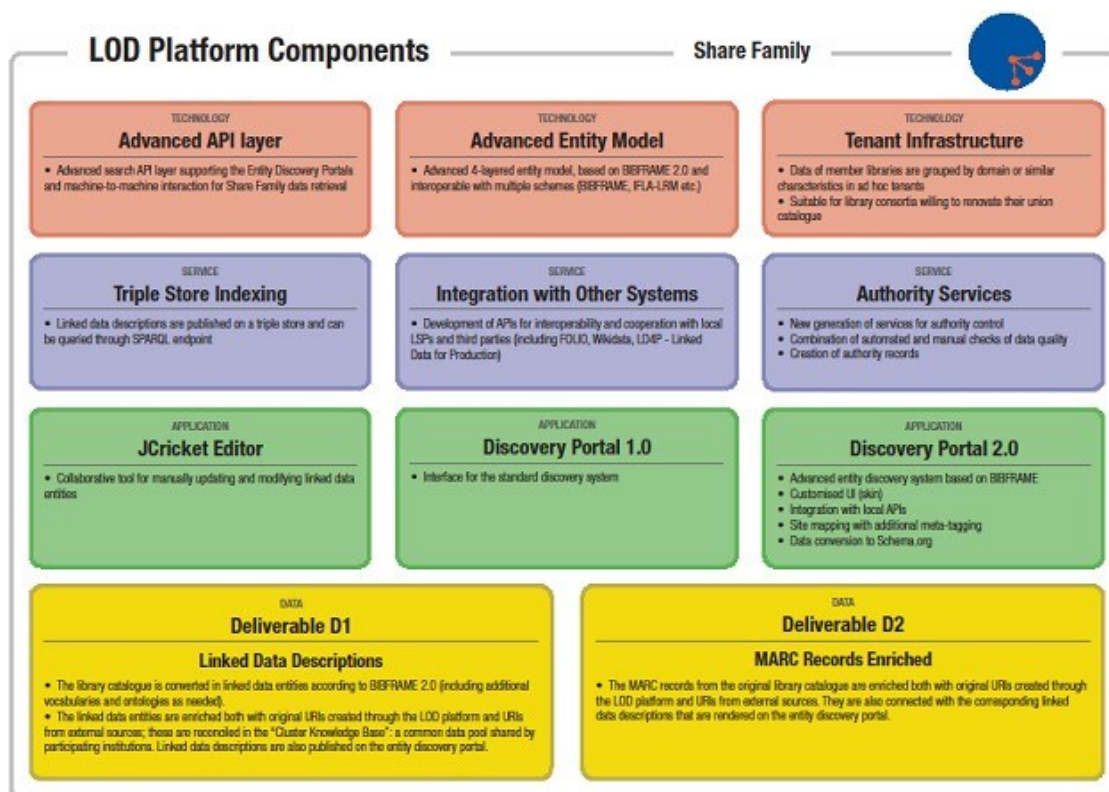


Figure 16: Overview of Share-VDE

The Library began to explore linked data in 2010. The British National Bibliography (BNB) was published as

linked open data the following year in collaboration with Talis and later TSO. However, the Library lacked the capacity to develop the service and in 2020 we took the decision to join the Share-Family. The Share Family is a collaborative community led by Casalini Libri and @Cult to develop a shared virtual discovery platform for Libraries. The “parents” are based in Italy, but it is a global family. Many of the members have been involved in Bibframe development through LD4P, etc. The Library was concerned that Share-VDE’s data model should support LRM/RDA data as well as Bibframe and has contributed to the development of the Share data model. This is still in development, but a major step was the introduction of svde:Opus to enable reconciliation of bf:Work with rda:Work and rda: Expression.



Figure 17: BNB home page on Share platform

BNB data has now been loaded to Share-VDE and BNB is the first member of the national bibliographies’ tenancy. The interface went live in June, shortly before I retired. The BNB has its own search page, but it is a part of the underlying knowledge base. The Share Family provides the potential to aggregate and cross search multiple national bibliographies for the first time.

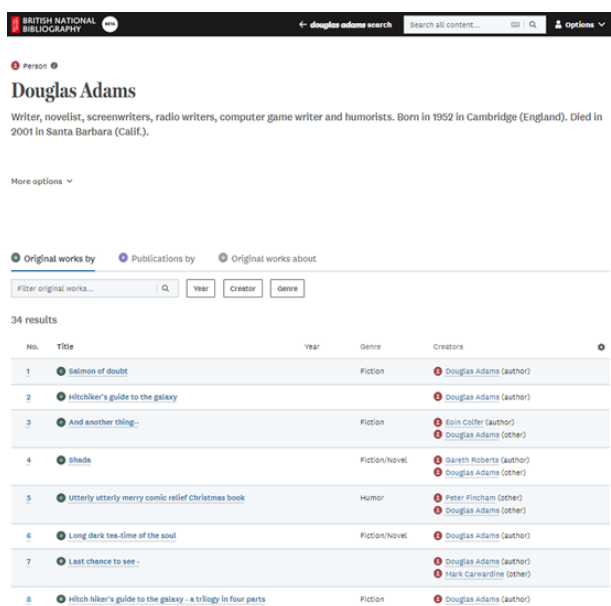


Figure 18: Example search results for original works by Douglas Adams

The Library expects to benefit from Share-VDE by learning from the practical experience our collaborators have gained through other linked data initiatives and by gaining access to a state of the art, extensible linked data infrastructure designed for library data. In Share VDE we will have a new channel for dissemination of the BNB, in aggregation with other national bibliographies.

Share-VDE also provides an opportunity to re-tool our metadata for the 21<sup>st</sup> Century. Our bibliographic metadata will be remodelled and clustered making it more compatible with entity-based data models, including the IFLA Library Reference Model, RDA: Resource Description and Access, and Bibframe. Our data will be enriched with URIs which will make it more effective in linked data environments. An entity-centred view of the BNB and subsequently the British Library catalogue will offer new perspectives for researchers.

### **Quality Assurance Metadata Measure (QAMM)**

I've talked a lot about good metadata, but what is good metadata? This is another issue that we have had to address over the years. Cataloguing is often seen as a luxury that can be cut back or done without when times are hard. Cataloguers are sometimes accused of gilding the lily or creating gold plated metadata. During lockdown I began to work with my Quality Assurance team to develop a new measure for metadata quality (QAMM), as part of a wider review of key performance indicators.

The following design principles were agreed.

- A method that is extensible across all collections and workflows: the method based on FRBR was too closely tied to MARC to be utilised by archives and MSS for example.
- A method that enables valid comparisons between collections and workflows: It's difficult to compare inputs from different workflows and standards, so we decide to compare outcomes.
- A method that enables quantification of the value added by the Library: What difference does our intervention make to discovery or collection management or efficiency.
- A method that accounts for stewardship of the Library's metadata assets: Enables us to identify the impact of different interventions over time.
- A method that identifies issues requiring remediation: Is sufficiently granular to enable us to identify and intervene where metadata falls short of the requirement.

We looked at a number of existing quality methodologies, but they tended to be MARC based and aimed at identification of candidate records for copy cataloguing, so we developed a different approach based on outputs, rather than inputs. An approach focused on what the description could deliver for the user rather than what the cataloguer had delivered in accordance with standards.

## Fitness for purpose

“Purpose and uses of the catalogues” in section 6 of *IFLA International Cataloguing Principles (ICP)*

- to **FIND** bibliographic resources in a collection as the result of a search using attributes or relationships of the entities
  - Single item (known item)
  - Set of resources by shared criteria
- to **IDENTIFY** a bibliographic resource or agent (that is, to confirm that the described entity corresponds to the entity sought or to distinguish between two or more entities with similar characteristics)
- to **SELECT** a bibliographic resource that is appropriate to the user's needs (that is, to choose a resource that meets the user's requirements with respect to medium, content, carrier, etc., or to reject a resource as being inappropriate to the user's needs)

- to **acquire or OBTAIN** access to an item described (that is, to provide information that will enable the user to acquire an item through purchase, loan, etc., or to access an item electronically through an online connection to a remote source); or to access, acquire, or obtain authority data or bibliographic data;
- to **navigate and EXPLORE**
  - within a catalogue, through the logical arrangement of bibliographic and authority data and the clear presentation of relationships among entities
  - beyond the catalogue, to other catalogues and in non-library contexts
- +
- to **MANAGE** the collection efficiently by supporting automation of processes and provision of management information

We established a number of use cases, based on IFLA's International Cataloguing Principles. It's not a coincidence that these are close to the FRBR and RDA user tasks. I added an extra task, called Manage to cover metadata that we need for internal purposes, including data exchange and management information.

## Example of BASIC description use cases

ID	Use case	Condition	Encoding	Status	PASS/FAIL	Additional Guidance
B01	FIND by the title or name of the entity described		MARC 245\$a	Mandatory		
B02	OBTAIN the item or access to the content using shelfmark or link		MARC 852;MARC 856	Mandatory		<b>Print:</b> Ensure that items for validation are obtained from collection storage or from Finishing output <b>Digital:</b> Validate on Explore the British Library to confirm that access to digital resources can be obtained. There may be several options for accessing the same publication.
B03	SELECT in accordance with restrictions on access or reuse		MARC 506;MARC 540	Mandatory if applicable		Rights metadata may be added in Aleph or in the Metadata Aggregator prior to publication to Explore the BL. It is therefore important to check EDR for digital resources. Born digital and digitized items, including Google books, will have a rights statement (Access restrictions) e.g. "No access restrictions", and Terms governing use: e.g. "Terms of use for Google books".
B04	SELECT by content, media or carrier type		MARC 336;337;338	Mandatory		A weekly fix routine is applied to records loaded to Aleph. Currently LDR 17=5 records are excluded and selection is limited to LDR 18=a. The routine will not assign any values if it is unable to assign content type. Collection Metadata receives a report of records that are updated and those that are selected but fail update.
B05	IDENTIFY or SELECT by attributes appropriate to the type of resource	See specific tabs	As specified	Mandatory if applicable		Check basic criteria for the relevant type of resource -e.g. serial, cartographic, music

The table above gives a zoomed in view of some basic use cases. A basic record should be sufficient to find a known item and obtain access to the content. The MARC coding is indicative not definitive and different encodings can be provided for other schemas. The QA team tested the basic concept, and we added additional contextual guidance to clarify ambiguities based on their feedback.

The measure was implemented in 2021 to supersede the previous “FRBR QM”. There are currently three applications for the measure.

### Corporate Quality Assurance measure (2021-)

- MARC Cataloguing is sampled and evaluated by the QA team.
- The intention is to roll it out to other systems and workflows over the next two years.

### Performance Management (2023-)

- Team leaders sample and evaluate cataloguers' work.

- Same methodology but the emphasis is on using QAMM as a tool for coaching.

#### Outsourcing (2022-)

- QA team applies the measure to sample data provided by prospective suppliers.

QAMM is used in conjunction with validation tools, such as MARC Report. We are also looking at the different levels as a means of evaluation of what level of intervention is necessary or justifiable for specific types of material. For example, it was possible to quickly process low use educational material at basic level and send it to shelf. A similar technique is being evaluated for travel guides (of which we still receive many through legal deposit). A basic record with the addition of a geographical heading should enable discovery.

#### Conclusion



**Figure 19. British Library Boston Spa aerial shot © British Library. Some of the buildings shown have been demolished to enable construction of a new high-density store.**

So, what has changed since 2023 as the Library celebrates its 50<sup>th</sup> anniversary? There is “One British Library” occupying just two sites, at St Pancras and Boston Spa. Much has been standardised. The collection is treated as a whole, in so far as possible, although there is still no single catalogue covering the whole. The Library has become a much more public facing institution. It holds major exhibitions, and its services are far more standardised and accessible. The Library aims to be inclusive and to be relevant to diverse audiences in the UK and abroad. I’ve visited many national libraries over the years, but I’ve never seen any that rivals St Pancras for public presence.

In 1987 cataloguing was a self-contained operation, even the cataloguing in publication programme (CIP) was carried out in house. Since then, the Library has collaborated widely on standards development and metadata creation. This has been essential to maintain (and increase) productivity in line with increasing legal deposit. The importance of metadata has obviously grown over that period, although the contributions of those who create it are not necessarily recognised or valued to the same extent.

These are perceptions that we have tried, with some success, to change through our metadata strategy. The key message of the Collection Metadata Strategy, first published in 2015, is that Metadata is an institutional asset. Metadata is as important to the institution as its collection, estate, and staff. The metadata represents

the labour of many decades and even centuries of work by many hands both within the Library and beyond. Creating accurate, timely metadata is an investment, which will repay itself over time. However, metadata does not mature over time. It needs to be refreshed and enhanced to stay useful by responding to changing user expectations and exploiting new technologies.

Pat Oddy has a profound influence on my career and on my understanding of the public value of what we do. It is appropriate to end with a quote from her book *Future libraries future catalogues*, in which she sums up the value professional cataloguers and the benefit they confer on the Library user:

*“Libraries provide access to the content of a collection in an organized and systematic way...The items in the collection are not only the organized output of human endeavour: the library also records and arranges these items in such a way as to give a further structural and organizational layer, which adds value to the whole and makes it far greater than the sum of its parts. This is a very long way from the provision of mere information and is at the heart of the professional values of the cataloguer.”*

*Pat Oddy Future Libraries future catalogues p.11  
London: Library Association Publishing, 1996.*

The substance of this message remains as true today, as it was 27 years ago, or in Panizzi's day, but the potential of modern technology means that the scope of our activity need no longer be limited to individual collections or a single library, but can be a common legacy for all.



## Further reading

British Library *Collection Metadata Strategy*

<https://www.bl.uk/collection-metadata/strategy-and-standards>

Alan Danskin (2020) "The Anglo-American Authority File: A PCC Story", *Cataloging & Classification Quarterly*, 58:3-4, 221-229, DOI: [10.1080/01639374.2019.1705952](https://doi.org/10.1080/01639374.2019.1705952)

Preprint: British Library Research Repository <https://bl.iro.bl.uk/concern/articles/0c3236f9-47d1-4d9b-bce9-7e7b0303cdd6>

Alan Danskin "Challenging legacies at the British Library" *Art Libraries Journal: Cataloguing Ethics Volume 48 Special Issue 2*, April 2023, pp. 38 - 42  
DOI: <https://doi.org/10.1017/alj.2023.4>

Preprint: British Library Research Repository <https://bl.iro.bl.uk/concern/articles/66261da8-ccb6-4072-b8d1-e48b96f67d1d>

Victoria Morris (2020) "Automated Language Identification of Bibliographic Resources", *Cataloging & Classification Quarterly*, 58:1, 1-27, DOI: [10.1080/01639374.2019.1700201](https://doi.org/10.1080/01639374.2019.1700201)

Preprint: British Library Research Repository <https://bl.iro.bl.uk/concern/articles/6c99ffcb-0003-477d-8a58-64cf8c45ecf5>

Janet Ashton & Caroline Kent (2017) "New Approaches to Subject Indexing at the British Library", *Cataloging & Classification Quarterly*, 55:7-8, 549-559, DOI: [10.1080/01639374.2017.1354345](https://doi.org/10.1080/01639374.2017.1354345)

Preprint: British Library Research Repository <https://bl.iro.bl.uk/concern/articles/50a09642-7356-45de-b6c4-5724fecbef7c>

Janet Ashton & Caroline Kent (2023) FAST: A Journey Toward Sustainability in Subject Indexing at the British Library, *Cataloging & Classification Quarterly*, 61:5-6, 525-534, DOI: [10.1080/01639374.2023.2238701](https://doi.org/10.1080/01639374.2023.2238701)

Pat Oddy *Future Libraries future catalogues*. London. Library Association. 1996 ISBN 1-85604-16-1

Sir Anthony Panizzi, *Rules for the compilation of the catalogue*. In: *Catalogue of printed books in the British Museum*. London. British Museum, 1841

(Also available online at Jeremy Norman's Historyofinformation website: <https://www.historyofinformation.com/detail.php?entryid=2426> (viewed 1/9/2023))

*Rules for compiling the catalogues of printed books, maps and music in the British Museum*. London: British Museum, revised edition, 1936.

"Purpose and uses of the catalogues" in section 6 of *IFLA statement of international cataloguing principles, 2016* [https://www.ifla.org/wp-content/uploads/2019/05/assets/cataloguing/icp/icp\\_2016-en.pdf](https://www.ifla.org/wp-content/uploads/2019/05/assets/cataloguing/icp/icp_2016-en.pdf)

UCL *Information through the ages* <https://www.ucl.ac.uk/module-catalogue/modules/information-through-the-ages-BASC0033>