

Cambridge University Press and Metadata

Concetta La Spada

Cambridge University Press, Cambridge, United Kingdom.

E-mail address: concetta.laspada@cambridge.org



Copyright © 2022 by Concetta La Spada. This work is made available under the terms of the Creative Commons Attribution 4.0 International License:

<http://creativecommons.org/licenses/by/4.0>

Abstract:

As we all know, good MARC records are necessary in order to attain good discoverability. If they are of poor quality, discoverability is seriously compromised and without them, discoverability can be almost impossible. The same goes for KBART lists and XML headers metadata.

Usually, metadata within non-library institutions is created automatically through various systems and then distributed to libraries and third parties but at the Cambridge University Press things are done differently.

We have worked on improving our metadata in the following ways:

- 1) *Manually enhancing the automated MARC records, using MarcEdit and Bibliofile and making these fully enhanced records available on our platform Cambridge Core*
- 2) *Creating KBART lists available to download on our platform*
- 3) *Distributing our metadata to library and library services suppliers at the same time*
- 4) *Keeping up to date with what international bodies like NISO and IFLA recommends in terms of metadata for publishing*
- 5) *Working with library services suppliers on quality control policies in order to make sure that only the best metadata is supplied to the libraries and institutions who use their services*

At Cambridge University Press, quality control within metadata is fundamental in order to ensure that resources are discovered by users. With this presentation, we will show how our metadata supply chain works, how our policies are applied and what big impact this has had within our company and in the way we collaborate with librarians and library services suppliers.

Keywords: metadata, cataloguing, publishing, MARC21, KBART

Cambridge University Press and Metadata

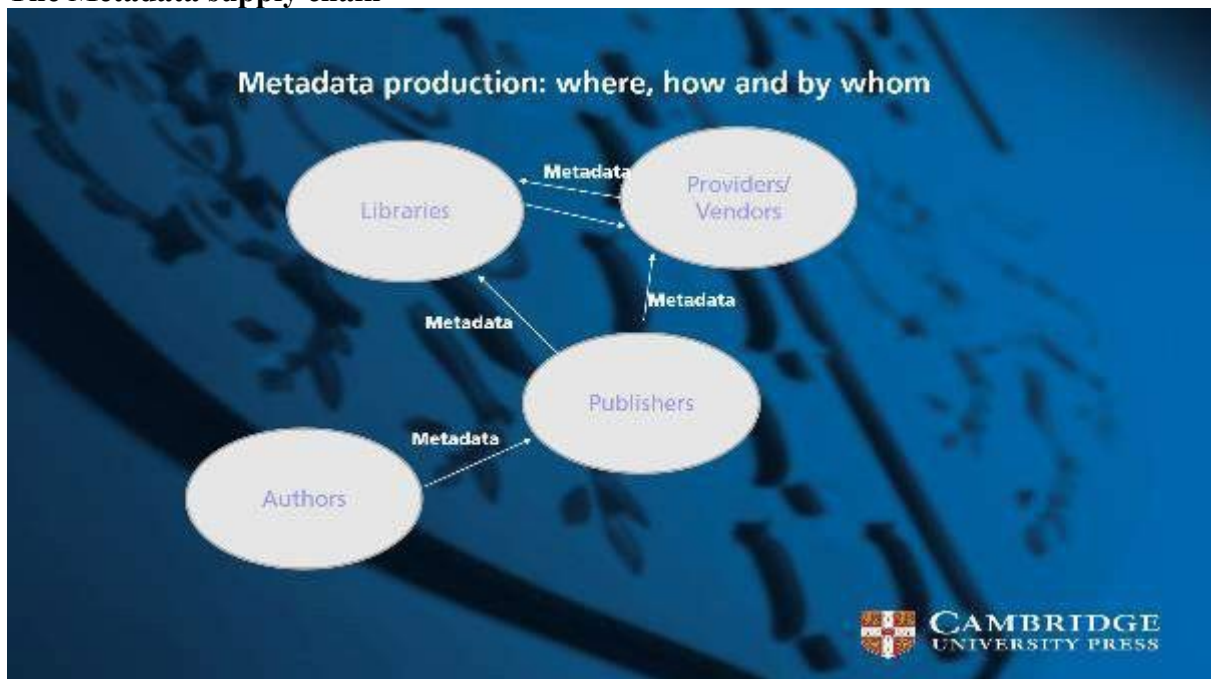
Concetta La Spada
Senior Metadata Librarian



Thank you for inviting me to attend the panel “Quality control for metadata”, I am glad to be able to discuss such topics with fellow librarians.

I am Concetta La Spada, Senior Metadata Librarian at Cambridge University Press. My job is to work on the metadata that we supply, particularly MARC records, and make sure that that metadata is of good quality.

The Metadata supply chain



Many of you attending today are academic librarians and cataloguers so you may be already aware of the dynamics within the Metadata Supply Chain, but for those of you who aren't here it is.

Metadata production begins with Authors and Publishers. Authors create the first metadata for their work and share this metadata with their publishers. Publishers then will re-use that metadata to create

various metadata formats. At CUP we distribute MARC records, KBART lists, XML and ONIX metadata. Publishers share that metadata with libraries and vendors/aggregators. Libraries then will also share this metadata with the “community zones” of the aggregators platforms and aggregators will make that metadata available to their users on their platforms. Other libraries will also send back to the publishers enriched metadata. One case is that of the Library of Congress (LC). CUP has a special agreement with the LC; they receive from CUP pre-publication metadata and will create print MARC records for CUP’s titles. So this means that about 90% of CUP MARC records contain data from the LC.

It is fundamental that all the various institutions involved in the metadata supply chain are aware of the importance of quality metadata and how bad quality metadata could affect resources usage and what challenges could bring to users.

Importance of metadata for discoverability

The screenshot shows the University of Cambridge LibrarySearch interface. The browser address bar displays 'http://search.lib.cam.ac.uk/'. The page title is 'THE IMPORTANCE OF METADATA TO IMPROVE DISCOVERABILITY'. The search results page shows 'No results found' for the query 'OUR FIRST AMBASSADOR TO CHINA'. The search criteria are: Author / Creator: Macartney, George; Publisher: Cambridge University Press; Place of publication: ; Publication date: ; Notes and contents: ; Series: Cambridge Library collection; Classmark: . The search options are set to 'All fields' and 'All locations'. A banner for 'Try LibrarySearch+ for articles, newspaper content and more ...' is visible. A book cover for 'OUR FIRST AMBASSADOR TO CHINA' is shown on the left side of the page.

This is an example of what happens if the metadata provided is not of good quality. Long time ago, I was looking for the above title on the University of Cambridge online library catalogue. I did a subject search, but nothing came up. Why? Because the MARC record for that title did not contain any subjects. This is something that it should not have happened. MARC records had to contain all the elements necessary to ensure full discoverability. This kind of situation bring disadvantages to everyone involved. To users, because they will lose time in try to find the resources they want; to librarians, because they will lose time in helping them; to publishers, because their resources will not be used.

MARC records and KBART lists

Cataloguing

The eBook records merged with the LC print records

No Subject headings

No authorized forms of authors' /editors' names and series titles.

Our systems create automate eBook MARC records using all the metadata available within these systems.

As you can see from the image above, these records are good enough records. Records with sufficient elements to ensure basic discoverability.

They contain: ISBNs, Title, Statement of responsibility, Names, Abstract, etc..

But there are various elements missing:

Library of Congress (LC) Classification numbers; Dewey numbers; LC authorized forms of names; and (most of all) Subject Headings.

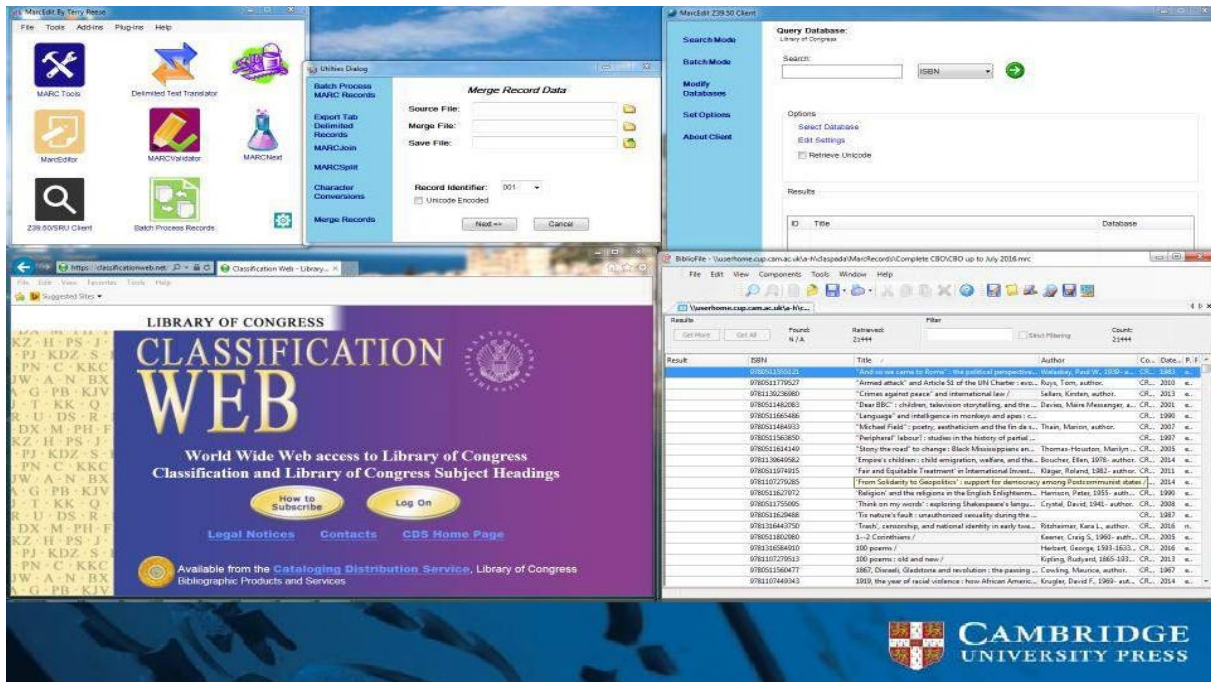
Before I started working at CUP, there wasn't a librarian working on my role and that meant that the MARC records distributed were only these automated ones. So, at the time, we had to rely on external librarians' expertise in order to ensure that our quality control policies were up to date with the international standards given by IFLA, Library of Congress and British Library.

But since September 2015, when I started working at CUP, things have changed. I am part of various libraries committees, including the CILIP Metadata & Discovery Group Committee. In this way, I keep up to date with what is going on within all the relevant bodies in terms of metadata and quality control. I review our policies and they way we create metadata and the kind of metadata standards we follow accordingly.

Every month, I enhance all the automated MARC records to full level; usually about 200 o 300 records, because CUP hosts, and provide MARC records for, Partners Publishers titles on our platform, Cambridge Core.

For part of these records (about 90%) I will use the data from the print LC records; these titles which do not have matching LC records I will catalogue from scratch. All the records I create are RDA-complaint.

This system works very well also because our data is free to anyone who needs it. From our platforms people can download both MARC records and KBART lists without having to log in.



I enhance MARC records using various cataloguing tools. MarcEdit for data manipulation, to merge and split records; and using the MarcValidator to ensure that there are no errors within the files. Bibliofile to modify single records. ClassificationWeb and Library of Congress Authorities to add classification numbers, subject headings and authorized forms of names and series.

In the end, all our MARC records will have the below elements. To ensure that discoverability is "covered" from every angle.

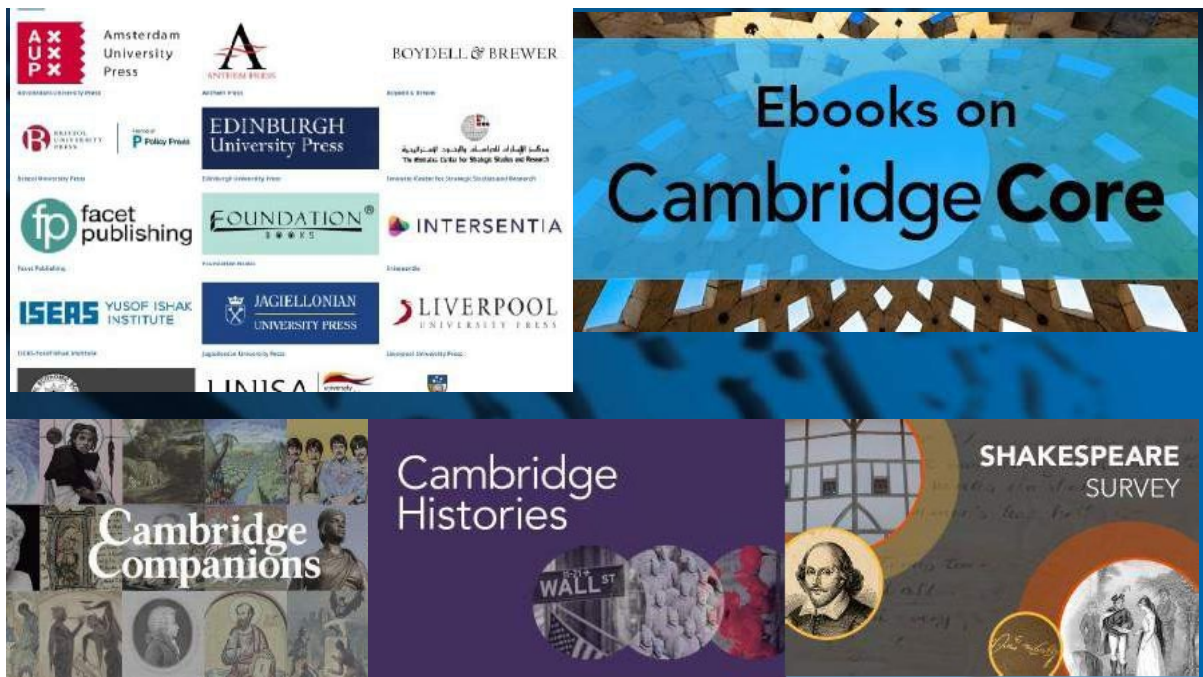
eBook MARC records at CUP

LC and Dewey Classification numbers (050 and 082 fields)
LC Subject Headings (600, 610, 611, 630, 650 and 651 fields)
LC Authorized forms of authors/editors' names (100 and 700 fields)

Also, in these records that require them,
Language and Geographical codes (041 and 043 fields)
LC Authorized forms of Series' titles (490 and 830 fields)
Corporate or Conference entries (110, 111, 610, 611, 710 and 711 fields)
Title variants (130, 240 and 246 fields)



And, as I was saying, I will do this for all the titles that are published on Cambridge Core.



I also create from scratch full records for CUP's journals, since our system does not create them automatically.

By the 10th of each month, I distributed all the previous month's eBook MARC records and XML metadata to all the aggregators at the same time. Our sales reps and support people will deliver it to librarians who get their metadata directly from us. In this way, I can ensure that the data reaches librarians at the same time and without delays.

publication_title	print_identifier	online_identifier	date_first	n	num	last	title_url
1 Why Aren't They Talking?: The Song Through Musical Fr...	978-1-108791-93-9	978-1-108866-86-6					https://doi.org/10.1017/9781108866866
2 A Criminology of Narrative Fiction	978-1-529208-05-4	978-1-529208-08-5					https://doi.org/10.46692/9781529208085
3 A History of Alexander the Great in World Culture	978-1-107167-09-8	978-1-316711-79-8					https://doi.org/10.1017/9781316711798
4 A History of Financial Technology and Regulation: From	978-1-107153-40-0	978-1-316597-73-6					https://doi.org/10.1017/9781316597736
5 A History of the Hausa Language: Reconstruction and Pe	978-1-009123-10-5	978-1-009128-07-0					https://doi.org/10.1017/9781009128070
6 A Logical Foundation for Potentialist Set Theory	978-1-108834-31-5	978-1-108992-75-6					https://doi.org/10.1017/9781108834315
7 A Philosopher Looks at Digital Communication	978-1-108986-81-6	978-1-108981-58-3					https://doi.org/10.1017/9781108986816
8 A Practical Guide to Basic Laboratory Andrology	978-1-009181-63-1	978-1-009181-64-8					https://doi.org/10.1017/9781009181631
9 A Social History of Literacy in Japan	978-1-785277-01-6	978-1-785277-02-3					https://doi.org/10.1017/9781785277016
10 Ability, Inequality and Post-Pandemic Schools: Rethinkin	978-1-447346-01-6	978-1-447347-01-9					https://doi.org/10.1017/9781447346016
11 Absorption and Theatricality: On Ghost Trio	978-1-009001-17-5	978-1-009001-16-8					https://doi.org/10.1017/9781009001175
12 Access to Justice and International Organisations: Court	978-1-108837-56-5	978-1-108946-37-7					https://doi.org/10.1017/9781108837565
13 Action Description in Interaction	978-1-108876-62-7	978-1-108673-41-4					https://doi.org/10.1017/9781108876627
14 Advanced Issues in the Green Economy and Sustainable	978-1-009097-98-7	978-1-009093-40-8					https://doi.org/10.1017/9781009097987
15 African American Literature in Transition, 1920-1930	978-1-108836-16-2	978-1-108992-40-3					https://doi.org/10.1017/9781108836162
16 African American Literature in Transition, 1930-1940	978-1-108837-55-5	978-1-108992-40-3					https://doi.org/10.1017/9781108837555
17 African Peacekeeping	978-1-108499-37-8	978-1-108499-38-5					https://doi.org/10.1017/9781108499378
18 African Marac Literature, Theory, and Value in the Twent	978-1-108489-28-7						https://doi.org/10.1017/9781108489287
19 Agents of the Hidden Imam: Forging Tuvlevi Shi'ism, 856	978-1-108838-39-3						https://doi.org/10.1017/9781108838393
20 Algorithms and the End of Politics: How Technology Sha	978-1-529215-31-*						https://doi.org/10.1017/9781529215310
21 All Too Human: Understanding and Improving our Relati	978-1-3167...						https://doi.org/10.1017/97813167...
22 Altered Earth: Getting the Anthropocene Right	978-1-...						https://doi.org/10.1017/9781...
23 Alzheimer's Disease Drug Developments: Research and P							https://doi.org/10.1017/978...
24 America's Energy Gamble: People, Economy and P							https://doi.org/10.1017/978...
25 An Introduction to the Philosophy of Spinoza							https://doi.org/10.1017/978...
26 Ancient Egypt in its African Context: F							https://doi.org/10.1017/978...
27 Anti-Constitutional Populism							https://doi.org/10.1017/978...
28 Antisocial Personality Dis							https://doi.org/10.1017/978...

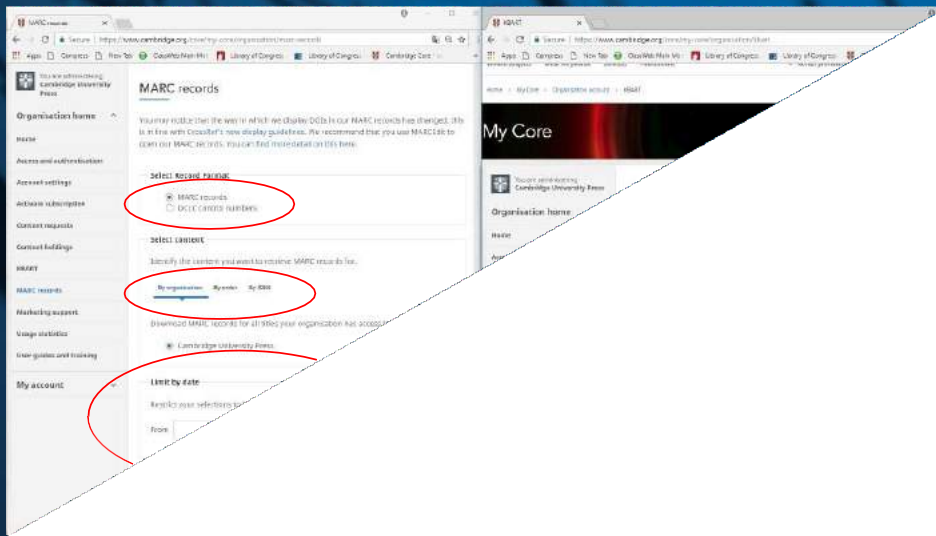
Our KBART lists, for both books and journals, follow the NISO Recommendations. Unlike the MARC records they are created and updated automatically by our system. KBART lists will be used by the aggregators to populate the matching collections on their platforms. When we create new collections/KBART lists we contact the aggregators to add them on their databases.

Metadata on Cambridge Core

The screenshot shows two browser windows from Cambridge Core. The left window displays the 'MARC records' page, where users can download bibliographic data. Red circles highlight the 'MARC records' link in the sidebar, the 'OCLC numbers' radio button under 'Desired output', the 'Open access books' radio button under 'Select content', and the 'Limit by date (optional)' section with 'From' and 'To' input fields. The right window displays the 'KBART title lists' page, where users can download KBART data. A red circle highlights the 'KBART title lists' link in the sidebar, and another red circle highlights a search box labeled 'Search and filter the results' on the main content area.

Anyone can download MARC records and KBART lists from the front-end pages on Cambridge Core. For MARC records they can refine their selection by collection/publisher and by date. For KBART lists, they can use a search box to search within the long list on that page.


Improvements to the automated systems for retrieving MARC records and KBART



If instead librarians have accounts on Cambridge Core. They can log in on their accounts and download MARC records and KBART lists according to their holdings.

Relationship with aggregators

CUP and Third Parties



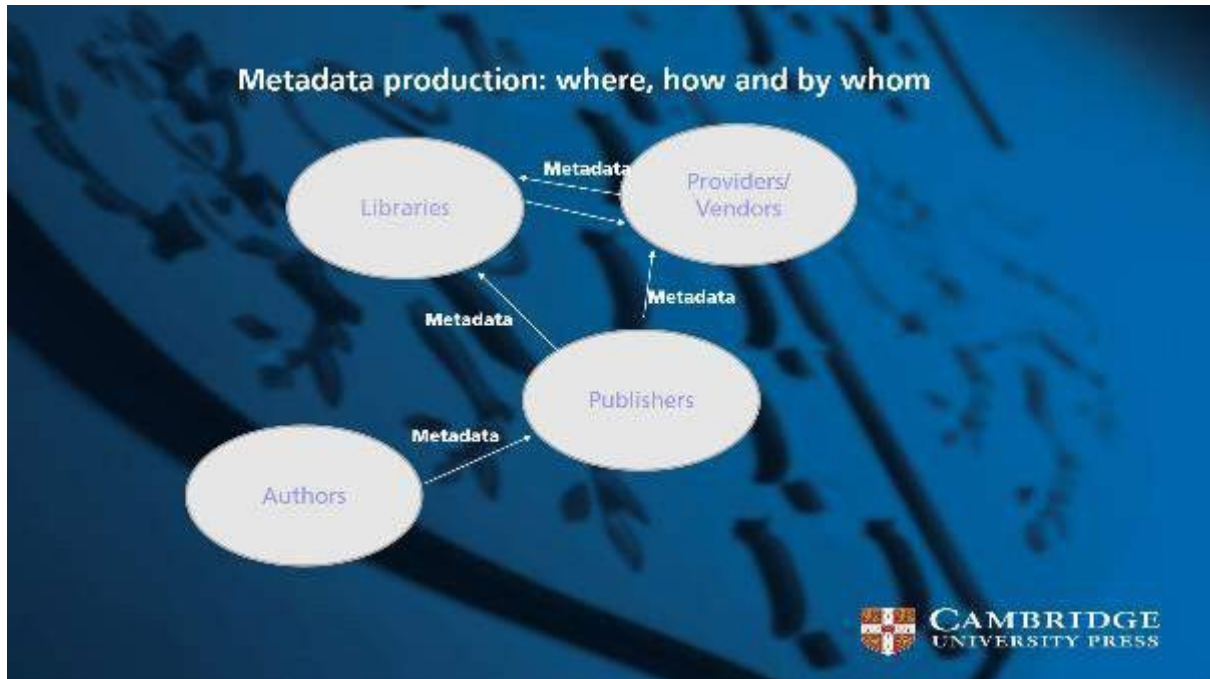
XML Metadata
eBook MARC records
KBART lists



We send our metadata to many aggregators and vendors. These are just a few. With all of them though, we have quarterly meetings in order to keep each other up to date with what's going in our companies. This has been fundamental because in this way not only we are up to date with

what the others are doing but also, we develop our knowledge on how their platforms work and how they use our metadata. This kind of collaboration is essential to make sure that the quality of the metadata we supply is maintained within the supply chain. We want our metadata to be used correctly.

Conclusion



I want to end my paper with the same image I have used at the beginning. This is because I want to stress the importance of collaboration between all the ones involved in the metadata production and distribution. Because it is only if we all work together towards the same goals that we can make sure that the quality of the metadata that we share among us is good and can bring the results that we want for our users. If we all do not produce metadata using the same standards and being knowledgeable of the same rules, then our metadata will not be effective. We all must work together.

References

1. MarcEdit can be found here <https://marcedit.reeset.net/downloads>
2. Here more information about Bibliofile <https://www.itsmarc.com/BiblioFileFAQ.asp>
3. ClassificationWeb <https://classweb.org/Menu/index.html>
4. Library of Congress Authorities, <https://authorities.loc.gov/>