

Examining Present Practices to Inform Future Metadata Use: An Empirical Analysis of MARC Content Designation Utilization

ABSTRACT

Libraries have a critical responsibility to organize materials and prepare them for access and use by end users. No group has been more central to this endeavor than library catalogers. No standard has been more central than the MARC record. Since its development more than 30 years ago, the MARC record has evolved into a complex encoding scheme providing rich content designation (in the form of fields, subfields, and indicators). Current MARC 21 specifications define nearly 2,000 fields/subfields available to catalogers to encode bibliographic data. The resulting MARC records reflect decisions by catalogers following standards, rules, and guidelines, and these records serve as artifacts of the entire cataloging enterprise. Surprisingly, there has been no programmatic effort to analyze MARC content designation utilization, and until recently, no current publicly available data on its utilization.

The results from a recent analysis of 400,000 MARC records conducted as part of an IMLS National Leadership Grant to establish a Z39.50 interoperability testbed indicated less than 50% of nearly 2,000 MARC 21 fields/subfields occurred even once in the records, and that only 36 of the fields/subfields accounted for approximately 80% of all use. These preliminary results have sparked interest by catalogers, managers of cataloging operations, standards developers, people involved in machine generation of metadata, and others. We are proposing a research project that builds upon the initial analysis to carry out a systematic analysis of MARC content designation use in large random samples of format-specific MARC 21 bibliographic records.

The overarching research question addressed by this project is: What is the extent of catalogers' use of content designation available in MARC 21? The project has three goals: 1) Provide empirical evidence to document MARC 21 content designation use; 2) Explore the evolution of MARC content designation for patterns of availability and adoption/use level; and 3) Investigate a methodological approach to understand the factors contributing to current levels of MARC content designation use and relationships with the cataloging enterprise.

The project has the following objectives:

- Develop and implement systematic methods, procedures, and software tools to produce reliable and valid analysis of MARC 21 content designation use
- Identify core elements of bibliographic records based on the analysis of format-specific samples and comparisons with existing recommendations for core records
- Document the evolution of MARC content designation to assess the availability of content designation at specific intervals of time and subsequent rates of adoption/use
- Develop a methodological approach to identify and understand factors contributing to catalogers' use of MARC content designation.

The project will result a number of high-impact deliverables of use and benefit to the library and other metadata communities:

- Documented methods and procedures, and open source software tools to conduct reliable and valid analyses of MARC 21 content designation, which can be used by individual libraries and adapted by other metadata communities
- Information on catalogers' use of MARC 21 content designation
- Identification of "core" elements in bibliographic records based on occurrence in the large samples of records and examination of initiatives recommending core bibliographic records
- A database application containing MARC 21 content designation specifications in structured, machine-processible form, for analyzing trends and patterns
- A methodology for identifying and understanding factors influencing the use of MARC 21 content designation.

An active dissemination effort will enable us to broadly share the results of this research. The research results will provide empirically based information to assist decision makers, including those assessing cataloging policies and practices, those involved with MARC maintenance and its evolution, and those developing metadata or other schemes for bibliographic data.