Benjamin Turkus

Kara Van Malssen

Digital Preservation

12 December 2013

'The Heavies Were All for Automation': Machine Readable Cataloging and the Bibliographic Framework Initiative

"The imagination of data is in some measure always an act of classification, of lumping and splitting, nesting and ranking, though the underlying principles at work can be hard to recover."

—Lisa Gitelman and Victoria Jackson, 'Raw Data' is an Oxymoron (2013)

"A classification of classificatory styles would be a good first step towards thinking systematically about distinctive styles of reasoning…The comparison of classifications as an index of other things that are happening in our society provides a small, provisional ladder of escape from the circle of self-reference."

—Mary Douglas, How Institutions Think (1986)

It is 1965. At a conference sponsored by the Council on Library Resources (CLR), the Association of Research Libraries' (ARL) Committee on Automation, and the Library of Congress (LC), representatives from "universities, research agencies, government agencies, and private industry" gather in Washington, D.C. to discuss a topic of increasing attention: the automation, or computerization, of library practice (Avram 3). With a mandate to "design and implement the procedures required to automate the cataloging, searching, indexing, and document retrieval functions" performed by librarians throughout the country, three employees of the Library of Congress—a reference librarian (Ruth Freitag), a cataloger (Kay Guiles), and an information systems analyst (Henriette Avram)—are given the daunting yet intriguing task to think as a computer would, "analyzing cataloging data from a machine processing point of view" (3). Their ultimate goal: to develop a stable method for converting the information held within the massive LC card catalog system into machine-readable form, creating records that would offer participating libraries the following services:

1) Computer-controlled photo-typesetting for catalog card printing;
2) Distribution of machine-readable catalog data to other libraries;
3) Production of special bibliographies and catalogs; and
4) Storage of records in electronic form for on-line interrogation (Snyder iii).

Though not the first attempt to automate library functions, Freitag, Guiles, and Avram's work would prove to be the longest lasting; the result of their efforts—the widely adopted Machine Readable Cataloging Format (MARC)—has been an unparalleled, though in some ways curiously unexamined success, quietly reigning

over the field of library science for nearly five decades.[1] How should we think of this hybridized (human/computer; analog/digital) moment of inscription, when the creators of MARC determined that 'readability' must function as a seesaw, balancing the needs of both the human *and* the technological components of this particular bibliographic system? Should we consider MARC as an industrializing and industrialized force, one that sought to radically eliminate the many forms of the human complexity that had previously governed library practice? In *Double Fold: Libraries and the Assault on Paper* (2001), Nicholson Baker urges us to follow this path, condemning MARC (while paying it a backhanded compliment) by describing it as "ungainly, cabalistically coded, but twenty years ahead of its time" (88). Are we better served to counter Baker's portrait, and instead focus our attention on MARC's development as a "communications format," something egalitarian, democratic, and distributed by its very nature?

Resisting the temptation to offer a recuperative study of MARC, getting entangled in debates regarding the contemporary value of MARC's latest iteration, MARC21 (1999), this essay will chart an alternative, historical course, exploring the contexts (social, political, economic, technological) in which MARC was developed.[2] By delving into this history—the history of a "classificatory style"—and by considering MARC as a way "data was imagined" by a particular community in a particular time, we will grant ourselves provisional escape from the "circle of self-reference" referred to by Mary Douglas at the opening of this essay (Douglas 108-109; Gitelman and Jackson 3). Temporarily setting ourselves apart, we can begin anew the process of exploring the "other things that are happening in our society"—the social order underpinning all systems of classification (108-109). And at the risk of appearing hopelessly unengaged with the challenges facing librarians and digital preservationists today, the primacy of historical work itself will be reasserted. If, to borrow from "Sustainable Economics for a Digital Planet: Ensuring Long-Term Access to Digital Information" (2010), we understand contemporary digital preservation to be a societal challenge in which "the needs and desires of the present day" are balanced with "those of the future" for the sake of cultivating a

---

[1] In "How MARC Has Changed: The History of a Format and Its Forthcoming Relationship to RDA [Resource Description and Access]" (2011), Michele Seikel and Thomas Steele trace library automation efforts to the early 1950s, detailing the pioneering work pursued at a number of institutions: the King County Library in Seattle, Washington; the Los Angeles County Public Library; the University of Illinois, Chicago; Florida Atlantic University; and the University of Toronto's Ontario New Universities Library Project. All of these efforts are worthy of greater critical attention, reminders that the MARC initiative is best understood as reflective of a wider societal drive for automation. Yet, as Samuel Snyder, coordinator of the LC's information systems 1964-1966, pointed out in his foreword to *ISS (Information Systems Specialist) Planning Memorandum Number 3: A Proposed Format for a Standardized Machine-Readable Catalog Record* (1965), "While some libraries have already begun to use automated cataloging techniques—and we have been happy to receive the benefits of their experience—the corresponding move at the Library of Congress is fraught with greater ramifications" (ii). Today, as the LC grapples with equally fraught decisions related to retiring MARC and replacing it with a Linked Data Model, it pays to recall how the institution handled earlier moments of dynamic technological change.

[2] See Appendix 1 for a partial timeline of MARC history.

"shared body of knowledge that will enable all of us to see farther," we must first recognize the essential need for archaeological work, for that excavating at the roots of our contemporary understandings, our contemporary "needs," and our contemporary "desires" (Rumsey 81). Following Jonathan Sterne's *MP3: The Meaning of a Format* (2012), we will place the history of MARC into a *general history of compression*, exploring some of the "big questions" that live inside a diminutive, diminishing format (17). Sterne elaborates, justifying his, and our, historical approach:

> As people and institutions have developed new media and new forms of representation, they have also sought out ways to build additional efficiencies into channels and to economize communication in the service of facilitating greater mobility. These practices often begin close to economic or technical considerations, but over time they take on a cultural life separate from their original, intended use (5).

Ultimately, by dwelling upon some of these big questions, we will search for answers: why has this important history—of a complex infrastructure of people, artifacts, and institutions, generating and sharing knowledge in new, interconnected ways—been relegated to the sidelines of 'computer networking' history; and, what are the consequences of such a dismissal?

This is the prehistory of modern librarianship, presented to us in a series of fascinating primary source materials. This is also a history largely forgotten, as many librarians have allowed formative, still-relevant works such as *ISS (Information Systems Specialist) Planning Memorandum Number 3: A Proposed Format for a Standardized Machine-Readable Catalog Record* (1965), *The MARC II Format: A Communications Format for Bibliographic Data* (1968), *MARC Manuals* (1968), *Guidelines For Library Automation: A Handbook for Federal and Other Libraries* (1972), and *MARC: Its History and Implications* (1975) to be shuttled off to off-site storage facilities.[3] Rather than behave as if the data management challenges of today are inherently and profoundly different from those faced by our counterparts in the past, we should begin mining these overlooked, abandoned materials, searching for potential insight into contemporary dilemmas. The "dominance over data" worldview that governed the creation and implementation of MARC continues to hold sway today, and much of the language of these texts calls to us with an eerily familiarity: they encourage us to "relieve humans of the drudgery" of data management, to make our information "more rapidly available" by "providing more access points to the data" (Avram 18). Yet, despite the symmetry of these past and present impulses, MARC does remain unique, particularly in its straddling of two different visions of what computers are all about—the "1960s notion of computers as centralized and centralizing calculation and management devices," and our more contemporary understanding of

---

[3] For a better sense of MARC's role in library automation, see Appendix 2 for Automated Library Flowcharts from *Guidelines for Library Automation*.

computers as "distributed communication tools" (Streeter 5). As a reshaping of the politics of information, MARC resists easy categorization, never fully aligning with the way many Americans felt about computers during the 1960s, summarized by Fred Turner in *From Counterculture to Cyberculture: Stewart Brand, The Whole Earth Catalog, and the Rise of Digital Utopianism* (2006): "Computers loomed as technologies of dehumanization, of centralized bureaucracy and the rationalization of social life, and, ultimately, of the Vietnam War" (2). Despite the profound social unrest of the era, the country's careening "from domestic problem to foreign crisis and back again," it is simply too easy to portray MARC as merely a homogenizing effort to usher in "conformity" to standards (Sterling and Kittross 2002: 407). By resisting this facile interpretation, and beginning to explore MARC's gray areas, we can see that MARC was a signal episode in what Lisa Gitelman and Victoria Jackson have termed the "acceleration of data in recent history"(2). This was a moment in which knowledge (bibliographic data) was compressed (into MARC records) for transmission (via magnetic tape), all for one overarching, all-important purpose: the creation of new, networked communities.

Indeed, it is the centralized/distributed ambiguity at the core of MARC that is perhaps its most interesting element. Despite its ubiquity, and near half century of dominance, in many ways MARC continues to hide in plain sight. In 2013, as the Library of Congress begins carefully extricating itself from MARC, pursuing the promise of Linked Data through the Bibliographic Framework (BIBFRAME) Initiative, the benefit of pausing briefly to consider the origins of MARC should be clear—contemporary information management problems *and* solutions both arise from earlier entanglements of humans and technologies (Owens, Beaton, and Langmead). Inspired by Thomas Streeter's *The Net Effect: Romanticism, Capitalism, and the Internet* (2011), MARC should be understood as "a kind of social philosophy in practice, as much a product of social visions as it was of technical and economic necessities" (3). Streeter elaborates:

> Contemporary computing…is in an important way the product of a gradual accumulation of social and cultural choices, choices among competing visions of computers' purposes and social capacities. These choices, in turn, typically rest on those collections of tacit assumptions that power social relations (8).

By blending the meanings attached to MARC, its technical functions, and the social uses to which it was put, we can begin re-envisioning MARC, considering it not as an inevitable consequence of the drive for automation that characterized the era, but rather as something worked out over time, in the context of considerable cultural debate.[4] Pushing past the simplistic notion of causality epitomized by one librarian's response to the 1970s survey that formed the backbone of *Guidelines for*

[4] This essay will regrettably refrain from in-depth analysis of individual MARC records. Roy Tennant, in "MARC Must Die" (2002), offers partial explanation: "There are only two kinds of people who believe themselves able to read a MARC record without consulting a stack of manuals: a handful of our [The California Digital Library's] top catalogers and those on serious drugs" (26). Though MARC is historically significant, it is arcane and, in some respects, incomprehensible in its details.

Library Automation—"the heavies were all for automation"—this more nuanced understanding allows us to recast MARC's "contradictions and misrecognitions"—its not entirely seamless implementation and its problematic, potentially universalizing agenda (Markuson et al. 12; Streeter 3).

While MARC should be considered as an organizational byproduct of a specific community assessing its needs within a particular social and historical moment, it must also be situated in broader terms, as an early example of 'computer networking,' an artifact of post-World War II efforts to manipulate technological advances to connect and share information assets across space and time. Emerging out of the same Cold War "military-industrial-university complex" that gave contemporaneous birth to the Internet, at times the list of individuals and organizations connected to the origins of MARC reads like a *Who's Who* of government security agencies: Avram, often referred to as the "Mother of MARC," spent years working at the National Security Agency (NSA); Samuel Snyder, the coordinator of the LC's information systems from 1964-1966, worked as a code-breaker at the Signal Intelligence Service (SIS); Verner Clapp, a longtime employee of the LC and president of the CLR during the MARC era, filled the board of directors at the CLR with a number of "extremely bright war scientists and CIA consultants;" the System Development Corporation, responsible for the publication *Guidelines for Library Automation* (1972), evolved out of the RAND Corporation (the private research wing of the Air Force), which made key software contributions to Project SAGE, the 1950s nuclear war early warning system that has been described as the "the first geographically distributed, online, real-time application of digital computers in the world" (Streeter 26; Baker 85; *IBM 100 "SAGE")*. For his part, Clapp was also responsible for perhaps the apotheosis of this alignment of library automation efforts and the "cold war visions that underwrote the Internet's early development": in the early 1960s, Clapp would hire J.C.R. Licklider, the now famous creator of ARPANET (the "Pentagon's precursor to the Internet"), to explore the effects of burgeoning technologies on librarianship in *Libraries of the Future* (1964) (Streeter 7; Baker 90). Licklider's work was likely an influence on the developers of MARC, whose *ISS Planning Memorandum Number 3* from the following year offers this seemingly indebted passage:

> The long-term objective of a time-shared system is the 'library of the future.'[5] Here we are concerned with hardware, large memories, on-line input and output devices, communications networks, etc. From the system point of view, there are problems of data organization, complexities of multi-programming in a real-time environment, dynamic allocation of storage, queuing theory, list processing, and many more (Avram, Freitag, and Guiles 6).

Though we should refrain from overemphasizing the connections between the pioneers of the Internet and those who were rethinking library operations in the early digital era, it is critical that both efforts be seen as attempts to form kinds of virtual communities (Baker "Discards" 73). Rather than consider MARC as simply an extension of the LC Card Catalog System, which began widely distributing uniform bibliographic catalog records in 1901, we must again return to the communicative function of MARC—its most prescient *and* overlooked quality. From the very beginning, Avram and her compatriots recognized that the future of librarianship would revolve around networking; in developing a common language, what they called a shared "communications format," these employees of the LC made a valuable contribution to library practice while simultaneously carving out a unique space for themselves in the highly gendered field of early computing (Avram, Knapp, and Rather 2).

Unlike the first generation of computer programmers—a group mostly comprised of women who were available to work during World War II, considered by many to be "better, and more conscientious computers, presumably because they were better at repetitious, clerical tasks"—the women leading the charge of the MARC initiative prompt us to add an interesting new layer to the historical narrative encapsulated by the title of Jennifer Light's 1999 article, "When Computers Were Women" (Chun 33; Light 455). While Light and other scholars have reclaimed the pathbreaking work of women, challenging the "male-centered" and "distorted" histories of technological development that have largely omitted the contributions of women, the work of Avram, Guiles, and Freitag directs us to push even harder, contesting not only the omission of women from computer history, but also the tacit power dynamics lurking behind stereotypical depictions of women working submissively as either "computers" or librarians (Light 475). These particular

---

[5] In the 1960s, the era of mainframe computers, time-sharing, or multiple users "sharing" the time (computational cycles) of a single computer, was a concept pursued at a number of institutions, notably MIT, UC Berkeley and Bell Laboratories. Time-sharing presented a host of technological hurdles, and, as Paul Ceruzzi describes in *A History of Modern Computing* (1998), the idea of time-sharing went "beyond the interactive nature of SAGE, which allowed for multiple users of one and only one data set, and beyond NASA's real-time systems, which restricted users to both specialized data sets and programming languages" (154). After much experimentation, the solution to time-sharing proved to be time itself: multi-programming was made possible by "the disparity between the few milliseconds (at least) between a typist's keystrokes and the ability of a computer to fetch and execute dozens, perhaps hundreds, of simple instructions" (155).

women challenge us to imagine a society capable of seeing individuals as more than one-dimensional cutouts. Far from subservient, and far from simply following orders, these women were firmly in charge of their work, firmly in charge of the development of their own descriptive language of "command and control" (Chun 29).

While one could object to the effusive terms often bandied about to describe either MARC's rapid development or its longevity, it is harder to deny MARC's utterly pervasive takeover of American librarianship. In "LC's Bibliographic Framework Initiative and the Attractiveness of Linked Data" (2013), Kevin Ford paints an extreme, though largely accurate picture of MARC, describing it in near-familial terms: "[MARC] permeates everything in the library community: it is embedded in the minds of most librarians…it is the butt of jokes; it is the topic of conversations; it is worried about; it is cared for; it is loved; it is hated—it is hard to envision life without MARC" (46). How did this 1960s effort to structure bibliographic data for machine readability so thoroughly infect contemporary library practice? While certainly not an impartial source, Avram's definition of MARC in *MARC: Its History and Implications* (1975) offers some intriguing clues into MARC's rampant spread: "MARC is an assemblage of formats, publications, procedures, people, standards, codes, programs, systems, equipment, etc. that has evolved over the years, stimulating the development of library automation and information networks" (31). Again, it is Avram's cybernetic prescience—her ability to foresee that an "assemblage" of human *and* technological actors would be needed to stimulate the continued evolution of bibliographic information—that creates an uncanny sort of symmetry, connecting the LC MARC initiative of the late 1960s-early 1970s to contemporary digital preservation efforts. Today, digital preservation is considered "successful" only when it accounts for both the human *and* technological components of a digital preservation system. As Brian Lavoie and Lorcan Dempsey explain in "Thirteen Ways of Looking at…Digital Preservation" (2004): "Preserving our digital heritage is more than just a technical process of perpetuating digital signals over long periods of time; it is also a *social and cultural process*…" (16-17, emphasis added). As the LC contemplates ways of phasing out MARC and ways of embracing the technological possibilities presented by Linked Data (neatly summarized by Ford as "publishing structured data over the same protocol used by the World Wide Web and linking that data to other data to enhance discoverability of more information"), it is essential for us to pause and consider the "social and cultural process[es]" that play critical roles in how we construe, or "imagine" data. As David Ribes and Steven J. Jackson explain in "Data Bite Man: The Work of Sustaining a Long-Term Study" (2013), "the work of producing, preserving, and sharing data reshapes the organizational, technological, and cultural worlds around them" (147). In this way, MARC's saturation of "everything in the library community" operates structurally as both cause and also effect. Daniel Rosenberg, in "Data Before the Fact," (2013), describes this as the "specifically rhetorical" function of data: though data "has no truth" ("false data is data nonetheless"), it does help us construct "our reality" (18; 37). The ways we conceive of data, like the ways we conceive of computing (and, like the ways we conceive of MARC) are not simply

the result of innocuous technological advances; they are, once again, the result a "gradual accumulation" of social and cultural choices (Streeter 8).

In an earlier moment, in 1965, Avram displayed an already firm grasp on the potential of computerization to reshape conceptions of data:

> One may regard the machine processing of cataloging data as providing a better means of doing essentially the same thing, with some added capabilities, or one may regard machine processing as providing a wide new range of possibilities. The assumption underlying the latter approach is that there is a new body of data which, once described and defined, *can be manipulated in new ways* (Avram, Freitag, and Guiles 2, emphasis added).

While it is perhaps Avram's characteristically circumspect approach that is most on display here, we can also see how MARC—an effort to manipulate bibliographic data, making it legible to computers by adding numerical tags to fixed- and variable-length data fields—at times seems to foreshadow the search for its own replacement. But before turning fully to the BIBFRAME Initiative, which offers the promise of even newer ways of manipulating bodies of data, we must ask: what was lost, and what was gained, in this earlier transition from handwritten, or individually typed catalog cards, to machine-readable, machine-processable MARC records?

Though a radical vision—a profoundly altering re-imagination of data—to contemporary audiences, the early MARC experiments likely appear either hopelessly antiquated and out-of-touch or entirely predictable and unsurprising. And to readers who have never known a world without computers, painstakingly detailed tomes such as *MARC Manuals* (1968) and *Guidelines for Library Automation* (1972) likely appear equally inconsequential, the relics of an earlier age. But by questioning this casual dismissal, and considering these texts as a corpus—an archive of thought—we can reassert MARC's cultural significance, its status as a signal moment in the both history of the "acceleration of data" and the history of electronic communication. While radical librarians or catalogers such as Sanford Berman (author of *Prejudices and Antipathies: A Tract on the LC Subject Heads Concerning People* (1971)) would likely object to the ideological work of MARC—that compression or elision that occurs when knowledge is transformed into grist for the mill of "data," and alternative modes of description are deemed immaterial, MARC's communicative function reminds us that a more measured, considered approach is needed. As Berman acolytes Bradley Dilger and William Thompson point out in "Ubiquitous Cataloging" (2008), "The catalog itself is a technological infrastructure that shapes what can and cannot be represented, not a transparent entity which passes information to users without value judgments" (44). At the same time, the limitations or spatial constraints of MARC records impose a sort of artificial constraint on thought itself. Nevertheless, the spread of MARC served to erode barriers, and bring new, previously unforeseen possibilities into existence. Nicholson Baker's mercurial treatment of MARC in *Double Fold* (2001) and "Discards" (1994) serves as testimony to the difficulty of reconciling MARC's

twinned (centralized/distributed; compressed/communicative) functions. With MARC, the establishment of greater bibliographic control will always sit somewhat uncomfortably alongside the format's more democratically inclined leanings; while in one moment Baker derides MARC, calling it "a daunting set of numbered fields and odd symbols…redolent of unfriendly first-generation database interfaces;" in the next he is cowed by its egalitarian open-endedness, offering, "What began mainly as a handy unilateral way of delivering the Library of Congress MARC files to member libraries turned into a highly democratic, omnidirectional collaboration among hundreds of thousands of once isolated documentalists" (73). As we move from the digital past of libraries toward the digital future of libraries, we must keep in mind that data, as Gitelman and Jackson tell us, "need to be understood as framed and framing, understood, that is, according to the uses to which they are put" (Gitelman and Jackson 5). MARC reminds us that while the uses of data may be divergent, the matter may ultimately be less *either/or*, more *and yes* (5).

Pivoting from MARC to the nascent BIBFRAME Initiative, it is immediately apparent that the future of bibliographic control will lie at least partially in the networked past of libraries. Surveying BIBFRAME-related literature, it appears that sometime between the "bold step" of the late 1960s and today, the library community retrenched, narrowing its focus by "tend[ing] to equate bibliographic control with the production of metadata solely for use within the library catalog" ("On the Record" 1998: 31). The reasons for this insularity are unclear; it was likely the result of a combination of factors: complacency; an overdeveloped bibliographic infrastructure not amenable to change; and ingrained cataloging practices focused on the physical item rather than the intellectual one (Tennant "Twenty-First" 175). BIBFRAME is portrayed as an effort to rectify these myopic tendencies, redesigning bibliographic data models as a means of rededicating the library community to the very idea of the "library of the future." By mirroring the "extraordinary achievement" of MARC, and by making "'the network' central, and interconnectedness commonplace," BIBFRAME will strive to accomplish the following objectives:

1) Differentiate clearly between conceptual content and its physical/digital manifestation(s);
2) Unambiguously identify information entities (e.g., authorities); and
3) Leverage and expose relationships between and among entities (*BIBFRAME FAQ*).

Based in part upon the Functional Requirements for Bibliographic Relationship (FRBR) entity relationship model (a "cataloging code or implementation"-agnostic conceptual model for bibliographic control), BIBFRAME will make use of the World Wide Web Consortium's (W3C) Resource Description Framework (RDF) to provide "all entities (resources), attributes, and relationships between entities (properties)" with unique identifiers as Web resources ("Bibliographic Framework as a Web of

Data" 2012: 9).[6] In clearer terms, BIBFRAME will mark a "shift from capturing and recording descriptive details about library resources to identifying and establishing more relationships between and among resources" (4). Tim Berners-Lee, inventor of the World Wide Web, and vocal proponent of Linked Data, boils this down even further: "It's not the documents, it is the things they are about which are important" (cited in "Web of Data" 6).

In "LC's Bibliographic Framework Initiative and the Attractiveness of Linked Data," Ford describes how RDF (a syntax which consists of relationships expressed in grammatical triples, i.e. a subject, a predicate, and an object) will contribute to this "linking" of data:

> Borrowing an analogy from English grammar, the parts of an RDF statement can be equated to those found in a basic linguistic sentence…the subject is a uniquely identified concept or thing (preferably with an HTTP URI, a uniform resource identifier), about which the statement is made. The other two parts are called the predicate (like a verb) and object. The predicate—also identified with a URI—records the relationship between the subject and object. The object may be identified with a URI or it may be a string (47).

By adapting to RDF, libraries will make bibliographic information more machine-readable, or, better yet, more machine-interpretable. In *Semantic Digital Libraries* (2009), Bill McDaniel and Sebastian Ryszard Kruk explain why this linking of data sets will prove a sea change in the structuring of the Internet: "As much as 80% of the information already stored on the Internet is unstructured data, locked up in documents, photos, html pages, and other formats….the interrelationships between all this information is implicit, largely locked up in our human languages and shared experience" (4). Linked Data will make explicit that which is implicit; in the case of libraries, the information held in MARC records will be "assimilated," "deconstructed," or "converted," transformed for use in a Web-scale environment (Tennant "Twenty-First Century" 176; Kroeger 882; "Web of Data" 8).

The extraction of this valuable data from MARC records will be a painstaking process, with widespread ramifications. But the need is clear. MARC is antiquated and unsuitable for contemporary use; as Roy Tennant explains, the library community needs a "broader, richer, more diverse set of tools, standards, and protocols" (Tennant "Twenty-First Century" 176). A partial list of the problems associated with overreliance on MARC includes:

1) No other community uses MARC, severely compromising its utility to other communities as a data transmission tool. MARC is not particularly

---

[6] "FRBR is a conceptual model for describing relationships between bibliographic entities, defined as *Work, Expression, Manifestation,* and *Item* (WEMI)" (Kroeger 882). BIBFRAME proposes a similar model, described by Eric Miller of Zepheira as "very light FRBR-esque" (Cited in Kroeger 882). Entities will be defined as *Creative Work (Work), Instance, Authority,* and *Annotation* (WIAA). ("Web of Data" 8).

interoperable: it is not easy to share information, particularly with non-library systems ("On the Record" 24).

2) MARC is rigid and riddled with irregularities, creating problems for catalogers and users (Tennant "MARC Must Die" 26).
3) MARC lacks the essential checks and balances needed to ensure appropriate granularity (the level of detail in an information set) (Tennant "MARC Must Die" 26).
4) MARC lacks the versatility and extensibility of more modern metadata infrastructures (Tennant "Twenty-First Century" 176)
5) MARC is difficult to learn: few programmers can or want to work with it (Krier "What's Wrong with MARC?").

Communication will play a critical role in this massive rethinking of library data management practice. In "Sustainable Economics for a Digital Planet" parlance, this is a transition that will require "mobilizing resources—human, technical, and financial—across a spectrum of stakeholders diffuse across space and time" (Rumsey 1). Interestingly though, it is the "time" portion of this equation that is often overlooked. Communication must incorporate some form of communion with our counterparts in the past. There is value to be found in looking to this earlier moment, when stakeholders representing varying interests worked together to achieve goals. Only by following this example will BIBFRAME, and any other digital preservation project, be able to reach its fullest potential.

We must remove the blinders separating us from the past, and recognize that our technological present is, in some sense, the realization of a "gradual accumulation" of social and cultural choices (Streeter 8). By investigating the social construction of MARC, we have uncovered the complex interplay of history, that push-pull tension between the arbitrary and the intentional, given shape, form, and sense of inevitability only in retrospect. As historians, as librarians, or as digital preservationists, we must strive, when we return to our contemporary vantage point, to challenge our comfortable perch in the present, questioning the decisions we make as we search for solutions to the "problems" posed by data. As we enter this "brave new metadata world," Tennant, who sounded the clarion call to replace MARC over a decade ago, helps us conclude on an appropriately contemplative note:

> "Having not been a part of the effort to create MARC those many decades ago, I cannot imagine what conditions fostered its birth. But in my ignorance I imagine that the opportunities created by computers inspired Henriette Avram and company to rise to the challenge of recreating our professional infrastructure in a revolutionary and farsighted way. We would do well to look to our past for the inspiration we need to create a future that our descendants will look back upon with a similar amazement" ("Twenty-First Century" 181).

Appendix 1: A MARC Timeline

1961: CLR (Council on Library Resources) funds a feasibility study to determine the possibility of automating LC (Library of Congress) functions

1963: The feasibility study, performed by consultant Gilbert King, is released, recommending automation

1964: CLR awards Lawrence Buckland of Inforonics, Inc. a contract to design a system for converting LC card catalog system to machine-readable form

1965: LC, CLR, and Committee on Automation of the Association of Research Libraries (ARL) hold automation conference (Jan.), LC taps Henriette Avram, Kay Guiles, and Ruth Freitag to build upon Buckland's efforts; First draft of a proposed format completed (Jun.); CLR awards the LC $130,000, funding the MARC Pilot Project (Dec.)

1966: Planning begins for MARC Pilot Project (Jan.); 16 participating libraries are selected (Feb.); First test tapes mailed to participants (Oct.); Weekly tape distribution begins (Nov.)

1967: MARC II is introduced at ALA (American Library Association) (Jun.); Ohio College Library Center opens

1968: Pilot Program (also known as MARC I) is completed, 50,000 records distributed (Jun.); "MARC Institutes" workshops held throughout the country

1969: MARC II becomes operational, with tape distribution service (Mar.)

1970: LC publishes formats for serials and maps (in accordance with original plan to specify MARC for material other than books)

1971: LC publishes format for films; MARC becomes an ANSI (American National Standards Institute) standard (Z.39.2-1971)

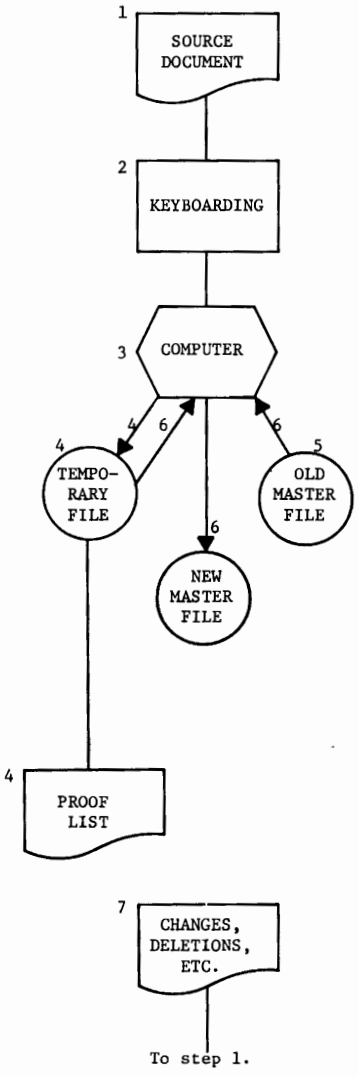
1973: LC publishes format for maps; MARC becomes an ISO (International Organization for Standardization) standard (ISO 2709-1973)

1975: LC publishes format for music and sound recordings

1999: USMARC (United States) and CAN/MARC (Canada) harmonized and republished as MARC21

2004: British Library adopts MARC21 as a replacement for UKMARC (United Kingdom)

Appendix 2: Library Automation Flowcharts from *Guidelines for Library Automation: A Handbook for Federal and Other Libraries* (1972)



1. The source document has the information on it that is to be put into machine-readable form. It could be a catalog work sheet, order form, etc. The data will be completely or partially tagged, coded, and fielded.

2. The information on the source document is punched up on equipment that can convert data to machine-readable form--e.g., keypunch, punched-paper-tape typewriter, on-line terminal, etc. Some tagging and fielding can be done by the keyboarder.

3. The records are read into the computer via a conversion reader (e.g., card reader, paper tape reader) and are edited and formatted by the computer. Both the data and the computer program for this operation are in core storage of the computer.

4. New records are usually put in a separate temporary file and are printed out for proofing and verification before the master file is updated.

5. The master file consists of all verified, active records organized in some logical sequence. This file is comprised of all the records that relate to some function or group of functions, e.g., master catalog file, master serial holdings file.

6. When the new records have been approved they are merged with the master file and a new updated file is written.

7. All actions to the file--correction, addition, deletions, are entered through the same process and new records, i.e., by keyboarding all or part of the record involved. This activity is called FILE MAINTENANCE.

8. Master files should be protected by back-up files and/or periodic listings of all contents. Tape and disc files are "cataloged" to identify name of file, approved users, date file written, etc.

9. File searches are formatted (the search request is the source document) keyboarded and input just as the data records are.

Figure II-16. Summary of File Maintenance Operations

Appendix 2: Library Automation Flowcharts from *Guidelines for Library Automation: A Handbook for Federal and Other Libraries* (1972)

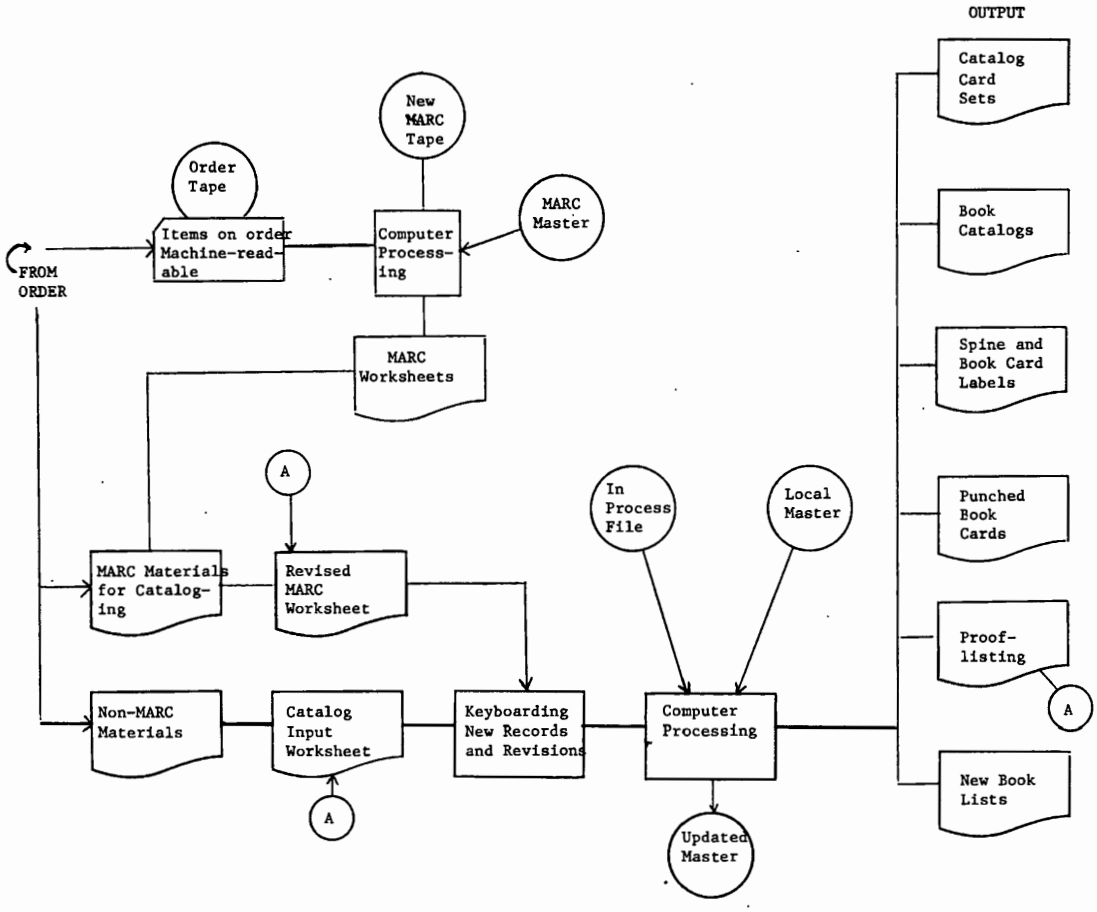


Figure II-23. Simplified Flowchart of Automated Cataloging System

Bibliography

Avram, Henriette. *MARC: Its History and Implications*. Washington, D.C.: Library of Congress, 1975.

Avram, Henriette, Ruth S. Freitag, and Kay D. Guiles. *ISS Planning Memorandum Number 3: A Proposed Format for a Standardized Machine-Readable Catalog Record*. Washington, D.C.: Library of Congress, 1965.

Avram, Henriette, John F. Knapp, and Lucia J. Rather. *The MARC II Format: A Communications Format for Bibliographic Data*. Washington, D.C.: Library of Congress, 1968.

Baker, Nicholson. "Discards." *The New Yorker*. April 4, 1994: 64-86.

----. *Double Fold: Libraries and the Assault on Paper*. New York: Randomhouse, 2001.

Berman, Sanford. *Prejudices and Antipathies: A Tract on the LC Subject Heads Concerning People*. Metuchen, New Jersey: The Scarecrow Press, Inc., 1971.

Ceruzzi, Paul E. *A History of Modern Computing*. Cambridge Massachusetts: The MIT Press, 1998.

Chun, Wendy Hui Kuyong. "On Software, or the Persistence of Visual Knowledge." *Grey Room*. Winter 2004, (18): 26-51.

Dilger, Bradley and William Thompson. "Ubiquitous Cataloging." *Radical Cataloging: Essays at the Front*. Ed: K.R. Roberto. Jefferson, North Carolina: McFarland and Company, 2008.

Douglas, Mary. *How Institutions Think*. Syracuse, New York: Syracuse University Press, 1986.

Ford, Kevin. "LC's Bibliographic Framework Initiative and the Attractiveness of Linked Data." *ISQ: Information Standards Quarterly*. Spring/Summer 2012, 24 (2/3): 46-50.

Gitelman, Lisa and Victoria Jackson. Introduction. *"Raw Data" Is an Oxymoron*. Ed:

Lisa Gitelman. Cambridge, Massachusetts: The MIT Press, 2013.

Lavoie, Brian and Lorcan Dempsey. "Thirteen Ways of Looking at…Digital Preservation." *D-Lib Magazine*. July/August 2004, (10) 7/8: 1-18.

The Library of Congress. "Bibliographic Framework as a Web of Data: Linked Data Model and Supporting Services." November 21, 2012.

----. *BIBFRAME: Frequently Asked Questions*.

The Library of Congress Working Group on the Future of Bibliographic Control. "On the Record." January 9, 2008.

Light, Jennifer. "When Computers Were Women." *Technology and Culture*. 1999, (40) 3: 455-483.

Markuson, Barbara Evans, Judith Wanger, Sharon Schatz, and Donald V. Black. *Guidelines for Library Automation: A Handbook for Federal and Other Libraries*. Falls Church, Virginia: System Development Corporation, 1972.

McDaniel, Bill and Sebastian Ryszard Kruk. Eds. *Semantic Digital Libraries*. New York: Springer Publishing, 2009.

Owens, Trevor. "Historicizing the Digital for Digital Preservation Education: An Interview with Alison Langmead and Brian Beaton. *The Signal*. Library of Congress, May 6, 2013.

Ribes, David and Steven J. Jackson. "Data Bite Man: The Work of Sustaining a Long -Term Study." *"Raw Data" Is an Oxymoron*. Ed: Lisa Gitelman. Cambridge, Massachusetts: The MIT Press, 2013.

Rosenberg, Daniel. "Data Before the Fact." *"Raw Data" Is an Oxymoron*. Ed: Lisa Gitelman. Cambridge, Massachusetts: The MIT Press, 2013.

Rumsey, Abby Smith. "Sustainable Economics for a Digital Planet: Ensuring Long-Term Access to Digital Information." *The Blue Ribbon Task Force on Sustainable Preservation and Access*. February, 2010.

Seikel, Michelle and Thomas Steele. "How MARC Has Changed and Its Forthcoming

Relationship with RDA." *Technical Services Quarterly*. July 2011, (28) 3: 322-334.

Snyder, Samuel. Foreword. *ISS Planning Memorandum Number 3: A Proposed Format for a Standardized Machine-Readable Catalog Record*. Washington, D.C.: Library of Congress, 1965.

Sterling, Christopher H. and John Michael Kittross. *Stay Tuned: A History of American Broadcasting*. Mahwah, New Jersey: Lawrence Erlbaum Associates, 2002.

Streeter, Thomas. *The Net Effect: Romanticism, Capitalism, and the Internet*. New York: New York University Press, 2011.

Tennant, Roy. "MARC Must Die." *Library Journal*. October 15, 2002: 26-27.

----. "A Bibliographic Metadata Infrastructure for the Twenty-First Century." *Library Hi Tech*. 2004, 22(2): 175-181.

Turner, Fred. *From Counterculture to Cyberculture: Stewart Brand, the Whole Earth Catalog, and the Rise of Digital Utopianism*. Chicago: The University of Chicago Press, 2006.