

Formato MARC: um paradigma a ser quebrado!
Elysio Mira Soares de Oliveira

“Na natureza não são os mais fortes nem os mais inteligentes que sobrevivem e sim aqueles com maiores capacidades de se adaptarem às mudanças” (Charles Darwin - 1859)

Meu primeiro contato com o "computador" se deu em 1973. Estava cursando o primeiro ano de Biblioteconomia e numa visita à biblioteca do Instituto de Matemática da EESC-USP, em São Carlos, conheci um catálogo composto por cartões perfurados. Completamente fora de todos os padrões que estávamos aprendendo na escola, mas muito funcional (os usuários adoravam). Em cada cartão havia somente uma linha (80 caracteres) com duas informações: entrada principal e localização. Havia uma gaveta com cartões de títulos, outra de assuntos e outra de autores. No final de cada dia, a bibliotecária inseria os novos cartões nas respectivas gavetas, em qualquer ordem e colocava todo o conteúdo da gaveta em uma máquina classificadora de cartões. A agilidade da "classificadora de cartões" na ordenação dos cartões me impressionou.

Na disciplina de Documentação, ministrada pelo saudoso Prof. Alfredo Américo Hamar, fomos inteirados sobre o formato de armazenamento e intercâmbio de registros bibliográficos criado pela Biblioteca do Congresso Americano (*Library of Congress*). Este formato permitia armazenar em fitas magnéticas, dados bibliográficos coletados a partir de cartões ou fitas perfuradas. Os cartões eram lidos em um leitor e os dados armazenados em fita magnética. Um software formatava estes dados para impressão das fichas catalográficas, iguais àquelas que éramos obrigados a datilografar e depois intercalar manualmente no fichário. Foi meu primeiro contato com o formato (*MARC - Catálogo Legível por Máquina*).

Ainda achava o método do Instituto de Matemática mais eficiente. Não via nenhuma outra inovação neste processo a não ser o de manter uma cópia do catálogo, armazenado na fita magnética, para possível intercâmbio. A rigidez das regras e padrões para a impressão das fichas catalográficas exigiam do bibliotecário, conhecimentos apurados no uso de espaçamentos e pontuações e obrigavam um alto grau de redundância na transcrição de elementos de dados para o desmembramento da ficha matriz em suas respectivas fichas secundárias.

Em 1978, quando passei a integrar a equipe encarregada da mecanização do catálogo da biblioteca do antigo IEA (*Instituto de Energia Atômica*), hoje IPEN (*Instituto de Pesquisas Energéticas e Nucleares*), me interei sobre um projeto da FGV-Rio da adoção do formato CALCO, uma variante brasileira do formato MARC.

Nesta ocasião a DIDC/IEA já tinha seus catálogos mecanizados utilizando outro formato, o ILS (Integrated Library System). Este formato utilizava as regras do AACR2 para a padronização dos elementos de dados e trazia uma novidade sobre o MARC, ele não estava preso às regras de espaçamento e pontuações requeridos pelo MARC e nem exigia redundância na transcrição de elementos de dados para o desdobramento das fichas catalográficas.

Por iniciativa e coragem da bibliotecária Terezine Arantes Ferraz, foi quebrado no Brasil um grande paradigma, o catálogo deixou de ser constituído de fichas e passou a ser

apresentado em papel contínuo, em um formato mais linear e econômico que aqueles recomendados pela AACR2 para as fichas catalográficas.

Neste mesmo ano, no IEA, tive o primeiro contato com um outro formato, o formato ANSI, desenvolvido pelo "American National Standard Institute", não mais para armazenamento de registros bibliográficos em fitas magnéticas para mecanização do processo de impressão de catálogos, como o MARC ou o ILS, mas sim para armazenamento do registro bibliográfico em uma nova mídia.

O novo catálogo "eletrônico" permitia que a informação de interesse fosse encontrada por qualquer palavra de qualquer elemento de dado ou por uma expressão booleana composta por conjuntos de palavras ou descritores digitados em uma linha de comando (prompt). Nesta nova mídia, a classificação (*CDU, Dewey, LC, etc.*) e a redundância na transcrição de dados, necessárias para a geração de entradas para os índices dos catálogos impressos, deixaram de ser necessárias para a localização da informação no catálogo. O conceito de entrada primária, secundárias, título uniforme, formas variantes do título, título para alfabetação, entre outros, perderam o sentido face aos novos recursos de acesso aos dados nesta nova mídia.

A IBM passa a comercializar o STAIRS©IBM (Storage and Information Retrieval System). Surgem os primeiros softwares de gestão de bases de dados, novos formatos de armazenamento de dados e com eles novas estruturas e formatos de registro bibliográfico.

O foco do processamento sofre uma mudança radical, os detalhes necessários para a descrição do suporte na catalogação convencional, voltada à produção de índices e catálogos impressos dão lugar à descrição de conteúdo. O registro bibliográfico é incrementado com resumos, sumários e descritores de conteúdo. Novas chaves de acesso são criadas, permitindo ao usuário definir à posteriori sua classe de interesse e obter a informação desejada de forma mais rápida e precisa.

As vantagens oferecidas por esta nova mídia, levaram os grandes sistemas de informações bibliográficas, a adotarem formatos mais eficientes para a alimentação de suas bases de dados. Órgãos do próprio governo norte-americano, como a USAEC (United States Atomic Energy Commission), que requeriam sistemas mais rápidos e precisos para o armazenamento e recuperação da informação, ignoraram o MARC e adotaram novos formatos como o ANSI adotado pela USAEC para a descrição bibliográfica do acervo de informações nucleares (*Nuclear Science Data Base*) e o formato ELHILL utilizado pela "National Library of Medicine" para a base de dados MEDLINE.

Bibliotecas e sistemas de informações de órgãos das Nações Unidas, face às facilidades apresentadas por esta nova mídia, também decidiram por padrões mais eficientes que aqueles criados para mecanização do processo de impressão de catálogos (*MARC, ILS, etc.*).

A concepção da catalogação tradicional, voltadas à produção de catálogos impressos começa a ser questionada. Algumas bibliotecas que ainda não haviam iniciado seu processo de automação passam a desenvolver sistemas fundamentados nestes novos formatos.

A eficiência do formato MARC, mundialmente difundido como formato de intercâmbio de registros bibliográficos em meio magnético e de mecanização do processo de impressão de fichas catalográficas, passa a ter sua eficiência questionada como formato de registro bibliográfico para composição de bases de dados de acesso em linha (*on line*).

As empresas de desenvolvimento de softwares comerciais se mantiveram alheias a estas discussões e persistiram na utilização do MARC, não pela sua eficiência, mas pelo simples fato de que seu mercado potencial se constituía de bibliotecas públicas e universitárias que já vinham produzindo seus catálogos mecanicamente a partir de registros armazenados em fitas magnéticas utilizando o MARC. Estas empresas encontraram nestas bibliotecas, um terreno fértil para o comércio de seus produtos. Já, entre as bibliotecas especializadas, as soluções fundamentadas no formato MARC não tiveram grandes mercados, pois estas preferiam soluções mais ágeis e eficientes.

Na década de 1970, a OIT (Organização Internacional do Trabalho), percebendo as facilidades desta nova mídia, inicia um projeto de desenvolvimento de um software de gestão de bases de dados voltada à gestão de dados bibliográficos. Em meados de 1975 este projeto é transferido para a Unesco e se dá o surgimento do CDS/ISIS (*Computerized Documentation System - Integrated Set for Information System*).

Surge um novo conceito para a gestão da informação, permitindo registros bibliográficos de tamanho variável, campo de dado repetitivo, acesso através de índice no formato de árvore balanceada (B-tree file), e exportação de registros em um padrão de intercâmbio, o ISO-2709.

Em 1972 o UNISIST desenvolve um manual de referência para registros legíveis por máquina, com recomendações submetidas à avaliação do curso de pós-graduação da "School of Librarianship and Information Science" da "University of Sheffield (UK)" e a grupos internacionais de bibliotecas e de agências de serviços de informação. Em 1974 estas recomendações foram incorporadas pelo WGBD (Working Group on Bibliographic Descriptions). Em 1976, a "British Library" em cooperação com a UNESCO, estabelece o UNIBID (*UNISIST International Centre for Bibliographic Description*) e publica a primeira edição do "UNISIST Reference Manual for Machine-Readable Bibliographic Description) com a colaboração de várias entidades: ICSU-AB (*Int'l Council of Scientific Unions Abstracting Board*), ISO (*International Standard Organization*), FID (*International Federation for Information and Documentation*), IFLA (*International Federation of Library Associations*), IATUL (*International Association of Technological University Libraries*), INIS (*International Nuclear Information System*), WIPO (*World Intellectual Property Organization*).

Estas recomendações foram logo assimiladas pelos sistemas de informação e bibliotecas especializadas das Nações Unidas (IAEA, OIT, Unesco, PAHO, FAO, CEPAL, CELADE, etc.).

No Brasil, em 1982, por iniciativa do GEADE/SEADE (*Grupo Especial de Análises Demográficas da Fundação SEADE*), dirigido pela socióloga Ana Maria Goldani e assessorados por Abel Packer, Sandra Acuña e Arthur Conning (*CELADE*), surge o SEADE/DOCPOP (*Sistema de Informação sobre População no Brasil*), o primeiro sistema

de informação no Brasil a adotar o CDS/ISIS e as recomendações do UNISIST para o formato do registro bibliográfico.

Com o advento dos minis-computadores, surge em 1985 o MINISIS, uma nova versão do CDS/ISIS. O MINISIS, desenvolvido e distribuído pelo IDRC (*CIID*) e logo adotado por várias bibliotecas, sistemas e centros de informação no Brasil e no exterior. No Brasil, o MINISIS foi adotado inicialmente pela Bireme/OPS/OMS para gestão da antiga base de dados IMLA (*Index Medicus Latinoamericano*) que convertida para o formato recomendado pela UNISIST foi rebatizada de LILACS (*Literatura Latino-Americana em Ciências da Saúde*); pelo IBICT para gestão do Catálogo Coletivo Nacional, pela Editora Abril Cultural e pela Promom Engenharia para gestão dos catálogos de seus centros de documentação.

Em 1986 a Unesco inicia a distribuição do MicroISIS para uso em microcomputadores. Dezenas de milhares de bibliotecas em todo o mundo passaram a adotar esta nova ferramenta para catalogação de seus acervos, ignorando o formato MARC e seguindo as recomendações do "UNISIST Reference Manual for Machine-Readable Bibliographic Description".

A Bireme/OPS/OMS desenvolve novas ferramentas para a gestão de bases de dados CDS/ISIS (*CISIS@Bireme*) e descentraliza a alimentação da base de dados LILACS, promovendo o uso do MicroISIS, dos utilitários CISIS e das recomendações da UNISIST no Brasil e em todos os países Americanos, através de programas de capacitação no uso desta nova tecnologia. Em apenas dois anos, estava estabelecida a maior rede cooperativa de intercâmbio de registros bibliográficos da América Latina, totalmente fundamentada em registros bibliográficos UNISIST.

Surge um novo componente, a internet e com ela o hipertexto, trazendo profundas alterações no processo de organização e transmissão do conhecimento. A catalogação, a descrição bibliográfica e os meios e padrões de armazenamento da informação, sofrem, mais uma vez, uma enorme transformação. Com os recursos da linguagem de marcação (*HTML*), o texto ganha a hiper-dimensão permitindo ao leitor caminhar nestas dimensões através de ligações pré-estabelecidas (*hiperlinks*).

Junto com o HTML surge o conceito de metadado, criando regras que possibilitam ao próprio autor assinalar os elementos de dados que descrevem a sua criação, permitindo que robôs criem automaticamente os pontos e acesso à massa de informação armazenada (catálogos). Surge o XML e em seguida a Bireme/OPS/OMS desenvolve o WWWis capaz de interpretar roteiros (*scripts*) escritos em XML IsisScript e interagir com bases de dados CDS/ISIS em ambiente WWW.

Em 1994, nasce em Ohio, Estados Unidos, um novo formato de descrição bibliográfica baseado em metadados, o Dublin Core, adotado pelo W3C (*The World Wide Web Consortium*) como formato "standard" para descrição da informação na internet. Hoje, os catálogos que consultamos através do Yahoo, Google, Altavista, etc. utilizam estes recursos.

Em maio de 2001, Martin Dillon(2) no artigo "The Role of Cataloging in the Future of Libraries" fala sobre o futuro das bibliotecas e especificamente sobre as mudanças nas técnicas, regras, formatos e métodos da catalogação face ao crescimento dos recursos

da internet e do surgimento de uma nova linguagem e de um novo formato de catalogação, o formato XML. Baseado na linguagem HTML e ideal para a catalogação.

Coincidentemente, naquele mesmo mês, o futuro a que ele se referia, já se fazia presente na biblioteca Dr. Romeu Ritter dos Reis em Porto Alegre, com o uso, em caráter experimental, de um sistema que permitia a gestão remota de serviços e catálogos da biblioteca através da Internet. Nascia ali o PHL (*Personal Home Library*) e a primeira experiência brasileira na automação de serviços e processos de bibliotecas através da Web, utilizando banco de dados CDS/ISIS, com registros baseados nas recomendações da UNISIST e servidos com o software WWWisis©Bireme

Já nesta época, havia a corrente de pensamento negativa sobre questões de segurança a respeito do uso da internet para controle de serviços e processos. Mas também, havia uma corrente contrária a este pensamento negativo e que mesmo ciente dos riscos, apostaram no uso desta mídia e começaram a desenvolver aplicações comerciais (*E-commerce*) e bancárias (*E-bank*).

Em alguns meses de teste, já se podia perceber o poder desta nova tecnologia onde muitos paradigmas começavam a ser quebrados. Esta nova concepção no tratamento da informação e as facilidades oferecidas por esta nova mídia, vieram proporcionar, não só aos bibliotecários, mas também, à profissionais de outras áreas, a possibilidade de organizarem seus acervos e serviços de forma simples, eficiente e intuitiva, se beneficiando de todos as facilidades do ambiente Web.

Se analisarmos atentamente, podemos perceber que no PHL, são tantos os novos conceitos incorporados e paradigmas quebrados, que até entre a maioria dos profissionais de informática se percebe a dificuldade em compreendê-los. O PHL não é um “software”, é uma aplicação Web que gera hipertextos a partir de informações bibliográficas armazenadas em bases de dados CDS/ISIS©Unesco e utilizando um formato baseado nas recomendações do UNISIST. Pode ser comparado ao Google, Altavista, ou Yahoo, com algumas diferenças, sendo uma delas, a de oferecer formulários que permitem modificar ou incrementar, em tempo real, seus conteúdos.

A eficiência dos novos formatos de armazenamento e intercâmbio de registros quebrou o paradigma do MARC. Podemos perceber que hoje, boa parte das redes e bibliotecas que mantêm bases de dados cooperativas não utilizam o formato MARC como formato de registro bibliográfico e nem como formato de intercâmbio.

Face esta breve reconstituição histórica do desenvolvimento dos meios e métodos de armazenamento e recuperação da informação, concluímos que é leviano indicar o MARC como sendo o único padrão aceito e recomendado para o registro bibliográfico em meio magnético. São muitos os padrões existentes, cada um deles com características sutis que os diferenciam entre si e os tornam mais ou menos eficientes dependendo de sua aplicação e do ambiente onde serão utilizados. Cabe ao bibliotecário avaliar criteriosamente sobre a ótica custo/benefício, qual o formato mais adequado para cada aplicação em particular.

Nossa experiência, ao longo destes anos, nos leva a concluir, com bastante segurança que hoje, para o intercâmbio de dados, pouco importa o nome ou a disposição do “container” do elemento de dado no registro bibliográfico, o importante é que o conteúdo

deste "container" siga padrões definidos e que cada registro seja constituído de um conjunto de "containers" que permitam sua descrição. E neste caso, graças o empenho da classe biblioteconômica mundial há um consenso unânime nas regras adotadas.

Veja abaixo as diferenças no registro do título de um documento em distintos formatos (*HTML, MARC, MARCXML, Dublin Core e UNISIST*) e note que, em cada um deles o título está encapsulado em "containers" distintos, mas todos trazem seu conteúdo transcrito de uma forma padrão (*AACR2*) e este padrão é que os tornam intercambiáveis entre si:

- a) Formato de metadado - HTML
<meta name="title">**A Guide to Growing Roses**</meta>
- b) Formato Dublin Core - XML
<dc:title>**A Guide to Growing Roses**</dc:title>
- c) Formato MARC (Library of Congress)
245 1 0 \$a **A Guide to Growing Roses**
- d) Formato MARCXML (Library of Congress)
<datafield tag="245" ind1="1" ind2="0">
<subfield code="a"> **A Guide to Growing Roses** </subfield>
</datafield>
- d) Formato MARC (Biblioteca Nacional)
245 12 |a **A Guide to Growing Roses**
- e) Formato MARC (Unicamp)
245 1 0 \a **A Guide to Growing Roses**
- f) Formato PHL - UNISIST
!v018!A **A Guide to Growing Roses**
- g) Formato PHL - XML
<datafield tag="18">**A Guide to Growing Roses**</datafield>

Nos últimos anos, tenho dedicado boa parte do tempo, convertendo bases de dados de bibliotecas que decidiram adotar o PHL como meio de gestão de seus serviços e catálogos. Centenas de bases de dados em Excell, Access, DBase, Paradox, Foxpro, etc., etc. foram convertidas para CDS/ISIS. Deste tempo, somente 5% foram investidos na conversão de formatos ou padrão do banco de dados. Os outros 95% do tempo foi investido na criação de rotinas para padronização dos elementos de dados daquelas bases cujos elementos de dados não seguiram as regras recomendadas pela AACR2. Portanto, podemos concluir que para o intercâmbio de registros bibliográficos, o mais importante é a padronização do elemento de dado (*metadado*) e não do formato adotado como "container" do registro bibliográfico. Pouco importa se a base de dados utiliza MARC, UNISIST, Dublin Core, ELHILL, ANSI, etc. O que realmente importa é que utilizem sempre as mesmas regras na transcrição dos seus elementos de dados.

Referências Bibliográficas

- 1) UNESCO. **UNISIST Reference Manual for Machine-Readable Bibliographic Descriptions of Research Projects and Institutions**. Paris: UNESCO, June 1980.
- 2) OLIVEIRA, Elysio Mira Soares de. **Sistema de documentação sobre população no Brasil**. São Paulo: s.n, jul. 1983. 6 p. Apresentado no Congresso Brasileiro de Biblioteconomia e Documentação, 12, Balneário de Camboriú - SC.
- 3) OLIVEIRA, Elysio Mira Soares de; PACKER, Abel Laerte. **Metodologia utilizada pelo SEADE / DOCPOP para a coleta e processamento de informações bibliográficas**. In: SEMINARIO SOBRE AUTOMACAO EM BIBLIOTECAS E CENTROS DE DOCUMENTACAO, 1, 1984, São José dos Campos. Resumo dos trabalhos. [S.l.]: [s.n.], 1984.
- 4) DILLON, Martin. **The Role of Cataloging in the Future of Libraries**. Ohio Valley Group of Technical Services Librarians Newsletter, v14, 2001.
- 5) MODESTO, Fernando. **Formato de intercâmbio: UNISIST**. São Paulo: ECA, 2005.
- 6) REINEHR, Hilda Maria Fiuza Abras; OLIVEIRA, Elysio Mira Soares de. **Personal home library: a biblioteca 24h no ar**. In: SEMINARIO NACIONAL DE BIBLIOTECAS UNIVERSITARIAS, 12, 2002, Recife. Anais. Recife, 2002
- 7) **Anglo-American Cataloguing Rules**. (<http://www.aacr2.org>).