

**Unesp - Universidade Estadual Paulista
Faculdade de Filosofia e Ciências
Campus de Marília**

XML na Ciência da Informação: uma análise do MARC 21

Marcos Antonio Siqueira

**Unesp - Universidade Estadual Paulista
Faculdade de Filosofia e Ciências
Campus de Marília**

XML na Ciência da Informação: uma análise do MARC 21

Dissertação apresentada a Faculdade de Filosofia e Ciências da Universidade Estadual Paulista, campus de Marília, ao Programa de Pós-Graduação em Ciência da Informação, Área de Concentração: Informação, Tecnologia e Conhecimento, como requisito para obtenção do título de mestre.

**Marcos Antonio Siqueira
Orientadora: Dra. Plácida L. V. A. da Costa Santos**

Marília / 2003

Dedico a minha esposa Karina, pois agora, está bem perto o dia em que juntos poderemos alcançar nosso objetivo maior, a minha mamãe Madalena pela grande e total dedicação e a minha orientadora Dr^a. Plácida pelo apoio, tanto na orientação, como em atitudes que são características apenas de uma grande amiga.

Agradecimentos

A todos os professores das disciplinas necessárias para obtenção dos créditos e aos funcionários da Universidade que ajudaram através das suas funções, especialmente na pós-graduação e no escritório de pesquisa. Um agradecimento especial para a Dr.^a Silvana Ap. Borsetti Gregório Vidotti e Dr.^a Mariângela Spotti Lopes Fujita, devido aos importantes apontamentos e sugestões feitos na qualificação que foram sem dúvida importantes para o aprimoramento do trabalho.

A todos os alunos do Curso de Biblioteconomia que sempre estiveram a disposição tirando minhas dúvidas sobre particularidades da Ciência da Informação, catalogação, classificação, metadados, MARC 21, Dublin Core, Bibliotecas Digitais, RDF, etc. Enfim, a todos que suportaram as minhas perguntas e que estavam sempre prontos a bater um bom papo sobre a ligação da Ciência da Computação com a Ciência da Informação. Neste mesmo sentido, agradeço também aos colegas do programa de mestrado.

Cono, proprietário da Única Informática em Assis SP, pelo apoio, em permitir minhas vindas para Marília, SP, ainda mesmo quando não tinha a certeza de entrar no programa de mestrado, e depois, por compreender que tinha que dedicar maior tempo para o meu mestrado.

Não posso esquecer dos professores da FEMA, especialmente o Almir pelo incentivo e interesse em me apresentar o mundo acadêmico no meu último ano da faculdade em Assis, SP, ao Guto pelo mesmo motivo e ao Begosso, professor que sempre admirei pela postura em sala de aula e qualidade no ensino, pelo apoio e incentivo nas minhas “idéias”, ainda enquanto aluno do curso.

SIQUEIRA, Marcos Antonio. **XML na Ciência da Informação: uma análise do MARC 21**. Marília : UNESP, 2003.

Resumo

No processo de catalogação, indispensável para a Ciência da Informação, ocorre a criação de formas de representação documentária descritiva, por isso, sua relação com as teorias de formas de representação estudadas pela Ciência Cognitiva. Essa relação é objeto de estudo dessa pesquisa, que se pautará na análise do formato de intercâmbio de dados bibliográficos Machine Readable Cataloging, MARC 21, em sua versão em Extensible Markup Language, XML, linguagem de marcação que prove mecanismos de comunicação de dados desenvolvida pela Ciência da Computação, caracterizando o referencial teórico da pesquisa a partir da interdisciplinaridade existente. O objetivo da pesquisa é destacar a importância da XML para a Ciência da Informação, a partir da própria estrutura da linguagem e de estudos sobre Formas de Representação, apresentando a versão MARC 21 em XML como ferramenta adequada para representação documentária. A proposição apresenta subsídios teóricos e metodológicos que destaca o MARC 21 em sua versão XML como uma relevante opção de representação documentária, e promove uma discussão sobre a necessidade de desenvolvimento de ferramentas de software e não de novos métodos de representação. Evidências técnicas sobre o MARC 21, XML e a indicação a versão MARC 21 em XML como método adequado para representação documentária embasarão a proposição. A partir de uma metodologia de análise, verificou-se que a necessidade de desenvolvimento de métodos e teorias para a fundamentação da representação documentária e também de implementação efetiva de ferramentas de software preocupada com a usabilidade da interface para otimização do processo de representação, ocasionam dificuldades na apresentação de uma catalogação de qualidade, especialmente, no que se refere a representação, descrição e intercâmbio de documentos eletrônicos disponíveis na Internet. Se obtém como resultados a conclusão de que em sua versão em XML, o MARC 21 destaca-se frente as iniciativas de padrões para a documentação de acervos específicos como estrutura de metadados, por entre outros motivos, ser o formato mais antigo, mais usado e maior base existente. Não estamos afirmando que não são necessários os estudos sobre metadados, entretanto, conclui-se que a discussão atual necessária é sobre o desenvolvimento de ferramentas de software com interfaces providas de usabilidade, que contemplem a solução para as questões de interoperabilidade entre os padrões já existentes.

Palavras-chaves: MARC 21, MARC 21 XML, XML, Catalogação, Formas de Representação, Metadados.

SIQUEIRA, Marcos Antonio. **XML na Ciência da Informação: uma análise do MARC 21**. Marília : UNESP, 2003.

Abstract

In the cataloguing process, indispensable for the Science of the Information appears the creation of forms of descriptive documentary representation, for that, its relation with the theories in representation ways studied by the Cognitive Science. That relation is the object of study of this research, that it will be ruled in the analysis of the format of exchange of bibliographical data Machine Readable Cataloging, MARC 21, in its version in Extensible Markup Language, XML, demarcation language that proves mechanisms of communication of data developed by the Science of the Computation, characterizing the theoretical referencial of the research starting from the existent interdisciplinary. The objective of the research is to detach the importance of XML for the Science of the Information, starting from the own structure of the language and of studies on Forms of Representation, introducing the version MARC 21 in XML as appropriate tool for documentary representation. The proposition presents theoretical and methodological subsidies that it detaches MARC 21 in its version XML as a relevant option of documentary representation, and it promotes a discussion about the need of development of software tools and not of new representation methods. Technical evidences about MARC 21 XML and the indication of the version MARC 21 in XML as appropriate method for documentary representation will base the proposition. Starting from an analysis methodology, it was verified that the need of development of methods and theories for the grounding of the documentary representation and also of effective implementation of software tools worried with the usability of the interface for optimization of the representation process, they cause difficulties in the presentation of a good cataloguing, especially, in what refers to the representation, description and exchange of available electronic documents in the Internet. It is obtained as results the conclusion that in its version in XML, MARC 21 stands out front the initiatives of patterns for the documentation of specific collections as metadata structure, for among other reasons, to be the oldest format, more used and larger existent base. We are not affirming that they are not necessary the studies on metadados, however, It is ended that the necessary discussion is about the development of software tools with provided interfaces of usabilidade, which contemplate the solution for the interoperability subjects among the patterns existent.

Keyword: MARC 21, MARC 21 XML, XML, Cataloguing, Forms of Representation, Metadata.

Lista de Abreviaturas

AACR - Anglo American Cataloging Rules

CI - Ciência da Informação

DTD - Definição de Tipo de Documento

DOM - Document Object Model

HTML - Hyper Text Markup Language

ISBD - International Standard Bibliographic Description

LC - Library of Congress

MARC - Machine Readable Cataloging

METS - Metadata Encoding & Transmission Standard

MS - Microsoft

OCLC - Online Computer Library Center

SGBD - Sistemas Gerenciadores de Bancos de Dados

XDR - XML-Data Reduced

XML - Extensible Markup Language

XSL - Extensible Stylesheet Language

Lista de Figuras

Figura 1. Uma ficha catalográfica	35
Figura 2. Ficha catalográfica ilustrando as formas de representação do MARC 21	44
Figura 3. Uma das formas de representação de um registro MARC 21	45
Figura 4. Arquivo texto com um registro MARC 21	46
Figura 5. Armazenar o significado das etiquetas, posições, indicadores e subcampos	54

Lista de Tabelas

Tabela 1. Entendendo o diretório de um registro MARC 21	55
Tabela 2. Particularidades da etiqueta 006	58
Tabela 3. Particularidades da etiqueta 007	59

Lista de Exemplos

Exemplo 1. Um registro MARC 21 em XML	85
Exemplo 2. Registro MARC 21 em XML resumido	88
Exemplo 3. Etiqueta 245 do MARC 21 em XML	88
Exemplo 4. Código fonte em ASP manipulando um arquivo XML	90

Sumário

1 Introdução	13
2 Formas de Representação: fundamentos para a catalogação	18
2.1 Formas de Representação segundo Donald Peterson	19
2.2 Perspectivismo de Donald Peterson	22
2.3 Aplicações práticas das teorias de Formas de Representação	24
3 Catalogação	31
3.1 A relação da Catalogação e as Formas de Representação	36
4 Machine Readable Cataloging - MARC 21	39
4.1 Aspectos técnicos do formato MARC 21	43
4.2 Considerações sobre o MARC 21 para um analista de sistemas	52
4.3 A catalogação na prática usando MARC 21	60
5 Extensible Markup Language - XML	69
5.1 Document Object Model - DOM	71
5.2 Document Type Definition - DTD	72
5.3 Extensible Stylesheet Language - XSL	75
5.4 Aplicações da XML	76
6 MARC 21 em XML	82
6.1 Considerações sobre o MARC 21 XML para um analista de sistemas	85

7 Considerações finais	93
8 Conclusão	100
9 Referências	104
10 Documentos consultados	109
Anexos	110
Anexo A - Documentação das tabelas sugeridas, que armazenam o significado das etiquetas, posições, indicadores e subcampos	111
Anexo B - Exemplo de um programa que importa dados bibliográficos desenvolvido na linguagem PHP	116
Anexo C - Elementos criados para corresponder as tags tradicionais do MARC 21: Schema MARC21slim.xsd	122

1 Introdução

A Ciência da Informação, CI, tem por objetivo estudar as propriedades da informação, tais como a natureza, gênese e efeitos (LE COADIC, 1996). Em outras palavras esta ciência analisa a construção, comunicação e uso da informação, verificando ainda a concepção dos produtos e sistemas que permitam esta análise, acrescentando ainda o armazenamento da informação.

As análises feitas pela Ciência da Informação são influenciadas pela utilização de um processo determinante da biblioteconomia, a catalogação. O processo de comunicação que se estabelece entre um indivíduo e a informação disponível em um item documentário armazenado em um acervo, ocorre a partir da preparação e organização padronizada da informação em uma forma de representação descritiva realizada pela catalogação (PEREIRA; SANTOS, 1998).

A catalogação é responsável pela criação de formas de representação documentária descritiva, por isso, sua ligação com as teorias de formas de representação estudadas pela Ciência Cognitiva. A criação de formas de representação na Ciência da Informação acaba sendo algo natural, visto que as necessidades informacionais dos usuários dos serviços oferecidos pela área são distintas.

A relação entre a catalogação, processo indispensável para a Ciência da Informação, e estudos sobre formas de representação, da Ciência Cognitiva, são objetos de estudo dessa pesquisa, que se pautará na análise do formato de intercâmbio de dados bibliográficos Machine Readable Cataloging, MARC 21, em sua versão em Extensible Markup Language, XML, linguagem de marcação que prove mecanismos de comunicação de dados desenvolvida pela Ciência da Computação. Essa interdisciplinaridade caracteriza o referencial teórico da pesquisa.

A informação é estudada em três diferentes perspectivas pela Ciência da Informação, fenômenos, processos e produtos e se propõe socializar a informação podendo gerar novos

conhecimentos. Nesse contexto, os processos têm sua importância destacada no que diz respeito à gestão da informação e podem ser divididos da seguinte forma:

- Produção da Informação;
- Organização da Informação;
- Tratamento da Informação;
- Recuperação da Informação;
- Disseminação da Informação;
- Uso da Informação.

A catalogação está presente em praticamente todos os processos, com um especial comprometimento no tratamento, organização e recuperação da informação. Deste ponto de vista podemos citar a utilização do formato de intercâmbio de dados bibliográficos e catalográficos entre bibliotecas, MARC 21, em departamentos de catalogação de documentos, onde são desenvolvidos métodos de controle de qualidade tanto da catalogação como dos registros importados e exportados em formato MARC 21 a fim de aumentar a confiabilidade das informações encontradas pelos usuários em uma eventual busca nos sistema de informação oferecido pela instituição.

Ainda com relação ao tratamento da informação, é importante destacar que este processo tem como finalidade a identificação, processamento e disponibilização do conteúdo informacional de diversificados suportes documentários. A busca do desenvolvimento teórico e metodológico, interdisciplinar, é necessária, e pode propiciar a elaboração de produtos documentários com especial influência na representação documentária. Destacada pela sua natural possibilidade de promover interdisciplinaridade entre diversas ciências, a Ciência da Informação contempla vários campos do conhecimento (LE COADIC, 1996). Nesse sentido, o uso das metodologias da Ciência da Computação proporcionaria um ambiente de enriquecimento para ambas as ciências, com a contribuição da Ciência Cognitiva.

Pinheiro e Loureiro (1995) afirmam que existe uma relação muito forte e quase natural, inclusive do ponto de vista histórico, entre a Biblioteconomia e Informática, ou seja, entre a Ciência da Informação e Ciência da Computação, na incorporação de técnicas computacionais e no uso e entendimento das tecnologias em informação.

Nesta pesquisa destaca-se a Ciência da Informação aprimorando a qualidade nas formas de representação documentária por meio da catalogação utilizando tecnologias da Ciência da Computação, e a Ciência da Computação utilizando-se das metodologias da Ciência da Informação no processo de recuperação e disseminação da informação.

Entretanto, a necessidade de desenvolvimento de métodos e teorias para a fundamentação da representação documentária e também de implementação efetiva de ferramentas de software para otimização do processo de representação ocasionam dificuldades na apresentação de uma catalogação de qualidade, especialmente, no que se refere a representação, descrição e intercâmbio de documentos eletrônicos disponíveis na Internet.

Nesse contexto, a proposição é apresentar subsídios teóricos e metodológicos destacando que o MARC 21 em sua versão XML caracteriza-se como uma relevante opção de representação documentária, e promover uma discussão sobre a necessidade de desenvolvimento de ferramentas de software e não de novos métodos de representação. Evidências técnicas sobre o MARC 21, XML e a indicação a versão MARC 21 em XML como método adequado para representação documentária embasarão a proposição.

O objetivo da pesquisa é destacar a importância da XML para a Ciência da Informação, a partir da própria estrutura da linguagem e de estudos sobre Formas de Representação, apresentando a versão MARC 21 em XML como ferramenta adequada para representação documentária.

O referencial teórico sobre formas de representação da área da Ciência Cognitiva está no capítulo 2, e visa a identificação de como diferentes formas de representação podem

auxiliar na realização de uma determinada tarefa. Esse objetivo no contexto da Biblioteconomia, mais especificamente da catalogação automatizada através do uso do formato de intercâmbio de dados bibliográficos e catalográficos MARC 21 em XML, apresenta-se como uma das formas de representação que auxiliam nas tarefas de recuperação e disseminação de itens documentários.

Este referencial estende-se à medida que é apresentado no capítulo 3 uma análise geral da Catalogação e um estudo da relação com as teorias das formas de representação, verificando no capítulo 4 detalhes do formato MARC 21, dando uma especial atenção para os aspectos técnicos de sua implementação em sistemas gerenciadores de bibliotecas e sua utilização nos departamentos e setores de catalogação, tendo em vista a preocupação com a qualidade dos registros importados e construídos.

Após esse embasamento teórico apresentam-se no capítulo 5 os aspectos técnicos da XML, bem como, um estudo das principais tecnologias que agregam valor aos objetivos e funções da linguagem de marcação XML, ao mesmo tempo em que se argumenta sobre as aplicações da linguagem e a relação existente entre formas de representação, catalogação e MARC 21. No capítulo 6 são apresentados detalhes também técnicos da versão MARC 21 em XML desenvolvida pela Biblioteca do Congresso Americano, Library of Congress – LC – com uma especial atenção a facilidade de desenvolvimento de um software de intercâmbio de dados bibliográficos e catalográficos baseado nessa versão.

Depois das considerações finais no capítulo 7 que ilustra o avanço do formato MARC 21 em XML apresentando dois possíveis novos produtos, conclui-se destacando o potencial da versão do MARC 21 em XML tendo como principais argumentos que o formato MARC é o mais antigo, mais usado e maior base existente (MARC21, 2002). Com sua versão em XML o MARC 21 destaca-se, frente às iniciativas de padrões para a documentação de acervos específicos, como estrutura de metadados. Conclui-se ainda, que a discussão necessária é sobre o desenvolvimento de ferramentas de software que contemplem essa solução e as questões de interoperabilidade entre os padrões já existentes.

Assim, a pesquisa sobre a XML na Ciência da Informação, enquanto um estudo do MARC 21, não está pautado na necessidade de desenvolvimento de metadados¹ ou métodos, mas em teorias para a representação documentária e implementação efetiva de ferramentas de software para otimização do processo de representação descritiva a partir da análise da versão do MARC 21 em XML como uma opção de qualidade como padrão de descrição de acesso para o processo de intercâmbio de dados bibliográficos e catalográficos. Nesse sentido, destaca-se a XML na Ciência da Informação, a partir de sua própria estrutura e da reflexão sobre as teorias de Formas de Representação.

¹ O termo Metadados é comumente definido como "dados sobre dados a partir de um dicionário digital de dados". Em outras palavras, metadado são informações sobre o dados, e tem como finalidade documentar e organizar de forma estruturada os dados com o objetivo de minimizar duplicação de esforços e facilitar a manutenção dos dados. São um conjunto de dados referenciais metodologicamente estruturado e codificados, conforme padrões internacionais, para localizar, identificar e recuperar pontos informacionais de documentos.

2 Formas de Representação: fundamentos para a catalogação

Iniciamos nossos estudos focalizando a pesquisa de Donald Peterson (1996) sobre formas de representação, incluindo as análises de Alves (1998) e Peschl (2002) sobre o tema, concluindo com aspectos práticos apresentados por Marcondes (2001) e Casa (1997), com o objetivo de apresentar a teoria de formas de representação com um fundamento para a catalogação como um processo de criação de formas de representação descritiva documentária.

Representar pode significar tornar alguma coisa presente novamente, do ponto de vista Etimológico, ou seja, a partir do “*estudo da origem e da evolução das palavras*” (HOUAISS; VILLAR, 2001, p. 1271). Segundo Pierce (apud Marcondes, 2001), para a Semiótica, representar significa a caracterização de um processo que envolve um objeto, algo que o representa e o efeito desta representação, quando o objeto não está presente, porém, representado na mente de um usuário. Ainda segundo Marcondes (2001, p. 64), representação é como “*um processo ocorrendo na mente de alguém, produzindo nesta mente algo distinto do objeto a que se refere*”.

Alves (1998, p. 26) diz que “*uma forma de representação é um par, consistindo de uma notação e uma ontologia² (A ontologia é, pois, uma interpretação para a notação)*”. É, portanto fato que uma forma de representação pode ser útil para um indivíduo, porém, a mesma, inútil para outro indivíduo. O autor define forma de representação “*como uma notação, com uma interpretação para essa notação*” (ALVES, 1998, p.25), e afirma ainda que, ontologia é “*um aparato conceitual ou um conjunto de aspectos do mundo que são relevantes para a realização da tarefa*” (ALVES, 1998, p. 26). Assim, podemos dizer que uma notação é uma representação, ou marcação por meio de símbolos de alguma coisa.

Os humanos fazem uso de formas de representações externas tais como: alfabetos, diagramas, desenhos, fluxogramas, linguagens de programação de computador, interfaces

² Ciência primeira, tem por objeto o objeto de todas as outras ciências, e como princípio o princípio que condiciona a validade de todas as outras. ABBAGNAMO (1998, p. 660).

de computador, símbolos, idioma, notas musicais etc. O trabalho de Peterson (1996) investiga a grande variedade de representações e sua aplicação, principalmente de representações externas. A aquisição de uma forma de representação externa apropriada pode ser vantajosa e ainda facilitar a compreensão, resolução de problemas, cálculo e principalmente, o crescimento de nosso conhecimento.

Quando estamos envolvidos em um processo de aprendizagem ou solução de um determinado problema, diferentes formas de representação podem nos auxiliar a manipular as dificuldades envolvidas. Diferentes formas de representação podem ainda ser usadas para apresentar um objeto ou conceito de diferentes pontos de vista (CASA, 1997). O aspecto positivo em usar diferentes formas de representação está em obter “*estruturas capazes de representar somente a informação relevante em um determinado momento*” (CASA, 1997). Podemos entender como estrutura neste contexto, algo capaz de dar suporte informacional.

À medida que se percebe a essência dos conceitos apresentados, nota-se a necessidade de identificar, avaliar e utilizar formas de representação. É essa necessidade que se destacará na análise da linguagem de marcação extensível, XML, no contexto da Biblioteconomia, mais especificamente da catalogação, no processo de intercâmbio de dados bibliográficos e catalográficos com a utilização do formato MARC 21.

2.1 Formas de Representação segundo Donald Peterson

A interação entre um ambiente e um sistema cognitivo humano é caracterizada pela necessidade de se resolver uma grande variedade de tarefas as quais encontram-se muitas vezes em níveis diferentes de dificuldades. O ambiente por si só representa um sistema dinâmico complexo, e o próprio ser humano também se constitui em um sistema dinâmico complexo. A questão está em estabelecer uma boa estabilidade dentro do sistema cognitivo e entre o ambiente (PESCHL, 2002).

A Ciência cognitiva e a psicologia cognitiva afirmam que para um indivíduo se comportar adequadamente em um ambiente qualquer, este, deve possuir alguma representação de pelo menos algumas partes deste ambiente. Estas representações são chamadas de representações internas. Esta idéia sugere que o mundo, ou partes deste, é representado na forma de símbolos, imagens mentais, redes semânticas etc. e um algoritmo ou algum mecanismo de manipulação opera estas representações. Estas manipulações resultam numa externalização de comportamento do indivíduo que poderia conduzir a estabilidade dentro do sistema cognitivo, entre o ser humano e o ambiente. Essa externalização resulta na chamada forma de representação externa, gerada a partir de ferramentas tais como lápis, canetas, impressoras etc (PESCHL, 2002).

É importante, portanto, lembrar que a utilização de formas de representações externas que vão desde alfabetos, diagramas, desenhos, interfaces computacionais, idiomas, notas musicais etc. Peterson (1996) analisa a grande variedade de representações, principalmente externas, e suas aplicações, pois a aquisição de uma forma de representação externa apropriada facilita o entendimento, influenciando no crescimento do conhecimento.

Nas ciências naturais modernas toda forma de visualização ou representação vívida é um exemplo da transformação de formas diferentes de representação em outras, para facilitar a compreensão ou percepção de um certo fenômeno. Apesar da complexidade do mundo e as diferenças entre tarefas e metas no sistema cognitivo, Peschl (2002) afirma que :

- Uma certa representação representa apenas uma parte da complexidade inteira;
- A forma escolhida de representação tem que ser apropriada à tarefa que o usuário da representação tem que realizar.

Investigar a relação entre as características representacionais de uma certa forma de representação com o ato de cumprir a tarefa dentro de um contexto particular apresenta um relevante assunto de pesquisa (PESCHL, 2002). Para o autor a forma de representação não apenas representa alguma realidade, vai além, pois, facilita tarefas e processos de cálculo e

manipulação, executados por usuários em tarefas particulares. Além disso, Peterson (1996) mostra que a epistemologia de formas de representação não é uma questão simples de endereçar a relação de correspondência entre inscrições e fatos, mas preocupa-se com as ações mentais e processos de sistemas cognitivos no ato de estender uma realidade pelo uso de representações.

Grande parte dos estudos de Peterson (1996) apresenta a relevância do uso de representações externas e a necessidade de se esquecer da idéia de encontrar um único método para representar o mundo, visto que, não se tem um acesso direto ao ambiente e o conhecimento sempre é um sistema relativo no senso que é construído para ser aplicado para a solução de um problema ou por gerar um comportamento que é específico para o usuário em uma tarefa interna e externa (PESCHL, 2002).

Em outras palavras, uma forma de representação não tem que conter e mostrar toda a possível informação sobre uma certa realidade, mas tem que prover a informação que é pertinente para a realização da tarefa. Conseqüentemente, uma representação pode omitir sistematicamente uma informação irrelevante para o propósito, e pode enfatizar ou duplicar informações pertinentes (PESCHL, 2002). Peterson (1996) parece interessado em identificar até que ponto uma forma de representação alternativa representa o mesmo domínio de realidade.

Para Peschl (2002), representações externas não têm que corresponder a estruturas de uma realidade com perfeição equivalente a fotografias, visto que a representação tem o objetivo de facilitar tarefas e processos. Afirmação essa, que destaca a importância de uma ferramenta para gerenciar formas de representação. Nessa perspectiva, a idéia tradicional de representações com perfeição equivalente a fotografias têm que ser abandonada a favor do conceito de olhar para representações como sistema que ajudam na estratégia de resolução da tarefa. Essa visão tem implicações na Inteligência Artificial e na Ciência Cognitiva, mas também nos estudos de interfaces para auxiliar na interação entre humano e computador.

A meta não é encontrar uma representação precisa, mas construir uma estrutura que apresente formas diferentes de representações com propósito específico de mudá-la de acordo com o problema a ser resolvido e de acordo com as habilidades cognitivas do usuário (PESCHL, 2002). A linguagem XML, como veremos no capítulo 5, em conjunto com o formato MARC 21, conforme as indicações do capítulo 6 dão condições para a construção de uma estrutura adequada no contexto da Ciência da Informação, uma vez que, há que se considerar as perspectivas da área.

2.2 Perspectivismo de Donald Peterson

Perspectivismo³ pode ser o estudo dos processos de adequação, envolvendo reajustes de uma determinada perspectiva adotada por um indivíduo em contextos específicos. Entretanto, uma adequação é sempre relativa, mas não se estabelece de modo arbitrário. Na escolha de uma perspectiva devemos levar em conta a sua adequação ao usuário, a uma situação determinada, a um processo, a uma ontologia e a uma meta específica.

O perspectivismo de Peterson (apud Alves, 1998) indica uma forma de relativismo⁴, que implica critérios de racionalidade. Alves (1998, p. 27) afirma que *“tal critério poderia ser o progresso da ciência rejeitando, porém, a distinção tradicional entre ciência e tecnologia”*, esse progresso ocorreria a partir de *“algum critério geral de racionalidade, pelo qual podemos escolher entre diferentes formas de representação”* (ALVES, 1998, p. 27).

³ Por *perspectivismo* Nietzsche entendeu a condição em virtude da qual “cada centro de força – e não só o homem – constrói todo o resto do universo partindo de si mesmo, ou seja, atribuindo ao universo dimensões, forma e modelo proporcionais à sua própria força” ABBAGNANO (1998, p. 759).

⁴ “Doutrina que afirma a relatividade do conhecimento, no sentido dado a essa expressão no século XIX, a saber: 1º como ação condicionante do sujeito sobre seus objetos de conhecimento; 2º como ação condicionante recíproca dos objetos de conhecimento. Este condicionamento duplo dos objetos de conhecimento foi primeiramente tomado como fundamento do relativismo por W. Hamilton, que, por um lado, insistia no fato de que todos os objetos existentes podem ser conhecidos apenas em relação com as faculdades humanas em condições ditadas por essas mesmas faculdades, e, por outro, na condicionalidade que os objetos de conhecimento exercem uns sobre os outros” ABBAGNANO (1998, p. 845).

Alves (1998, p. 30) aponta que “*utilizando-se o critério do desenvolvimento tecnológico, numa visão interdependente entre ciência e tecnologia, estamos em condição de ver o Perspectivismo de Donald Peterson um processo perfeitamente racional e adequado*”.

Analisar como é o desenvolvimento do conhecimento é o objetivo de Peterson (1996) e, uma das maneiras é “*através da aquisição e aplicação de uma forma apropriada de representação*”. O conhecimento se desenvolve a partir de uma interação entre quem busca o conhecimento com as formas de representação, ou notações disponíveis, fala-se de um conhecimento que gera ação, ou seja, processos são dirigidos por uma forma de representação que guia as decisões de um indivíduo. Não se quer apenas conhecer as coisas, antes, atuar sobre as coisas, muitas vezes, através de notações.

O Perspectivismo de Peterson como aponta Alves (1998, p. 26),

[...] é uma posição essencialmente metodológica. Não se está interessado em saber se nossas crenças representam o mundo, mas o que se quer saber é qual o papel das formas de representação nas ações mentais de cálculo e processamento requeridas para gerar crenças, derivar soluções e realizar tarefas. É nessa medida, que o estudo das formas de representação é essencialmente multidisciplinar.

O perspectivismo de Peterson (1996) pode levar a resposta do papel das formas de representação nas ações mentais. Evidentemente, porém, este estudo deve ser influenciado pela multidisciplinaridade natural que este assunto pode gerar.

O processo de obter conhecimento envolve diversas dimensões. Ajustar as circunstâncias, aos usuários, a ontologia e as tarefas envolvidas constituem essas dimensões, sendo esta última a mais importante. No uso de uma notação para ajuste em uma tarefa, se quer saber a relevância desta notação.

Portanto, visto que um indivíduo pode levar tempo para identificar o que é relevante, outra ontologia pode surgir. Esse assunto sugere como problema comum, “*saber se uma ontologia é apropriada a uma tarefa e, também, se uma notação é apropriada a uma*

ontologia” (ALVES, 1998, p. 26). Um exemplo oportuno da Biblioteconomia é a habilitação do usuário para compreender as informações representadas catalograficamente, visto que não se pode afirmar que qualquer usuário consegue dar significado a essa forma de representação descritiva documentária e identificar um documento. Isso pode ocorrer com outras opções de representação, tais como, citações e a referências bibliográficas.

Para o autor, “*o tratamento adequado das formas de representação é subjetivista*” (ALVES, 1998, p. 27), justificada pela ressalva de Peterson (1996) que aponta a ocorrência de diversos fatores na abordagem perspectivista durante o processo de aquisição do conhecimento. Todavia é clara a necessidade de uma avaliação das formas de representação disponíveis, a começar pelo próprio fato de que é muito difícil que uma forma de representação satisfaça todos os desejos e necessidade.

Uma mesma forma de representação pode não satisfazer os desejos de todos usuários, podendo ser útil para um indivíduo, mas não para outro, assim identificar o que podemos chamar de mínimo aspecto relevante em uma forma de representação parece ser um estudo necessário e relevante.

Nesse sentido, analisar a teoria das formas de representação como fundamento teórico para verificar a utilidade de uma representação bibliográfica face a uma determinada tarefa se justifica na avaliação das aplicações práticas dos padrões de representações catalográficas e de modo especial, do formato de intercâmbio de dados bibliográficos e catalográficos MARC 21 em XML conforme destaca o capítulo 6.

2.3 Aplicações práticas das teorias de Formas de Representação

Para que a base teórica de formas de representação seja entendida, apresentamos dois estudos sobre os aspectos práticos da teoria. No primeiro estudo identificamos a afirmação de Marcondes (2001), que as formas de representação da informação, influenciam decisivamente na atual economia da informação ou conhecimento, especialmente, com a chamada explosão informacional. Deste ponto de vista, “*identificação e avaliação se*

tornam atividades essenciais para a utilização de recursos informacionais” (MARCONDES, 2001, p. 61).

No trabalho de informar, a representação é tão trivial que sua relevância é freqüentemente ignorada. As tecnologias em informação inicialmente foram usadas para identificar, localizar e acessar documentos em papel armazenados em unidades de informação. Entretanto, a identificação de documentos relevantes tornou-se um assunto sério devido à quantidade de documentos que podem ser recuperados nos mais diversos tipos de suporte.

O autor apresenta a idéia de que a representação da informação, no contexto da economia da mesma, serve como um dispositivo de inferência sobre a relevância do recurso informacional para as necessidades de informação do usuário que a interpreta (MARCONDES, 2001). A forma de representar a informação a ser recuperada influenciará na decisão do usuário em acessá-la, pois este, identificou, avaliou sua relevância e custo-benefício a partir da forma de representação da informação apresentada, por exemplo, no resultado de uma busca.

A produção de formas de representações de recursos informacionais apresenta uma complexidade relevante por requerer do profissional da informação a habilidade de identificar as possíveis necessidades de informação dos leitores, seguida da criação de um registro catalográfico e/ou bibliográfico, um resumo, que intermediará o acesso, identificação e avaliação do usuário ao documento original.

Neste ponto, é importante lembrar que, para a Semiótica, segundo Pierce (apud Marcondes, 2001), representar é um processo que envolve um objeto, algo que o representa e o efeito desta representação, quando o objeto não está presente, mas, representado na mente de um usuário, por outro lado, Marcondes (2001, p. 64), afirma que representação é *“um processo ocorrendo na mente de alguém, produzindo nesta mente algo distinto do objeto a que se refere”*.

A busca de informações pode ser considerada um processo semiótico, visto que envolve a interpretação e a manipulação de formas de representações, como exemplo, as referências bibliográficas e a catalogação. Assim, no artigo de Marcondes (2001, p. 66) encontramos,

Belkin diz que, em situações de busca de informações, o usuário se encontra em um “*estado anômalo de conhecimento*” que deve ser preenchido com informação/conhecimento procurado para se tornar coerente.

Essa situação cria a motivação necessária para a busca de informações e fornece os parâmetros para avaliar as representações, por exemplo, referências bibliográficas e fichas catalográficas, permitindo inferência no processo de forma de representação no contexto da busca de informações. A representação deve levar em conta dois aspectos :

- Ser abrangente do ponto de vista cognitivo;
- Ser sintética do ponto de vista da energia gasta pelo usuário.

Marcondes (2001, p. 67) afirma que:

a representação deve ser tão rica cognitivamente de modo a permitir a um usuário inferir o conteúdo do documento, de maneira a permitir-lhe decidir acerca da validade de obter e realizar a leitura do documento completo. Simultaneamente, este processo deve implicar uma significativa economia de energia em comparação com a obtenção e leitura do documento original.

A grande quantidade de documentos disponíveis nas estruturas convencionais, nos suportes eletrônicos e digitais e na estrutura da Internet, como um ambiente de publicação e de republicação de documentos, gerou o desenvolvimento de ferramentas de busca automáticas como uma consequência. Nesse contexto, a exigência de produtos informacionais que permitam a representação, o armazenamento, o intercâmbio e o acesso às informações dinâmicas estruturadas e representadas das mais diversas formas, ocorreu naturalmente.

No ambiente da World Wide Web, WWW, as atividades de descobrir, identificar, avaliar e usar as informações já não têm sido realizados apenas por usuários, mas, automaticamente por softwares, o que requer que as formas de representações de recursos informacionais sejam compreensíveis para computadores e para usuários.

A linguagem XML apresenta-se como uma solução tecnicamente mais viável para a representação documentária, devido à iniciativa da Library of Congress, LC, desenvolvedora do formato de intercâmbio de dados bibliográficos e catalográficos MARC 21, ter criado sua versão em XML. A aceitação e utilização da linguagem XML na comunidade computacional também é uma vantagem, pois, são analistas de sistemas e programadores, juntamente com bibliotecários, que desenvolvem os sistemas de gerenciamento de bibliotecas.

O trabalho de Casa (1997) aborda o desenvolvimento de ambientes computacionais com múltiplas formas de representação, o que segundo o autor, diferentes formas de representação podem auxiliar a manipular as dificuldades envolvidas em um processo de aprendizagem ou solução de um determinado problema. Isso ocorre, pois, diferentes formas de representação são usadas para apresentar um objeto ou conceito de diferentes pontos de vista. Mas tudo isso terá sentido, apenas se existir uma ferramenta para gerenciar “*estruturas capazes de representar somente a informação relevante em um determinado momento*” (CASA, 1997, p. 210).

Segundo Casa (1997), dois são os critérios para estabelecer o nível de especialização de um usuário:

- Verificar se a forma de representação selecionada para uma determinada tarefa é adequada.
- Identificar como são usadas as características informais contidas nas formas de representação. Características informais são “*propriedades estruturais que não fazem parte do formalismo propriamente dito*” (CASA, 1997, p. 210).

Até o final da década de 1990, não existiam pesquisas que analisassem a utilização de diversas formas de representação no raciocínio, na Ciência Cognitiva. Os estudos até então, verificavam a utilidade de uma única forma de representação frente a uma tarefa, chegando em alguns casos propor mudanças nessas, para aprimorar sua eficiência e eficácia. Em

alguns casos foi reconhecido que diferentes formas de representação tinham sua utilidade, mas não foram integrados ao modelo formal de raciocínio. (CASA, 1997).

A Ciência Cognitiva reconhece a importância das representações externas na forma de raciocinar. Um modelo cognitivo deve levar em conta a existência de um mecanismo mental usado para raciocinar com base em representações externas, que poderá ser usado nos estudos realizados na interação entre homem e computador.

Em outras palavras, as interfaces poderão estar preparadas para usar diferentes formas de representação em diversas situações. Criada essa necessidade, e do ponto de vista computacional, os estudos posteriores e necessários levam a criação de softwares que poderão dar suporte a esses ambientes. Segundo Casa (1997, p. 211), dois problemas devem ser verificados:

- Um problema de *designer* para propor arquitetura de software para ambientes com múltiplas representações, e
- Um problema cognitivo para compreender a maneira como empregamos múltiplas formas de representação para resolver problemas e como poderíamos fazer uso de um ambiente computacional onde representações alternativas estão disponíveis.

A questão básica para Casa (1997, p. 213) é analisar “*em que situação uma interface oferecendo representações alternativas pode ser útil*”. Desta forma parece mais lógico fazer uma simples comparação de produtividade e qualidade de resultados obtidos a partir da análise dos resultados dos usuários que usam apenas uma forma de representação e os que usam várias formas de representação.

O autor afirma, por exemplo, que para analisar o sucesso da resolução de um problema não se pode necessariamente estar baseado em dados quantitativos como, por exemplo, o tempo. Principalmente se se concordar que um indivíduo poderá compreender melhor ou mesmo ter uma visão mais clara de um problema se lhe apresentarem representações deste, em múltiplas formas. Casa (1997, p. 214) afirma que “*escolher e produzir uma*

representação adequada é muitas vezes o estágio mais complexo do processo de solução de problemas”.

No processo de avaliação de formas de representação, não se pode desconsiderar a familiaridade do usuário com uma representação já pouco ou muito conhecida, afinal, *“representações com as quais a comunidade de usuários está familiarizada tem uma chance maior de serem usadas com sucesso”* (CASA, 1997, p. 215), ou seja, para afirmar que uma determinada forma de representação é adequada para uma determinada tarefa, verifica-se inicialmente se esta, é familiar ao usuário, ou ao menos garanta que esta seja o mais próximo possível do que chama de familiar ao usuário.

Por exemplo, uma ficha catalográfica convencional é familiar ao bibliotecário, mas não ao usuário comum, sem conhecimentos biblioteconômicos. Entretanto, a questão relevante é entender que o fato do registro estar em XML, no formato MARC 21, não compromete sua representação. Pelo contrário, estando o registro bibliográfico e catalográfico neste formato, é possível apresentá-lo de acordo com a base de conhecimento do usuário, independente do tipo de instituição produtora do registro ou detentora do documento, conforme veremos no capítulo 5.

Experimentos mostram que usuários iniciantes tendem a usar uma única forma de representação por terem dificuldades em entender ou produzir outras representações, entretanto, se a forma de representação apresentada deixar o usuário tão à vontade a ponto deste “arriscar” usar outras formas, esta demonstra ser relevante. Assim, ambientes computacionais projetados para auxiliar na transição de uma forma de representação para outra, poderia sem dúvida ser muito útil.

Um ambiente computacional pode ajudar na manipulação de diversas formas de representação, como por exemplo, na editoração para construção e gerenciamento de representações, na manutenção da consistência em representações após alterações ou interações, no auxílio da utilização das representações de acordo com o estilo cognitivo do

usuário e tornando possível as adaptações quando surgirem novas representações (CASA, 1997).

Nesse contexto, agentes inteligentes seriam encarregados de gerenciar as características estruturais das representações, independentemente da informação que carregam. Neste caso, os agentes podem ser definidos como programas de computadores que teriam como principais características a possibilidade de executar tarefas sem a intervenção periódica, ou não, de um usuário, e ainda com a possibilidade de se comunicar com outros agentes e trocar informações a respeito de representações externas (RIBEIRO JÚNIOR, 2001).

Para o desenvolvimento desse ambiente computacional, é necessário desenvolver um modelo cognitivo que leve em consideração características existentes entre o relacionamento de pessoas com diferentes formas de representação, para aplica-lo no desenvolvimento de um modelo de interação entre a pessoa e o computador.

Não se tem a pretensão de desenvolver um agente, porém, estes agentes poderiam encontrar na XML o suporte e a estrutura para armazenar a informação necessária para seu funcionamento.

Na análise sobre formas de representação, percebe-se que uma evidente relação pode-se estabelecer com a linguagem de marcação extensível XML. Através de informações estruturadas segundo a metodologia da XML, é possível representar a mesma informação de diferentes formas, nesse sentido, a relação da XML com a catalogação, e as formas de representação será apontada no próximo capítulo.

3 Catalogação

O processo de catalogação pode ser identificado como a construção de um meio de comunicação, de um instrumento de ligação entre o usuário e o documento. Para que essa comunicação aconteça de forma coerente, Mey (1995, p. 7) destaca que *"a catalogação deve possuir as seguintes características: integridade, clareza, precisão, lógica e consistência"*. A autora define catalogação da seguinte forma (MEY, 1995, p.5):

Catalogação é o estudo, preparação e organização de mensagens codificadas, com base em itens existentes ou passíveis de inclusão em um ou vários acervos, de forma a permitir interseção entre as mensagens contidas nos itens e as mensagens internas dos usuários.

Podemos entender que a descrição padronizada de um documento de forma resumida, que amplia as formas de acesso de maneira que o documento possa ser localizado propiciando sua recuperação, identificação e a disseminação da informação contida nele, e ainda tornando o item documentário único caracteriza o processo ou objetivo da catalogação. A padronização citada esta baseada em regras seguidas por vários códigos, sendo que o mais disseminado é o AACR, Anglo American Cataloging Rules em sua 2ª edição. Tais regras têm o objetivo de construir uma forma de representação dos documentos para auxiliar no seu acesso.

O documento catalogado pode estar em diversos tipos de suportes como, por exemplo, papel, fita magnética, CD-ROM, pedra, papiro, pergaminho, vinil, madeira etc. A catalogação geralmente é realizada em uma unidade de da catalogação em uma biblioteca ou centros de informação, com o objetivo de representar descritivamente um documento para sua utilização. De acordo com Mey (1995, p.38):

A catalogação deve individualizar os itens, de forma que não sejam confundidos entre si, reunir itens por suas semelhanças, estabelecendo relações entre si, e finalmente, permitir a localização de um item específico em acervo determinado.

A catalogação exige que a instituição gaste recursos financeiros, tempo e recursos humanos, especializados, num processo às vezes repetitivo que deveria ser feito uma única

vez. O livro de Barbosa (1978, p.72-73) expressa este ponto de vista de maneira mais interessante:

Qual o preço de uma catalogação ? qual o tempo gasto na elaboração de uma ficha? São perguntas com varias respostas, pois a catalogação é uma atividade de múltiplos aspectos e verificação. Depende de vários fatores: tipo de biblioteca, tipo de registro desejado, completo ou simplificado; e um particular, de pessoal qualificado que tenha, principalmente, conhecimentos técnicos, de língua estrangeira e cultura geral. Acrescente-se a tudo isto a dificuldade na análise e interpretação das páginas de rosto cuja diagramação, na maioria das vezes, omite dados fundamentais o que provoca excessiva demora nas pesquisas essenciais a descrição correta de uma obra e na determinação exata de sua autoria principal.

Diante dessas inquietações percebe-se que um ponto determinante para o sucesso da catalogação é o recurso humano especializado. A ausência de bibliotecários catalogadores qualificados poderá comprometer a representação e recuperação das informações contidas nos itens documentários. Santos (1995, p.35) apresenta a questão do seguinte modo:

É preciso concentrar esforços para que a disciplina Representação Descritiva deixe de ser encarada como um recurso meramente técnico da Biblioteconomia, em que o treino da redação de fichas catalográficas serão suficientes para capacitar um profissional ao desempenho de sua função em um sistema de informação. Providências imediatas são necessárias para modernizar e expandir o preparo do bibliotecário no que se refere a catalogação. É importante que se pense por na reformulação do conteúdo e da metodologia da disciplina de Representação Descritiva, para ser assimilada de forma coerente pelo aluno. É preciso que ele compreenda o papel da catalogação e conheça sua relação e inter-relação no contexto informacional de um sistema. É necessário, enfim, se desmistificar a idéia de que o fazer da catalogação seja um produto absoluto e completo que se encerra nele mesmo.

No que diz respeito à qualidade da catalogação, percebemos na literatura uma preocupação com as *"possíveis dúvidas quanto à descrição do registro bibliográfico à luz do AACR2"* (DIAS, 1999, p. 34) e com a necessidade de *"se tomar decisões corretas diante das inúmeras dúvidas, surgidas no processo de catalogação"* (DIAS, 1999, p. 2).

As tentativas de agilização do processo de catalogação possibilitaram a criação de três sistemas de alimentação de catálogos e bases de dados bibliográficos e catalográficos:

- Catalogação na publicação: discutida teoricamente por Charles Jewett em 1850 e apresentada como um programa em 1948 por Ranganathan como catalogação pré-natal, em que editores são motivados a participar de modo que os documentos publicados circulem já catalogados por centrais especializadas;
- Catalogação cooperativa: idealizada também por Charles Jewett em 1850, nesse programa bibliotecas participantes catalogam em cooperação e se responsabilizam pela padronização, normalização e distribuição dos registros para as instituições co-operantes;
- Catalogação centralizada: efetivada em 1901 pela LC como uma central responsável pela catalogação e distribuição das representações as instituições associadas.

Esses programas desenvolvidos com o objetivo de criar mecanismos que possibilitassem a catalogação de um documento uma única vez, permitindo a otimização do processo e a redução de custos.

A trajetória histórica da catalogação mostra que esta, sempre utilizou as tecnologias vigentes em cada época em busca de métodos mais eficientes e econômicos para facilitar a representação, a recuperação e a disseminação da informação. Atualmente a informação caracteriza-se como insumo básico para uma sociedade que exige agilidade informacional, e como vimos, um dos propósitos da catalogação é possibilitar em tempo menor o acesso e a utilização da informação em um documento, através dos meios de comunicação disponíveis.

Mas a história da catalogação tem seu início com a escrita, desde os primitivos suporte, ou seja, a partir da possibilidade de registrar os pensamentos em um suporte informacional, aparece a necessidade e a preocupação em recuperar de alguma maneira o que foi produzido. Os processos de descrição e organização evoluíram até as estruturas atuais de representação e organização dos catálogos.

Os catálogos foram determinantes para as bibliotecas, pois segundo Mey (1995, p. 9):

Catálogo é um canal de comunicação estruturado, que veicula mensagens contidas nos itens, sobre os itens, de um ou vários acervos, apresentado-as sob forma codificada e organizada, agrupadas pôr semelhanças, aos usuários desse(s) acervo(s).

As formas primitivas de catálogos foram descobertas em escavações hititas, datadas de 1300 a.C., sendo em sua forma tabletes de argila contendo informações, identificando o número do tablete numa série, o título e escriba. Em 650 a.C, novamente são encontrados tabletes registrando título, número do tablete, as primeiras palavras do próximo tablete, nome do possuidor, nome do escriba e um selo, na biblioteca assíria, do rei Assurbanipal. Os mosteiros foram os preservadores, copistas e catalogadores de livros na Idade Média, sendo que no início eram feitos apenas inventários dos acervos. O primeiro catálogo considerado como tal foi à lista do convento de Saint Martin (VIEIRA, 2003).

O governo da Revolução Francesa estabeleceu o primeiro código nacional de catalogação no século XVIII, por isso, se deve a este o uso de catálogos em fichas, pois, a falta de papel fez com que cartas de baralho fossem usadas. Já no século XIX, Panizzi estabelece suas 91 regras, aprovadas pelo parlamento britânico e influenciando decisivamente a Biblioteconomia inglesa e americana. Otlet e La Fontaine desenvolveram os primeiros passos para o Controle Bibliográfico Universal, iniciando com o levantamento e registro das publicações editadas no mundo (VIEIRA, 2003).

Os códigos nacionais de catalogação tornaram-se populares no século XX, porém, a proposta de padronização ocorre com o surgimento do Anglo American Cataloguing Rules, AACR, e o Código da Vaticana. A evolução do processo de construção de formas de representação catalográfica é marcada pela segunda edição do AACR e pela iniciativa da Biblioteca do Congresso Americano, Library of Congress – LC, com o desenvolvimento do projeto Machine Readable Cataloging, MARC, conforme apresenta o próximo capítulo.

A análise dos sistemas de alimentação de catálogos – a Catalogação na Publicação, a Catalogação Cooperativa, e a Catalogação Centralizada – mostra que o objetivo é prover um meio de realizar o trabalho de representação descritiva em uma única vez, a partir de um processo que garanta a qualidade de representação de um documento de modo descritivo com interesse de prover a sua posterior recuperação e uso.

Outro ponto que chama a atenção é a tradicional forma de representação documentária por meio da ficha catalográfica. Um catálogo de fichas de uma biblioteca é notavelmente reconhecido pela característica de fornecer uma grande quantidade de informação em um pequeno espaço, e sua principal relevância está na sua estrutura padronizada e flexível, conforme a Figura 1. A dificuldade na substituição dessa forma de representação é percebida pelos desenvolvedores de catálogos online, não no aspecto técnico, mas de aceitação de outras representações:

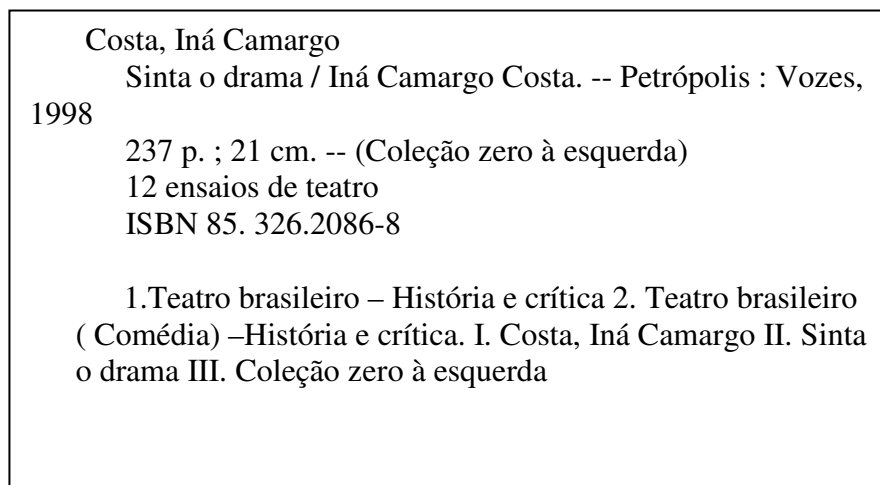


Figura 1. Uma ficha catalográfica.

A ficha catalográfica como forma de representação documentária é o resultado, entre outras coisas, do uso das tecnologias disponíveis na época de seu desenvolvimento. Com o uso do computador nos processos biblioteconômicos surgem novas estratégias de desenvolvimento catalográfico.

A funcionalidade prática de uma ficha catalográfica se destaca no processo de identificação e recuperação do documento, visto que a mesma apresenta uma forma de representação descritiva do documento. Na prática, o usuário pode chegar até o documento após fazer uma pesquisa nas fichas catalográficas disponíveis no catálogo da unidade de informação. Em uma unidade de informação automatizada, a disponibilidade ocorre em catálogos on-line.

A existência das fichas catalográficas deve ocorrer durante o processo de transição e de assimilação do usuário de uma forma de representação para outra, da manual para a digital. Esta pesquisa, entretanto, preocupa-se em destacar a nova versão do MARC 21 em XML por sua capacidade de prover diferentes formas de representação documentária e informacional, a partir da necessidade do usuário.

Como veremos no capítulo 5, sobre XML, prover uma estrutura para diferentes formas de representação de uma informação, é um dos aspectos marcantes dessa linguagem, portanto, se um registro catalográfico em MARC 21 foi desenvolvido de acordo com as especificações da XML, ele pode ser apresentado de diferentes formas.

Cabe destacar que os três sistemas de alimentação de catálogos – a Catalogação na Publicação, a Catalogação Cooperativa, e a Catalogação Centralizada – podem beneficiar-se da computação para sua realização, em especial no uso otimizado do formato de intercâmbio de dados bibliográficos e catalográficos MARC 21, objeto de estudo do próximo capítulo.

3.1 A relação da Catalogação e as Formas de Representação

A relação entre a teoria sobre formas de representação e catalogação ocorre ao considerar que o processo de catalogação tem como objetivo criar um meio de comunicação entre o usuário e o documento, por isso, a produção de formas de representação acaba sendo natural a medida que as necessidades dos usuários definem-se de diferentes modos, por

outro lado, as formas de representação não precisam conter todas informações sobre uma certa realidade.

O fato de uma forma de representação prover a informação que é pertinente para a realização de uma determinada tarefa, justifica sua padronização. O uso de códigos na catalogação para o desenvolvimento de uma forma de representação bibliográfica e catalográfica aceita universalmente é baseada em regras, sendo recomendado o uso do Código de Catalogação Anglo-Americano, 2^a. edição - AACR2. Tais regras têm por objetivo facilitar a construção da representação de documentos para a alimentação de bases de dados e catálogos.

A catalogação é uma forma de representação sucinta e padronizada de um item documentário, que tem como objetivo ampliar as formas de acesso a um documento facilitando tarefas e os processos de localização de documentos e informações. Por isso, destacamos a importância do desenvolvimento de métodos e processos que garantam a qualidade das formas de representação para a catalogação e a sua aceitação. Historicamente essa tentativa foi marcada pelo uso dos catálogos, como meios de comunicação estruturados, que apresentam informações contidas nos documentos, sobre os documentos, caracterizados por uma forma de representação codificada e organizada, agrupadas pôr possíveis necessidades dos usuários.

O propósito da análise da teoria de formas de representação sobre o enfoque dos objetivos da catalogação é oferecer subsídios para a construção de estruturas que tornem possíveis apresentar de modo consciente as formas de representações de acordo com o problema a ser resolvido e de acordo com as habilidades cognitivas do usuário. Esse propósito está direcionado para identificar a relação existente entre a catalogação, formas de representação e a hipótese de que a versão do MARC 21 em XML se apresenta como uma opção de metadados, pois, oferece uma estrutura que permite a representação de dados bibliográficos de modo a serem lidos por usuários e por máquinas.

Ao destacar a importância da XML para a Ciência da Informação o objetivo é apresentar a relação entre essa linguagem de marcação extensível e a catalogação como forma de representação descritiva documentária, baseado-se na própria estrutura da XML. Especificamente será apresentado o diferencial da versão em XML do MARC 21 como argumento principal para afirmar que, atualmente, a discussão sobre representação descritiva deveria estar voltada para o desenvolvimento de ferramentas de software e para a construção de sistemas que garantam, ou que possibilitem, a interoperabilidade⁵ entre os sistemas disponíveis para a representação documentária. No próximo capítulo, a discussão se pautará na importância do MARC 21 para a catalogação.

⁵ É comum encontrar na literatura definições diversas para o termo interoperabilidade, visto que para defini-lo, é mais apropriado inserir a definição em um contexto específico. Podemos definir interoperabilidade entre bancos de dados, softwares, hardwares, ambientes web, arquiteturas de redes e vários outros contextos. Assim, no contexto da pesquisa apresentada, interoperabilidade é a capacidade de compartilhamento de informações entre softwares, independente da estrutura de armazenamento dos dados usada em seu banco de dados, ou seja, a estrutura de armazenamento dos dados não impede a troca de informações entre instituições.

4 Machine Readable Cataloging – MARC 21

Diante das discussões em torno da necessidade de metadados, métodos e teorias e da pouca implementação efetiva de ferramentas de software para contemplar o processo de descrição e intercâmbio de dados bibliográficos e catalográficos no sentido de gerar níveis mais elevados de qualidade na catalogação, especialmente, dos documentos eletrônicos disponíveis na web, acreditamos ser relevante a análise do formato MARC em sua nova versão como uma forma adequada de representação documentária.

A versão MARC 21 em XML pode ser considerada como uma opção de metadados para a documentação de acessos específicos superior aos seus, por assim dizer concorrentes, em razão do seu formato inicial ser o mais amplamente utilizado pela comunidade biblioteconômica, e ser a estrutura de representação da maior base bibliográfica e catalográfica disponível atualmente no mundo (MARC21, 2002).

O formato MARC na sua estrutura original apresenta-se de modo altamente estruturado a partir de padrões internacionais e está estruturado por meio de etiquetas e tem como similares os formatos Text Encoding Initiative (TEI), Cultural Heritage Sources and Objects (CIMI) e Encoded Archival Description (EAD). Já o formato MARC em XML apresenta-se como similar ao Dublin Core, planilha Internet Anonymous FTP Archive (IAFA), Request For Comments (RFC1807), Summary Objects Interchange Format (SOIF) e Lightweight (LDIF), como formatos estruturados a partir de padrões emergentes, e estão estruturados por campos (DEMPSEY; HEERY, 1997).

A diferença entre o formato MARC em sua estrutura original criada em 1960 e o formato atual em XML está na forma em que os dados bibliográficos e catalográficos são organizados para que a comunicação, ou troca de informações, entre computadores ocorra. Ou seja, como veremos em mais detalhes, as etiquetas, indicadores e subcampos continuam sendo fundamentais para a interoperabilidade entre softwares de catalogação.

Outro aspecto é a grande abrangência descritiva que o formato oferece e que atualmente está disponível em XML, principal linguagem de marcação usada por outras iniciativas, assim, sua capacidade descritiva e o volume de registros disponíveis o torna indiscutivelmente superior como padrão para representação documentária, especialmente em sua versão XML, assunto do capítulo 6.

Com o objetivo de reduzir custos, tempo, recuperar e compartilhar informações já catalogadas sobre itens documentários, a catalogação vem usando as diversas tecnologias disponíveis ao longo de cada época como as tecnologias semimecanizadas tais como a máquina de escrever, estêncil etc, as tecnologias mecanizadas como as máquinas xerográficas, tabuladores, entre outras, e atualmente as tecnologias automatizadas, representadas pelos computadores.

Nos anos de 1950 os Estados Unidos utilizavam sistemas mecanizados para a produção de catálogos e a administração de empréstimos de livros nas bibliotecas, com o aumento da produção de documentos nas instituições acadêmicas e conseqüentemente das atividades de pesquisa, foi necessário prover meios para que os serviços biblioteconômicos fossem agilizados, e com isso, atender as exigências dos usuários. De acordo com Barbosa (1978, p. 196),

A necessidade de prover serviços em maior profundidade de forma mais rápida e um maior número de usuários, bem como o aumento quantitativo dos materiais tradicionais, acrescentado ao aparecimento de novas formas de materiais, levaram as bibliotecas dos países desenvolvidos a optar pelo uso de computadores para processamento de suas operações internas. Por meio de processos simplesmente manuais tornava-se impossível garantir o tratamento técnico atualizado das coleções e o atendimento, em tempo hábil, aos usuários. Exigia-se, portanto, um melhor nível de serviços.

Com a popularização dos computadores nos anos de 1960 a contribuição para o avanço da tecnologia e da ciência em geral foi inevitável, o que no campo da Biblioteconomia, permitiu a automação de bibliotecas e o desenvolvimento de um formato de catalogação legível de intercâmbio de dados por máquinas. Como visto, os formatos são utilizados para

promover um intercâmbio de dados bibliográficos de forma que os dados organizados dentro de um padrão possam ser interpretados por computadores.

Um formato caracteriza-se por uma estrutura na qual os dados que descrevem o documento são inseridos. Essa estrutura contém campos que indicam como cada um dos dados referentes ao documento serão inseridos. Os formatos têm como objetivo descrever as informações do documento de tal modo que um computador possa interpretá-lo, criando um cenário em que qualquer usuário, em qualquer unidade de informação, recupere o documento.

Depois de diversas pesquisas, segundo Souza (2002), o formato Machine-Readable Cataloging – MARC, foi o primeiro formato de intercâmbio de dados criado para a catalogação automatizada. Isso ocorreu na década de 1960, nos Estados Unidos e foi desenvolvido pela Library of Congress – LC. No início o formato MARC foi desenvolvido para uso da LC, mas acabou sendo implantado em outras instituições para ser testado. De acordo com Barbosa (1978, p. 199):

- O MARC começou como um projeto piloto distribuindo os registros da L.C., em forma legível por computador, para as bibliotecas participantes. O projeto piloto resultou em :
- a) um formato padrão para intercâmbio de todas as formas de dados bibliográficos;
 - b) um conjunto generalizado de programas para a criação de registros bibliográficos em forma legível por computadores na LC., e
 - c) o desenvolvimento de um serviço de distribuição para tais registros.

O formato apresentava como objetivo, de acordo com a autora, “*testar os benefícios e problemas da produção centralizada de registros catalográficos legíveis por computador e distribuí-los às bibliotecas usuárias*” (BARBOSA, 1978, p. 202). Além deste objetivo, Barbosa (1978, p. 205), ainda apresenta a finalidade do formato e amplia os objetivos do MARC:

A finalidade do MARC não é apenas a de facilitar a circulação dos dados catalográficos (inclusive em plano internacional) usando uma linguagem comum em sistema de informação tão flexível que se preste às mais diversas exigências de apresentação formal de documentos; consiste, também, numa tentativa mais funcional da análise das unidades de

informação contidas numa ficha catalográfica, permitindo controlá-las e recuperá-las o mais rapidamente possível.

O sistema MARC tem por objetivos:

- a) proporcionar um registro bibliográfico central para o uso da LC;
- b) proporcionar uma base de dados para os serviços bibliográficos norte-americanos; e
- c) fornecer à comunidade internacional informações bibliográficas.

As evoluções do projeto MARC podem ser consideradas notáveis, visto que a idéia passou por modificações até este ser aceito como um formato que tornava possível o intercâmbio de dados bibliográficos e catalográficos entre bibliotecas através dos computadores, sendo determinante para os sistemas de catalogação de alimentação de catálogos e bases de dados: a catalogação na publicação, a catalogação centralizada e especialmente a catalogação cooperativa.

Sua evolução é marcada pela elaboração do formato MARC II, fruto de um exame detalhado no projeto piloto MARC. Mais completo que o projeto piloto, MARC II originou vários outros, conforme atestado por Barbosa (1978, p. 212):

Por ter sido considerado como uma linguagem padrão para intercâmbio de informações bibliográficas, o MARC II passou a interessar outros países que, com algumas alterações, adotaram-no na compilação de suas bibliografias nacionais e serviços centralizados de catalogação.

Atualmente, trabalham no mesmo esquema a Inglaterra, com o UK/MARC; a França, com o MONOCLE; o Canadá, com Canadian MARC; a Bélgica, com MARC/BR; a Espanha, com o IBERMARC; o México, com o MARC/México; o Brasil, com o CALCO; e ainda, o INTERMARC, elaborado por representantes da França, Bélgica, Holanda, Suíça e Grã-Bretanha, visando ao intercâmbio de dados catalográficos.

Temos ainda referências ao MARCAL, MARC para a América Latina; UNIMARC, iniciativa internacional com o objetivo de ser utilizado entre vários países; Common Communication Format, CCF; IBICT, união do UNIMARC com o CCF e o considerado mais atualizado USMARC, pois foi o único que passou por revisões e aperfeiçoamentos.

Não temos o objetivo de destacar as diferenças existentes entre as versões derivadas da evolução do projeto MARC, mas podemos notar que o projeto inicial se consolidou com o

USMARC, na década 1980 e posteriormente no MARC 21 no final da década de 1990, sendo o número 21 uma referência ao atual século. Sua evolução ilustra muito bem sua abrangência e potencialidade. MARC 21 é o resultado da integração entre USMARC e o CAN/MARC, um projeto MARC canadense (MARC21, 2002).

4.1 Aspectos técnicos do formato MARC 21

Nesta breve introdução a respeito do projeto MARC, percebemos que um registro MARC 21 tem como objetivo tornar um registro bibliográfico legível por um computador. Legível por máquina deve ser entendido como o processo que um computador irá executar para ler e interpretar os dados bibliográficos e catalográficos (FURRIE, 2002).

Por outro lado, um registro bibliográfico ou registro catalográfico contém informações que tradicionalmente são encontradas em uma ficha catalográfica conforme apresentado na Figura 1 do capítulo 3. Deste modo, o primeiro passo é entender o significado da forma de representação de um registro MARC 21.

Duas formas de representação de um registro MARC 21 para o usuário serão apresentadas, e como o próprio estudo das formas de representação indicam, ambas as formas contém informações que são encontradas em uma ficha catalográfica convencional (FURRIE, 2002).

É importante destacar que na própria utilização do formato MARC foi projetado um ambiente computacional para auxiliar na transição de uma forma de representação para outra, considerando a familiaridade do usuário com uma representação já construída, conforme foi abordado na seção 2.3 do capítulo 2 sobre formas de representação.

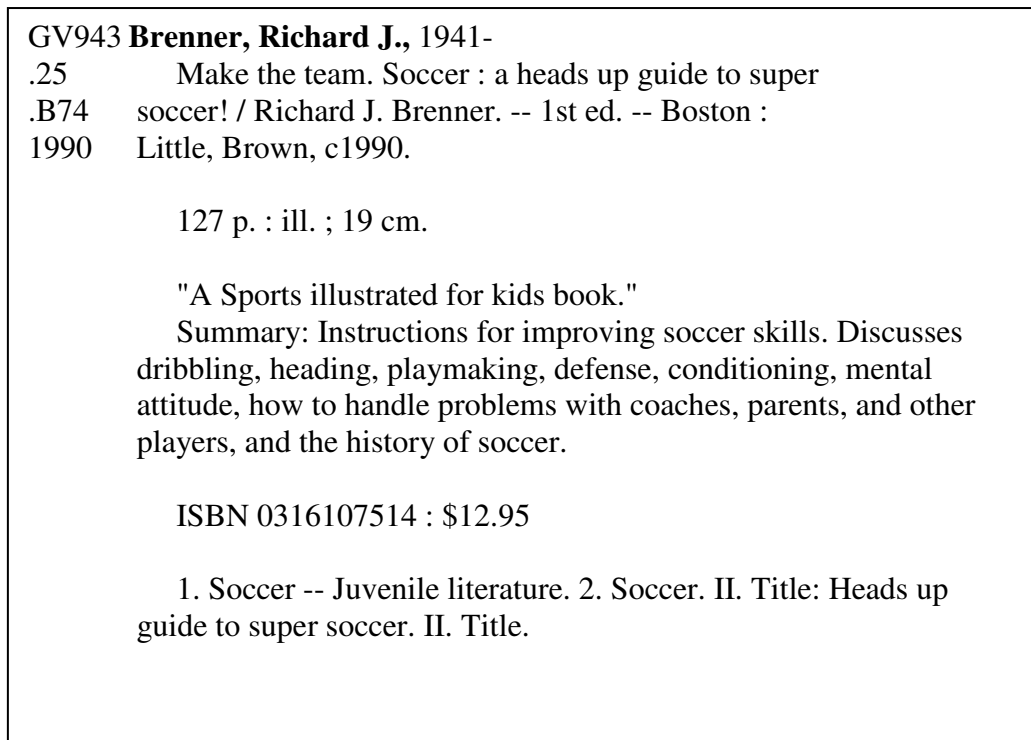


Figura 2. Ficha catalográfica ilustrando as formas de representação do MARC 21.

Como ponto de partida apresentamos a seguir um exemplo de um registro MARC 21 conforme é tradicionalmente apresentado para o usuário do sistema. É importante destacar que eventualmente ocorre uma variação entre as interfaces, tais como métodos diferentes de oferecer ajuda ou confecção da tela do software, entretanto, as informações apresentadas na Figura 3, estarão presentes em qualquer sistema que importe registros bibliográficos e catalográficos no formato MARC 21:

Líder	01041cam 2200265 a 4500
Nº. de Controle	001 ###89048230
Iden. Nº. Controle	003 DLC
Ult. Intervenção	005 19911106082810.9
Campos Fixos	008 891101s1990 maua j 001 0 eng
Nº. Controle LC	010 ## \$a ###89048230
ISBN	020 ## \$a 0316107514 : \$c \$12.95
ISBN	020 ## \$a 0316107506 (pbk.) : \$c \$5.95 (\$6.95 Can.)
Fonte de Catalog.	040 ## \$a DLC; \$c DLC; \$d DLC
Nº Chamada LC.	050 00 \$a GV943.25; \$b .B74 1990
Classif. Dewey	082 00 \$a 796.334/2; \$2 20
Nome Pessoal	100 1# \$a Brenner, Richard J., ; \$d 1941-
Título	245 10 \$a Make the team. \$p Soccer : \$b a heads up guide to super soccer! / \$c Richard J. Brenner.
Título Variante	246 30 \$a Heads up guide to super soccer
Edição	250 ## \$a 1st ed.
Imprenta	260 ## \$a Boston : \$b Little, Brown, \$c c1990.
Desc. Física	300 ## \$a 127 p. : \$b ill. ; \$c 19 cm.
Nota Geral	500 ## \$a "A Sports illustrated for kids book."
Nota Resumo	520 ## \$a Instructions for improving soccer skills. Discusses dribbling, heading, playmaking, defense, conditioning, mental attitude, how to handle problems with coaches, parents, and other players, and the history of soccer.
Assunto:	650 #0 \$a Soccer \$v Juvenile literature.
Assunto:	650 #1 \$a Soccer.

Figura 3. Uma das formas de representação de um registro MARC 21

Esta forma de representação é destacada por três aspectos, conforme pode ser comparada com a Figura 3. Primeiro, o leitor pode perceber que cada informação é precedida por um número formado sempre de três caracteres, por exemplo, 005, 010, 100 e 245. Estes são denominados usualmente como etiquetas MARC, e cada uma tem seu significado, por exemplo, 245 contém o título do documento.

Segundo, logo após as etiquetas encontramos os indicadores, compostos de dois caracteres, como no caso da etiqueta 100, os caracteres #1 caracterizam-se como indicadores, ou por uma seqüência de caracteres, como no caso da etiqueta 008, que neste caso não são denominados indicadores. Em ambos os casos cada caractere tem seu significado.

O terceiro e último ponto, são os caracteres encontrados após o símbolo \$, como por exemplo, na etiqueta 100, tem-se os caracteres \$a, sendo o \$, identificador de subcampos e o “a” o subcampo propriamente dito. Esses caracteres, assim como no caso dos indicadores, têm significados distintos.

A forma de representação apresentada na Figura 3 é utilizada na interface com o usuário no processo de intercâmbio de dados a representação apresenta-se como mostra a Figura 4. (FURRIE, 2002):

```
01041cam 2200265 a 4500001002000000000300040002000
50017000240080041000410100024000820200025001060200
04400131040001800175050002400193082001800217100003
20023524500870026724600360035425000120039026000370
04023000029004395000042004685200220005106500033007
30650001200763^###89048230#/AC/r91^DLC^19911106082
810.9^891101s1990####maua###j#####000#0#eng##^##$
a###89048230#/AC/r91^###$a0316107514 :$c$12.95^##$a
0316107506 (pbk.) :$c$5.95 ($6.95 Can.)^##$aDLC$cD
LC$dDLC^00$aGV943.25$b.B74 1990^00$a796.334/2$220^
10$aBrenner, Richard J.,$d1941-^10$aMake the team.
$pSoccer :$ba heads up guide to super soccer! /$cR
ichard J. Brenner.^30$aHeads up guide to super soc
cer.^##$a1st ed.^##$aBoston :$bLittle, Brown,$cc19
90.^##$a127 p. :$bill. ;$c19 cm.^##$a"A Sports ill
ustrated for kids book."^##$aInstructions for impr
oving soccer skills. Discusses dribbling, heading,
playmaking, defense, conditioning, mental attitud
e, how to handle problems with coaches, parents, a
nd other players, and the history of soccer.^#0$aS
occer$vJuvenile literature.^#1$aSoccer.^
```

Figura 4. Arquivo texto com um registro MARC 21

Entretanto, cabe esclarecer que apesar da figura 4 apresentar a ilustração do registro em várias linhas, um registro MARC 21 é composto por uma única linha com todos os caracteres que o representam. No processo de intercâmbio de um registro MARC 21 os dados da figura 4 serão utilizados pelo sistema da biblioteca no processo de importação e de exportação de registros.

As formas de representações, conforme referencial teórico apresentado no capítulo 2 devem estar adaptadas para as tarefas que se pretende realizar contendo em alguns casos as

mesmas informações, de múltiplas formas, como o que se pode perceber, ao visualizar, a mesma informação, em três formas de representação, conforme as figuras 2,3 e 4.

Na figura 2 uma ficha catalográfica convencional, na figura 3 um registro MARC 21 na sua interface para o usuário e por último o registro em forma de representação para leitura de máquina na figura 4. No capítulo 6, esse mesmo registro será apresentado no formato MARC 21 em XML.

Um estudo mais detalhado do formato MARC 21, baseado nas informações encontradas no site oficial do projeto (MARC21, 2002), no trabalho de Furrie (2002) e de Ferreira (2002) será apresentado a partir de três elementos que formam um registro MARC 21, ou seja, a estrutura do registro, a indicação de conteúdo e os dados contidos nos registros.

A estrutura do registro é baseada na American National Standard for Bibliographic Information Interchange, ANSI Z39.2, e na Documentation – Format for Bibliographic Information Interchange on Magnetic Tape, ISO 2709. Os principais formatos de intercâmbio de dados internacionais utilizam estas normas.

A indicação de conteúdo é composta de códigos e convenções determinadas para identificar os dados encontrados no registro MARC 21, que por outro lado, permite a manipulação do registro. Esses códigos e convenções são definidos no MARC 21 e são compostos por informações referentes às etiquetas, indicadores e subcampos.

O conteúdo dos elementos que compõem um registro MARC 21 são as informações propriamente ditas, ou seja, os dados da ficha catalográfica. Essas informações geralmente são apresentadas conforme padrões externos, tais como, International Standard Bibliographic Description, ISBD, Anglo American Cataloging Rules, AACR, e Library of Congress Subjects Headings, LCSH, e outras convenções em alguns casos criadas pela instituição construtora do registro. Nos estudos dos elementos do MARC 21 outra divisão pode ser feita, agora, composta pelo Líder, Diretório e Campos Variáveis.

O Líder corresponde aos dados que fornecem informações detalhadas que tornam possível o processamento do registro. Composto por uma seqüência de 24 caracteres alfanuméricos cada qual com seu significado correspondente, determinado pela sua posição na seqüência. Os 24 caracteres compõem o primeiro campo de um registro MARC 21. Compare as figuras 3 e 4.

O Diretório é como um índice gerado pelo sistema, composto por uma série de entradas que trazem informações referentes à localização e o tamanho de cada etiqueta no registro MARC 21. O diretório é determinante para que o processamento do registro seja feito de forma correta. Cada notação possui 12 caracteres, iniciando com os campos variáveis, seguidas pelas etiquetas em ordem crescente. Um estudo mais pormenorizando sobre o Diretório está disponível logo à frente a partir de um estudo da Tabela 1.

Posteriormente surgem os campos de dados variáveis, organizados em ordem crescente, a partir do primeiro caractere da etiqueta. A seqüência do armazenamento dos campos de dados variáveis se apresenta necessariamente na ordem das entradas no Diretório. Quanto aos campos repetitivos, estes são diferenciados pela localização dos campos respectivos no registro. O fim do Diretório é determinado pelo caractere ASCII 1E16, ou seja, o caractere ^ (FURRIE, 2002).

Os Campos Variáveis são os dados organizados correspondentes a cada etiqueta, indicados por três caracteres numéricos, que estão registrados no Diretório e correspondem a cada campo. Existem dois tipos de campos variáveis, ou seja, campos de controle variável e campos de dados variáveis.

Os campos de controle variável são os campos 00X, e se caracterizam por não possuírem indicadores e subcampos. Estruturalmente são diferentes dos campos de dados variáveis por conterem um único dado ou uma série de dados de tamanho fixo que são identificados pela sua posição ou localização da seqüência de caracteres. As etiquetas de controle variável são:

- 001 – Número de controle;
- 003 – Identificador do número de controle;
- 005 – Data e hora da última intervenção;
- 006 – Campo de ligação;
- 007 – Campos fixos de descrição física;
- 008 – Campos fixos de dados.

Os campos de dados variáveis são agrupados em blocos de acordo com o primeiro caractere da etiqueta e estruturalmente podem conter indicadores e subcampos. Com algumas exceções, podem identificar a função do dado no registro. O restante da etiqueta identifica o tipo da informação do campo. A lista de etiquetas deste campo variável é muito extensa, e pode ser encontrada em detalhes no site do projeto MARC 21 (MARC21, 2002). Para idéia geral da potencialidade do formato apresenta-se o resumo a seguir:

- 010 ao 048 – Contêm números de controle e ligação, números e códigos relativos ao documento descrito no registro;
- 050 ao 088 – Contém números de classificação e chamada relacionados com o documento descrito no registro;
- 100 ao 130 – Contém o nome ou o título uniforme usado como entrada principal no registro bibliográfico;
- 210 ao 247 – Contém o título do item descrito no registro bibliográfico, e títulos variantes que também podem ser aplicados ao documento. A partir desses campos podem ser gerados pontos de acesso e exibir notas para os diversos títulos;
- 250 ao 263 – Contém informações sobre edição, impressão, endereço e outros detalhes sobre a fonte de publicação. Dados relacionados a formas específicas do material que se aplica ao documento descrito no registro;
- 300 ao 362 – Contém características físicas, frequência de publicação e preço. Também são registradas nestes campos informações sobre

disseminação do documento e status de segurança do dado bibliográfico relativo ao item;

- 400 ao 490 – Contém um autor ou título usado como série no registro bibliográfico;
- 500 ao 535 – Contém notas ligadas ao aspecto do documento descrito no registro que não são específicos de nenhum tipo particular de material ou controle. Notas especializadas são usadas quando o acesso ao dado é necessário e a nota é introduzida por uma palavra ou frase distinta;
- 600 ao 658 – Contém os cabeçalhos de assunto ou termos de acesso que possibilita acesso adicional a um documento descrito no registro. Os termos são definidos de acordo com a base de autoridade ou thesaurus;
- 700 ao 755 – Contém um nome, título ou termo que possibilita acessar o documento descrito no registro e que não tenha sido fornecido nos campos 1XX, 20X, 24X, 4XX, 6XX ou 8XX;
- 760 ao 787 – Contém dados referentes a itens relacionados, especificando o relacionamento existente;
- 800 ao 840 – Contém o nome ou título usado como entrada secundária de série quando a série principal foi fornecida na etiqueta 490, menção de série não determinada, ou etiqueta 500, nota geral;
- 940 ao 949 – Contém informações locais que interessam apenas a instituição proprietária do documento descrito no registro.

Os intervalos apresentados não significam necessariamente que a seqüência é exata, por exemplo, o conjunto de etiquetas do intervalo 100 ao 130 é composto pelas etiquetas 100, 110, 111 e 130. Este resumo tem como objetivo demonstrar a potencialidade do formato à medida que conduz o leitor a verificar sua abrangência.

Estruturalmente tudo começa a partir de identificada a etiqueta, o próximo passo é determinar se a mesma é composta por uma seqüência de caracteres que tem seu significado dependente da sua posição, como por exemplo, o Líder, as etiquetas 006, 007 e 008 ou se é composto por indicadores e sub-campos, estes dois últimos serão apresentados em mais detalhes.

O indicador corresponde as duas primeiras posições no início de cada campo de dados variáveis, e tem como objetivo interpretar ou complementar os dados contidos no campo. Os valores dos indicadores podem são numéricos entre 0 a 9, e são interpretados independentemente, ou seja, o seu significado está ligado a etiqueta em questão. O indicador 1, pode significar uma informação relacionada a etiqueta 100 e outra para 245. Alguns indicadores definem os conteúdos da etiqueta de modo mais detalhado, outros informam sobre a fonte de dados ou especificam o texto a preceder a etiqueta quando este é exibido ou impresso.

Os subcampos representam o nível inferior de designação de conteúdo nos registros MARC 21. E são compostos por dois caracteres, sendo um deles o delimitador representado por um caractere ASCII, que geralmente é o #, ou \$ e o outro pode ser uma letra minúscula ou números, por exemplo, \$a. Assim como no caso dos indicadores, os subcampos têm seu significado ligado diretamente a etiqueta, ou seja, \$a pode ter um significado para a etiqueta 100 e outro para a 245, e devem ser definidos separadamente para cada etiqueta, sendo sua ordem especificada por padrões de conteúdo de dados, como nas convenções de catalogação. Os subcampos geralmente ajudam a interpretar as etiquetas, e podem permitir que estas sejam manipuladas por razões especiais.

Nesse apanhado técnico sobre o formato MARC 21, teoricamente pode-se dizer que todas as etiquetas e subcampos podem ser repetitivas, porém, a natureza dos dados às vezes impede a repetição. Por exemplo, um registro pode conter apenas uma etiqueta 100, por sua vez esta, deve conter apenas um subcampo \$a, nome pessoal, mas pode conter mais de um subcampo \$c, títulos e outras palavras associadas ao nome.

4.2 Considerações sobre o MARC 21 para um analista de sistemas

Podemos dizer que o conceito de múltiplas formas de representação para programadores de computador não é novo. O trabalho desse profissional exige que este tenha em mente a necessidade de independentemente de como os dados estejam armazenados, prover uma interface amigável ao seu usuário. É comum ocorrer em programas de computadores o preenchimento de várias tabelas do banco de dados a partir de uma única tela apresentada ao usuário. Um cadastro de clientes, por exemplo, ilustra a idéia de que as formas de representações precisam apresentar apenas o necessário para a realização de uma tarefa.

Em outros casos, uma simples digitação de uma nota fiscal de entrada de produtos no sistema, ou em uma única tela apresentada ao usuário, pode desencadear uma série de processamentos que podem incluir a atualização do estoque, alteração de preços, contas a pagar, fluxo de caixa etc. e de acordo com a tarefa que esteja sendo realizada, não é necessário apresentar todas as informações.

No caso da construção de registros catalográficos legíveis por máquinas, segundo Furrie (2002), provavelmente apenas programadores de computadores poderiam ter interesse e contato com a forma de representação de um registro MARC 21 conforme a figura 4. Entretanto, essa forma de representação caracteriza a essência do MARC 21, pois, é neste formato que os registros bibliográficos são apresentados ao programa de computador, e este, deve estar preparado para interpretar e processar os dados.

No módulo de catalogação, uma ferramenta de software, da maioria dos sistemas gerenciadores de bibliotecas é possível encontrar a opção de importar e exportar dados bibliográficos e catalográficos em formato MARC 21, entretanto, a partir de agora, tem-se o objetivo de entender como o software deve navegar pelos dados a fim de poder interpretá-los e processá-los e assim realizar a importação dos dados. Cabe lembrar que apesar de uma maior atenção para o processo de importação, as informações aqui apresentadas poderão auxiliar o analista de sistemas no desenvolvimento de um programa que também seja capaz

de exportar os dados bibliográficos, ou seja, gerar um arquivo texto no formato encontrado na figura 4.

Basicamente o programador deverá desenvolver um programa de computador que seja capaz de navegar caractere por caractere dando-lhes significado. Para isso, será necessária a criação de uma base de dados que armazene as etiquetas e suas respectivas posições, no caso do Líder, etiquetas 006, 007 e 008, ou indicadores e subcampos no caso das outras etiquetas. Na prática o programador criará tabelas no sistema gerenciador de banco de dados usado pelo seu cliente, tais como, Oracle, Microsoft SQL-Server, MySQL ou PostgreSQL, que possibilitará compreender o significado de cada caractere.

É importante destacar que o MARC 21 é um formato ou método, que informa como um registro bibliográfico e catalográfico deve ser organizado para que sua importação ou exportação ocorra com sucesso, sendo assim, como o registro é armazenado em um sistema gerenciador de banco de dados é de responsabilidade de cada administrador de banco de dados, conhecidos também como DBA, ou seja, as estruturas criadas para armazenar os dados bibliográficos e catalográficos possivelmente serão diferentes, visto que cada DBA usará a técnica que mais lhe agrada. É verdade que para esse trabalho, existe um método chamado de normalização de dados, que por sua vez é dividido em cinco estágios, e o DBA não é obrigado a chegar ao último. Veja uma sugestão destas tabelas criadas no Microsoft ACCESS 2000. A documentação completa das tabelas pode ser consultada no anexo A:

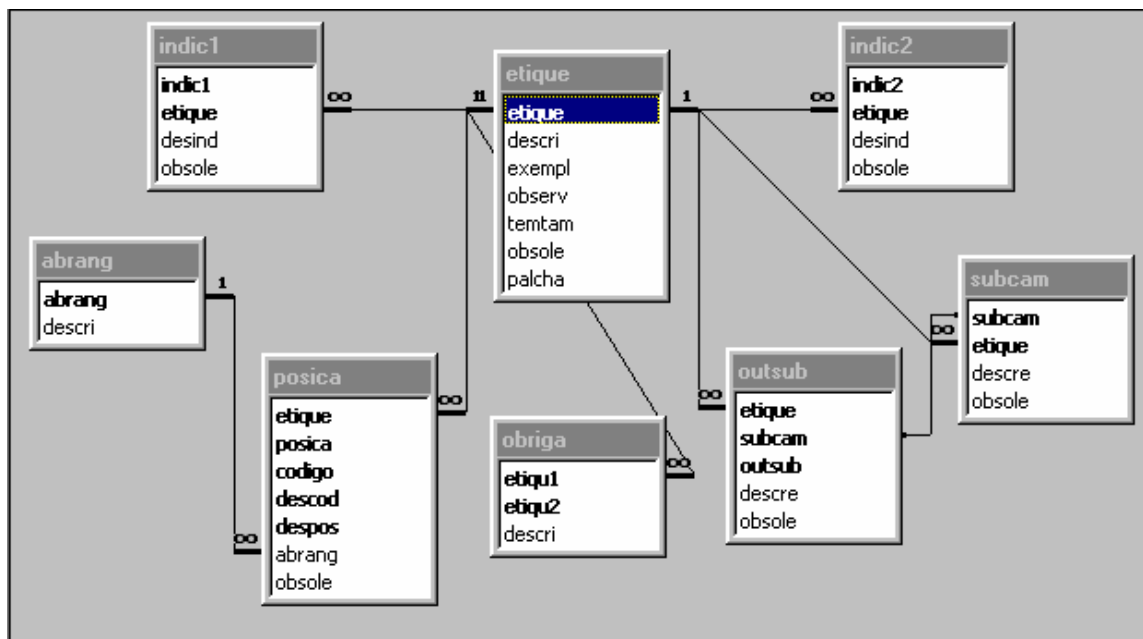


Figura 5. Armazenar o significado das etiquetas, posições, indicadores e subcampos

Com uma base de dados como esta, preenchida devidamente de acordo com as informações encontradas em MARC21 (2002), ou Ferreira (2002), o programador tem a principal ferramenta para poder desenvolver um programa capaz de processar os dados bibliográficos e catalográficos. O próximo passo é saber como navegar nos registros conforme figura 4.

O programa de computador desenvolvido deverá realizar a leitura dessa forma de representação, figura 4, levando em conta que os primeiros 24 caracteres correspondem ao Líder, e, cada caractere tem um significado de acordo com a posição que ocupa na seqüência, que de acordo com a figura 4 termina em 4500, portanto, o primeiro ponto que justifica a criação da base apresentada na figura 5.

Logo em seguida o programa de computador iniciará a leitura do Diretório, que é formado por conjuntos de doze caracteres sendo os três primeiros referentes à etiqueta, os quatro posteriores referente à quantidade de caracteres que compõe a informação da etiqueta e os cinco últimos referentes ao início do registro tendo como ponto de partida o primeiro caractere após o final do diretório, iniciando em zero (FURRIE, 2002).

O diretório é finalizado por um caractere ASCII 1E16, que neste caso é o ^, e termina no último conjunto de doze caracteres, **650001200763^**. Os delimitadores são essenciais para que o processamento seja feito corretamente, portanto, o programa terá que ser capaz de identificar os diversos delimitadores que podem surgir ao longo de um registro MARC 21, por exemplo, #, ^, \$, / e \ etc. A seguir uma tabela referente a figura 4:

Etiqueta	Tamanho	Inicia em	Etiqueta	Tamanho	Inicia em
001	0020	00000	100	0032	00235
003	0004	00020	245	0087	00267
005	0017	00024	246	0036	00354
008	0041	00041	250	0012	00390
010	0024	00082	260	0037	00402
020	0025	00106	300	0029	00439
020	0044	00131	500	0042	00468
040	0018	00175	520	0220	00510
050	0024	00193	650	0033	00730
082	0018	00217	650	0012	00763

Tabela 1. Entendendo o diretório de um registro MARC 21.

Após o termino do diretório iniciam-se os dados referentes a cada etiqueta, portanto, estarão na mesma ordem que as etiquetas foram apresentadas no diretório, e neste caso, sem a necessidade de indicar novamente a etiqueta, visto que os dados correspondem às mesmas.

A tabela informa que a etiqueta 008 tem 41 caracteres compondo as informações que esta traz, iniciando coincidentemente, no 41º caractere, a partir do final do diretório. Se o leitor contar até quarenta e um, começando do zero, e partindo do final do diretório, **650001200763^**, sendo o ^ igual a zero, encontrará as informações correspondentes à etiqueta 008, a saber, **^891101s1990####maua####j#####000#0#eng##**, que tem 41 caracteres, e, cada um tem o seu significado segundo a posição que ocupa na seqüência (MARC21, 2002), (FURRIE, 2002).

Em uma análise de uma etiqueta que tem indicadores e subcampos, tem-se a etiqueta 010 com 24 caracteres compondo as informações que esta traz, iniciando, no 82º caractere, a partir do final do diretório. Se o leitor contar até oitenta e dois, começando do zero, e partindo do final do diretório, **650001200763^**, sendo o ^ igual a zero, encontrará as informações correspondentes à etiqueta 010, a saber, **^##\$a###89048230#/AC/r91**, que tem 24 caracteres, sendo que, o primeiro caractere corresponde ao delimitador, os dois posteriores aos indicadores, que no caso significam vazio, depois o subcampo “a”, identificado com seu delimitador, \$a, que também tem o seu significado, sendo o restante as informações correspondentes ao subcampo “a” da etiqueta 010 (MARC21, 2002), (FURRIE, 2002). Outros subcampos são precedidos por um delimitador, por exemplo, \$.

Os dois exemplos apresentados aqui devem ser suficientes para perceber como o programa de computador deverá comportar-se. Cabe esclarecer que o significado das etiquetas, indicadores, subcampos e posições podem ser encontrados em MARC21(2002) ou Ferreira (2002), o objetivo é mostrar como desenvolver um programa de computador que interprete e consiga processar corretamente a representação dos dados bibliográficos a partir do formato MARC 21.

Para mostrar como podem ocorrer sérias conseqüências se o analista de sistemas não estiver atendo as considerações do formato, apresenta-se um breve estudo nos aspectos técnicos das etiquetas 006 e 007, visto que os significados dos códigos contidos no registro estarão associados com a posição do caractere dentro da seqüência, o que requer atenção deste profissional.

Este estudo mostra características que se não facilmente percebidas por analistas de sistemas ou bibliotecários, podem gerar alguns problemas na validação de dados. Justifica-se com esse estudo, a importância da qualidade dos dados nos processos de importação e exportação dos registros bibliográficos e catalográficos (MARC21, 2002).

Na etiqueta 006, por exemplo, ocorrem situações em que a mesma posição armazena o mesmo código, porém, com significados diferentes. A posição 16 para material textual com código [c] significa histórias em quadrinhos, tratando-se de ficção, por outro lado, na mesma posição 16 para periódicos, com o código também [c], pode significar Cirílico tratando-se de alfabeto original ou escrita do título. Ainda na mesma posição 16, o código [c] pode significar também arte ou reprodução, tratando-se de tipo de material visual (FERREIRA, 2002, p. 7, 19 e 21), (MARC21-006, 2002).

Outro ponto relevante é a possibilidade de colocar ou não uma seqüência de códigos com um significado único. Nas posições 16-17 que tratam de características especiais do formato para mapas, isso ocorre. O que por outro lado poderia gerar um problema conforme descreve-se a seguir (FERREIRA, 2002, p. 12), (MARC21-006, 2002).

As posições 01-02 para música que tratam de forma de composição, em conjunto, veja bem, em conjunto podem ter os códigos [an], o código [a] na posição 01 e o código [n] na posição 02, que significa canção patriótica. Entretanto, na mesma posição 01 para periódico que agora trata de frequência, o mesmo código [a], observe que é apenas o código [a], significa anual. Já na posição 02 ainda para periódico, o código [n], apenas o código [n], significa irregular normalizado quando trata de regularidade (FERREIRA, 2002, p. 12 e 16), (MARC21-006, 2002). Rapidamente percebe-se o erro que isso pode ocasionar se o analista de sistema ou o bibliotecário não reparar esses pormenores.

Evidentemente, estes são apenas exemplos para ilustrar o assunto, pois fatos como estes ocorrem em diversos lugares do formato na etiqueta 006 o que poderia gerar a perda de um documento levando em conta que em uma instituição os mesmos dados referem-se a um periódico e em outra, a uma música como mostra o último exemplo.

Exemplos de particularidades da etiqueta 006:

Posição	Tipo de Documento	Código / Valor	Tipo de Significado
Posição 16	Material Textual	c – Histórias em quadrinhos	Ficção
Posição 16	Periódico	c – Cirílico	Alfabeto original ou Escrita do título
Posição 16	Material Visual	c – Arte (reprodução)	Tipo de Material Visual
Posição 01-02	Música	an – Canção patriótica	Forma de composição
Posição 01	Periódico	a – Anual	Trata de frequência
Posição 02	Periódico	n – Irregular	Regularidade

Tabela 2. Particularidades da etiqueta 006.

Na etiqueta 007 observou-se também que além de uma mesma posição armazenar o mesmo código, porém com significados diferentes, podem ocorrer casos em que o mesmo significado é tratado em posições diferentes. Pode-se perceber que tanto a posição 04 para micro-formas como a posição 07 para material projetável trata de dimensões (FERREIRA, 2002, p. 32 e 33), (MARC21-007, 2002).

O contrário pode ser verdadeiro, ou seja, a mesma posição armazena o mesmo código, mas com significados diferentes. Um exemplo é a posição 04 para arquivos de computador com código [a] que significa 3 1/2 pol. tratando de dimensões, por outro lado, a mesma posição 04 para globo com o código também [a], significa papel tratando-se de suporte físico (FERREIRA, 2002, p. 26 e 28), (MARC21-007, 2002).

Ainda, a posição 04 pode tratar de dimensões para arquivos de computador e ter o código [a] significando 3 1/2 pol., porém, pode ocorrer também que a mesma posição 04 para micro-formas ainda tratando de dimensões e com o mesmo código [a], signifique outro

valor - 8mm. (FERREIRA, 2002, p. 26 e 33), (MARC21-007, 2002). Tem-se, portanto, a mesma posição, com o mesmo código, mas com valores diferentes, sendo o tipo de documento que esta sendo catalogado que determina o significado.

Outra questão observada foi a posição 09 para Micro-formas tratar de cor, e a posição 03 para arquivos de computador também tratar de cor (FERREIRA, 2002, p. 25 e 34), (MARC21-007, 2002).

Exemplos de particularidades da etiqueta 007:

Posição	Tipo de Documento	Código / Valor	Tipo de Significado
Posição 04	Micro-formas		Dimensões
Posição 07	Material Projetável		Dimensões
Posição 04	Globo	a – Papel	Meio Físico
Posição 04	Micro-formas	a – 8mm	Dimensões
Posição 04	Arquivos de computador	a – 3 ½ pol	Dimensões

Tabela 3. Particularidades da etiqueta 007.

Mais uma vez é necessário deixar claro que estes são apenas exemplos para ilustrar o assunto, pois fatos como estes ocorrem em diversos lugares do formato, principalmente nas etiquetas 006, 007, 008 e no próprio Líder. Outro ponto desconsiderado aqui é o fato de, em alguns casos, uma etiqueta estar ligada com outra, ou seja, se a etiqueta X está sendo usada para descrever o item documentário, isso obrigatoriamente faz com que a etiqueta Y também apareça na importação dos dados. O estudo apresentado até aqui mostra a importância da criação das tabelas para armazenar o significado das etiquetas, posições, indicadores e subcampos sugeridos anteriormente na figura 5.

Como último exemplo, destaca-se uma análise efetuada no software Aleph 500, quanto a sua qualidade e determinantes de desempenho, salientando que este produto é considerado

um dos melhores sistemas gerenciadores de biblioteca (GUSMÃO, 2001). Apesar das versões mais atuais do software ter resolvido as questões apresentadas na pesquisa de Gusmão (2001, p. 92), a quantidade de problemas encontrados no sistema sob o ponto de vista de profissionais bibliotecários, foi praticamente a mesma para os módulos de catalogação, de circulação e nos relatórios, segundo o resultado do trabalho do autor. Sobre a avaliação do sistema, destaca-se aqui o módulo de catalogação em seus itens 1, 2 e 3:

ITEM AVALIADO: Subsistema de Catalogação

>>> 1- Não está com o USMARC atualizado.

>>> 2- Não faz consistência de sub-campos ou campos com tags de dados fixos (ex. 008 e LDR).

>>> 3- Não faz consistência/validação dos valores dos campos em geral.

4- As Autoridades não seguem um padrão de funcionamento on line na atualização dos registros bibliográficos.

5- Incompatibilidade da versão atual do USMARC com a versão 11.5 do ALEPH 500.

6- Não há ordenação dos sub-campos e dos campos no salvamento do registro. Gerando problemas na criação de ACCs.

7- Falta segurança, é preciso reiniciar a máquina para o módulo “esquecer a senha anterior”.

8- Nível de Autorização – impossibilidade de definir níveis de autorização distintos, para alteração por campos.

9- Problemas com diacríticos.

Assim, espera-se ter oferecido uma contribuição para compreensão dos aspectos relacionados com o desenvolvimento de um programa de computador que seja capaz de interpretar e entender registros bibliográficos em formato MARC 21. No capítulo 6, ao descrever as características do formato MARC 21 em XML será percebido o avanço que a LC propiciou com o desenvolvimento dessa versão, tanto do ponto de vista de facilidade no desenvolvimento, como também nas novas oportunidades de desenvolvimento que podem surgir.

4.3 A catalogação na prática usando MARC 21

A importação e exportação de dados bibliográficos e catalográficos no processo de catalogação têm importância determinante pelo menos em dois aspectos: qualidade e

produtividade. Os três sistemas de alimentação de catálogos e bases de dados bibliográficos e catalográficos – a catalogação na publicação, a catalogação centralizada e a catalogação cooperativa – podem beneficiar-se da importação de dados bibliográficos e catalográficos do ponto de vista da qualidade, sendo a catalogação cooperativa um bom programa para ilustrar esse fato. Nesse sistema o bibliotecário catalogador aumenta a qualidade da catalogação a medida que passa a ter acesso aos dados bibliográficos que provavelmente já foram analisados por diversos outros especialistas que conseqüentemente auxiliaram na correção de possíveis incoerências.

A produtividade é indiscutível a partir do momento em que compara-se o tempo gasto com o processo de catalogação de um documento sem a utilização da importação de dados. Na importação de dados, como mais detalhadamente será apresentada, o bibliotecário catalogador divide seu trabalho basicamente em dois estágios: uma busca automatizada pelo registro desejado em uma base de dados e a conferência deste registro após a importação, que em muitos sistemas ocorre a partir de um simples clique no mouse do computador.

Nesse ponto apresenta-se brevemente aos bibliotecários que não tem contato com o formato MARC 21, e especialmente aos analistas de sistemas interessados em desenvolver sistemas para bibliotecas, como é o trabalho de um catalogador que tem a sua disposição um sistema gerenciador de bibliotecas, particularmente no módulo de catalogação, a opção de importar dados bibliográficos a partir de um registro MARC 21.

Antes, a qualidade na catalogação deve ser destacada, quando feita a partir de um registro MARC 21. Essa análise pode ser realizada em dois aspectos, e segundo Dias (1999, p. 3) a *“prática da catalogação não pode mais estar dissociada do formato bibliográfico USMARC, sob pena de formarmos catálogos particulares, isolados, sem possibilidade de intercâmbio automatizado com outras instituições.”*

O primeiro aspecto a ser considerado é referente ao trabalho do próprio catalogador, que pode ser resumido a capacidade do bibliotecário em usar de maneira correta uma das suas principais ferramentas de trabalho, o AACR2. Como já visto, Dias (1999, p. 2) afirma que muitas vezes o catalogador precisará “*tomar decisões corretas diante das inúmeras dúvidas, surgidas no processo de catalogação*” e geralmente as dúvidas são referentes “*à descrição do registro bibliográfico à luz do AACR2*” (DIAS, 1999, p. 34). Portanto, a qualidade da catalogação nesse ponto de vista, está ligada ao interesse do profissional pelo seu trabalho, que pode ser estimulado pela instituição a partir da motivação a educação continuada.

O segundo aspecto pode ser mais fácil de controlar, comparado ao primeiro que pode depender de variáveis tão indeterminadas, por exemplo, as personalidades humanas, no caso, o sujeito institucional, o bibliotecário. Estamos falando aqui das ciências exatas, e a qualidade na catalogação vista agora deste ponto de vista, se resume em saber se o módulo de catalogação do sistema gerenciador de bibliotecas usado tem a capacidade de fazer uma validação dos dados contidos em um registro MARC 21, ou seja, verificar entre outras coisas, se as etiquetas, posições, indicadores e subcampos apresentados existem no formato e estão sendo usadas corretamente.

Apesar de parecer óbvia a necessidade deste requisito no módulo de catalogação, muitos sistemas no mercado não oferecem-no. Uma parte do resultado do trabalho de Gusmão (2001, p. 92), quanto a relação de problemas no módulo de catalogação, demonstra que o sistema gerenciador de bibliotecas Aleph 500, considerado um dos melhores do mercado, “*não faz consistência/validação dos valores dos campos em geral*”, assim pode-se dizer que a qualidade dos serviços na catalogação aumentará em sistemas desenvolvidos com estruturas que disponibilizem a validação.

Em seu trabalho, disponível na Internet, Balby (2003) apresenta alguns softwares, que segundo a autora, fazem validações. Entretanto um simples teste no módulo de catalogação do software desejado será mais adequado. Por exemplo, após o software apresentar a tela

com o resultado da importação de um registro MARC 21, altere algumas etiquetas com indicadores, subcampos e informações incorretas e solicite a gravação do registro. Se o registro for gravado, o sistema não faz validação alguma.

Cabe ressaltar que a validação aqui apresentada refere-se a estrutura do registro MARC 21, entretanto, não se pode garantir que a digitação do conteúdo foi correta, por exemplo, a etiqueta 245 refere-se ao título da obra, mas o sistema não tem como verificar se o título foi corretamente digitado.

A validação estará, por exemplo, verificando no registro MARC 21 se o caractere “s” apresentado na posição 7 da etiqueta 008 realmente existe, ou, verificará se a etiqueta 100 realmente pode ter o caractere “7” como indicador 1, o caractere “3” como indicador 2 e ainda, se o subcampo “w” existe nesta etiqueta. Ainda segundo catalogadores, quando utiliza-se uma determinada etiqueta, em alguns casos, obrigatoriamente utiliza-se uma outra, por exemplo, quando a etiqueta 490 é usada, informação de série, sua complementação está nas informações da etiqueta 830, entrada secundária de série.

Depois deste breve comentário sobre qualidade, apresenta-se uma descrição da seqüência de trabalho de um catalogador para visualizar a utilidade, importância e abrangência do formato de intercâmbio de dados MARC 21 e alguns aspectos que o módulo de catalogação de um sistema gerenciador de bibliotecas deve conter. A seqüência apresentada aqui está baseada na oportunidade de ter presenciado os trabalhos executados por catalogadores.

Ao catalogar um documento, o primeiro passo é fazer uma pesquisa nas bases bibliográficas disponíveis e localizar o item, geralmente as bases são da Library of Congress, LC, da Online Computer Library Center, OCLC ou da rede Bibliodata. Quando localizado, o catalogador importa o registro, verifica a qualidade dos dados ao mesmo tempo em que vai corrigindo e adaptando-o para a sua realidade, salvando assim um novo registro. Para importar o registro o catalogador deve ter antes identificado o tipo do documento, por exemplo, livro, periódico ou mapa, e escolhido a planilha adequada no seu

sistema para trabalhar com o registro MARC 21. Detalhes sobre essa planilha serão apresentados a seguir.

Se o registro não for encontrado em nenhuma base disponível, o catalogador faz então a representação descritiva escolhendo a planilha adequada conforme o tipo do documento. Visto a disponibilidade de pelo menos três programas de catalogação – a catalogação na publicação, catalogação cooperativa e catalogação centralizada – a seqüência apresentada aqui foi resumida com o objetivo de respeitar as particularidades da cada programa, entretanto, o programa de catalogação cooperativa, parece ser o que melhor se adaptaria a rotina de importação de dados apresentada.

A rotina de importação pode ser verificada a seguir. Cabe destacar que está se referindo ao processo de catalogação em um ambiente que disponibilize uma estrutura computacional, e por sua vez, torna possível a importação de dados bibliográficos usando o formato MARC 21 na seguinte rotina:

- Realiza uma busca em bases de dados

[ENCONTROU]

- Bibliotecário Catalogador faz a importação;
- Bibliotecário Catalogador seleciona uma planilha compatível com o tipo de documento;
 - Essa seleção pode ser feita automaticamente dependendo do sistema usado;
- O sistema verifica a estrutura dos dados;
- Bibliotecário Catalogador verifica a qualidade dos dados;
- Bibliotecário Catalogador digita as informações institucionais;
- Bibliotecário Catalogador grava o novo registro na base.

[NÃO ENCONTROU]

- Bibliotecário Catalogador seleciona uma planilha compatível com o tipo de documento;
- Bibliotecário Catalogador efetua a digitação dos dados com qualidade;
- Grava o novo registro na base.

A relação entre o catalogador, o sistema e o MARC 21 acontece a partir da utilização das planilhas. Estas mesmas são o conjunto de etiquetas selecionadas pelo catalogador para representar um determinado tipo de documento. Respeitando suas particularidades, a maioria dos módulos de catalogação dos sistemas gerenciadores de bibliotecas apresenta uma interface para que os chamados analistas bibliotecários montem suas planilhas. Essas planilhas são criadas uma única vez, uma para cada tipo de documento, e depois modificadas de acordo com a necessidade.

Visto que o formato MARC 21 é muito abrangente, e às vezes encarado como complexo com conotação negativa, a criação destas planilhas tem como objetivo escolher quais etiquetas e seus respectivos indicadores, subcampos ou posições serão usados para representar um determinado documento, neste mesmo momento, podem também ser criadas as regras de validação se o sistema assim permitir. Sendo assim, quando o catalogador importar um registro bibliográfico, ele identifica o tipo de documento e escolhe a planilha adequada para trabalhar o registro. É possível que essa identificação também seja automática.

Sugere-se um resumo que tem o objetivo de identificar as principais características do módulo de catalogação de um bom sistema gerenciador de bibliotecas, ao mesmo tempo, que serve como uma breve lista de requisitos para um analista de sistemas desenvolver um bom software de importação de dados MARC 21:

- Importa e exporta registros bibliográficos no formato tradicional do MARC 21;
- Importa e exporta registros bibliográficos no formato XML do MARC 21;
- Faz validação da estrutura do registro MARC 21;
- Faz validação das possíveis relações existentes entre etiquetas;
- Possibilita a criação de planilhas para cada tipo de documento;
- Cria regras de validação diversificada para as etiquetas;

- Escolha automática pelo sistema da planilha apropriada no ato da importação;
- Interface que ajuda a entender as etiquetas, posições, indicadores e sub-campos;
- Interface que ajuda na catalogação do registro bibliográfico à luz do AACR2.

Importar e exportar registros bibliográficos no formato XML do MARC 21 é um dos requisitos de um bom software. Este assunto na verdade é ponto chave da pesquisa e estará sendo detalhado no capítulo 6.

A maior crítica que encontramos na literatura sobre o MARC 21, é a afirmação de que o formato seria muito complexo. O que ocorre é que a estrutura do formato apresenta uma possibilidade grande de detalhamento para uma descrição bibliográfica e catalográfica, ou seja, o formato foi desenvolvido para representar qualquer tipo de documento, seja ele um livro, mapa, periódico, globo, CD-ROM, arquivos de computador, fitas de áudio ou vídeo, recursos disponíveis na Web, etc, do modo mais detalhado possível.

Na prática tem-se um número relevante de etiquetas, indicadores, subcampo e posições a disposição para criar formas de representação, para um registro bibliográfico e catalográfico, de um item documentário independentemente de seu formato. Deste modo, sua complexidade está na sua abrangência. Uma outra questão, é que muitos confundem a utilização obrigatória ou não do AACR2 juntamente com o MARC 21. Lembramos que se tratam de dois instrumentos distintos, o AACR2 é um código de catalogação para descrição de dados bibliográficos e catalográficos e o MARC 21 é um formato de intercâmbio de dados bibliográficos e catalográficos, entretanto, visto que a principal preocupação da catalogação como forma de representação é a padronização e a qualidade dos registros, pode-se afirmar que a utilização conjunta de ambos é justificável.

A complexidade aparente que pode existir na utilização do formato MARC 21 pode ser resultado de interfaces de softwares desprovidas de usabilidade e a falta de entendimento do que venha ser a verdadeira utilidade das planilhas de captação de dados conforme já apresentado ao abordarmos a rotina de importação de um bibliotecário catalogador.

Quanto à interface, pode-se dizer que se um software for bem desenvolvido, os detalhes técnicos apresentados referentes ao MARC 21 acabam sendo resolvidos. Por exemplo, basta lembrar como era o procedimento para copiar, apagar, renomear etc, de um arquivo quando o sistema operacional comum da época era o MS-DOS, vários comandos deveriam ser decorados, entretanto, com o lançamento do MS-Windows 3.X, esses comandos foram em muitos casos substituídos por um clicar com o botão direito do mouse no arquivo e seleccionar a opção desejada. Um simples pressionar da tecla F1 já apresenta uma ajuda eficiente para os usuários.

A criação de planilhas de utilização do formato MARC 21, visa basicamente a seleccionar as etiquetas, indicadores, subcampos e posições apropriadas para uma forma de representação de um determinado documento. Essa função, muitas vezes, é realizada por um profissional especializado conhecido, em alguns casos, como analista bibliotecário. Desta forma, o fato de o formato MARC 21 apresentar um número relevante de etiquetas, indicadores, subcampos e posições, não significa que a forma de representação de um documento deva conter todas elas. Por exemplo, Viana (2002) apresenta uma lista mínima para representar informações em uma página web:

(R)=Repetitivo - (NR)=Não-Repetitivo - (OB)=Obrigatório

>>> **LDR - Líder (NR)(OB)**

>>> **008 - Informações de tamanho fixo (NR)(OB)**

040 - Código da Fonte Catalogadora na LC (NR)

041 - Código do Idioma da Publicação (R)

044 - Código do País de Publicação (NR)

100 - Autor Pessoal (NR)

110 - Entidade (NR)

111 - Evento (NR)

>>> **245 - Título (NR)(OB)**

246 - Forma Variante do Título (R)

256 - Características arquivo de computador (NR)

>>> **260 - Imprensa (\$aLocal, \$bEditora e \$cData) (R)(OB)**

4xx - Série (400,410,411,440) (R)
505 - Nota de Conteúdo (R)
506 - Nota de Restrição de Acesso (R)
508 - Nota de Créditos (pela criação do site) (R)
516 - Tipo de Arquivo de Computador ou Nota de Arquivo de Computador(R)
520 - Nota de Resumo (R)
>>> **650 - Assunto (R)(OB)**
700 - Entrada Secundária pelo Nome Pessoal (R)
710 - Entrada Secundária pelo Nome de Entidade (R)
711 - Entrada Secundária pelo Nome de Evento (R)
>>> **856 - Acesso Eletrônico - LINK (R) (já que é uma página)**

Podemos notar que segundo Viana (2002), são necessários 23 etiquetas, sendo efetivamente obrigatórias, apenas 5, visto que o Líder é obrigatório em qualquer descrição e é criado automaticamente pelo computador. De acordo com Prado (2002), o necessário são apenas 8 etiquetas, contudo, o que se quer destacar é que o formato pode ser considerado complexo, mas não com uma conotação negativa, visto que também no caso da criação das planilhas, uma boa interface de computador dotada de um eficiente mecanismo de ajuda pode auxiliar até mesmo um leigo a selecionar as etiquetas apropriadas.

Diante destes detalhes, o problema está em implementar bons produtos e interfaces que possam colocar em prática toda a flexibilidade, funcionalidade e particularidades do formato MARC 21. Tal necessidade encaminha para uma análise mais detalhada da XML, como linguagem apropriada para o desenvolvimento de representações em MARC 21.

5 Extensible Markup Language – XML

A melhor forma de entender o que pode ser feito com a Extensible Markup Language, XML, é fazer uma comparação com sua antecessora Hyper Text Markup Language, HTML. Atualmente, a maior parte dos sites disponíveis na Internet foram desenvolvidos usando a HTML. Esta linguagem é uma metodologia não proprietária para criação de páginas web. HTML define disposição da página, tipos de letra, elementos gráficos e ligações ou links de hipertexto para outros documentos web por meio de códigos ou tags dentro do texto.

A principal diferença entre HTML e XML, é que esta última, permite que sejam criados códigos ou tags de acordo com a necessidade do usuário (DÉCIO, 2000). Um exemplo de um código em HTML é apresentado, e nele destaca-se a tag <p> que desloca um determinado conteúdo informacional em um novo parágrafo. A tag <p> é fixa, ou seja, não pode ser modificada por um usuário comum, mas apenas pelo comitê responsável pela HTML, o W3C (W3C, 2002):

```
<html>
  <body>
    <p>A XML é uma linguagem desenvolvida em 1996, pelo comitê W3C</p>
  </body>
</html>
```

A mesma informação, agora estruturada segundo os padrões da XML, também definida pelo W3C (W3C, 2002) que contém o mesmo conteúdo informacional:

```
<!xml>
  <texto>
    <linguagem>XML</linguagem>
    <autor>W3C</autor>
    <data_desenvolvimento>1996</data_desenvolvimento>
  </texto>
```

Como podemos notar, a estrutura anterior, composta de tags criadas a partir da necessidade de um usuário comum, torna o documento mais “inteligente” por tornar possível identificar o significado das informações. O autor pode ler o documento assim: Qual a linguagem ? XML; Qual o autor ? W3C; Qual a data de desenvolvimento ? 1996. O destacável é que do ponto de vista de um computador, este, desde que equipado com uma linguagem, também poderá ser capaz de interpretar as informações de forma mais eficiente (BOSAK; BRAY, 2002), estes autores tiveram papéis cruciais no desenvolvimento da XML.

Segundo Furgeri (2001), alguns aspectos são necessários para que a XML se torne uma realidade. O autor afirma que cada um segmento da indústria ou grupo com interesses em comum, deve montar sua estrutura padrão, ou seja, criar a sua própria linguagem de marcação a partir da XML, o que provavelmente, tenha sido feito pela LC na adaptação do MARC 21 para XML, criando assim um padrão na indústria bibliográfica, até porque, se cada usuário individualmente começar a criar o seu próprio padrão, na prática não teremos padrão.

O autor afirma ainda que devido à característica da XML em permitir a definição de tags de acordo com as necessidades do usuário, as ferramentas de pesquisas terão que se adaptar a essa nova realidade. As interfaces de pesquisas terão que exibir as opções de um modo útil, de forma que as possibilidades de pesquisa estejam disponíveis aos usuários.

Atualmente no mercado já estão disponíveis diversos padrões para marcação de documentos que podem ser criados a partir da estrutura que a XML oferece. Um exemplo é o MARC XML. Outros como a MML - Mathematical Markup Language utilizada para descrição de fórmulas matemáticas, CDF - Channel Definition Format, usada na área de tecnologia Pash e HCML - Health Care Markup Language, linguagem de marcação usada na área médica para planos de saúde (FURGERI, 2001).

HTML diz como o documento será apresentado, a XML dá um passo a frente neste sentido, ou seja, além de fazer o mesmo, pode por assim dizer, especificar o significado das

informações que um documento contém. No caso da LC, o arquivo em XML não foi feito apenas para a apresentação do documento, mas para dizer o significado dos dados. Em outras palavras, um programa de computador poderia usar o XML da LC para converter dados de um banco de dados MS-Access 97 para outro, por exemplo, o Oracle. (LIBERTY; KRALEY, 2001), conforme abordagem do capítulo 6.

Pode-se afirmar que a XML é uma linguagem para criar padrões de comunicação entre sistemas de computadores, o que permitirá a integração tanto da base de dados como de arquiteturas, hardwares e métodos de programação usados, favorecendo a interoperabilidade. Entretanto, a XML sozinha não é nada, ou seja, um arquivo de computador com informações estruturadas segundo a metodologia XML só terá seu valor prático se outras tecnologias estiverem sendo usadas em conjunto, como por exemplo, as três siglas que destacam o potencial desta linguagem: DOM, DTD e XSL.

5.1 Document Object Model – DOM

Uma das principais características da XML é apresentar as informações de maneira organizada, hierarquizada, permitindo assim uma manipulação mais eficiente dos dados comparados a outros métodos semelhantes aos objetivos da linguagem. O DOM é o responsável pela manipulação, ou seja, o Document Object Model disponibiliza os métodos para a manipulação de um arquivo XML e conseqüentemente das informações.

Segundo Kopelevitch (2002a), o DOM *“nos permite processar, criar e editar documentos XML, utilizando as principais linguagens do mercado (PHP, Perl, ASP, VisualBasic, etc)”*. Para que possa ficar claro para o leitor, linguagens aqui, refere-se à ferramenta usada pelos programadores de computador para desenvolver sistemas de computadores, tais como gerenciadores de bibliotecas. Em alguns casos essas linguagens necessitam de um software externo que a permita executar uma tarefa específica, por exemplo, acesso aos principais gerenciadores de banco de dados do mercado, tais como Oracle, MS SQL-Server, IBM DB2, etc. O DOM é o software externo que permite a manipulação de um arquivo XML.

Dentre os software que fazem o papel de DOM disponíveis no mercado, temos o Microsoft XML Parser, que foi desenvolvido para carregar, criar e manipular documentos XML, entre outras funções, utilizando as linguagens já citadas. Kopelevitch (2002a) afirma que atualmente existem três versões de XMLDOM: MSXML 2.0, MSXML 2.6, MSXML 3.0 e MSXML 4.0, este último em fase de testes. Resumindo, com o Document Object Model, DOM, se pode manipular qualquer tipo de informação que esteja na árvore do documento XML usando qualquer das principais linguagens de programação disponível atualmente.

5.2 Document Type Definition – DTD

Relembrando que um dos objetivos do processo de catalogação que dispõe da utilização do MARC 21 é melhorar e até mesmo garantir a qualidade dos registros bibliográficos, a Document Type Definition, DTD, assume uma grande importância neste ponto do trabalho. Kopelevitch (2002b) afirma que esta seria “*uma forma de validar o documento XML*”, ou seja, um arquivo DTD contém tags ou códigos que define como deve ser construído um arquivo XML.

Validar, aqui, pode ser entendido como verificar se a informação contida dentro de um arquivo XML esta correta e aceitável, visto do ponto de vista estrutural, ou melhor, um arquivo DTD é usado para especificar quais elementos, tags ou atributos são permitidos no arquivo XML, até mesmo verificando o local em que podem surgir. Do ponto de vista da qualidade dos dados, quando usado um arquivo DTD junto com um arquivo XML, existirá um validador natural, por assim dizer.

A denominação de validador natural apresenta-se adequada, pois, montar um processo de validação de arquivos XML usando o DTD é muito simples, por exemplo, o próprio software usado para navegar pela Internet pode ser capaz de validar um arquivo XML. O Microsoft Internet Explorer, por exemplo, pode fazer isso após uma simples atualização no site da empresa. E esse tipo de validação em outras estruturas parecem ser muito mais complexas, uma prova desta afirmação, é lembrar que a maioria dos softwares que fazem

importação de dados usando o formato MARC 21 tradicional, não oferecem esse mecanismo como visto no resultado do trabalho de Gusmão (2001).

Outra questão importante é que qualquer programador de computador “*poderia montar um documento XML apenas olhando para um DTD*” (KOPELEVITCH, 2002b), portanto, o DTD pode ser uma regra para criação, até mesmo automática, de um arquivo XML. As tags ou códigos do arquivo DTD podem ficar dentro ou fora do arquivo XML, mas geralmente, é usado um arquivo DTD externo para economizar espaço no arquivo XML (KOPELEVITCH, 2002b).

Cada elemento, ou tag, de um documento XML é definido da seguinte forma: `<! ELEMENT nome_tag (conteudo_tag)>`. O `conteudo_tag`, pode ser definido de maneiras diferentes, dependendo da situação, entretanto, basicamente o `conteudo_tag` mostra o tipo de informação que o `nome_tag` pode apresentar, ou se esta, tem outras tags vinculadas (SILVA, 2001). O exemplo a seguir mostra o código necessário para definir o tipo de informação que o `nome_tag` pode apresentar: `<! ELEMENT leader (#PCDATA)>`. A tag `leader` foi definida para apresentar um conteúdo do tipo texto. O documento XML correspondente seria: `<leader>01329cam 2200313 a 4500</leader>`.

Quando a definição tem o objetivo de mostrar tags vinculadas, ou seja, `nome_tag` com outra tag vinculada, usa-se escrever um código da seguinte forma:

```
<! ELEMENT datafield (subfield)>
<! ELEMENT subfield (#PCDATA)>
```

O `datafield` tem uma tag vinculada chamada `subfield` que pode apresentar um conteúdo do tipo texto. O documento XML correspondente:

```
<datafield tag="010" ind1="" ind2="">
  <subfield code="a">89048230</subfield>
</datafield>
```

Uma tag XML pode ter atributos, como nas tags datafield e subfield, ou seja, as palavras tag, ind1, ind2 e code são denominadas atributos de uma tag XML. Para definir um atributo para uma tag XML, escreve-se o seguinte código: <!ATTLIST nome_tag nome_atributo tipo_atributo>. Apresentam-se a seguir exemplos baseados na tag <datafield tag="010" ind1="" ind2="">:

```
<!ATTLIST datafield tag #REQUIRED>
```

```
<!ATTLIST datafield ind1>
```

```
<!ATTLIST datafield ind1>
```

No exemplo, a palavra #REQUIRED, informa que o preenchimento do atributo tag é obrigatório. Nota-se que a definição de um DTD para um documento XML pode ser muito simples, entretanto, existem diversos detalhes que podem ser usados tanto na definição de tags como de atributos de um DTD, podendo existir definições mais complexas, e qualquer situação pode ser prevista ao se definir um DTD.

Para realizar a associação entre um documento XML e um DTD basta fazer a referência ao arquivo DTD antes do primeiro elemento do documento que no exemplo a seguir é a tag <EXEMPLO>:

```
<? xml version = "1.0"?>
```

```
<!DOCTYPE EXEMPLO SYSTEM meuDTD.dtd>
```

```
<EXEMPLO>
```

```
  <SAUDACAO>Seja bem vindo</SAUDACAO>
```

```
</EXEMPLO>
```

Na validação do arquivo XML, as informações necessárias para o processo estarão no arquivo meuDTD.dtd. Como apresentado, pode-se colocar no próprio corpo do documento XML a definição do seu DTD, entretanto, isso não é apropriado dado aos objetivos de um arquivo DTD.

Além da DTD, tem-se a disposição as XDR, XML-Data Reduced, esquemas que também são usadas para validar a estrutura e conteúdo de um arquivo XML. A XDR é muito similar ao DTD, em termos de função, pois a sintaxe é completamente diferente, porém, segundo Kopelevitch (2002c) “*muito mais eficiente*”. Com a XML-Data Reduced é possível definir os tipo de tag, e assim também como no caso da DTD, as tags podem ficar tanto no documento XML como fora dele (KOPELEVITCH, 2002c). Assim, desenvolver um processo de validação de dados em arquivos XML é simples do ponto de vista técnico e estrutura dos dados.

É necessário destacar que a validação feita a partir de um arquivo DTD valida estrutura e não conteúdo, ou seja, o que se pode validar é se um determinado registro contém dados no formato que se esperava, e não se o dado é o que se esperava. Por exemplo, pode-se garantir que o título foi digitado, mas não se realmente era o título, ou ainda, pode-se verificar se o dado informado no campo data de publicação é realmente uma data, mas não se garante que seja a data correta.

5.3 Extensible Stylesheet Language – XSL

De acordo com Kopelevitch (2002d) a XSL “*é a folha de estilos do XML. O XSL serve para transformar os dados do XML em HTML, para serem visualizados corretamente no navegador*”. Além disso, o XSL pode ser usado também para fazer buscas específicas num arquivo XML, criar tags, criar estruturas condicionais e tudo mais do ponto de vista de formas de representação da informação contida em um arquivo XML.

A partir de um arquivo XML podem derivar várias formas de representação para as informações contidas neste arquivo, de acordo com a necessidade do usuário. Por isso que afirma-se, que do ponto de vista computacional, e no contexto da Ciência da Informação, os estudos e assuntos levantados no capítulo 2 sobre formas de representação são bem representados na interação entre os arquivos XML e XSL.

Na verdade, visto que em um arquivo XML existem apenas os dados organizados de maneira hierárquica e agrupados de forma a tentar dar significado para estes, trabalhar com diversas formas de representação para estes dados é simples, e existem várias possibilidades, por exemplo, com o uso de uma linguagem de programação e o DOM, como visto ainda a pouco, isso é possível. Entretanto, a XSL parece ter sido construída exatamente para este fim, por isso destaca-se nesta pesquisa (XSL, 2002).

A XSL é dividida em duas partes, uma para transformar o documento XML em outro tipo de documento, a outra, para definir objetos de formatação para apresentação gráfica (ESTEVEES; SANTOS; GUIMARÃES, 2002a):

- XSLT – linguagem de transformação
- XSLFO – linguagem de formatação de objetos.

O XSL está implementado nos browsers, software usado para navegar pela web, através da sua linguagem de transformação XSLT. O XSLT é uma linguagem baseada na própria XML, portanto, neste arquivo estão definidas transformações que podem ocorrer nos elementos contidos em um arquivo XML, e seu funcionamento se dá por percorrer o documento, verificando para cada nível se existe uma transformação associada. Caso exista ela é aplicada. A XSLFO é responsável por definir os objetos de formatação para sua apresentação gráfica. (ESTEVEES; SANTOS; GUIMARÃES, 2002a).

5.4 Aplicações da XML

A apresentação dos últimos assuntos teve como objetivo discorrer brevemente sobre os objetivos da XML e conseqüentemente o potencial desta linguagem de marcação. Portanto, apesar de recente, a XML tem atualmente uma importante participação na área das tecnologias de informação, com uma presença marcante, na maioria das vezes, nos sistemas baseados em tecnologias web que tenham sido desenvolvidos recentemente. Além disso, apresentam-se as possíveis aplicações da XML na biblioteca, desta forma, destacando sua importância para a Ciência da Informação.

Dentre os vários exemplos de implementação de soluções XML que se pode encontrar na web estão especialmente os sites que realizem operações bancárias, leilões on-line ou pesquisa indexada de documentos, entretanto, o comércio eletrônico é um o exemplo mais marcante. As principais soluções de e-commerce desenvolvidas, optaram por usar o potencial da XML. O principal objetivo deste tipo de aplicação é prover uma integração entre os diversos sistemas e tecnologias freqüentemente disponíveis neste ambiente, por isso, usam a XML como linguagem de comunicação, solucionando este problema (ESTEVEES; SANTOS; GUIMARÃES, 2002a).

Em grandes corporações esta linguagem de marcação também pode ser muito útil, por exemplo, para auxiliar na realização de tarefas que se repetem freqüentemente, como a geração de relatórios estatísticos. Com a utilização da XML, o relatório em si, só é produzido uma vez, a partir deste ponto, sua visualização pode ser possível em qualquer meio que for desejado, por exemplo, na tela de um computador pessoal, telefone celular, computadores de bolso etc. Exemplificando, Esteves, Santos e Guimarães (2002b) afirmam:

Ao separar os dados (conteúdo) da apresentação (aparência), o XML possibilita a criação de templates separados para vários suportes, desde a televisão interactiva aos telemóveis. Permite ainda alterar rapidamente a aparência de todas as páginas de um site, bastando para isso alterar o template respectivo.

Nos portais, a participação da XML traz igualmente grandes vantagens para a equipe de programadores e analistas de sistemas, pois em um arquivo XML pode conter os conteúdos do portal, bastando criar um processo de personalização de uma página em função do perfil do usuário, através de folhas de estilo, XSL, por exemplo, ou por aplicação de filtros na pesquisa na base de dados. Como mostra a idéia de Bosak e Bray (2002), a intenção é escrever uma única vez e publicar quando, como, onde e quantas vezes quiserem. Mas é no contexto das bibliotecas digitais, que a utilização desta linguagem é com certeza, um passo importante neste novo paradigma da estruturação e organização da informação que gera conhecimento, e está é o principal interesse desta pesquisa.

Atualmente o ambiente de desenvolvimento da Biblioteconomia e da Ciência da Informação está cada vez mais automatizado, ao mesmo tempo, ocorre uma grande produção documental nunca antes presenciada, e em diversos formatos digitais, tais como, MS-Word, Wordpad, textos ASCII, HTML, PDF etc, e o tradicional impresso em papel. Este contexto resulta em grandes dificuldades de armazenamento, organização e disseminação da informação. Com o surgimento de redes locais e da própria Internet, a circulação de documentos tornou-se uma operação trivial, por isso, foi necessário estudar novas formas de trabalhar com a informação documental (ESTEVES; SANTOS; GUIMARÃES, 2002c).

Esteves, Santos e Guimarães (2002c) afirmam que:

na área das bibliotecas, surgem os formatos MARC (Machine Readable Catalogue) para responder às necessidades de informatização de catálogos. No entanto, estes formatos não conseguem responder no contexto da Internet, pois não possuem uma linguagem de fácil aplicação por qualquer utilizador e que, ao mesmo tempo, possa ser interpretada pelos browsers. O XML aparece, no contexto da internet, para facilitar a difusão da informação documental. Como possui uma semântica própria, descreve a estrutura e conteúdo do documento, não a sua formatação, tornando-se, por isso, "revolucionário" em relação ao HTML que apenas possibilita a formatação dos dados no que respeita à sua apresentação gráfica, não fornecendo nenhum conteúdo semântico.

Cabe ressaltar que os mesmos se referem ao formato tradicional do MARC, ou seja, aquele basicamente composto de uma seqüência de caracteres, muitas vezes chamada de “uma tira MARC 21” armazenada em um arquivo texto, apresentado no capítulo 4. A solução apresentada, XML, é sem dúvida nenhuma muito apropriada, e como será apresentado com mais detalhes no próximo capítulo, a LC desenvolveu uma versão do MARC 21 em XML, assim, o assunto mostra o potencial desta solução, ponto foco da pesquisa.

A XML, por si só, não é a solução, mas sim a ferramenta que irá gerar a solução e essa afirmação está baseada nas características técnicas da própria linguagem, como já apresentado. Por isso, a solução é desenvolver um padrão para apresentar os dados bibliográficos e catalográficos em XML, e foi exatamente o que a LC fez na versão XML

do MARC 21. Antes, porém, alguns detalhes técnicos que mostram as principais diferenças do HTML e XML dentro do contexto ou necessidades da Biblioteconomia e Ciência da Informação são apresentados.

A formatação de documentos em HTML é limitada, dentro das necessidades da Biblioteconomia, devido às marcas desta, serem fixas. A transformação de documentos em formato proprietário para HTML necessita de uma conversão, o que frequentemente resulta na perda de algumas características do texto. O usuário também fica limitado pelo conjunto de tipos de formatação disponível, por exemplo, é impossível apresentar fórmulas matemáticas ou químicas com símbolos em HTML, sem inserir uma imagem externa ao documento. Além disso, com o HTML não é possível manter uma base de índices de documentos para depois realizar pesquisas temáticas (ESTEVEVES; SANTOS; GUIMARÃES, 2002c).

O surgimento do XML tem como objetivo trazer solução a estes problemas, e se apresenta como um possível padrão para a Ciência da Informação, para isso, será necessário que os browsers façam as adaptações necessárias para implementar todas as funcionalidades que a XML oferece. Isso não deve demorar muito, visto que, a própria comunidade da Ciência da Computação já adotou a XML como padrão de comunicação e percebeu a necessidade da interoperabilidade entre sistemas.

A Extensible Markup Language, XML, tem como principal diferencial à possibilidade de criar suas próprias marcações, uma linguagem de criação de marcas ou tags que pode ser extensível de acordo com a necessidade de um usuário. O usuário aqui deve ser entendido como no mínimo um grupo de indivíduos com interesses comuns, visto que de um modo individual, a principal vantagem da linguagem, a criação de padrões, seria perdida. Deste modo a linguagem se apresenta como uma importante ferramenta para auxiliar nas possibilidades e potencialidades de qualquer área.

Na Ciência da Informação não é diferente, pois, o mais importante e efetivamente usado formato de intercâmbio de dados bibliográficos, o MARC 21, conta com sua versão em XML. Entretanto, existem outras iniciativas que beneficiam a área, como por exemplo, o Metadata Encoding & Transmission Standard, METS, que pretende ser um padrão de metadados para suprir as necessidades das bibliotecas digitais, também uma iniciativa da LC, e a XML promete ser um padrão para criação de padrões na CI.

Todas as possibilidades e potencialidades que a XML oferece é resultado da sua própria estrutura, ou seja, sendo uma linguagem de marcação que desde sua criação já se preocupou em fazer uma distinta divisão entre informação, interface e processamento, XML, XLS, DTD e DOM, respectivamente, sendo os dois últimos usados para processamento, temos uma excelente ferramenta. Essa distinta divisão é meta atual na Ciência da Computação, sendo na verdade uma necessidade diante das inúmeras opções que a área oferece, sendo determinante a existência de uma solução que disponibilize uma interação entre tecnologias.

A importância dos arquivos XML, XSL, DTD e da tecnologia DOM, e a natural interação existente entre estes, acabam sendo importantes também para os estudos de formas de representação apresentados inicialmente. A relação entre esses arquivos pode realizar o que os estudos de formas de representação indicam como apresentar apenas a informação relevante para a realização da tarefa. No arquivo XML podem estar todos os dados, porém, na apresentação ao usuário, o arquivo XSL, o programa, através do DOM, poderá selecionar apenas as informações que se enquadram no perfil do usuário. Quanto a integridade das informações, o arquivo DTD encarrega-se disso.

Tendo em mente o potencial da linguagem XML apresentado neste capítulo, comparado com os estudos sobre formas de representação, capítulo 2, e os objetivos da catalogação vistos no capítulo 3, acredita-se que no contexto da Ciência da Informação, da Biblioteconomia e especialmente da catalogação, a XML apresenta-se como ferramenta

relevante para colocar em prática as propostas das pesquisas realizadas quanto às formas de representação.

Uma das críticas mais severas a utilização do formato MARC 21 está relacionada a incapacidade de utilizá-lo em ambiente Web, entretanto, descrever um recurso eletrônico com o MARC 21 nunca foi problema, como vimos nos exemplos de Viana (2002) e Prado (2002). Talvez o problema seja a falta de conhecimento do formato, e neste caso um usuário comum não tem a responsabilidade de conhecê-lo, mas no ambiente da Biblioteconomia, mais especificamente da catalogação, conhecer o MARC 21 não é opcional, e sim uma obrigação, mesmo que não se tenha o interesse de implantá-lo.

Utilizar o formato no ambiente web, em uma página na Internet com as etiquetas no próprio corpo do documento, como ocorre com outras iniciativas ou com a utilização da tag `<META NAME="" CONTENT="">` da própria HTML era uma dificuldade que foi solucionada com a versão em XML do MARC 21 conforme será demonstrado no próximo capítulo.

6 MARC 21 em XML

A grande diferença entre o formato tradicional do MARC 21 e a sua versão em XML está na estrutura usada para organizar os dados bibliográficos e catalográficos. Tradicionalmente marcado por uma seqüência de caracteres contidos em um arquivo texto, a versão em XML do MARC 21 apresenta uma estrutura mais organizada, hierárquica, exatamente como a principal característica da linguagem XML. Deste modo, a essência do que foi apresentado no capítulo 4, sobre o formato MARC 21 será aproveitado. Destacam-se especialmente as informações apresentadas sobre o significado e funções dos elementos MARC: etiquetas, posições, indicadores e subcampos, que continuam sendo importantes e usados quanto as suas funções.

Visto que toda a documentação sobre o assunto está devidamente disponível na Internet, (MARC21-XML, 2002), não se pretende ser repetitivo. Sendo assim, este capítulo tem como principal objetivo apresentar a iniciativa da LC em desenvolver a versão em XML e como foco descrever como será o trabalho do programador de computador com essa nova versão, destacando especialmente, a facilidade, potencialidade e até mesmo as possíveis novas possibilidades de utilização desta iniciativa.

Mantido atualmente nos escritórios da LC o projeto desenvolveu uma estrutura para trabalhar com dados do MARC em um ambiente XML (MARC21-XML, 2002). Esta estrutura é flexível e extensível para permitir que os usuários trabalhem com dados do MARC nas maneiras específicas a suas necessidades. O projeto contém muitos componentes tais como estrutura de dados, listas de estilos de apresentação de dados e ferramentas do software desenvolvidas e mantidas pela Biblioteca do Congresso, LC. Baseado no site oficial do projeto, MARC21-XML (2002), detalhes são apresentados a seguir.

O MARC XML poderia ser usado potencialmente para representar um registro completo do MARC em XML, outra possibilidade é ser usado como uma estrutura da extensão a METS, Metadata Encoding & Transmission Standard. O METS é um padrão para um metadado

descritivo, administrativo e estrutural para codificar os objetos em uma biblioteca digital, desenvolvida usando a linguagem XML. O padrão é mantido no escritório do desenvolvimento da rede e dos padrões do MARC da Biblioteca do Congresso, e está sendo desenvolvido como uma iniciativa da Federação da Biblioteca Digital.

A estrutura do MARC 21 em XML suporta todos os dados codificados na forma tradicional e sua arquitetura é baseada em componentes, sendo extensível, ou seja, permite que os programadores de computador distribuam partes diferentes do software para construir soluções feitas sob encomenda. Estas são algumas das vantagens do MARC XML.

Segundo o MARC21-XML (2002), uma limitação identificada no projeto mostra que as validações do MARC não são reforçadas pela sua própria estrutura contida no arquivo XML, mas por um software externo. Entretanto, isso não se apresenta como problema, pois como apresentado no capítulo 5, este software, o DOM que manipulará um arquivo DTD, responsável pela validação, pode ser facilmente incorporado aos browsers após uma atualização opcional disponível no site das principais empresas, e as principais linguagens de programação contam com a capacidade de interagir com este software, tornando o desenvolvimento de um programa de validação muito simples em comparação com o trabalho necessário para validar um registro MARC 21 em seu formato tradicional.

O DTD do MARC 21 em XML define todos os elementos que podem constituir um registro bibliográfico neste formato. Atualmente encontram-se disponíveis na Internet, (MARC21-XML, 2002), os arquivos DTDs responsáveis pelas definições de todos os elementos de um registro MARC 21 em XML. Para se ter uma idéia dessa documentação, os arquivos mrcaxmlfile.dtd e mrcbxmlfile.dtd, juntos, correspondem a mais de 600 páginas de definições dos elementos MARC 21 XML.

Quanto à apresentação dos dados, destaca-se, que uma vez, o registro MARC foi convertido em XML, sua apresentação é possível a partir da criação de um stylesheet, ou listas de estilos de XML, para selecionar os elementos do MARC que serão apresentados e para

aplicar a marcação apropriada. Stylesheet são listas de apresentação ou modelos de apresentação predefinidos em uma linguagem de programação que podem ser usados e influenciados por determinadas circunstâncias. Um exemplo é a XLS apresentada no capítulo 5. Na prática, tem-se um determinado registro completo em MARC 21 XML e, de acordo com a necessidade, pode ser representado como uma referência bibliográfica, ficha catalográfica, uma listagem ou catálogo bibliográfico.

A maioria das conversões de dados pode ser escrita como transformações de XML, com muitas ferramentas disponíveis na web, algumas gratuitas, como o próprio programa disponível na página do projeto, na LC, desenvolvido na linguagem PERL. Para transformações mais complexas dos dados, as ferramentas de software que lêem o MARC XML deverão ser desenvolvidas, entretanto, esse processo de desenvolvimento é muito mais simples comparado com a versão tradicional, a de 1960. A validação da estrutura é realizada através de uma ferramenta de software, externa a estrutura, e a LC disponibiliza também um validador que fornecerá três níveis possíveis de validação:

- Validação básica de XML de acordo com a estrutura do MARC XML;
- Validação de etiquetas MARC21, incluindo os campos e subcampos;
- Validação do índice do registro do MARC, por exemplo, valores codificados, datas, e épocas.

Utilizando a XML como a estrutura para registros do MARC 21, os programadores de computador podem desenvolver mais facilmente suas próprias ferramentas para criar, manipular, e converter dados do formato. Como se pode notar, na web, e especialmente na página oficial do projeto, MARC21-XML (2002) e MARC21 (2002), encontram-se programas de computadores que auxiliam o programador no seu trabalho.

Segundo Viana (2002), a versão em XML poderia substituir o formato tradicional do MARC 21 com alguns ganhos, isso porque, a XML permite a descrição padronizada de diferentes metadados. Portanto, a iniciativa da LC tem como objetivo possibilitar que sistemas que utilizam metadados não MARC 21 possam convertê-los em XML e depois

para MARC 21. Para entender melhor, pense, por exemplo, em um sistema que utiliza como gerenciador de banco de dados o WinISIS para armazenar os registros bibliográficos. Seria possível transformar os dados do banco de dados em XML compatível com o MARC 21 em XML (ISIS, 2003), (MEDLANE, 2003) e (oss4lib, 2003).

Ainda de acordo com o autor, outra possibilidade é a conversão de dados de sistemas que utilizam MARC 21, como o Aleph 500, por exemplo, em registros para o formato XML, uma vez que a XML vem sendo adotada por muitas empresas para descrição de metadados por ser um padrão mais difundido, mais flexível e mais fácil de codificar e trabalhar eletronicamente do que o formato tradicional do MARC 21. Se uma instituição ou desenvolvedora de software decidir usar a XML, seu cliente, que já possui uma grande base de dados em MARC 21 poderá facilmente converter os dados, ou vice-versa.

6.1 Considerações sobre o MARC 21 XML para um analista de sistemas

Quanto ao programa que o analista de sistemas terá de desenvolver para interpretar registros bibliográficos em formato MARC 21 em XML, é muito simples, especialmente comparado com a forma tradicional. Essa facilidade é clara até mesmo para os leigos em computação. Veja um exemplo completo de um registro MARC 21 em XML logo abaixo, e compare, do ponto de vista de estrutura, com a figura 4 que contém o formato tradicional do MARC 21:

```
<?xml version="1.0" encoding="UTF-8" ?>
<collection xmlns="http://www.loc.gov/MARC21"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xsi:schemaLocation="http://www.loc.gov/MARC21
http://www.loc.gov/standards/marcxml/schema/MARC21.xsd">

<record type="Bibliographic">

<leader>01329cam 2200313 a 4500 </leader>

<controlfield tag="001">3873627 </controlfield>
<controlfield tag="005">19911106082810.9 </controlfield>
<controlfield tag="008">891101s1990 maua j 000 0 eng
</controlfield>

<datafield tag="035" ind1="" ind2="">
<subfield code="9">(DLC) 89048230 </subfield>
</datafield>
```

```

<datafield tag="906" ind1="" ind2="">
  <subfield code="a">7</subfield>
  <subfield code="b">cbc</subfield>
  <subfield code="c">orignew</subfield>
  <subfield code="d">1</subfield>
  <subfield code="e">ocip</subfield>
  <subfield code="f">19</subfield>
  <subfield code="g">y-gencatlg          </subfield>
</datafield>

<datafield tag="955" ind1="" ind2="">
  <subfield code="a">pc14 to he00 11-01-89; he06 11-01-89; he00 11-03-
89; fa00 11-06-89; fa05 11-22-89; he03 11-28-89; CIP ver. he06 08-08-90
</subfield>
</datafield>

<datafield tag="010" ind1="" ind2="">
  <subfield code="a"> 89048230          </subfield>
</datafield>

<datafield tag="020" ind1="" ind2="">
  <subfield code="a">0316107514 :</subfield>
  <subfield code="c">$12.95          </subfield>
</datafield>

<datafield tag="020" ind1="" ind2="">
  <subfield code="a">0316107506 (pbk.) :</subfield>
  <subfield code="c">$5.95 ($6.95 Can.)          </subfield>
</datafield>

<datafield tag="040" ind1="" ind2="">
  <subfield code="a">DLC</subfield>
  <subfield code="c">DLC</subfield>
  <subfield code="d">DLC          </subfield>
</datafield>

<datafield tag="042" ind1="" ind2="">
  <subfield code="a">lcac          </subfield>
</datafield>

<datafield tag="050" ind1="0" ind2="0">
  <subfield code="a">GV943.25</subfield>
  <subfield code="b">.B74 1990          </subfield>
</datafield>

<datafield tag="082" ind1="0" ind2="0">
  <subfield code="a">796.334/2</subfield>
  <subfield code="2">20          </subfield>
</datafield>

<datafield tag="100" ind1="1" ind2="">
  <subfield code="a">Brenner, Richard J.,</subfield>
  <subfield code="d">1941-          </subfield>

```

```

</datafield>

<datafield tag="245" ind1="1" ind2="0">
  <subfield code="a">Make the team.</subfield>
  <subfield code="p">Soccer :</subfield>
  <subfield code="b">a heads up guide to super soccer! /</subfield>
  <subfield code="c">Richard J. Brenner. </subfield>
</datafield>

<datafield tag="250" ind1="" ind2="">
  <subfield code="a">1st ed. </subfield>
</datafield>

<datafield tag="260" ind1="" ind2="">
  <subfield code="a">Boston :</subfield>
  <subfield code="b">Little, Brown,</subfield>
  <subfield code="c">c1990. </subfield>
</datafield>

<datafield tag="300" ind1="" ind2="">
  <subfield code="a">127 p. :</subfield>
  <subfield code="b">ill. ;</subfield>
  <subfield code="c">19 cm. </subfield>
</datafield>

<datafield tag="500" ind1="" ind2="">
  <subfield code="a">"A Sports illustrated for kids book."</subfield>
</datafield>

<datafield tag="520" ind1="" ind2="">
  <subfield code="a">Instructions for improving soccer skills. Discusses
dribbling, heading, playmaking, defense, conditioning, mental attitude,
how to handle problems with coaches, parents, and other players, and the
history of soccer.
</subfield>
</datafield>

<datafield tag="650" ind1="" ind2="0">
  <subfield code="a">Soccer</subfield>
  <subfield code="x">Juvenile literature. </subfield>
</datafield>

<datafield tag="650" ind1="" ind2="1">
  <subfield code="a">Soccer. </subfield>
</datafield>

<datafield tag="740" ind1="0" ind2="">
  <subfield code="a">Heads up guide to super soccer. </subfield>
</datafield>

<datafield tag="991" ind1="" ind2="">
  <subfield code="b">c-GenColl</subfield>
  <subfield code="h">GV943.25</subfield>

```

```

    <subfield code="i">.B74 1990</subfield>
    <subfield code="p">0002016704A</subfield>
    <subfield code="t">Copy 1</subfield>
    <subfield code="w">BOOKS          </subfield>
</datafield>

</record>
</collection>

```

Exemplo 1. Um registro MARC 21 em XML

Diante deste registro bibliográfico MARC 21 completo, compreende-se o grande avanço que a versão em XML do formato apresenta. A própria estrutura em que a informação é apresentada já surpreende em comparação com o formato tradicional, devido à organização e a facilidade de fazer uma interpretação. A surpresa ocorre em razão da principal particularidade da XML, a criação de tags de acordo com a necessidade, como se verifica ao analisar uma versão resumida da estrutura a pouco apresentada, as tags criadas pela LC:

```

<collection>
  <record type="">
    <leader> </leader>
    <controlfield tag="00X"> </controlfield>
    <datafield tag="XXX" ind1=" " ind2=" ">
      <subfield code="x"> </subfield>
    </datafield>
  </record>
</collection>

```

Exemplo 2. Registro MARC 21 em XML resumido

A seguir um exemplo de apenas uma etiqueta. A etiqueta 245, praticamente obrigatória em todo registro MARC 21 por trazer as informações sobre o título do documento:

```

<datafield tag="245" ind1="1" ind2="0">
  <subfield code="a">
    Make the team.
  </subfield>

  <subfield code="p">
    Soccer :
  </subfield>

```



```

<subfield code="b">
  a heads up guide to super soccer! /
</subfield>

<subfield code="c">
  Richard J. Brenner.
</subfield>
</datafield>

```

Exemplo 3. Etiqueta 245 do MARC 21 em XML

A partir dos exemplos apresentados identificam-se as tags criadas pela LC com o objetivo de utilizar as vantagens da linguagem XML no do formato MARC 21. Todas elas têm como característica a tag inicial, < >, e a tag final </ >. São elas:

- <collection> : Início do registro, descrevendo informações sobre a coleção de documentos que será apresentada, por exemplo, a localização;
- <record> : Informa o tipo de registro que será apresentado, livro, mapa, etc.;
- <leader> : O Líder do registro MARC 21;
- <controlfield> : Informações que se enquadram nas etiquetas 00X, por exemplo, 006, 007 e 008;
- <datafield> : Informações que se enquadram nas demais etiquetas, as que utilizam subcampos, por exemplo, 010, 100 e 245;
- <subfield> : Os subcampos.

Nota-se que a própria forma de representação de um registro MARC 21 em XML é mais simples para o analista de sistemas desenvolver um programa que importa dados neste formato, portanto, para finalizar, tem-se pelo menos quatro aspectos relevantes disponíveis a partir da linguagem XML:

- A linguagem XML já é de conhecimento dos programadores de computadores;

- As ferramentas existentes a disposição dos programadores de computador já trazem recursos que facilitam em muito a recuperação das etiquetas, posições, indicadores, subcampos e seus respectivos valores;
- A própria estrutura da XML provê um meio de validação dos dados através dos arquivos associados DTD;
- Um registro bibliográfico MARC 21 no formato XML possibilita diversas formas de representação para o mesmo conteúdo informacional, quando usado, por exemplo, o arquivo XSL.

Os dois primeiros aspectos relevantes beneficiam o programador de computador, especialmente o segundo. Visto que as ferramentas existentes a disposição dos programadores dispõem de recursos para manipular um arquivo XML, as principais linguagens de programação no mercado incorporaram essa funcionalidade também, tornando assim, mais fácil a recuperação das etiquetas, posições, indicadores e subcampos e seus respectivos valores. Deste modo, a maioria das preocupações técnicas, que um analista de sistemas teria para manipular um registro MARC 21 em seu formato tradicional, são dizimadas.

Para ilustrar, um exemplo de um programa desenvolvido na linguagem de programação Microsoft ASP que através da utilização do DOM, que pode capturar a descrição das tags de um arquivo XML com seus respectivos valores:

```
<% // Conecta com o DOM da Microsoft
    set xmlDoc = server.createObject( "MSXML2.DOMDocument" )
    xmlDoc.async = false

    // Acessa o arquivo XML, aqui denominado registroMARC21.xml
    xmlDoc.load ( server.mappath( "registroMARC21.xml" ) )

    // Executa uma sequência de leituras, passando por todas
    as tags do arquivo XML.
    for each i in ObjDOM.selectNodes("ROOT/CADASTRO/*")

        // Recupera o nome da tag
        response.write "<b>Nome da Tag:</b> " & i.nodeName & " <br>"

        // Recupera o valor da tag
        response.write "<b>Valor da Tag:</b> " & i.text & "<br><br>"
```

```
next  
set xmlDoc = nothing %>
```

Exemplo 4. Código fonte em ASP manipulando um arquivo XML

Cabe destacar que este programa não está preparado para efetivamente realizar a importação do registro MARC 21 em XML, porém, visto que este é capaz de identificar as tags e seus respectivos valores, tem-se a essência do programa. Portanto, a versão do MARC 21 em XML não dispensa a utilização das tabelas sugeridas no capítulo 4, sobre o MARC 21, no item 4.2, ilustrada na figura 5. Como visto, estas tabelas armazenam o significado das etiquetas, posições, indicadores e subcampos. Contudo, a comparação com o anexo B que traz um exemplo de um programa que importa dados bibliográficos e catalográficos do MARC 21 no formato tradicional, desenvolvido na linguagem de programação PHP, pode ser surpreendente, especialmente no que diz respeito a simplicidade e tamanho do código fonte.

O terceiro aspecto, validação dos dados através dos arquivos associados DTD, é um recurso muito útil para um arquivo XML, ou seja, cada tag mostrada no Exemplo 1. Exemplo de um registro MARC 21 em XML, pode ser definido em um arquivo DTD, o que significa que a tag `<datafield tag="991" ind1="" ind2="">`, ainda a pouco apresentada pode ser validada automaticamente pelo próprio browser de Internet e verificar se, por exemplo, quando a tag for igual à "991", pode existir um subcampo "b" conforme mostrado pela tag `<subfield code="b">c-GenColl</subfield>` e assim por diante.

Este é um recurso importante, visto que muitos sistemas de importação de dados baseados nos processamentos tradicionais antes apresentados, não fazem nenhum tipo de validação, e a XML, por assim dizer, tem um validador natural, o que motiva o programador até mesmo a desenvolver um validador mais sofisticado. O chamado aqui de validador, corresponde ao programa externo de validação mencionado anteriormente, mas que, na prática é muito simples utilizar, por exemplo, o Internet Explorer da Microsoft, a partir de uma atualização

simples no site da empresa pode ser capaz de validar um arquivo XML através do arquivo DTD associado, inclusive um registro MARC 21 em XML.

O último aspecto é relevante, por abrir novas possibilidades de uso para o MARC 21. Visto que um registro bibliográfico MARC 21 no formato XML, possibilita diversas formas de representação para a mesma informação, por exemplo, quando usado o arquivo XSL em parceria, podemos pensar na utilização do registro MARC 21 em XML para até mesmo produzir um recurso de apresentação de resumos e notas do documento pela Internet, evidentemente quando preenchidas as etiquetas necessárias.

Outra possibilidade seria usar o próprio registro MARC 21 em XML para efetuar a busca. Resumindo, a partir do preenchimento mais abrangente das etiquetas do formato e com a possibilidade de produzir diversas formas de representação das informações contidas no registro MARC 21 em XML, novos produtos informacionais podem ser idealizados. A utilização do MARC 21 em XML poderá até mesmo ampliar a funcionalidade deste formato, pois, sua potencialidade poderá ser explorada também em ambiente web.

Enquanto ocorrem várias discussões sobre a marcação apropriada para a informação bibliográfica na Web, Johnson (2001) afirma que o formato MARC 21 continuará a ser útil principalmente com o esforço no desenvolvimento de uma conversão para XML. O formato MARC será relevante no futuro, e seu uso será explorado na web para acesso à informação bibliográfica e catalográfica, entretanto, apesar de um começo promissor, mais pesquisas e investigações são necessárias, especialmente, no que diz respeito a produtos capazes de efetivamente colocar em prática a iniciativa da LC, a utilização da versão em XML para o MARC 21.

7 Considerações finais

A junção do que foi apresentado no capítulo 4, sobre MARC 21 e no capítulo 5, sobre XML, comparada com as informações apresentadas no capítulo 6 sobre a versão MARC 21 em XML acabam sendo a essência, ou aspecto relevante na utilização do referencial teórico sobre formas de representação como um fundamento para a catalogação.

O merecido destaque deve ser também atribuído à iniciativa da biblioteca do congresso americano, LC, por primeiramente desenvolver o principal formato de intercâmbio de dados bibliográficos e cataflográficos efetivamente e eficazmente usado no mundo, e atualmente, demonstrar estar atenta e apresentar uma versão que utiliza o mais novo e promissor método de prover a interoperabilidade entre sistemas, a XML.

Um importante ponto é explicar que a versão em XML do MARC acaba criando uma nova necessidade nos sistemas gerenciadores de bibliotecas. Atualmente estes sistemas estão preparados para fazer a importação e exportação de registros bibliográficos e catalográficos usando o formato tradicional do MARC 21, a seqüência de caracteres, contudo, a utilização da XML exige que os sistemas passem a disponibilizar a opção de realizar o intercâmbio de dados a partir desta linguagem, com o formato MARC 21.

A medida que os sistemas gerenciadores de bibliotecas passarem a disponibilizar opção de importar um registro MARC 21 a partir de um arquivo XML, nos padrões da LC apresentados aqui, e as instituições passarem a usar, acreditamos na tendência de desaparecimento do formato tradicional de 1960.

Dois fatores contribuirão para isso. O conhecimento prévio da linguagem de marcação XML que os profissionais da informática têm e a facilidade de desenvolvimento que a versão em XML do MARC 21 oferece, em comparação com a necessidade dos analistas de sistemas em compreender a seqüência de caracteres do formato tradicional. O segundo fator está relacionado com a iniciativa da LC em procurar prover uma tecnologia capaz de tornar possível o intercâmbio de dados de qualquer base bibliográfica, armazenado em formatos

diversos, para o MARC 21, ou seja, a linguagem de marcação XML está sendo usada inicialmente como uma ponte entre formatos de dados diversos, o que é um avanço natural e inevitável. No futuro, os métodos que possibilitarão a interoperabilidade entre diversos formatos de intercâmbio de dados bibliográficos e catalográficos terão no MARC 21 em XML uma opção relevante de metadados.

Acredita-se ainda, que atualmente a maioria dos softwares para biblioteca não disponibilizam uma ferramenta de intercâmbio de dados bibliográficos e catalográficos por não conhecerem o MARC 21, e pela complexidade exigida para o desenvolvimento na versão tradicional. Com a versão em XML do MARC 21, o desenvolvimento fica indiscutivelmente simples para um analista de sistemas atualizado nas tendências tecnológicas da Ciência da Computação. Por outro lado, tornar conhecida a importância do MARC 21 para a Biblioteconomia deve ser uma responsabilidade da área na formação de seus sujeitos institucionais e na disseminação de suas ferramentas de trabalho.

Apesar de a preocupação desta pesquisa estar centrada na questão da importação de dados, outra questão relevante diz respeito a exportação de um registro MARC 21. Com a versão tradicional do MARC 21, a dificuldade técnica para desenvolver um software que importe os dados é a mesma para exportação. Com o MARC 21 em XML a facilidade de produzir um software para exportar é indiscutivelmente simples, ou seja, não é possível estabelecer parâmetros de comparação com a versão tradicional do MARC 21 dado ao avanço alcançado na versão em XML.

Visto que a linguagem de marcação XML já foi incorporada pela Ciência da Computação, os principais sistemas gerenciadores de banco de dados, SGBD's disponíveis no mercado, por exemplo, Oracle, IBM DB2 e MS-SQL Server, contam com a particularidade de a partir da base armazenada, criar um arquivo XML. Isso na prática substitui a necessidade de desenvolver um programa para exportar os dados por alguns simples cliques e configurações no SGBD usado pela instituição ou biblioteca.

A potencialidade do MARC 21 destaca-se na possibilidade de ampliação na sua versão em XML. Em outras palavras, é possível pensar em novos produtos informacionais a partir da versão do MARC 21 em XML, ou mais especificamente, que a linguagem XML poderá se constituir em uma importante ferramenta para os interesses da Ciência da Informação.

Para ilustrar, é possível pensar em dois novos produtos. Primeiro, a descrição de documentos eletrônicos, efetuada no próprio corpo da página Web e, segundo, o desenvolvimento de sistemas de busca mais eficientes. Pensar nesses produtos também ajuda a embasar a afirmação de que a questão está em produzir sistemas com interfaces mais amigáveis, e, mais do que nunca a interdisciplinaridade natural existente entre a Ciência da Informação e a Ciência da Computação solicitou tanta atenção.

Quanto ao primeiro produto, de acordo com Prado (2002), para utilizar o campo 856 do MARC para disponibilizar texto integral da produção docente da UDESC na Internet foram selecionados oito etiquetas. Veja as etiquetas necessárias segundo Prado (2002):

```

Campo IND ----- Dados
100 -- 01 ---- Esteves, Paulo Cesar Leite.
245 -- 10 --- Universidade do Estado de Santa Catarina - UDESC : \b uma
universidade de qualidade / \c Paulo Cesar Leite Esteves, Raimundo
Zumblick. -
260 ----- Florianópolis, \c 1997.
300 ----- 27 p. -
650 ---- 04 ---- Universidades - \x Qualidade.
650 ---- 04 ---- Gestão pela Qualidade Total.
700 ---- 01----- Zumblick, Raimundo.
856      ---      40      ---      \q      Texto      html      \u
http://www.udesc.br/reitoria/bu/pd001.htm

```

O interessante é que este exemplo de etiquetas para descrever um documento eletrônico tem o objetivo de disponibilizar o documento na íntegra na Web, o que com a utilização do formato MARC 21 tradicional haveria dificuldade na descrição do documento feito no seu próprio corpo, como no caso das tags <META> da HTML ou outras iniciativas. Contudo tecnicamente falando, não seria impossível que um editor de HTML, software usado para a criação de páginas Web, oferecesse uma interface orientando o usuário na descrição da página, sem que este, entendesse o formato MARC 21.

Para melhor compreensão, imagine um editor de HTML que durante o processo inicial de configuração da sua futura página Web solicita em telas amigáveis o título, autor, assunto, resumos, notas etc, enfim, todas as informações necessárias para compor uma forma de representação descritiva do documento. Essas informações na verdade, sem que o usuário entenda exatamente o funcionamento serão usadas pelo editor HTML para inserir em uma tag <META> a versão tradicional do MARC 21, por exemplo, <META NAME="MARC21" CONTENT="01041cam 2200265 a 4500001002000000003000400020005001700024008004100041010002400082020002500 1060200044001310400018001750500024001930820018002171000032002352450087002 7246003600354250001200390260003700402300002900439500004200468520022000510 650003300730650001200763^###89048230#/AC/r91^DLC^19911106082810.9^891101s1 990#####maua###j#####000#0".... >, ou seja, a tradicional seqüência de caracteres que representa um registro MARC. Com a versão em XML do MARC 21 podemos imaginar outras possibilidades.

Estando essa informação no corpo do documento, o desenvolvimento do segundo produto se justifica, ou seja, um outro programa seria utilizado para realizar a catalogação desta página. Seria mais ou menos como o funcionamento do reconhecido site de busca Google, www.google.com.br, ou seja, através de um programa chamado robô de busca essa ferramenta de busca pode navegar pela Internet e catalogar atualmente qualquer página, com essa idéia em prática, o trabalho desse software seria até mesmo simplificado e conseqüentemente mais eficiente, visto que o robô de busca teria apenas que ler um registro MARC 21 que estaria em um lugar específico. Visto que o MARC 21 disponibiliza etiquetas para os vários aspectos de um documento, a busca baseada nestas, a partir de uma interface amigável, tornaria as buscas mais eficientes.

Para descrever de maneira eficiente um documento eletrônico no seu próprio corpo, o problema não está na falta de metadados e sim na possibilidade de uma junção entre o interesse do usuário em catalogar a sua página, com as funcionalidades que as ferramentas de editoração de HTML disponíveis deveriam oferecer.

Mas, além disso, com a versão do MARC 21 em XML e um software dotado de uma interface mais amigável, introduzir o formato MARC 21 no corpo do documento é mais simples, ou seja, não seria necessário usar a tag <META> da HTML para introduzir a seqüência de caracteres tradicional do MARC 21, como já apresentado, mas apenas fazer uma referência a um arquivo externo em XML, por exemplo, <META NAME="MARC21" CONTENT="http://www.biblioteca.com.br/ID9876.xml">, que corresponde ao documento visualizado, ou até mesmo abandonar a HTML e produzir o documento todo em XML.

O processo acaba ficando mais simples visto que a utilização das tags <META> da HTML poderia ser substituída pelas tags XML do MARC, ou adaptadas para trabalharem em conjunto. Apesar do objetivo da versão em XML do MARC não ser apenas o de descrever o documento eletrônico no próprio corpo, podemos notar que se antes isso não era feito, o problema não era o tradicional formato do MARC 21.

Ao solicitar a exibição do código fonte dos principais navegadores para Internet comprova-se que atualmente não se tem o costume de catalogar um documento eletrônico, nem mesmo com os recursos mais simples da HTML, e quando ocorre, a descrição é elaborada com objetivos econômicos somente.

Quanto ao desenvolvimento de sistemas de busca mais eficientes basta uma comparação com o funcionamento do Google. Como pode ser confirmado na página web da empresa, o buscador usa os chamados robôs de busca que tem seus critérios para catalogar a página. Basicamente, o robô analisa inicialmente se existe a tag <META> da HTML para palavras chave, posteriormente verifica se tem outras páginas que apontam para a que está prestes a ser catalogada. Quando não existe a tag <META> da HTML o robô executa um procedimento de tentar identificar as palavras chave da página.

Este parece ser o problema das buscas, pois, muitas vezes as palavras-chave usadas não são apropriadas, nem quando inserida pelo usuário e nem quando selecionadas pelo robô de busca. Portanto, com a inserção das tags XML do MARC 21 através do uso de um editor de

HTML com interface amigável, possivelmente auxiliaria no desempenho dos robôs de busca. Até mesmo as interfaces dos buscadores, provavelmente seriam diferentes dos conhecidos atualmente, visto que mais opções de busca estariam disponíveis a partir de uma provável catalogação séria e confiável.

Por isso a necessidade do desenvolvimento de produtos que incorporem e facilitem o uso do MARC 21 em XML, e da produção de componentes de software que englobem as funcionalidades da XML sem dúvida trariam benefícios para a Ciência da Informação, em especial para os profissionais responsáveis pela catalogação em suas instituições, provendo assim um maior controle da qualidade nos sistemas de gerenciamento de bibliotecas.

Entretanto, uma das principais preocupações deste componente de software deverá ser a interface com o usuário, especialmente, quanto à possibilidade de oferecer múltiplas formas de representação de um registro bibliográfico e catalográfico.

Ao analisar um registro MARC 21 na versão XML, percebe-se que manipular e processar essa estrutura torna-se simples, visto que, existem inúmeras ferramentas no mercado que auxiliam o analista de sistemas neste sentido. A partir deste ponto de vista, parece claro que o uso do recurso, MARC 21 em XML, é essencial para garantir a qualidade e segurança, mostrando ser uma solução para o intercâmbio dos dados bibliográficos e catalográficos.

Desta forma, concluí-se que a linguagem de marcação XML se apresenta como uma importante ferramenta para alcançarmos os objetivos da Ciência da Informação, ou seja, esta linguagem influenciará na construção, comunicação e uso da informação, auxiliando até mesmo na concepção de produtos e sistemas que permitam o tratamento adequado das propriedades da informação.

Em sua versão XML, o MARC 21 continuará exercendo um papel determinante na catalogação, e em outras, ao contemplar as funcionalidades que esta linguagem oferece. Finalizando, parece claro que os problemas vivenciados nos processos de catalogação, de

recuperação e de disseminação da informação não estão diretamente relacionados à ausência de teorias, métodos ou metadados, e sim a ausência de produtos informacionais adequados que englobem as iniciativas. O trabalho em conjunto da Ciência da Informação com a Ciência da Computação será decisivo para a mudança desta realidade.

É evidente que o interesse em catalogar corretamente um documento, especialmente o eletrônico, é determinante, todavia, não se pode esquecer das limitações financeiras, pessoais e até mesmo tecnológicas.

8 Conclusão

Ao longo da descrição da pesquisa foram apresentadas considerações e afirmações relevantes, nesse momento pretende-se organizar as conclusões que foram apresentadas, tornando disponíveis as informações que entendemos ser uma contribuição para a área da Ciência da Informação como também para a Ciência da Computação.

A conclusão mais relevante e empolgante foi poder perceber que os atuais problemas da catalogação de documentos, especialmente de documentos eletrônicos, não se devem a falta de métodos, metadados ou teorias, mas sim, a falta de desenvolvimento efetivo e eficaz de softwares com interfaces amigáveis que contemplem as diversas iniciativas. Entende-se ser muito importante estar alerta para este fato, como um possível tema de pesquisas futuras.

Avaliar as questões referentes às formas de representação mostrou ser muito importante, em especial, na análise de uma área que tradicionalmente tem suas formas de representações, as fichas catalográficas, citações, referências, catálogos bibliográficos etc. Assim, verificou-se que não se tratava da proposição de formas de representação, pois se entende que as atuais suprem as necessidades, mas sim verificar como criar essas representações da melhor forma possível a partir da mesma base bibliográfica e catalográfica, e em conformidade com as tecnologias disponíveis. Nesse sentido, outra vertente que se apresenta como estudos futuros está em enumerar os aspectos relevantes em uma forma de representação adequada de itens documentários em sua estrutura descritiva, analítica e ordenadora.

Como exemplo, a importância de uma ficha catalográfica está no processo de identificação e recuperação do documento, visto que a mesma apresenta uma forma de representação descritiva resumida e padronizada do documento. Isso na prática ocorre quando o usuário localiza o documento de sua necessidade a partir de uma pesquisa nas fichas catalográficas disponíveis em catálogos. Entretanto, em uma unidade de informação automatizada, o processo de busca é alterado, ou seja, as fichas catalográficas tradicionais não são mais consultadas, mas uma outra forma de representação da mesma, via computador. Nas

unidades de informação automatizadas, ocorreu uma mudança na forma de representação do documento, o mesmo apresenta-se de formas opcionais, basicamente na estrutura AACR2, MARC e referências bibliográficas.

Em outras palavras, e lembrando da afirmação de Peterson (1996) que temos que esquecer a idéia de encontrar um único método para representar o mundo, no contexto tecnológico, o uso de diversas formas de representação só será possível e útil se existir uma ferramenta capaz de gerenciar as diversas representações, e a XML apresenta-se atualmente como um componente determinante para o desenvolvimento desta ferramenta.

Contudo, a importância da XML para a pesquisa se destaca no sentido de que esta linguagem foi desenvolvida para possibilitar a criação de formatos de intercâmbio de dados, a interoperabilidade, ou seja, a XML por si só nada representa se não for utilizada para a criação de um formato. Sendo assim, encontrar uma descrição eficiente para o intercâmbio de dados bibliográficos e catalográficos seria a parte mais importante e complexa se não houvesse o formato de intercâmbio de dados bibliográficos MARC 21.

O formato de intercâmbio de dados MARC 21 em sua versão tradicional, por si só pode ser usado como ferramenta para a catalogação de documentos eletrônicos e o entendimento que o formato destaca-se entre os metadados desenvolvidos para a descrição de acervos específicos como TEI, CIMI, EAD e ICPSR, incentivou a investigação da versão do formato em XML.

O formato MARC 21 é o mais amplamente utilizado pela comunidade biblioteconômica, portanto, a maior base bibliográfica disponível atualmente no mundo, ou seja, o maior volume de descrições bibliográficas e catalográficas disponíveis atualmente se encontram no formato MARC 21. Sendo assim, a versão MARC 21 em XML pode ser considerada como uma opção de metadados ótima, quando inserida nesse contexto.

A grande abrangência descritiva que o formato oferece, e que atualmente está disponível em XML, principal linguagem de marcação usada por outras iniciativas, é outro aspecto relevante. Em outras palavras, a única coisa em comum que outras iniciativas parecem ter em comparação com o MARC 21 é a linguagem XML, o volume de dados disponíveis em MARC 21 e a capacidade descritiva, é indiscutivelmente superior no formato, especialmente em sua versão XML, ou seja, as funcionalidades que podem ser dadas ao MARC 21 em XML, parecem não ser encontradas ou reunidas em um único outro formato.

Infelizmente, à falta de conhecimento do formato MARC 21 tem gerado distorções no entendimento sobre sua abrangência e complexidade. O formato sem dúvida apresenta uma estrutura complexa, mas em nenhum momento isso pode ser encarado como negativo, e uma boa interface de um software resolveria a questão, considerando ainda que qualquer item documentário pode ser catalogado, em detalhes, pelo formato, devido a sua abrangência. Cabe destacar que não é obrigatória a utilização de todas as etiquetas do MARC 21 para representar um documento, ou seja, com apenas cinco ou oito etiquetas é possível desenvolver uma forma de representação aceitável para um documento eletrônico, conforme apresentado por Viana (2002) no capítulo 4, item 4.3, ou por Prado (2002) no capítulo 7.

Outra questão que chama a atenção, é a validação da importação de um registro MARC 21 XML, ou seja, essa validação é simples de ser implementada e até mesmo mais confiável, chegando a ponto de se afirmar que a estrutura disponível, arquivos DTDs, motiva a implementação de validadores mais sofisticados. Cabe ressaltar que a validação sempre será feita na estrutura do registro MARC 21, ou seja, não se pode garantir a qualidade da digitação, por exemplo, na etiqueta 245 que se refere ao título da obra, não se tem como verificar se a digitação do título foi correta.

Deste modo verifica-se que a relação entre a XML e a catalogação como forma de representação descritiva documentária é caracterizada pela versão em XML do MARC 21, visto que a importância do formato para a catalogação é indiscutível. Dado às

características técnicas da XML, contempla-se a possibilidade de ampliação da representação documentária a partir do uso estratégico das tecnologias dependentes da XML.

Visto que a criação de interfaces amigáveis para os módulos de catalogação dos sistemas gerenciadores de bibliotecas é destacada na pesquisa, entende-se que a Ciência da Computação pode auxiliar, na análise da usabilidade de software, uma das disciplinas da engenharia de software.

As questões de usabilidade de software estão relacionadas à classificação da qualidade da interação entre humano-computador. Esse conceito pode ser definido tradicionalmente como uma conjunção de cinco atributos: Facilidade de Aprendizado, Eficiência de Uso, Retenção, Minimização de Erros e Satisfação (HEURÍSTICAS, 2001).

Pressman (1995) define a usabilidade de uma interface como o esforço para aprender, operar, preparar a entrada e interpretar a saída de um programa, ou como ele mesmo resume: *“algo amigável ao usuário”*. Entendendo as definições apresentadas afirma-se que o termo Usabilidade pode ser definido como aquilo que encontramos em um sistema quando este, é fácil de usar e aprender.

O mais importante é concluir que a maior base bibliográfica disponível no mundo está construída no formato MARC 21, e este se apresenta atualmente na opção de utilização de uma das linguagens computacionais de marcação mais aceita e discutida, a XML – MARC 21 XML. Por isso, a implementação efetiva de ferramentas de software com interfaces amigáveis e que contemplem as funcionalidades do MARC 21 em XML, poderão constituir-se em um grande avanço para a catalogação de documentos bibliográficos, em especial, de documentos eletrônicos. Nesse sentido, levando-se em conta que o sucesso de outros formatos também dependerá do desenvolvimento de softwares com interfaces amigáveis, recomenda-se a utilização da “roda” já inventada.

9 Referências

ABBAGNANO, N. **Dicionário de filosofia**. 2. ed. São Paulo : Martins Fontes, 1998.

ALVES, E. H. **Observações sobre o perspectivismo de Donald Peterson**. In: Encontro com as ciências cognitivas. Organização de Maria Eunice Quilici Gonzáles e Mariana Claudia Broens. Marília : Unesp-Marília-Publicações, 1998, v. 2, p. 25-31.

BALBY, C. N. **Sistemas gerenciadores de biblioteca**. Disponível em <http://utopia.com.br/~simone/sistbib/tabelas/tabela_7.html> Acesso em 25 abr. 2003.

BARBOSA, A. P. **Novos rumos da catalogação**. Organização, revisão e Atualização de Elza Lima e Silva Maia. Rio de Janeiro: BNG/Brasilart, 1978. 246p. (Coleção biblioteconômica, documentação, ciência da informação).

BELKIN, N. J., ODDY, R. N., BROOKS, H. M. **ASK for information retrieval**. **Journal Documentation**, 38, n. 2, p.61-71, Part I, n.3, p.145-65, 1982. Part II.

BOSAK, J; BRAY, T. XML and the Second-Generation Web. **Scientific American**. Disponível em <<http://www.sciam.com/article.cfm?articleID=0008C786-91DB-1CD6-B4A8809EC588EEDF>> Acesso em 6 nov. 2002.

CASA, M. E. **Ambientes computacionais com múltiplas formas de representação**. In: Encontro com as ciências cognitivas. Organização de Maria Eunice Quilici Gonzáles [et. al.] 2. ed. rev. e ampl. Marília : Faculdade de Filosofia e Ciências, 1997, p. 209-220.

DÉCIO, O. C. **Guia de consulta rápida XML**. São Paulo: Novatec, 2000, p. 96.

DEMPSEY, L. HEERY, R. **A review of metadata**: a survey of current resource description on formats. Revision 10 mar. 1997. Disponível em <www.ukoen.ac.uk/metadata/desire/overview> Acesso em 27 jun. 2003.

DIAS, M. do R. I. **Catalogação e qualidade**: breve estudo. Marília: UNESP/CGB, 1999. 53 p.

ESTEVEVES, A.; SANTOS, L.; GUIMARÃES, P. **XML nas Bibliotecas Digitais: Introdução**. Disponível em <http://www.bibliosoft.pt/projectoxml/_introducao.htm> Acesso em 6 nov. 2002a.

ESTEVEVES, A.; SANTOS, L.; GUIMARÃES, P. **XML nas Bibliotecas Digitais: Polemica**. Disponível em <http://www.bibliosoft.pt/projectoxml/_polemica.htm> Acesso em 6 nov. 2002b.

ESTEVEVES, A.; SANTOS, L.; GUIMARÃES, P. **XML nas Bibliotecas Digitais: Standard**. Disponível em <http://www.bibliosoft.pt/projectoxml/_standard.htm> Acesso em 6 nov. 2002c.

FERREIRA, M. M. **MARC21 - Formato Condensado para Dados Bibliográficos**. Tradução e adaptação de MARC21: Concise for Bibliographic Data da Network Development and Marc Standards Office - Library of Congress - USA – 2. ed. Marília : Unesp-Marília-Publicações, 2002, 247 p. v1.

FURGERI, Sergio. **Business to Business: Aprenda a desenvolver aplicações**. São Paulo: Érica, 2001, 243p.

FURRIE, B. **Understanding MARC Bibliographic: Machine-Readable Cataloging**. Disponível em <<http://www.loc.gov/marc/umb/>> Acesso em 15 out. 2002.

GNUTECA. **Sistema aberto de gestão de bibliotecas**. Disponível em <<http://gnuteca.codigoaberto.org.br>> Acesso em 27 jul. 2002.

GUSMÃO, A. O. M. **Avaliação da qualidade e determinantes de desempenho do Aleph 500 em bibliotecas universitárias brasileiras**. 2001. 213 f. Dissertação (Mestrado em Ciência da Informação) - Universidade Federal da Paraíba, João Pessoa.

HEURÍSTICAS. **Análise Heurística de Usabilidade**. Disponível em <<http://www.dcc.ufmg.br/~aarj/usabilidade/heur.html>> Acesso em 27 set. 2001.

HOUAISS, A.; VILLAR, M. S. **Dicionário Houaiss da Língua Portuguesa / Antônio Houaiss e Mauro de Salles Villar, elaborado no Instituto Antônio Houaiss de Lexicografia e Banco de dados da Língua Portuguesa S/C Ltda. - Rio de Janeiro : Objetiva, 2001, 2925 p.**

ISIS. **X Reunión Regional de CDS/ISIS**. Disponível em
<<http://infolac.ucol.mx/eventos/cds-isis/programa.htm>> ou
<<http://infolac.ucol.mx/eventos/cds-isis/BIREME/X-ISIS-0.ppt>> Acesso em 8 jun. 2003.

JOHNSON, B. C. XML and MARC: Which is “Right” ? **Cataloging & Classification Quarterly**. Published by The Haworth Information Press. v. 32, n. 1, 2001.

KOPELEVITCH, A. **Introdução ao DOM**. Disponível em
<<http://www.htmlstaff.kit.net/xml/xml23.html>> Acesso em 21 out. 2002a.

KOPELEVITCH, A. **Introdução ao DTD**. Disponível em
<<http://www.htmlstaff.kit.net/xml/xml31.html>> Acesso em 21 out. 2002b.

KOPELEVITCH, A. **Introdução ao XDR**. Disponível em
<<http://www.htmlstaff.kit.net/xml/xml37.html>> Acesso em 21 out. 2002c.

KOPELEVITCH, A. **Introdução ao XSL**. Disponível em
<<http://www.htmlstaff.kit.net/xml/xml50.html>> Acesso em 21 out. 2002d.

LE COADIC, Yves-François. **A ciência da informação**. Tradução de Maria Yêda F.S de Figueiras Gomes. Brasília, DF : Briquete de Lemos/Livros, 1996, 119p.

MARC21. **MARC Standards**. Disponível em <<http://www.loc.gov/marc/>> Acesso em 22 jan. 2002.

LIBERTY, J; KRALEY, M. **Aprendendo a desenvolver documentos XML para a web**. Tradução de Flávia Cruz, Revisão Técnica de Álvaro Rodrigues Antunes. São Paulo: Makron Books, 2001, 274p.

MARC21-006. **Informações sobre a etiqueta 006**. Disponível em
<<http://lcweb.loc.gov/marc/bibliographic/ecbdctr.html#mrcb006>> ou
<<http://lcweb.loc.gov/marc/bibliographic/ecbdlist.html>> Acesso em 24 jan. 2002.

MARC21-007. **Informações sobre a etiqueta 007**. Disponível em <<http://www.loc.gov/marc/bibliographic/ecbd007s.html>> ou <<http://lcweb.loc.gov/marc/bibliographic/ecbdlist.html>> Acesso em 24 jan. 2002.

MARC21-XML. **MARC in XML**. Disponível em <<http://www.loc.gov/marc/marcxml.html>> Acesso em 12 jul. 2002.

MARCONDES, C. H. Representação e economia da informação. **Ci. Inf.**, jan./abr. 2001, vol.30, no.1, p.61-70. Disponível em <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0100-19652001000100008&lng=pt&nrm=iso> Acesso em 11 set. 2002.

MEY, E. S. A. **Introdução à catalogação**. Brasília: Briquet de Lemos/Livros, 1995, 123 p.

MEDLANE. **MEDLANE XMLMARC**: a project of Lane Medical Library, Stanford University Medical Center. Disponível <<http://xmlmarc.stanford.edu>> Acesso em 8 jun. 2003.

OSS4LIB. **oss4lib**: open source systems for libraries. Disponível em <<http://www.oss4lib.org/projects>> Acesso em 8 jun. 2003.

PEREIRA, A. M.; SANTOS, P. L. V. A. da C. O uso estratégico das tecnologias em catalogação. **Cad. F.F.C Marília**, v.7, n. 1/2, p. 121-131, 1998.

PESCHL, M. F. **Representing Representations**. In: *Forms of Representation: an interdisciplinary theme for cognitive science*. Donald Peterson (ed). 208pp. Exeter UK: Intellect Books, 1996. ISBN 1-871516-34-X. Disponível em <<http://www.chass.utoronto.ca/epc/srb/srb/representing.html>> Acesso em 27 ago. 2002.

PETERSON, D. (Ed.) **Forms of representation**: an interdisciplinary theme for Cognitive Science. Wiltshire: Cromwell Press, 1996.

PINHEIRO, L. V. R; LOUREIRO, J. M. M. **Traçados e limites da ciência da informação**. **Ci. Inf.**, 1995, vol.24 no.1. Disponível em <<http://www.ibict.br/cionline/240195/24019506.pdf>> Acesso em 16 mai. 2002.

PRADO, N. S. **Utilizando o campo 856 do MARC para disponibilizar texto integral da produção docente da UDESC na internet.** Disponível em

<<http://www.puccamp.br/~biblio/transinformacao/old/vol10n2/pag104.html>> Acesso em 15 out. 2002

PRESSMANS, R. S. **Engenharia de software** / Roger S. Pressman ; tradução José Carlos Barbosa dos Santos ; revisão técnica José Carlos Maldonato, Paulo Cesar Masiero, Rosely Sanches. - São Paulo : Makron Books, 1995, 1056 p.

RIBEIRO JÚNIOR, D. I. **Agentes inteligentes como mediadores na recuperação de informação.** 2001. 113 f. Dissertação (Mestrado em Ciência da Informação) - Universidade Estadual Paulista, Faculdade de Filosofia e Ciências, Marília, SP.

SANTOS, P. L. V. A. C. Processos de representação descritiva e seu ensino nos cursos de graduação em biblioteconomia. Marília : **Cadernos da FFC**, 1995. v. 4, n. 1, p. 36-39.

SILVA, O. J. **XML: aplicações práticas.** São Paulo : Érica, 2001, 276 p.

SOUZA, T. B; CATARINO, M. E; SANTOS, P. C. **Metadados: Catalogando dados na Internet.** Disponível em <<http://www.puccamp.br/~biblio/tbsouza92.html>> Acesso em 23 jan. 2002.

VIANA, M. M. M. **Guia de Biblioteconomia na Internet.** Disponível em <www.sobresites.com/biblioteconomia> ou <miquepoa@terra.com.br> Acesso em 10 ago. 2002

VIEIRA, K. C. **Processamento técnico: uma perspectiva histórica.** Disponível em <snbu.bvs.br/snbu2000/docs/pt/doc/poster004.doc> Acesso em 12 mai. 2003.

W3C. **World Wide Web Consortium.** Disponível em <<http://www.w3.org/> 2002> Acesso em 6 nov. 2002.

XSL. **The Extensible Stylesheet Language (XSL).** Disponível em <<http://www.w3.org/Style/XSL/>> Acesso em 28 out. 2002.

10 Documentos consultados

CAFÉ, L; SANTOS, C. D; MACEDO, F. Proposta de um método para escolha de software de automação de bibliotecas. **Ci. Inf.** v.30 n.2 Brasília maio/ago. 2001. Disponível em <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0100-19652001000200009&lng=pt&nrm=isso> Acesso em 23 jan. 2002.

CÔRTE, A. R, [et al]. Automação de bibliotecas e centros de documentação: o processo de avaliação e seleção de softwares. **Ci. Inf.** v. 28 n.3 Brasília set./dez. 1999. Disponível em <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0100-19651999000300002&lng=pt&nrm=isso> Acesso em 23 jan. 2002.

DIAS, T. M. Pergamum – Sistema informatizado da biblioteca da PUC/PR. **Ci. Inf.** v.27 n.3 Brasília Set. 1998. Disponível em <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0100-19651998000300010&lng=pt&nrm=isso> Acesso em 23 jan. 2002.

GONÇALVES, E. M. S, [et al]. Informatização da informação: a experiência do Sistema de Bibliotecas da Universidade Federal do Rio Grande do Sul. **Ci. Inf.** v.27 n.1 Brasília 1998. Disponível em <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0100-19651998000100014&lng=pt&nrm=isso> Acesso em 23 jan. 2002.

KRAEMER, L. L. B. **Metadados: estudo de sua aplicação no tratamento de recursos virtuais e análise de um projeto do programa Prossiga do IBICT.** 2001, 138 f. Dissertação (Mestrado) Programa de Pós-graduação em Tecnologia, Centro Federal de Educação Tecnológica do Paraná, Curitiba.

LIMA, G. Â. B. Softwares para automação de bibliotecas e centros de documentação na literatura brasileira até 1998. **Ci. Inf.** v.28 n.3 Brasília set./dez. 1999. Disponível em <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0100-19651999000300009&lng=pt&nrm=isso> Acesso em 23 jan. 2002.

Anexos

Anexo A - Documentação das tabelas sugeridas, que armazenam o significado das etiquetas, posições, indicadores e subcampos

Destacamos que estas tabelas além de poderem ser usadas pelo programa de computador para auxiliar na interpretação dos dados bibliográficos contidos em um registro MARC 21, podem também ser usadas para a implementação de um sistema de ajuda ao bibliotecário. Com os dados destas tabelas devidamente preenchidos, o programador terá a principal ferramenta para produzir um programa do tipo help desk para os usuários do sistema.

Cabe relembrar novamente que o MARC 21 é um formato ou método, que informa como um registro bibliográfico deve ser organizado para que sua importação ou exportação ocorra com sucesso, sendo assim, como um registro bibliográfico é armazenado em um sistema gerenciador de banco de dados é de responsabilidade de cada administrador do banco de dados, conhecidos também como DBA, ou seja, a estrutura criada para armazenar os dados bibliográficos possivelmente será diferente visto que cada DBA usará a técnica que mais lhe agrada. Para esse trabalho, existe um método chamado de normalização de dados, que por sua vez é dividido em cinco estágios, que o DBA não é obrigado a seguir completamente.

Tabela: **abrang**

Tipos de documentos que podem ser armazenados no MARC21. Ex: Livros (BK), Música (MU) etc.

Colunas

<u>Nome</u>	<u>Tipo</u>	<u>Tamanho</u>
abrang	Texto	2
	Código do tipo de documento a ser catalogado.	
descri	Texto	50
	Descrição do documento a ser catalogado.	

Tabela: **etique**

Etiquetas.

Colunas

<u>Nome</u>	<u>Tipo</u>	<u>Tamanho</u>
etique	Texto	5
Etiqueta		
descri	Texto	100
Descrição do significado da Etiqueta		
exempl	Texto	255
Exemplo da utilização da etiqueta completa, incluindo seus indicadores e subcampo.		
observ	Memorando	-
Explicação detalhada da etiqueta		
temptam	Sim/Não	1
Se for verdadeiro, significa que trabalha com tamanho de campo, e não com subcampos e indicadores. Ex: 006 a 008		
obsole	Sim/Não	1
Se verdadeiro : etiqueta obsoleta		
palcha	Texto	255
255 caracteres para informar palavras chaves para auxiliar a interface de busca. Separar com espaço		

Tabela: **indic1**

Indicadores 1 da etiqueta

Colunas

<u>Nome</u>	<u>Tipo</u>	<u>Tamanho</u>
indic1	Texto	1
Indicador 1 da etiqueta relacionada		
etique	Texto	5
Etiqueta que o indicador esta relacionado		
desind	Texto	50
Descrição do significado do indicador quando relacionado com uma etiqueta		
obsole	Sim/Não	1
Se verdadeiro : indicador obsoleto		

Tabela: **indic2**

Indicadores 2 da etiqueta

Colunas

Nome	Tipo	Tamanho
indic2	Texto	1
Indicador 2 da etiqueta relacionada		
etique	Texto	5
Etiqueta que o indicador esta relacionado		
desind	Texto	50
Descrição do significado do indicador		
obsole	Sim/Não	1
Se verdadeiro : indicador obsoleto		

Tabela: **obriga**

Tabela que informa as relações de obrigatoriedade existentes entre etiquetas do MARC 21.

Colunas

Nome	Tipo	Tamanho
etiqu1	Texto	5
Primeira etiqueta para realizar a relação de obrigatoriedade.		
etiqu2	Texto	5
Segunda etiqueta para realizar a relação de obrigatoriedade.		
descri	Texto	255
Descrição do motivo da obrigatoriedade de relação		

Tabela: **outsub**

Significado dos códigos que em alguns casos são inseridos em subcampos.

Colunas

Nome	Tipo	Tamanho
etique	Texto	5
Etiqueta		
subcam	Texto	1
Subcampo associado com a etiqueta		
outsub	Texto	7
Outro código relacionado com o subcampo		
descre	Texto	50
Descrição do significado do código relacionado com o subcampo		
obsole	Sim/Não	1
Se verdadeiro : indicador obsoleto		

Tabela: posica

Para o Líder e etiquetas 006 até 008 que trabalha com posições e não indicadores e subcampos.

Colunas

<u>Nome</u>	<u>Tipo</u>	<u>Tamanho</u>
etique	Texto	5
Etiqueta		
posica	Texto	10
Posição(ões) do código dentro das etiquetas 006 a 008 e Líder		
codigo	Texto	10
Código(s) que preenche a posição informada		
descod	Texto	80
Descrição do significado do código especificado		
despos	Texto	80
Descrição do significado da posição especificado		
abrang	Texto	2
Código do tipo do documento que esta sendo catalogado : Usar especialmente para a etiqueta 008		
obsole	Sim/Não	1
Se verdadeiro : posição obsoleta		

Tabela: subcam

Subcampos da etiqueta.

Colunas

<u>Nome</u>	<u>Tipo</u>	<u>Tamanho</u>
subcam	Texto	1
código do subcampo		
etique	Texto	5
etiqueta que o subcampo esta relacionado		
descre	Texto	50
Descrição do significado do subcampo relacionado com uma etiqueta		
obsole	Sim/Não	1
Se verdadeiro : subcampo obsoleto		

Anexo B - Exemplo de um programa que importa dados bibliográficos desenvolvido na linguagem PHP

Segundo GNUTECA (2002):

```

<?
/*****
/* Importa Arquivo ISO para intercâmbio de acervos          */
/* Linguagem PHP                                           */
/* Autor: Pablo Dall'Oglio e João Alex Fritsch            */
/* Última alteração em 17 Agosto 2001 por Pablo          */
/*****

include_once '../gtk/util/GnutecaDialog.class';
include_once '../gtk/util/GnutecaGrid.class';
include_once 'EntryBox.php';
include_once '../../../../miolo/miolo.conf';
include_once '../../../../miolo/miolo.class';
include_once '../db/GnutecaMIOLO.class';

function LogError($mensagem, $ArquivoISOaImportar, $registro)
{
    $logFileError = fopen("$ArquivoISOaImportar.log", "a"); // arquivo para
controle de erros
    echo("$mensagem \n");
    fwrite($logFileError, "$mensagem \n $registro \n");
    fclose($logFileError);
}

function CheckDelimitador($aprocurar, $delimiter)
{
    $fl_possui=false;
    for ($n=0; $n<strlen($aprocurar); $n++)
        if (substr($aprocurar,$n, 1) == $delimiter)
            $fl_possui=true;

    return $fl_possui;
}

function TrataRegistro($registro, $delimitadorDeCampo, $delimitadorDeSubCampo,
$ArquivoISOaImportar)
{
    global $BusinessMaterial;

    $tamanhoFisicoDoRegistro = strlen($registro);
    $tamanhoLogicoDoRegistro = substr($registro,0,5);
    $numeroDeIndicadores = substr($registro,10,1);
    $tamanhoDoTamanho = substr($registro,20,1);
    $tamanhoDaPosicaoInicial = substr($registro,21,1);

    $slider = substr($registro,0,24);
    $n = 24;
    $i = 1;

    while (substr($registro,$n,1) != $delimitadorDeCampo)
    {
        $codigoDaEtiqueta = substr($registro,$n, 3); $n +=3;
        $tamanhoDaEtiqueta = substr($registro,$n, $tamanhoDoTamanho); $n +=
$tamanhoDoTamanho;
        $posicaoDaEtiqueta = substr($registro,$n, $tamanhoDaPosicaoInicial); $n +=
$tamanhoDaPosicaoInicial;

        $etiqueta["$codigoDaEtiqueta-$i"][1] = $posicaoDaEtiqueta; // Posicao
Relativa
        $etiqueta["$codigoDaEtiqueta-$i"][2] = $tamanhoDaEtiqueta; // Tamanho do
Campo
        if ($codigoDaEtiqueta == '010')
            $GuardarEtiqueta = $i;

        $i ++;
    }
}

```

```

}

$i = $GuardarEtiqueta;
echo "aqui \n";
if (($etiqueta["010-$i"]==null)) // Não possui númeroDaObra
{ LogError("Registro inconsistente: lider= $lider (Sem Obra)",
$ArquivoISOaImportar, $registro ); }

else if ($tamanhoFisicoDoRegistro != $tamanhoLogicoDoRegistro)
{ LogError("Registro inconsistente: lider= $lider (Tamanhos incompatíveis)",
$ArquivoISOaImportar, $registro); }

else
{
    $numeroDaObra = substr($registro, $etiqueta["010-$i"][1] + $n +1,
$etiqueta["010-$i"][2] -1);
    echo "010-$i\n";

    echo("Importando N° Obra: " . $numeroDaObra . " ... \n");
    $n++;

    $NovoMaterial = new GnutecaMaterial();
    $NovoMaterial->NumeroDaObra = $numeroDaObra;
    $NovoMaterial->Etiqueta = '000';
    $NovoMaterial->Indicador1 = ' ';
    $NovoMaterial->Indicador2 = ' ';
    $NovoMaterial->Subcampo = 'a';
    $NovoMaterial->Conteudo = $lider;

    $BusinessMaterial->NovoMaterial($NovoMaterial);

    foreach ($etiqueta as $PreCodigoDaEtiqueta=>$CadaEtiqueta)
    {
        $codigoDaEtiqueta = substr($PreCodigoDaEtiqueta,0,3);
        $sufixoDaEtiqueta = substr($PreCodigoDaEtiqueta,4);

        $conteudo = substr($registro, $CadaEtiqueta[1] + $n, $CadaEtiqueta[2] -1);
        TrataCampo($numeroDaObra, $codigoDaEtiqueta, $conteudo,
$delimitadorDeSubCampo, $numeroDeIndicadores, $sufixoDaEtiqueta);
    }
}

function TrataCampo($numeroDaObra, $codigoDaEtiqueta, $conteudo,
$delimitadorDeSubCampo, $numeroDeIndicadores, $sufixoDaEtiqueta)
{
    global $BusinessMaterial;
    $i = 0;

    $indicador1 = ($numeroDeIndicadores==2) ? substr($conteudo,0,1) : ' ';
    $indicador2 = ($numeroDeIndicadores==2) ? substr($conteudo,1,1) : ' ';
    $conteudo = substr($conteudo, $numeroDeIndicadores);

    if (substr($conteudo,0,1) != $delimitadorDeSubCampo)
        $conteudo = $delimitadorDeSubCampo . 'a' . $conteudo; // Corrigidos Registros
Sem início SubCampo

    for($n=0; $n<=strlen($conteudo); $n++)
    {
        $letrinha = substr($conteudo,$n,1);

        if ($letrinha==$delimitadorDeSubCampo)
        { $i ++; $subcampo[$i] =""; }
        else
        { $subcampos[$i] .= $letrinha; }
    }

    foreach($subcampos as $subcampo)
    {

```

```

$codigoDoSubCampo = substr($subcampo,0,1);
$conteudoDoSubCampo = substr($subcampo,1);

$NovoMaterial = new GnutecaMaterial();
$NovoMaterial->NumeroDaObra = $numeroDaObra;
$NovoMaterial->Linha = $sufixoDaEtiqueta;
$NovoMaterial->Etiqueta = $codigoDaEtiqueta;
$NovoMaterial->Indicador1 = $indicador1;
$NovoMaterial->Indicador2 = $indicador2;
$NovoMaterial->Subcampo = $codigoDoSubCampo;
$NovoMaterial->Conteudo = $conteudoDoSubCampo;
$BusinessMaterial->NovoMaterial($NovoMaterial);
}
}

function Importa()
{
    global $ArquivoISOaImportar, $DelimitadorDeRegistro,
           $DelimitadorDeCampo, $DelimitadorDeSubCampo;

    $delimitadorDeRegistro = chr(trim($DelimitadorDeRegistro->get_text()));
    $delimitadorDeCampo = chr(trim($DelimitadorDeCampo->get_text()));
    $delimitadorDeSubCampo = chr(trim($DelimitadorDeSubCampo->get_text()));

    $filename = $ArquivoISOaImportar->get_text();

    $fd = @fopen("$filename", "r");
    if ($fd)
    {
        while (!feof ($fd))
        {
            $buffer='';
            $buffer = fgets($fd, 80);

            $buffer = ereg_replace("\n", '', $buffer);
            $buffer = ereg_replace(chr(13), '', $buffer);
            $registro .= $buffer;

            if (CheckDelimitador($buffer, $delimitadorDeRegistro))
            {
                TrataRegistro($registro, $delimitadorDeCampo, $delimitadorDeSubCampo,
$filename);
                $registro = '';
            }
        }
        fclose ($fd);
    }

    Gtk::main_quit();
}

function FechaJanela()
{
    Gtk::main_quit();
}

if (strtoupper(substr(PHP_OS, 0, 3)) == 'WIN')
    dl('php_gtk.dll');
else
    dl('php_gtk.so');

$MIOLO = new GnutecaMIOLO('/usr/local/bis');
$MIOLO->Authenticate('postgres', 'postgres');

$BusinessMaterial = $MIOLO->GetBusiness('gnuteca', 'Material');

$window = &new GtkWindow;

```

```

$window->connect('delete-event', 'FechaJanela');
$window->set_title('Preparação de Ambiente - Importação');
$window->set_border_width(0);
$window->set_default_size(480, 100);
$window->set_uposition(80, 80);

$ vbox = &new GtkVBox();
$window->add($vbox);
$ vbox->show();

    $ hbox = &new GtkHBox();
    $ vbox->add($ hbox);
    $ hbox->show();
    $window->realize();

        $ArquivoISOaImportar = &new GtkEntry();
        $ hbox->pack_start(CreateEntryBox($ArquivoISOaImportar, 20, true, 'Arquivo',
2, 15, 20, GTK_BUTTONBOX_SPREAD), true, false, 10);

        $DelimitadorDeRegistro = &new GtkEntry();
        $ hbox->pack_start(CreateEntryBox($DelimitadorDeRegistro, 10, true,
'Delimitador Registro', 2, 15, 20, GTK_BUTTONBOX_SPREAD), true, false, 10);

    $ hbox = &new GtkHBox();
    $ vbox->add($ hbox);
    $ hbox->show();
    $window->realize();

        $DelimitadorDeCampo = &new GtkEntry();
        $ hbox->pack_start(CreateEntryBox($DelimitadorDeCampo, 10, true,
'Delimitador Campo', 2, 15, 20, GTK_BUTTONBOX_SPREAD), true, false, 10);

        $DelimitadorDeSubCampo = &new GtkEntry();
        $ hbox->pack_start(CreateEntryBox($DelimitadorDeSubCampo, 10, true,
'Delimitador SubCampo', 2, 15, 20, GTK_BUTTONBOX_SPREAD), true, false, 10);

    $ hbox = &new GtkHBox();
    $ vbox->add($ hbox);
    $ hbox->show();
    $window->realize();

        $button = &new GtkButton('Importa');
        $button->connect('clicked', 'Importa', $ctree);
        $ hbox->pack_start($button);
        $button->show();

$window->show();

Gtk::main();
?>

```


**Anexo C - Elementos criados para corresponder as tags tradicionais do
MARC 21: Schema MARC21slim.xsd**

Schema **MARC21slim.xsd**

schema location: <http://www.loc.gov/standards/marcxml/schema/MARC21slim.xsd>
 targetNamespace: <http://www.loc.gov/MARC21/slim>

Elements	Complex types	Simple types
<u>collection</u>	<u>collectionType</u>	<u>controlDataType</u>
<u>record</u>	<u>controlFieldType</u>	<u>controltagDataType</u>
	<u>dataFieldType</u>	<u>idDataType</u>
	<u>leaderFieldType</u>	<u>indicatorDataType</u>
	<u>recordType</u>	<u>leaderDataType</u>
	<u>subfieldatafieldType</u>	<u>recordTypeType</u>
		<u>subfieldcodeDataType</u>
		<u>subfieldDataType</u>
		<u>tagDataType</u>

schema location: <http://www.loc.gov/standards/marcxml/schema/xml.xsd>
 targetNamespace: <http://www.w3.org/XML/1998/namespace>

Attr. groups
[xmlGlobalAttributes](#)

element **collection**

diagram	<p>The diagram shows a class collectionType (highlighted in yellow) containing an element record. A dashed line connects the collection element to the collectionType class. The collection element is described as a top level container element for 0 or many records. The record element is described as a top level container element for all of the field elements which compose the record.</p>					
namespace	http://www.loc.gov/MARC21/slim					
type	<u>collectionType</u>					
children	<u>record</u>					
attributes	Name	Type	Use	Default	Fixed	Annotation
	id	idDataType	optional			
annotation	documentation collection is a top level container element for 0 or many records					
source	<pre><xsd:element name="collection" type="collectionType" nillable="true" id="collection.e"> <xsd:annotation> <xsd:documentation>collection is a top level container element for 0 or many records</xsd:documentation> </xsd:annotation> </xsd:element></pre>					

element **record**

diagram	<p>record is a top level container element for all of the field elements which compose the record</p>																		
namespace	http://www.loc.gov/MARC21/slim																		
type	<u>recordType</u>																		
children	<u>leader</u> <u>controlfield</u> <u>datafield</u>																		
used by	complexType <u>collectionType</u>																		
attributes	<table border="1"> <thead> <tr> <th>Name</th> <th>Type</th> <th>Use</th> <th>Default</th> <th>Fixed</th> <th>Annotation</th> </tr> </thead> <tbody> <tr> <td>type</td> <td>recordTypeType</td> <td>optional</td> <td></td> <td></td> <td></td> </tr> <tr> <td>id</td> <td>idDataType</td> <td>optional</td> <td></td> <td></td> <td></td> </tr> </tbody> </table>	Name	Type	Use	Default	Fixed	Annotation	type	recordTypeType	optional				id	idDataType	optional			
Name	Type	Use	Default	Fixed	Annotation														
type	recordTypeType	optional																	
id	idDataType	optional																	
annotation	documentation record is a top level container element for all of the field elements which compose the record																		
source	<pre><xsd:element name="record" type="recordType" nillable="true" id="record.e"> <xsd:annotation> <xsd:documentation>record is a top level container element for all of the field elements which compose the record</xsd:documentation> </xsd:annotation> </xsd:element></pre>																		

complexType **collectionType**

diagram	<p>record is a top level container element for all of the field elements which compose the record</p>					
namespace	http://www.loc.gov/MARC21/slim					
children	record					
used by	element collection					
attributes	Name id	Type idDataType	Use optional	Default	Fixed	Annotation
source	<pre><xsd:complexType name="collectionType" id="collection.ct"> <xsd:sequence minOccurs="0" maxOccurs="unbounded"> <xsd:element ref="record"/> </xsd:sequence> <xsd:attribute name="id" type="idDataType" use="optional"/> </xsd:complexType></pre>					

complexType **controlFieldType**

diagram	<p>MARC21 Fields 001-009</p>					
namespace	http://www.loc.gov/MARC21/slim					
type	extension of controlDataType					
used by	element recordType/controlfield					
facets	whiteSpace preserve					
attributes	Name id tag xml:space	Type idDataType controltagDataType xsd:NCName	Use optional required	Default	Fixed preserve	Annotation
annotation	documentation MARC21 Fields 001-009					
source	<pre><xsd:complexType name="controlFieldType" id="controlfield.ct"> <xsd:annotation> <xsd:documentation>MARC21 Fields 001-009</xsd:documentation> </xsd:annotation> <xsd:simpleContent></pre>					

```

<xsd:extension base="controlDataType">
  <xsd:attribute name="id" type="idDataType" use="optional"/>
  <xsd:attribute name="tag" type="controltagDataType" use="required"/>
  <xsd:attribute ref="xml:space" fixed="preserve"/>
</xsd:extension>
</xsd:simpleContent>
</xsd:complexType>

```


complexType dataFieldType

diagram						
namespace	http://www.loc.gov/MARC21/slim					
children	subfield					
used by	element recordType/datafield					
attributes	Name	Type	Use	Default	Fixed	Annotation
	id	idDataType	optional			
	tag	tagDataType	required			
	ind1	indicatorDataType	required			
	ind2	indicatorDataType	required			
annotation	documentation MARC21 Variable Data Fields 010-999					
source	<pre> <xsd:complexType name="dataFieldType" id="datafield.ct"> <xsd:annotation> <xsd:documentation>MARC21 Variable Data Fields 010-999</xsd:documentation> </xsd:annotation> <xsd:sequence maxOccurs="unbounded"> <xsd:element name="subfield" type="subfielddataFieldType"/> </xsd:sequence> <xsd:attribute name="id" type="idDataType" use="optional"/> <xsd:attribute name="tag" type="tagDataType" use="required"/> <xsd:attribute name="ind1" type="indicatorDataType" use="required"/> <xsd:attribute name="ind2" type="indicatorDataType" use="required"/> </xsd:complexType> </pre>					

element dataFieldType/subfield

diagram						
namespace	http://www.loc.gov/MARC21/slim					
type	subfielddataFieldType					
facets	whiteSpace preserve					
attributes	Name	Type	Use	Default	Fixed	Annotation
	id	idDataType	optional			
	code	subfieldcodeDataType	required			
source	<pre> <xsd:element name="subfield" type="subfielddataFieldType"/> </pre>					

complexType leaderFieldType

diagram						
namespace	http://www.loc.gov/MARC21/slim					
type	extension of leaderDataType					
used by	element recordType/leader					
facets	whiteSpace preserve pattern [\d][5][\dA-Za-z]{1}[\dA-Za-z]{1}[\dA-Za-z]{3}(2) (2)[\d][5][\dA-Za-z]{3}(4500)					
attributes	Name	Type	Use	Default	Fixed	Annotation
	id	idDataType	optional			
	xml:space	xsd:NCName			preserve	
annotation	documentation MARC21 Leader, 24 bytes					
source	<pre> <xsd:complexType name="leaderFieldType" id="leader.ct"> <xsd:annotation> <xsd:documentation>MARC21 Leader, 24 bytes</xsd:documentation> </xsd:annotation> <xsd:simpleContent> <xsd:extension base="leaderDataType"> <xsd:attribute name="id" type="idDataType" use="optional"/> <xsd:attribute ref="xml:space" fixed="preserve"/> </xsd:extension> </xsd:simpleContent> </xsd:complexType> </pre>					


complexType recordType

diagram						
namespace	http://www.loc.gov/MARC21/slim					
children	<u>leader</u> <u>controlfield</u> <u>datafield</u>					
used by	element <u>record</u>					
attributes	Name	Type	Use	Default	Fixed	Annotation
	type	recordTypeType	optional			
	id	idDataType	optional			
source	<pre> <xsd:complexType name="recordType" id="record.ct"> <xsd:sequence minOccurs="0"> <xsd:element name="leader" type="leaderFieldType"/> <xsd:element name="controlfield" type="controlFieldType" minOccurs="0" maxOccurs="unbounded"/> <xsd:element name="datafield" type="dataFieldType" minOccurs="0" maxOccurs="unbounded"/> </xsd:sequence> <xsd:attribute name="type" type="recordTypeType" use="optional"/> <xsd:attribute name="id" type="idDataType" use="optional"/> </xsd:complexType> </pre>					

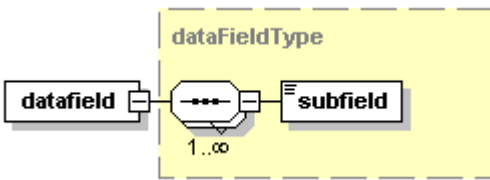
element recordType/leader

diagram						
namespace	http://www.loc.gov/MARC21/slim					
type	<u>leaderFieldType</u>					
facets	whiteSpace preserve pattern [\d]{5}[\dA-Za-z]{1}[\dA-Za-z]{1}[\dA-Za-z]{3}(2)) (2))[\d]{5}[\dA-Za-z]{3}(4500))					
attributes	Name	Type	Use	Default	Fixed	Annotation
	id	idDataType	optional			
	xml:space	xsd:NCName			preserve	
source	<pre> <xsd:element name="leader" type="leaderFieldType"/> </pre>					

element recordType/controlfield

diagram						
namespace	http://www.loc.gov/MARC21/slim					
type	controlFieldType					
facets	whiteSpace preserve					
attributes	Name	Type	Use	Default	Fixed	Annotation
	id	idDataType	optional			
	tag	controltagDataType	required			
	xml:space	xsd:NCName			preserve	
source	<code><xsd:element name="controlfield" type="controlFieldType" minOccurs="0" maxOccurs="unbounded"/></code>					

element recordType/datafield

diagram						
namespace	http://www.loc.gov/MARC21/slim					
type	dataFieldType					
children	subfield					
attributes	Name	Type	Use	Default	Fixed	Annotation
	id	idDataType	optional			
	tag	tagDataType	required			
	ind1	indicatorDataType	required			
	ind2	indicatorDataType	required			
source	<code><xsd:element name="datafield" type="dataFieldType" minOccurs="0" maxOccurs="unbounded"/></code>					

complexType **subfielddataType**

diagram						
namespace	http://www.loc.gov/MARC21/slim					
type	extension of subfieldDataType					
used by	element dataFieldType/subfield					
facets	whiteSpace preserve					
attributes	Name	Type	Use	Default	Fixed	Annotation
	id	idDataType	optional			
	code	subfieldcodeDataType	required			
source	<pre><xsd:complexType name="subfielddataType" id="subfield.ct"> <xsd:simpleContent> <xsd:extension base="subfieldDataType"> <xsd:attribute name="id" type="idDataType" use="optional"/> <xsd:attribute name="code" type="subfieldcodeDataType" use="required"/> </xsd:extension> </xsd:simpleContent> </xsd:complexType></pre>					

simpleType **controlDataType**

namespace	http://www.loc.gov/MARC21/slim					
type	restriction of xsd:string					
used by	complexType controlFieldType					
facets	whiteSpace preserve					
source	<pre><xsd:simpleType name="controlDataType" id="controlfield.st"> <xsd:restriction base="xsd:string"> <xsd:whiteSpace value="preserve"/> </xsd:restriction> </xsd:simpleType></pre>					

simpleType **controltagDataType**

namespace	http://www.loc.gov/MARC21/slim					
type	restriction of xsd:string					
used by	attribute controlFieldType/@tag					
facets	whiteSpace preserve pattern 00[1-9A-Za-z]{1}					
source	<pre><xsd:simpleType name="controltagDataType" id="controltag.st"> <xsd:restriction base="xsd:string"> <xsd:whiteSpace value="preserve"/> <xsd:pattern value="00[1-9A-Za-z]{1}"/> </xsd:restriction> </xsd:simpleType></pre>					

simpleType **idDataType**

namespace	http://www.loc.gov/MARC21/slim
type	xsd:ID
used by	attributes <u>collectionType/@id recordType/@id leaderFieldType/@id controlFieldType/@id dataFieldType/@id subfielddatafieldType/@id</u>
source	<pre><xsd:simpleType name="idDataType" id="id.st"> <xsd:restriction base="xsd:ID"/> </xsd:simpleType></pre>

simpleType **indicatorDataType**

namespace	http://www.loc.gov/MARC21/slim
type	restriction of xsd:string
used by	attributes <u>dataFieldType/@ind1 dataFieldType/@ind2</u>
facets	whiteSpace preserve pattern [dA-Za-z]{1}
source	<pre><xsd:simpleType name="indicatorDataType" id="ind.st"> <xsd:restriction base="xsd:string"> <xsd:whiteSpace value="preserve"/> <xsd:pattern value="[dA-Za-z]{1}"/> </xsd:restriction> </xsd:simpleType></pre>

simpleType **leaderDataType**

namespace	http://www.loc.gov/MARC21/slim
type	restriction of xsd:string
used by	complexType <u>leaderFieldType</u>
facets	whiteSpace preserve pattern [d]{5}[dA-Za-z]{1}[dA-Za-z]{1}[dA-Za-z]{3}(2)) (2) [d]{5}[dA-Za-z]{3}(4500)
source	<pre><xsd:simpleType name="leaderDataType" id="leader.st"> <xsd:restriction base="xsd:string"> <xsd:whiteSpace value="preserve"/> <xsd:pattern value="[d]{5}[dA-Za-z]{1}[dA-Za-z]{1}[dA-Za-z]{3}(2)) (2) [d]{5}[dA-Za-z]{3}(4500)"/> </xsd:restriction> </xsd:simpleType></pre>

simpleType **recordTypeType**

namespace	http://www.loc.gov/MARC21/slim
type	restriction of xsd:string
used by	attribute recordType/@type
facets	enumeration Bibliographic enumeration Authority enumeration Holdings enumeration Classification enumeration Community
source	<pre><xsd:simpleType name="recordTypeType" id="type.st"> <xsd:restriction base="xsd:string"> <xsd:enumeration value="Bibliographic"/> <xsd:enumeration value="Authority"/> <xsd:enumeration value="Holdings"/> <xsd:enumeration value="Classification"/> <xsd:enumeration value="Community"/> </xsd:restriction> </xsd:simpleType></pre>

simpleType **subfieldcodeDataType**

namespace	http://www.loc.gov/MARC21/slim
type	restriction of xsd:string
used by	attribute subfielddataType/@code
facets	whiteSpace preserve pattern [dA-Za-z!"#\$%&'()*+,-./:;<=>?{}_!~\[\]\{1}
source	<pre><xsd:simpleType name="subfieldcodeDataType" id="code.st"> <xsd:restriction base="xsd:string"> <xsd:whiteSpace value="preserve"/> <xsd:pattern value="[dA-Za-z!"#\$%&'()*+,-./:;<=>?{}_!~\[\]\{1}"/> </xsd:restriction> </xsd:simpleType></pre>

simpleType **subfieldDataType**

namespace	http://www.loc.gov/MARC21/slim
type	restriction of xsd:string
used by	complexType subfielddataType
facets	whiteSpace preserve
source	<pre><xsd:simpleType name="subfieldDataType" id="subfield.st"> <xsd:restriction base="xsd:string"> <xsd:whiteSpace value="preserve"/> </xsd:restriction> </xsd:simpleType></pre>

simpleType **tagDataType**

namespace	http://www.loc.gov/MARC21/slim
type	restriction of xsd:string
used by	attribute <u>dataFieldType/@tag</u>
facets	whiteSpace preserve pattern (0([1-9A-Z][1-9a-z]){1}[\dA-Z]{1}) ((([1-9A-Z][1-9a-z]){1}([\dA-Z][\da-z]){2})
source	<pre><xsd:simpleType name="tagDataType" id="tag.st"> <xsd:restriction base="xsd:string"> <xsd:whiteSpace value="preserve"/> <xsd:pattern value="(0([1-9A-Z][1-9a-z]){1}[\dA-Z]{1}) ((([1-9A-Z][1-9a-z]){1}([\dA-Z][\da-z]){2})"/> </xsd:restriction> </xsd:simpleType></pre>

attributeGroup **xml:xmlGlobalAttributes**

namespace	http://www.w3.org/XML/1998/namespace					
attributes	Name	Type	Use	Default	Fixed	Annotation
	x:lang	xsd:language				
	x:space	xsd:NCName		preserve		
	x:base	xsd:anyURI				
source	<pre><xsd:attributeGroup name="xmlGlobalAttributes"> <xsd:attribute ref="x:lang"/> <xsd:attribute ref="x:space"/> <xsd:attribute ref="x:base"/> </xsd:attributeGroup></pre>					