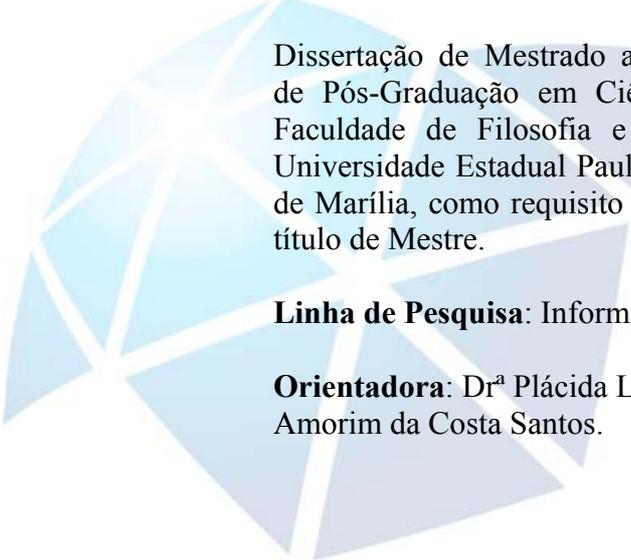


ADRIANA NASCIMENTO FLAMINO

**MARCXML: um padrão de descrição para recursos
informacionais em *Open Archives***

ADRIANA NASCIMENTO FLAMINO

**MARCXML: um padrão de descrição para recursos
informacionais em *Open Archives***



Dissertação de Mestrado apresentada ao Programa de Pós-Graduação em Ciência da Informação da Faculdade de Filosofia e Ciências – FFC – da Universidade Estadual Paulista – UNESP – Campus de Marília, como requisito parcial para obtenção do título de Mestre.

Linha de Pesquisa: Informação e Tecnologia

Orientadora: Dr^a Plácida Leopoldina Ventura Amorim da Costa Santos.

Marília – SP
2006

Flamino, Adriana Nascimento

F579m MARCXML: um padrão de descrição para recursos
informacionais em *Open Archives* / Adriana Nascimento Flamino.
– Marília: UNESP, 2006
164 f : il. ; 30 cm

Dissertação (Mestrado em Ciência da Informação) – Faculdade
de Filosofia e Ciências – Universidade Estadual Paulista, Marília,
2006.

1 Arquivos Abertos. 2 MARC 3. XML 4. MARCXML
5. Metadados 6. Comunicação Científica.

I Autor. II Título.

CDD 025.4

ADRIANA NASCIMENTO FLAMINO

MARCXML: um padrão de descrição para recursos informacionais em *Open Archives*

Dissertação de Mestrado apresentada ao Programa de Pós-Graduação em Ciência da Informação da Faculdade de Filosofia e Ciências – FFC – da Universidade Estadual Paulista – UNESP – Campus de Marília, como requisito parcial para obtenção do título de Mestre.

Linha de Pesquisa: Informação e Tecnologia

Orientadora: Dr^a Plácida Leopoldina Ventura Amorim da Costa Santos.

Data de defesa : 27 / 06 / 2006

BANCA EXAMINADORA:

Nome: Plácida Leopoldina Ventura Amorim da Costa Santos

Titulação: Prof^a. do Programa de Pós-Graduação em Ciência da Informação, Faculdade de Filosofia e Ciências, UNESP – Campus de Marília.

Nome: Silvana Aparecida Borsetti Gregório Vidotti

Titulação: Prof^a. do Programa de Pós-Graduação em Ciência da Informação, Faculdade de Filosofia e Ciências, UNESP – Campus de Marília.

Nome: José Fernando Modesto da Silva

Titulação: Prof^o do Departamento de Biblioteconomia e Documentação, Escola de Comunicação e Artes, USP – São Paulo.

Local: Universidade Estadual Paulista – UNESP
Faculdade de Filosofia e Ciências
Campus de Marília

À minha família, com amor.

Agradecimentos

A todos aqueles que contribuíram para o desenvolvimento desta pesquisa:

À Professora Plácida Leopoldina Ventura Amorim da Costa Santos;

À Professora Silvana Aparecida Borsetti Gregorio Vidotti;

À Coordenadoria de Aperfeiçoamento de Pessoal de Nível Superior (CAPES);

Aos Professores do Programa de Pós-Graduação em Ciência da Informação do campus de Marília-UNESP;

À Professora Mariângela Spotti Lopes Fujita;

À Professora Maria Eunice Quilici Gonzalez do Programa de Pós-Graduação em Filosofia do campus de Marília-UNESP;

Ao Professor José Fernando Modesto da Silva;

Aos colegas de Mestrado;

Às bibliotecárias do Campus da UNESP de Marília;

Às bibliotecárias e funcionários do Museu de Zoologia da USP (MZUSP);

A todos vocês, muito obrigada por terem colaborado com esta pesquisa, cada um com o seu jeito especial de ser, viver, criticar, elogiar e contribuir.

“A quem nos dirigimos ao escrever uma tese? Ao examinador? A todos os estudantes ou estudiosos que terão oportunidade de consultá-la depois? Ao vasto público dos não especializados? Devemos imaginá-la como um livro, a andar nas mãos de milhares de leitores, ou como uma comunicação erudita a uma academia científica? [...]. Digamos então que uma tese é um trabalho que, por razões ocasionais, se dirige ao examinador, mas presume que possa ser lida e consultada, de fato, por muitos outros, mesmo estudiosos não versados naquela disciplina” (Umberto Eco).

“Devemos aprender durante toda a vida, sem imaginar que a sabedoria vem com a velhice” (Platão).

“Só sabemos com exatidão quando sabemos pouco; com os conhecimentos vem a dúvida” (Goethe).

“Sem a livre concorrência de pensamentos não pode haver desenvolvimento científico. E sem liberdade de pensamento não pode haver uma livre concorrência de pensamentos científicos” (Karl Popper).

FLAMINO, A. N. **MARCXML**: um padrão de descrição para recursos informacionais em *Open Archives*. 2006. 164f. Dissertação (Mestrado em Ciência da Informação) – Faculdade de Filosofia e Ciências, Universidade Estadual Paulista – UNESP, Marília. 2006.

RESUMO

A comunicação científica vem sofrendo consideráveis alterações tanto no seu processo quanto na sua estrutura e filosofia. As iniciativas de arquivos abertos (*open archives*) e de acesso livre (*open access*) estão contribuindo significativamente para a desconstrução do modelo tradicional de disponibilização e acesso de comunicação científica e para a construção de um novo modelo desagregado e interoperável, mais justo e eficiente para disseminar os resultados de pesquisa e assim, o conhecimento gerado pelas comunidades científicas. Entretanto, devido aos avanços das tecnologias de informação e comunicação, não somente a estrutura e o fluxo da comunicação científica vem sofrendo consideráveis alterações, como também o próprio conceito e suporte dos documentos científicos. Isto tem gerado a necessidade do desenvolvimento de ferramentas para a otimização dos processos de organização, descrição, intercâmbio e recuperação de informações, além da preservação digital, dentre outros. Destaca-se que o Formato MARC tem permitido por décadas a descrição e o intercâmbio de registros bibliográficos e catalográficos às instituições, favorecendo o acesso aos conteúdos informacionais contidos em diversos acervos. No entanto, com o crescimento exponencial de informações e da geração de documentos (sobretudo digitais), têm-se exigido maior flexibilidade e interoperabilidade entre os diversos sistemas informacionais disponíveis. Neste cenário, a linguagem de marcação XML é apresentada como um dos desenvolvimentos atuais que tem como propósito facilitar e otimizar o gerenciamento, armazenamento e transmissão de conteúdos via Internet, sendo incorporada por diversos setores e áreas do conhecimento por sua facilidade de manuseio e flexibilidade operacional. Diante disso, realizou-se um estudo exploratório de análise teórica, identificando a adequação do formato MARCXML na construção de formas de representação descritiva para recursos informacionais em arquivos abertos, como um padrão de metadados complexo e flexível, que possibilitará a interoperabilidade entre sistemas de informação heterogêneos, além de suas vantagens e flexibilidades na transferência de registros bibliográficos e catalográficos e no acesso às informações. Como resultado desta pesquisa, considera-se que o MARCXML é um formato adequado para descrição de dados numa estrutura complexa. Conclui-se que a medida que aumenta a complexidade dos documentos nos repositórios e *open archives*, mais se justifica uma estrutura de metadados, como a do formato MARCXML, que suporte a descrição das especificidades dos recursos informacionais, uma vez que esta iniciativa não está e nem estará se restringindo a documentos científicos, mas se expandindo a outros tipos de recursos informacionais cada vez mais complexos e específicos, demandando também uma descrição apropriada para a especificidade das entidades bibliográficas.

Palavras-chave: Arquivos Abertos. MARC. XML. MARCXML. Metadados. Comunicação Científica.

FLAMINO, A. N. **MARCXML**: um padrão de descrição para recursos informacionais em *Open Archives*. 2006. 164f. Dissertação (Mestrado em Ciência da Informação) – Faculdade de Filosofia e Ciências, Universidade Estadual Paulista – UNESP, Marília. 2006.

ABSTRACT

The scientific communication is suffering considerable alterations so much in its process as in its structure and philosophy. The open archives and open access initiatives are contributing significantly for the undoing of the traditional model of scientific communication and for the construction of a new disaggregated model and with interoperability, fairer and efficient to disseminate the research results and like this, the knowledge generated by the scientific communities. However, due to the progresses of the information and communication technologies, not only the structure and the flow of the scientific communication is suffering considerable alterations, as well as the own concept and support of the scientific documents. This has been generating the need of the development of tools to optimize the organization, description, exchange and information retrieval processes, besides the digital preservation, among others. Highlight that the MARC format it has been allowing per decades the description and the exchange of bibliographical and cataloging registrations to the institutions, favoring the access to the contents informacionais contained in several collections. However, with the exponential growth of information and of the documents generation (above all digital), this has been demanding larger flexibility and interoperability among the several information systems available. In this scenery, the XML markup language is presented as one of the current developments that has as purpose to facilitate and to optimize the administration, storage and transmission of contents through Internet, it being incorporate for several sections and areas of the knowledge for the handling easiness and operational flexibility. Front to that, an exploratory study of theoretical analysis was accomplished, identifying the adaptation of the MARCXML format in the construction in ways of descriptive representation for information resources in open archives, as a complex and flexible standard of metadata, that will make possible the interoperability among information systems heterogeneous, besides the access to the information. As result of this research, It's considered that MARCXML is an appropriate format for description of data in a complex structure. It's ended that the measure that increases the complexity of the documents in the repositories and open archives, plus it's justified a structure of metadata, as the MARCXML format, that support the description of the specificities of the informational resources, once this initiative is not and nor it will be if restricting to scientific documents, but expanding the other types of informational resources more and more complex and specific, also demanding an appropriate description for the specificities of the bibliographical entities.

Keywords: Open Archives. MARC. XML. MARCXML. Metadata. Scientific Communication

LISTA DE ILUSTRAÇÕES

QUADRO 1: Funções da comunicação científica no sistema tradicional de periódicos acadêmicos	36
QUADRO 2: Funções da comunicação científica no novo modelo desagregado	36
FIGURA 1: Exemplo pedido/resposta	44
FIGURA 2: Esquema Funcional do OA mais atual	46
FIGURA 3: Modelo dos Arquivos Abertos: Características	47
FIGURA 4: Tipologia do documento científico	57
FIGURA 5: Estrutura hierárquica de um artigo científico	59
FIGURA 6: Exemplo de marcação de conteúdo de um texto científico de acordo com a estrutura hierárquica	62
FIGURA 7: Representação sugestiva da Ciência da Informação	68
FIGURA 8: Formato de entrada	85
FIGURA 9: Formato de intercâmbio	87
FIGURA 10: Exibição com etiqueta	87
FIGURA 11: Exibição formatada	88
FIGURA 12: Entidades do “Grupo 1” dos FRBR	100
FIGURA 13: Exemplos de documentos em XML	106
FIGURA 14: Exemplo de uma DTD interna	109
FIGURA 15: Exemplo de uso de uma DTD externa	110
FIGURA 16: Exemplo de um arquivo DTD	110
FIGURA 17: Vários estilos de apresentação a partir de um mesmo documento XML	112
FIGURA 18: Filosofia da XML	114
FIGURA 19: Proposed namespace	116
FIGURA 20: O ciclo do aprisionamento	117
FIGURA 21: Exemplo de um registro DTD XML para MARC.	123
FIGURA 22: Esquema MARCXML (simplificado)	126
FIGURA 23: Exemplo de um registro MARCXML	126
FIGURA 24: Exemplo de uma coleção (registros múltiplos)	127
FIGURA 25: Conversão MARC 21 (2709) de/para MARC 21 (XML)	131
FIGURA 26: Exemplo de um registro de livro em MODS	135
FIGURA 27: Exemplo de registro MADS	139

LISTA DE SIGLAS

AACR	Anglo-American Cataloguing Rules
ANSI	American National Standards Institute
BOAI	Budapest Open Access Initiative
CAPES	Coordenação de Aperfeiçoamento de Pessoal de nível Superior
CQL	Common Query Language
CSDGM	Content Standard for Digital Geospatial Metadata
CSS	Cascading Style Sheets
DC	Dublin Core
DTD	Document Type Definition
FAPESP	Fundação de Amparo à Pesquisa do Estado de São Paulo
FRANAR	Functional Requirements and Numbering for Authority Records
FRBR	Functional Requirements for Bibliographic Records
FTP	File Transfer Protocol
GILS	Government Information Locator Service
HTML	HyperText Markup Language
HTTP	Hyper Text Transfer Protocol
IBICT	Instituto Brasileiro de Informação em Ciência e Tecnologia
IFLA	International Federation Library Associations and Institutions
ISBD	International Standard Bibliographic Description
ISO	International Standardization Organization
JSCAACR	Joint Steering Committee for Revision of AACR
LSP	Linked Systems Project
LC	Library of Congress
MADS	Metadata Authority Description Schema
MARC	Machine Readable Cataloging
MARXML	Machine Readable Cataloging - Extensible Markup Language
MEDLINE	Stanford University's Lane Medical Library
METS	Metadata Encoding and Transmission Standard
MODS	Metadata Object Description Schema
NISO	National Information Standards Organization
NLM	National Library of Medicine
OAI	Open Archives Initiative

OAI-PMH	Open Archives Initiative-Protocol for Metadata Harvesting
OAIS	Open Archival Information System
OCLC	Online Computer Library Center
OPAC	Online Public Access Catalog
OSI	Open Society Institute
RDA	Resource Description and Access
RLG	Research Libraries Group
RTF	Rich Text Format
SGML	Standard Generalized Markup Language
SRW	Service Web of Search and Retrieval
SRU	Service URL of Search and Retrieval
TCP/IP	Transmission Control Protocol / Internet Protocol
UNESP	Universidade Estadual Paulista
Xlink	XML Linking Language
XML	Extensible Markup Language
Xpath	XML Path Language
Xpointer	XML Pointer Language
XSL	eXtensible Stylesheet Language
XSLFO	XSL Formating Object
XSLT	XSL Transformation
W3C	World Wide Web Consortium
WLN	Western Library Network
ZING	Z39.50 International: Next Generation

SUMÁRIO

1 INTRODUÇÃO	12
2 OPEN ARCHIVES: um novo modelo de publicação para a comunicação científica	22
2.1 Iniciativa de Arquivos Abertos (<i>Open Archives</i>)	29
2.2 Protocolo OAI-PMH	37
2.3 Repositórios Institucionais Digitais	47
2.4 O documento científico e as estruturas textuais no ambiente digital	51
3 INFORMAÇÃO, CIÊNCIA DA INFORMAÇÃO E FORMAS DE REPRESENTAÇÃO	64
3.1 Representação da Informação	69
3.1.1 Formas de Representação	72
4 MARC e XML	76
4.1 MARC	83
4.1.1 PROTOCOLO Z39.50	88
4.1.2 NORMA ISO 2709	93
4.1.3 AACR; FRBR e RDA	96
4.2 XML	103
5 MARCXML para a OAI	121
5.1 MODS	134
5.1.1 MADS e METS: Metadados associados	139
6 CONSIDERAÇÕES FINAIS	147
REFERÊNCIAS	153
BIBLIOGRAFIA CONSULTADA	162
LISTA DE LINKS	164

1 INTRODUÇÃO

“A verdadeira viagem do descobrimento não consiste em buscar novas paisagens mas novos olhares”.

Marcel Proust

Antes de iniciar vale salientar que, essa dissertação começou a dar frutos antes mesmo de ser finalizada. Partes dela foram apresentadas em simpósios, conferências e publicadas em anais de eventos ¹.

Nessa seção apresenta-se o universo desta pesquisa, abordando sua origem, justificativa, problema, hipótese, proposição, objetivos e metodologia, além de como está organizada a sua estrutura em seções.

A presente pesquisa intitulada **MARXML: um padrão de descrição para recursos informacionais em *Open Archives***, inserida na linha de pesquisa Informação e Tecnologia, do Programa de Pós-Graduação em Ciência da Informação do Campus de Marília–SP, financiada parcialmente pela Coordenação de Aperfeiçoamento de Pessoal de nível Superior – **CAPES**, é apresentada como Dissertação de Mestrado, um dos requisitos para a obtenção do grau de Mestre em Ciência da Informação, da Faculdade de Filosofia e Ciências da UNESP – Campus de Marília.

¹ FLAMINO, Adriana Nascimento; SANTOS, Plácida Leopoldina Ventura Amorim da Costa; FUJITA, Mariângela Spotti Lopes. Uma breve reflexão sobre documento, estruturas textuais e a xml nos repositórios institucionais digitais. In: FUJITA, Mariângela SPOTTI LOPES et al. (Org.). **A dimensão social da Biblioteca digital na organização e acesso ao conhecimento: aspectos teóricos e aplicados**. São Paulo: Departamento Técnico do SIBi/USP; IBICT, 2005. 2 v. p. 172-196. ISBN 857314032-1. Disponível em: <<http://bibliotecas-cruesp.usp.br/3sibd/docs/flamino194.pdf>>. Acesso em: 02 dez. 2005.

GARCIA, Rodrigo Moreira et al. Reflexões filosóficas sobre a informação no contexto cibernético e biológico. In: PAGNI, P. (Org.). **Universidade e contemporaneidade: produção do conhecimento e formação profissional – coletânea de textos do VI Simpósio em Filosofia e Ciência**. Marília: FFC/Marília/UNESP – Comissão Permanente de publicações, 2005. CD-ROM. ISBN 858673828-X.

FLAMINO, Adriana Nascimento; SANTOS, Plácida Leopoldina Ventura Amorim da Costa. Open Archives: um novo modelo para a comunicação científica. In: COSTA, Sely Maria de Sousa et al. (Ed.). **1ª Conferência Iberoamericana de Publicações Eletrônicas no Contexto da Comunicação Científica**, 2006, Brasília, 1ª CIPECC, p. 211-216. Disponível em: <<http://portal.cid.unb.br/CIPECCbr/papers.php>> Acesso em: 05 jun. 2006.

As inovações tecnológicas, principalmente aquelas ligadas à informação e à comunicação, a todo instante promovem mudanças que atingem o fazer das pessoas e os processos das organizações, além de dinamizar, cada vez mais, o fluxo de informações. Segundo Carvalho (2003, p. 76),

As informações geradas atualmente estão, cada vez mais, sendo armazenadas no formato digital. As vantagens desta forma de armazenamento de informação são muitas, destacando-se, entre elas, a flexibilidade oferecida para a sua recuperação e a possibilidade de armazenamento e veiculação em diferentes tipos de mídia.

Ao empregar as tecnologias de informação e comunicação, as instituições acadêmicas estão criando novos serviços e aperfeiçoando os já existentes. Essas instituições, como parte integrante da Sociedade da Informação, acompanham de perto os processos de desenvolvimento econômico, social e tecnológico, e sendo formadoras de profissionais e produtoras de pesquisas, têm como algumas de suas principais funções a responsabilidade de assegurar a preservação intelectual dos seus membros e a disseminação deste capital intelectual através de suas unidades de informação, utilizando-se para isso as tecnologias de informação e comunicação.

As bibliotecas, centros de documentação, museus, arquivos entre outros, considerados unidades de informação, têm sido ao longo de sua história organizadoras e transmissoras eficientes de conhecimento. Hoje a revolução digital faz com que o modo em que tratamos a informação, no aspecto documental, adquira uma posição central, pois, determina não somente um novo modo de transmissão da informação, como também leva consigo uma proposta de novos objetivos, novas funções e novos modelos organizacionais (AGUSTÍN LACRUZ, 1998).

Nesse cenário, a tendência para os próximos anos é de que a iniciativa dos arquivos abertos ganhe maior visibilidade, uma vez que sua promoção e, conseqüentemente,

sua utilização será vista como um dos indicadores significantes da qualidade acadêmica de uma instituição, visto que, os repositórios institucionais digitais (uma das modalidades de arquivos abertos) têm, como alguns dos seus objetivos, preservar a produção intelectual dos seus membros, a rápida comunicação científica entre os pares e a credibilidade da instituição.

Atualmente é visível a necessidade e o uso cada vez mais constante e intenso de recursos abertos, produtos de código-fonte aberto ou sistemas “*open source*”, nas mais diversas áreas. Essa inovação tecnológica é hoje considerada, por muitos, uma das ferramentas mais importantes e indispensáveis na sociedade da informação. E o momento exige cada vez mais o uso dessas tecnologias de fontes abertas, de baixo custo e de licença pública, uma vez que elas dispensam custos com licença reduzindo muito mais os investimentos dos usuários dessas tecnologias, além de contribuir para o futuro da preservação informacional.

Tais inovações vêm ao encontro da filosofia biblioteconômica e dos objetivos da Ciência da Informação: criar mecanismos para que as informações disponíveis estejam acessíveis e que um maior número de pessoas tenham acesso, no momento certo, na hora certa e com o menor custo possível, de preferência a custo zero. Adicionalmente, as instituições ao utilizarem produtos *open source* se livram do aprisionamento tecnológico/comercial. Item imprescindível na atual economia da informação.

A sociedade atual, a Sociedade da Informação, está agora sob um novo paradigma, o do acesso, em que “é mais interessante ter condições de acessar a informação, o conteúdo do documento, do que ter o próprio documento (posse) em mãos, uma vez que isto se torna desnecessário frente às tecnologias digitais e virtuais que rompem barreiras geográficas” (DZIEKANIAK, 2004, p. 44).

Os grandes debates, discussões e iniciativas na atualidade giram em torno do acesso às informações, ou mais especificamente, ao acesso livre a essas informações. Para Rodrigues (2004, p.25),

o debate e as iniciativas relacionadas com o acesso à literatura científica tem crescido de forma significativa nos últimos anos. No quadro desse debate, a aspiração e exigência de Acesso Livre ao conhecimento produzido pelos investigadores e acadêmicos tem conquistado cada vez mais defensores e adeptos, dentro e fora do mundo universitário.

De uma forma simples, o acesso livre significa a disponibilização livre na Internet de literatura de caráter acadêmico ou científico, permitindo a qualquer pessoa ler, copiar, distribuir, imprimir, pesquisar ou referenciar o texto integral dos documentos (RODRIGUES, 2004). Para que o acesso livre seja possível é preciso utilizar tecnologias de fontes abertas como, softwares livres, protocolos, padrões e outros, também livres.

O surgimento do movimento de acesso livre se deu devido às crises no tradicional sistema de comunicação científica causadas, como por exemplo, pelo aumento exponencial dos títulos de periódicos e a diminuição significativa das assinaturas destas revistas científicas, devido ao expoente custo em obtê-las. Os profissionais da informação, conscientes da crise designada “crise dos periódicos” e das graves conseqüências que as limitações ao acesso à literatura produziam ao próprio sistema científico e, ao mesmo tempo, a generalização da utilização da Internet e da Web, acompanhada por uma maior compreensão das suas potencialidades e aplicações na publicação científica, contribuíram de forma significativa para o surgimento de diversas iniciativas que estão na base do atual movimento do Acesso Livre como, OAI, BOAI, *open sources*, arquivos abertos (*open archives*), repositórios institucionais dentre outros (RODRIGUES, 2004).

Os repositórios institucionais como sistemas abertos de informações que servem para armazenar, preservar e difundir a produção intelectual de uma ou várias instituições universitárias, podem ser criados e mantidos de forma individualizada ou cooperativa, utilizando-se de uma das múltiplas plataformas de softwares livres, padrões e protocolos também livres, atualmente disponíveis (RODRIGUES, 2004).

É evidente que o ambiente informacional evoluiu, assim como os suportes e os conteúdos informacionais. A tendência atual é de que o bibliotecário trate de informação cada vez mais em meio digital, em acervos digitais e de acesso ao conteúdo completo dos documentos. De acordo com Dziekaniak (2004, p. 52) “a produção da informação foi elevada exponencialmente e sua forma de disseminação passou a ser o espaço virtual”. Atualmente, textos digitais completos estão disponíveis, e milhares se não milhões deles estão a caminho. A necessidade de descrições de alta qualidade é visível.

Os profissionais da Ciência da Informação, da Tecnologia da Informação e inclusive Lingüistas, além de outros profissionais, estão reunidos em muitos projetos na busca incessante e extremamente urgente por métodos satisfatórios no tratamento de informações distribuídas pela Internet (DZIEKANIAK, 2004).

A sociedade da informação exige, cada vez mais, profissionais da informação que tenham conhecimento no tratamento, armazenamento e recuperação de novas mídias e acervos: os digitais (DZIEKANIAK, 2004).

Ciente de toda essa realidade, possibilidades e necessidades vigentes, percebe-se que com os avanços das tecnologias de informação e comunicação, o crescente uso da Internet, o aumento dos bancos de dados, dos Arquivos Abertos (*open archives*), dos repositórios institucionais digitais, entre outros, geraram a necessidade do desenvolvimento de ferramentas para a otimização dos processos de organização, descrição, recuperação, intercâmbio de informações, preservação digital, dentre outros.

Neste sentido, destacamos que o formato de intercâmbio MARC tem permitido às instituições o intercâmbio de dados bibliográficos e catalográficos por décadas, favorecendo o acesso aos conteúdos informacionais contidos em diversos acervos. Mas, o crescimento exponencial de informações e da geração de documentos, tem exigido uma maior flexibilidade e interoperabilidade entre os sistemas informacionais disponíveis.

Nesse contexto, a Ciência da Computação procura, por meio do desenvolvimento constante de tecnologias, facilitar o gerenciamento, o armazenamento e a transmissão de dados de modo digital e, de forma especial, criar mecanismos facilitadores para a disponibilização e acesso às informações na *World Wide Web*, via Internet, utilizando, neste momento, a linguagem XML que tem sido incorporada por diversos setores e áreas do conhecimento por sua facilidade de manuseio e flexibilidade operacional.

Buscando adequar o MARC à nova filosofia da comunicação científica baseada no livre acesso, visto que ele é uma estrutura de representação bibliográfica completa que atende a necessidade de padronização de metadados e facilita a interoperabilidade entre sistemas de informação devido a sua arquitetura organizacional, surgiram várias iniciativas de adaptação do MARC para o ambiente da XML, pois somente a estrutura MARC não efetiva a interoperabilidade no atual cenário digital. É preciso outros serviços acoplados a esta estrutura para garantir e efetivar a interoperabilidade entre sistemas de informação heterogêneos.

Diante disso, buscou-se resposta para a pergunta: **O formato de intercâmbio de dados bibliográficos e catalográficos MARCXML apresenta-se como um padrão adequado para a representação descritiva dos recursos informacionais nos *Open Archives*?**

Visto que, de acordo com a literatura estudada, o MARC é uma estrutura de descrição documentária legível por máquina e a XML é uma linguagem computacional de marcação de dados, portanto, coisas diferentes mas que se complementam no sentido de que,

o formato MARC serve como um recipiente no qual a informação descritiva é armazenada, enquanto que a XML proporciona, com inúmeras vantagens, o transporte desta informação em ambiente digital, acreditou-se como hipótese que: **Com a implantação efetiva do formato MARCXML como uma estrutura de representação de recursos informacionais da área da Ciência da Informação, as questões do intercâmbio de dados e da interoperabilidade entre sistemas de informação, em especial na implantação dos Arquivos Abertos, ficariam mais eficientes considerando-se que o MARC, como um padrão de metadados completo e flexível, garante a qualidade do registro enquanto que a XML garante a interoperabilidade entre os sistemas de informação heterogêneos, possibilitando, sobretudo, a otimização dos processos de recuperação da informação.**

O interesse por esse estudo partiu da pesquisa de Iniciação Científica intitulada MARC21 e XML como ferramentas para a consolidação da Catalogação Cooperativa Automatizada: uma revisão de literatura ², subsidiada pela Fundação de Amparo à Pesquisa do Estado de São Paulo – **FAPESP**. Ao término da referida pesquisa pode-se perceber a importância e a necessidade de se estudar o formato de intercâmbio bibliográfico e catalográfico MARC em XML uma vez que já existem grandes empreendimentos sendo feitos neste sentido por instituições conceituadas como *Library of Congress – LC (MARCXML)*, *Stanford University's Lane Medical Library – MEDLINE (XMLMARC)*, entre outras, ficando demonstrado desta forma, sua importância como um padrão de metadados complexo ³ e flexível para a construção de formas de representação para os sistemas informacionais. No entanto, a título de delimitação para estudo, o presente trabalho se restringiu a estudar somente o padrão para metadados MARCXML, desenvolvido pela LC, por esta ser a

² O relatório final de bolsa de Iniciação Científica, processo nº 02/12156-0 foi aprovado pela FAPESP em 15/03/2004 e o mesmo resultou em um capítulo de livro: SANTOS, P. L. V. A. da C.; FLAMINO, A. N. MARC21 e XML como ferramentas para a consolidação da Catalogação Cooperativa Automatizada: uma revisão de literatura. In: VIDOTTI, S. Ap. B. Gregório (Coord). **Tecnologia e conteúdos informacionais: abordagens teóricas e práticas**. São Paulo: Polis, 2004, p. 114-138.

³ Complexo num sentido positivo em que abrange ou encerra muitos elementos ou partes.

instituição mantenedora do esquema a partir de seu escritório de desenvolvimento de redes e padrões MARC.

Justifica-se esta pesquisa por se tratar de um tema atual e necessário para a área da Ciência da Informação e em especial para a Biblioteconomia no que diz respeito à avaliação das formas de representação para o tratamento adequado das informações registradas e o seu livre acesso. E também, pela necessidade de estudos como esse na referida área, inclusive para que os profissionais da informação, principalmente o bibliotecário, tenham maior conhecimento e domínio sobre temas atuais que norteiam a sua profissão como, formatos de descrição para intercâmbio de metadados, protocolos, arquivos abertos, repositórios institucionais digitais, interoperabilidade entre sistemas heterogêneos de informação, preservação digital, desenvolvimento de coleções digitais e outros procedimentos envolvidos nos processos e fluxos da informação.

Inclusive, ao relacionar temas que deveriam merecer a atenção dos pesquisadores da Ciência da Informação nos próximos anos, Robredo (2003) inclui os estudos de portabilidade e compatibilidade de sistemas e bases de dados, por exemplo conversão MARC em linguagens de marcação da família dos HTML dinâmicos, como a XML, além de estudos e aplicações de novas formas de preservação dos registros da informação e do conhecimento, o que demonstra a pertinência do presente estudo.

Assim, propõe-se com esta pesquisa analisar a adequação do padrão de metadados MARCXML para a descrição de recursos informacionais em Arquivos Abertos, uma vez que, as estruturas numéricas do MARC (e a sua complexidade e completude), conhecidas universalmente, juntamente com a linguagem XML, possibilitariam uma interoperabilidade mais efetiva entre os sistemas existentes, alcançando, assim, maior cooperação entre as instituições.

O objetivo geral ou final desta pesquisa é divulgar à comunidade de profissionais da informação a adequação do formato MARCXML para a descrição de recursos informacionais nos Arquivos Abertos como um padrão de metadados completo e flexível que possibilitará a interoperabilidade entre sistemas de informação heterogêneos. Como objetivos específicos, realizar uma análise teórica sobre o tema em questão, identificando as vantagens e flexibilidades do formato MARCXML na representação de registros bibliográficos e catalográficos e no acesso e recuperação das informações.

Para a elaboração do referencial teórico utilizou-se de fontes bibliográficas primárias, secundárias e terciárias. Compreende-se por fontes bibliográficas primárias, obras produzidas com a interferência direta do autor da pesquisa como, livros, teses, dissertações, relatórios técnicos, normas técnicas, artigos científicos e outros. Como fontes secundárias, documentos produzidos a partir de documentos primários e são representadas por exemplo pelas enciclopédias, tratados, manuais, dicionários e outros. Como fontes terciárias documentos que têm por função guiar o usuário para as fontes primárias e secundárias como, bibliografias, catálogos coletivos, guias de literatura, diretórios, índices e outros (MUELLER, 2000).

O levantamento bibliográfico dos documentos primários, secundários e terciários para análise foi feito nas seguintes fontes bibliográficas da área de Ciência da Informação: Livros, periódicos, anais e relatórios; Bases de dados nacionais e internacionais, textuais e referenciais, disponíveis em: Portal de Periódicos da CAPES, Portal Bibliotecas UNESP, Portal da Pesquisa, Unibibli WEB, ERL, Biblioteca Eletrônica do CRUESP, dentre outras, Arquivos Abertos, Repositórios Institucionais e Internet. Também buscou-se levantar informações nos eventos da área realizados durante o período da pesquisa. A pesquisa foi realizada através de leituras, análises e fichamento dos documentos selecionados, nos idiomas inglês, espanhol e português.

Com o intuito de alcançar os objetivos desta pesquisa, estruturou-se a dissertação em 5 seções, incluindo esta que aborda a origem, justificativa, problema, hipótese, proposição, objetivos e metodologia da referida pesquisa. As seguintes seções estão assim organizadas:

2 OPEN ARCHIVES: um novo modelo de publicação para a comunicação científica: apresenta um breve histórico, conceitos, a filosofia dos novos meios de comunicação científica, a Iniciativa de Arquivos Abertos, Protocolo para coleta de metadados, provedores de dados e de serviços, Repositórios Institucionais Digitais, o documento científico e as estruturas textuais.

3 INFORMAÇÃO, FORMAS DE REPRESENTAÇÃO E A CIÊNCIA DA INFORMAÇÃO: faz uma breve abordagem a respeito da polêmica questão sobre os conceitos de informação, formas de representação e a Ciência da Informação.

4 MARC e XML: Aborda o conceito de metadados, o formato MARC, a linguagem de marcação XML e as ferramentas tecnológicas associadas.

5 MARCXML para a OAI: apresenta o histórico e os conceitos dos formatos MARCXML, MODS, MADS e METS.

6 CONSIDERAÇÕES FINAIS: relata os resultados alcançados por esta pesquisa.

A seguir apresenta-se a seção 2: *Open Archives*: um novo modelo de publicação para a comunicação científica.

2 *OPEN ARCHIVES*: um novo modelo de publicação para a comunicação científica

“Conhecimento não divulgado é como palco com cortina fechada; Como rádio sem som; Como um quadro num quarto escuro; Como céu sem estrelas”.

Livis Dzelve

Aborda-se nesta seção, o novo cenário da comunicação científica mediante as inovações tecnológicas e a filosofia de acesso livre.

A ciência se fundamenta “no consenso dos cientistas, e os autores se destacam pela frequência com que são lidos e citados, portanto procuram ampla divulgação para seus trabalhos” (CAMPELLO, 2000, p.26). Diante disto, na ciência, é fundamental a comunicação, a circulação e o intercâmbio de idéias, e a essa troca de informações entre cientistas denomina-se comunicação científica (RODRIGUES; LIMA; GARCIA, 1998). De acordo com Garvey (1979) citado por Mueller (1995, p. 64), a comunicação científica “inclui todas as atividades associadas com a produção, disseminação e uso da informação, desde a hora em que o cientista teve a idéia da pesquisa até o momento em que os resultados de seu trabalho são aceitos como parte integrante do conhecimento científico”. Para Meadows (1999, p. 161) “a realização de pesquisas e a comunicação de seus resultados são atividades inseparáveis”. Portanto, existem diversos métodos pelos quais a comunidade científica pode tomar conhecimento de pesquisas e a importância concernente a um método pode mudar no decorrer do tempo (MEADOWS, 1999).

A publicação científica “assume variadas formas e utiliza diversos canais. Quando registrada em veículos formais como, livros, periódicos ou meios eletrônicos, produz a literatura científica” (MUELLER, 1995; RODRIGUES; LIMA; GARCIA, 1998, p. 148).

A invenção da imprensa possibilitou o desenvolvimento de uma comunidade científica que se apóia em processos de produção e disseminação de conhecimentos através de

canais formais e informais. Ou seja, entre os cientistas e seu público estão os canais pelos quais eles se comunicam. Estes podem ser formais, como as páginas impressas e as redes de computadores ou informais, como a transmissão de informação oral. Este paradigma sofreu poucas mudanças no longo período que se seguiu à invenção da imprensa. No entanto, os livros e periódicos como canais de difusão da informação científica, passaram por mudanças notáveis ao longo do tempo e em ritmos cada vez mais intensos, com a disponibilização dos documentos em suportes digitais (MEADOWS, 1999; ARAUJO; SOUZA, 2004).

Neste sentido, as mudanças sempre estiveram presentes na comunicação científica. Apenas aconteciam com relativa lentidão quando os principais veículos de comunicação eram a palavra escrita e impressa. Os novos meios de comunicação criados pelas tecnologias de informação e comunicação aceleraram esse ritmo e agora a comunicação científica passa por um período de rápida evolução. No entanto, o problema básico continua sendo como um dado veículo pode ser melhor utilizado para atender as necessidades de comunicação de uma determinada comunidade científica (MEADOWS, 1999).

No universo da comunicação científica, Vannevar Bush (1945), um dos expoentes da era dos computadores, apontava o tempo perdido entre a redação do trabalho científico, a sua publicação e posterior assimilação e avaliação pelos pares, exemplificando com as leis da genética de Mendel, perdidas por uma geração, porque sua publicação não alcançou os que seriam capazes de entendê-las e de estendê-las (SENA, 2000).

Foi Mendel quem enunciou as leis da hereditariedade. Seu artigo publicado em 1865, no qual explicava sua experiência de cruzamento de plantas, permaneceu ignorado por 35 anos. Darwin e outros naturalistas repetiram uma série de experiências que Mendel já tinha anunciado, mas que eles desconheciam. Somente em 1900 um botânico totalmente por acaso, folheando a própria revista, chegou a ver o artigo de Mendel. Como a filosofia já nos disse, os povos que desconhecem sua história estão condenados a repeti-la (SMIT, 1987).

A informação científica é um dos elementos fundamentais para o desenvolvimento científico-tecnológico de uma sociedade. Consiste de um “[...] processo contínuo em que a informação científica contribui para o desenvolvimento científico, e este, por sua vez, gera novos conteúdos realimentando todo o processo” (IBICT, 2006a).

No entanto,

Os cientistas – autores e revisores dos artigos – em sua maioria são mantidos por instituições que financiam as pesquisas. Essas instituições são as mesmas que compram os periódicos que publicam os trabalhos produzidos por seus próprios pesquisadores, e muitas vezes não conseguem adquiri-los devido aos preços elevados (CORREA DA CRUZ et al, 2003, p. 49).

Essa crise no tradicional sistema de publicação científica em que, artigos, trabalhos científicos, resultados de pesquisas etc. são publicados em periódicos especializados e o acesso à informação ocorre por intermédio de assinaturas das publicações por bibliotecas ou por pesquisadores interessados, tem exigido a busca por alternativas de divulgação e acesso à comunicação científica e as tecnologias de informação e comunicação têm influenciado na definição dessas opções.

Foram necessárias décadas de avanços tecnológicos até encontrar condições favoráveis para a questão do acesso à literatura científica com o surgimento e consolidação dos *open archives*, arquivos on-line de acesso público, também definidos como diretórios existentes em um computador que podem estar abertos para o acesso via FTP ou HTTP, armazenando uma coleção de artigos ou de metadados de documentos armazenados em outros locais (SENA, 2000).

A *Open Archives Initiative*¹ (OAI, 1999) é um exemplo disto. “Esta iniciativa estabelece, além de padrões de interoperabilidade, alguns princípios e ideais como, o uso de

¹ www.openarchives.org/

open source e o *open access* à informação. Surge, então, o paradigma do acesso livre à informação” (IBICT, 2006a).

Acesso livre significa a disponibilização livre na Internet de literatura de caráter acadêmico ou científico (em particular os artigos de revistas científicas), permitindo a qualquer utilizador ler, descarregar (*download*), copiar, distribuir, imprimir, pesquisar ou referenciar o texto integral dos documentos (IBICT, 2006b; RODRIGUES, 2004).

O Acesso Livre aplica-se primariamente à versão final (após peer-review) de artigos de revistas (postprints), mas também inclui versões não revistas (preprints) que os investigadores queiram divulgar para alertar sobre novos resultados, estabelecer a primazia, etc. O Acesso Livre não se aplica a livros sobre os quais os autores pretendam obter receitas, ou textos não acadêmicos, como notícias ou ficção. O Acesso Livre pode aplicar-se naturalmente a todos os trabalhos dos quais os autores não esperem pagamento. Esses textos podem incluir monografias especializadas em determinado tópico, comunicações a conferências e congressos, teses e dissertações, relatórios técnicos, working papers, relatórios governamentais, etc (RODRIGUES, 2004, slide 5).

As origens do Acesso Livre à informação acadêmica e técnico-científica estão representadas na consciência das limitações e contradições do tradicional sistema de comunicação científica, combinada às atuais possibilidades tecnológicas. Desta forma, o movimento de acesso livre objetiva maximizar o impacto das pesquisas, possibilitando o acesso aos seus resultados, além de reassumir o controle do sistema de comunicação da ciência (RODRIGUES, 2004).

Diversas foram as iniciativas anteriores, mas foi em 1998 que se consolidou o atual movimento do ‘*Open Access*’.

A criação da Scholarly Publishing and Academic Resources Coalition (SPARC) pela Association of Research Libraries (ARL), o lançamento do Fórum da *American Scientist* e o início de uma vaga de “Declarações de Independência” são alguns dos acontecimentos que assinalam a afirmação deste movimento em 1998 (RODRIGUES, 2004, p. 28).

Em 1999, foi lançada a *Open Archives Initiative* (OAI), com o objetivo de criar uma plataforma simples para permitir a interoperabilidade e a busca de publicações científicas de diversas disciplinas.

Em dezembro de 2001 realizou-se uma reunião em Budapeste, promovida pelo *Open Society Institute* (OSI) para discutir a questão do acesso à literatura científica, da qual resultou um dos mais importantes documentos e iniciativas do movimento do Acesso Livre, conhecida como **Budapest Open Access Initiative** (BOAI). A BOAI estabeleceu o significado e âmbito do Acesso Livre. É “[...] uma declaração de princípios, uma definição de estratégia e uma afirmação de empenhamento. É por isso um dos mais importantes documentos do movimento do Acesso Livre” (IBICT, 2006b).

Em 2003, firma-se mais o movimento de acesso livre pois “[...] assumem ainda maior relevância os diversos documentos, iniciativas e tomadas de posição de sociedades científicas e organizações governamentais sobre a problemática do acesso à informação científica e técnica” (RODRIGUES, 2004, p. 29). No dia 11 de Abril de 2003 reuniram-se na sede do Howard Hughes Medical Institute diversas personalidades (cientistas, editores, bibliotecários etc.) ligadas à informação na área biomédica com o objetivo de estimular a discussão sobre a forma de concretizar, tão rapidamente quanto possível, ações para atingir o objetivo de assegurar o acesso livre à literatura científica.

A Declaração **Bethesda Statement on Open Access Publishing** contém uma definição de Acesso Livre, bem como conclusões e recomendações de grupos de trabalho sobre Organismos e instituições financiadoras de pesquisa e desenvolvimento, Bibliotecas e editores, Sociedades científicas e investigadores (IBICT, 2006b). Em 22 de outubro de 2003, foi assinado por representantes de várias e importantes instituições científicas europeias, entre as quais a Sociedade *Max-Planck* (Alemanha) e o *Centre National de la Recherche Scientifique* (França), a **Declaração de Berlim sobre o Acesso Livre ao Conhecimento nas**

Ciências e Humanidades, apoiando o *Open Access* e o depósito em arquivos de acesso livre, e afirmando que irão encorajar os seus investigadores e bolsistas a depositar os seus trabalhos em pelo menos um repositório (IBICT, 2006b).

Estas últimas são as três declarações fundamentais (também conhecidas pelas 3 B's; Budapeste, Bethesda e Berlim) relacionadas com o acesso livre (IBICT, 2006b). Em dezembro de 2003 a **Declaração de Princípios e Plano de Ação da Reunião de Cúpula Mundial sobre a Sociedade da Informação**, documento aprovado em Genebra, apoia explicitamente o acesso livre à informação científica. No Plano de Ação recomenda-se o encorajamento de iniciativas para facilitar o acesso, incluindo o acesso livre e economicamente viável às revistas e aos livros, e a arquivos abertos de informação científica (IBICT, 2006b).

No dia 30 de Janeiro de 2004, em Paris, os ministros da Ciência e Tecnologia de 34 estados membros da OCDE tornaram pública uma declaração (***Declaration on Access to Research Data from Public Funding/OCDE***) reconhecendo que o acesso livre maximiza o valor do investimento público na ciência e tecnologia e incentivando os governos a procurarem regulamentações e políticas que promovam o intercâmbio internacional dos resultados da investigação científica de forma livre (IBICT, 2006b). No dia 11 de Outubro de 2004, na *Royal Society of Edinburgh*, o *Scottish Science Information Strategy Working Group* apresentou a sua Declaração de Acesso Livre (***Scottish Declaration on Open Access***). Este grupo de trabalho tem estudado assuntos ligados à implementação de repositórios institucionais em universidades escocesas e em outros centros de investigação.

No Brasil, em setembro de 2005 o IBICT² (Instituto Brasileiro de Informação em Ciência e Tecnologia) apresenta o **Manifesto Brasileiro de Apoio ao Acesso Livre à Informação Científica** e, em 2 de dezembro de 2005 pesquisadores, professores,

² <http://www.ibict.br>. Acesso em: 3 fev. 2006.

bibliotecários e representantes de organizações da sociedade civil divulgaram a **Carta de São Paulo**, documento em defesa da disponibilização e do acesso público por meios digitais à literatura científica derivada de pesquisas financiadas direta ou indiretamente com recursos públicos.

Em 17 de fevereiro de 2006, é publicada no Diário Oficial a Portaria de nº 13 de 15 de Fevereiro de 2006 a qual a Fundação Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – **CAPES** –

[...] considerando as manifestações do Conselho Técnico-Científico verificadas no ano de 2005, indicando que a produção científica discente é um relevante indicador da qualidade dos programas de mestrado e doutorado, **não aferível apenas através da publicação seletiva nos periódicos especializados** [...] (BRASIL, 2006, grifo nosso),

estabelece a instalação e manutenção de ‘arquivos digitais’ para o acesso livre de teses e dissertações como um dos critérios para o acompanhamento, avaliação e reconhecimento dos programas de Pós-Graduação. E afirma que, “o financiamento de trabalho com verba pública, sob forma de bolsa de estudo ou auxílio de qualquer natureza concedido ao Programa, induz à obrigação do mestre ou doutor apresentá-lo à sociedade que custeou a realização [...]” (BRASIL, 2006).

É o momento do *open access*. Acesso livre como direito, desafio e responsabilidade. Os argumentos a favor do *open access* poderão ser desde os pontos éticos mais elevados (o direito humano sobre o conhecimento) até a vaidade humana (quanto mais as pessoas lerem mais conhecido o autor ficará, será mencionado, citado mais vezes do que quando divulgado em meios tradicionais como periódicos científicos impressos e outros) .

Como visto, o benefício público que o movimento de ‘*Open Access*’ por meio de bibliotecas digitais de Teses e Dissertações, periódicos de acesso livre, repositórios

institucionais, entre outros *Open Archives*, é a possibilidade de divulgação dos resultados de pesquisas através da ‘distribuição eletrônica’, via Internet, com a eliminação de barreiras de acesso à literatura científica, contribuindo deste modo, para o desenvolvimento das pesquisas, para o enriquecimento da educação e para o avanço da ciência.

Atualmente vive-se um período de transição, e os meios de comunicação dominantes estão mudando, da forma impressa para a forma digital. O que vem comprovando isso são os novos modelos desagregados e interoperáveis de comunicação científica como os periódicos de acesso livre, repositórios institucionais, entre outros, frutos das iniciativas de *open archives* e *open access*.

2.1 Iniciativa de Arquivos Abertos (*Open Archives*)

Os arquivos abertos ou repositórios de documentos eletrônicos estão surgindo como alternativa para a rápida comunicação de resultados de pesquisas entre os pares. Estes arquivos trazem mudanças significativas nos modelos tradicionais de comunicação científica.

McGarry, já em 1999 dizia que,

A formação de redes é uma das mais importantes questões com que hoje se defronta a comunidade bibliotecária e de informação. A convergência da tecnologia da informática com as comunicações afeta a criação, gestão e uso da informação de modo inédito desde a introdução da imprensa de tipos móveis (McGARRY 1999, p. 122).

E ainda que, “o recurso das redes está se tornando um meio de publicação formal; isto é verdade principalmente na área das pesquisas acadêmicas e industriais” (McGARRY, 1999, p. 124).

Os arquivos abertos são repositórios que podem armazenar versões eletrônicas preliminares de documentos científicos, os chamados *preprints*, como também os artigos aceitos para publicação pelo processo de revisão tradicional pelos pares, os chamados *reprints*, além de possibilitar anotações, atualizações e comentários por parte dos autores e pelos pares respectivamente.

“Antes de haver arbitragem e publicação, eles são *preprints*, propriedade do autor. Depois da arbitragem, aceitação e publicação, eles são (em virtude do acordo de copyright), usualmente, propriedade do editor e chamados *reprints*” (HARNARD, 1994 citado por SENA, 2000, p. 72).

A Iniciativa de Arquivos Abertos freqüentemente usa termos técnicos muito específicos, e que de certo modo às vezes deixa desconcertados aqueles que estão fora da comunidade de OAI. Mas na realidade a especificidade de certos termos usados dentro desta comunidade ajuda a explicar a arquitetura de arquivos abertos. ‘Pré-print’ é um dos conceitos centrais surgidos da experiência de Los Alamos, significando um documento não revisado pelos pares disponível em formato eletrônico antes da formal publicação impressa revisada pelos pares. Por extensão, um ‘arquivo Pré-print’ é essencialmente um arquivo de documentos não revisados pelos pares em formato eletrônico, disponibilizados antes da formal publicação revisada pelos pares (HUNTER; GUY, 2004).

“Definições do que deve ser arquivado ou disponível, com esta tecnologia no entanto mudaram, desde a criação do Arquivo de Los Alamos em 1991. A OAI fala agora sobre ‘eprints’, por exemplo, com um conceito totalmente diferente” (HUNTER; GUY, 2004, p. 2, tradução nossa).

Desde então há outras categorias de documentos que podem ser arquivados e disponibilizados de um modo semelhante, a terminologia foi então ampliada para acomodá-los. Conseqüência disto são as categorias de documentos ‘postprints’ e ‘eprints’. Postprints

(como podemos presumir) são documentos os quais os metadados de coleta estão disponíveis depois do processo de revisão pelos pares. Coletivamente, eles são todos ‘eprints’: ‘eprint’ é definido na prática como o termo coletivo para todos os outros itens imprimíveis (HUNTER; GUY, 2004).

O termo **arquivo** neste contexto das publicações eletrônicas não está relacionado com o conceito tradicional de arquivo para preservação e conservação e sim como um depósito para armazenar documentos científicos de texto completo na forma digital (BARRUECO; COLL, 2003). “E o termo **aberto** se refere ao ponto de vista da arquitetura do sistema. Trata-se de definir interfaces que facilitem a disponibilidade de conteúdos procedentes de uma variedade de provedores. Abertura tão pouco significa gratuidade ou acesso ilimitado a essas informações” (BARRUECO, COLL, 2003, p. 2, tradução nossa).

O mais conhecido destes arquivos é o **arXiv.org**³ criado por Paul Ginsparg em 1991 e mantido pelo Laboratório Nacional de Los Alamos, Novo México, que se tornou um repositório global de artigos nas áreas da física, matemática, ciência da computação e ciências não-lineares (SENA, 2000). “Com o aumento desses tipos de repositórios disponíveis na Web, sentiu-se a necessidade de criar uma estrutura técnica e estabelecer padrões de tecnologias de informação e comunicação que viabilizassem a interoperabilidade entre eles” (CAFÉ; LAGE, 2002, p. 4).

Então, a *Open Archives Initiative* - OAI foi criada com a missão de desenvolver e promover padrões de interoperabilidade, ou seja, padrões compatíveis entre diferentes sistemas e/ou plataformas, para facilitar a difusão eficiente de conteúdos na Internet, permitindo o intercâmbio de vários formatos bibliográficos entre diferentes computadores utilizando-se de um mesmo protocolo (BARRUECO ; COLL, 2003).

³ <http://www.arxiv.org/>.

De acordo com Sompel e Lagoze (2000), a Iniciativa de Arquivos Abertos surgiu devido ao número crescente de arquivos eletrônicos de preprints e eprints existentes. Muitos desses arquivos começaram como veículos informais para a disseminação de resultados preliminares e não revisados pelos pares, “literatura cinzenta”. E vários deles desenvolveram-se dentro de um meio específico para compartilhar resultados de pesquisas entre os profissionais de uma determinada área do conhecimento.

Segundo Hunter e Guy (2004), a origem da Iniciativa de Arquivos Abertos – OAI pode ser localizada muito antes dos anos 1990, e da criação do Arquivo de Pré-print de Los Alamos por Paul Ginsparg. Isto no entanto é considerado como o exemplo precursor da história dos arquivos de pré-prints, e Paul Ginsparg é reconhecido internacionalmente como um dos líderes na área de novas arquiteturas de publicação científica. ArXiv (como o arquivo é conhecido) “tem evoluído para um repositório global de resultados de pesquisas não revisadas pelos pares, em uma variedade de áreas de física”, mas também incorpora matemática, ciências não-lineares e informática. O ArXiv foi para a Universidade de Cornell em 2001.

A Iniciativa de Arquivos Abertos possui dois objetivos principais, segundo Hunter; Guy, (2004, p. 2): a rapidez da comunicação científica, e o livre acesso para as comunidades interessadas nestes recursos.

Foi em uma importante reunião iniciada por Paul Ginsparg, Rick Luce e Herbert Van de Sompel e originalmente chamada de “Universal Preprint Service meeting”, presidida em Santa Fé no ano de 1999, que a Iniciativa de Arquivos Abertos foi estabelecida (HUNTER; GUY, 2004). Nesta reunião eles definiram que a meta da OAI seria :

contribuir de maneira concreta com a transformação da comunicação científica. O veículo proposto para esta transformação é a definição de técnicas e suportes de aspectos organizacionais de uma estrutura de publicação científica aberta na qual

podem ser estabelecidas ambas as categorias, gratuita e comercial (SOMPEL; LAGOZE, 2000, p.1, tradução nossa).

Na visão dos autores, o estabelecimento da OAI se constitui na combinação de princípios organizacionais e especificações técnicas para obter um nível mínimo mas potencialmente elevado e altamente funcional de interoperabilidade entre os arquivos de eprints científicos.

Os arquivos abertos apresentam mudanças no modelo tradicional de comunicação científica e constituem-se em mais um desafio para o sistema de periódicos científicos.

Os estudiosos Sompel e Lagoze (2000, p.1, tradução nossa) destacam que:

- O crescimento explosivo da Internet proporcionou aos cientistas um meio de comunicação de acesso quase universal que facilita o compartilhamento imediato de resultados.
- A rapidez dos avanços na maioria dos campos científicos fez com que a lentidão do modelo de publicação tradicional seja um impedimento para o compartilhamento acadêmico.
- A transferência integral dos direitos de autor para o publicador funciona freqüentemente como um impedimento ao autor científico cuja preocupação principal é a disseminação abrangente dos resultados.
- A implementação atual de revisão pelos pares – uma característica essencial da comunicação científica - é muito rígida e às vezes influencia para impedir que novas idéias apareçam, favorece artigos de instituições prestigiosas, e causa atrasos devido a publicações impróprias, inadequadas.
- O desequilíbrio entre a alta dos preços das assinaturas e a diminuição destas, ou melhor, a estagnação dos orçamentos das bibliotecas, estão ocasionando uma crise econômica para as bibliotecas de pesquisas.

Os autores ainda complementam dizendo que, a proposta de arquivos de e-prints pode ser vista como um modelo mais justo e eficiente para disseminar resultados de pesquisas. E destacam que o desafio é aumentar o impacto dos arquivos de e-prints com

revisão pelos pares, essencial para a comunicação científica e esse é o foco da iniciativa de Arquivos Abertos.

Hunter e Guy (2004) em uma conferência em Londres, demonstraram ao público como trabalhos submetidos em um arquivo aberto estavam disponíveis mundialmente no dia da submissão e retratam que esse é um prêmio valioso para a comunidade acadêmica e destacam que os altos preços estabelecidos pelos publicadores de periódicos acadêmicos reforçam a difusão da idéia de eprints na comunidade acadêmica.

Conforme Ziman (1984, p. 84), citado por Targino (2000, p. 15), “A ciência é conhecimento público, disponível livremente para todos”. “Os resultados da pesquisa não pertencem ao cientista, mas à humanidade. Constituem produto da colaboração social e como tais devem ser partilhados com todos, sem privilegiar segmentos ou pessoas” (TARGINO, 2000, p.15).

Para a autora,

Os direitos de propriedade na ciência reduzem-se, conforme Merton (1973), ao reconhecimento da autoria. Por sua vez, a única maneira pela qual um cientista pode requerer para si a autoria de descobertas e o reconhecimento dentre os pares é tornar público seu trabalho. As descobertas científicas devem ser automaticamente comunicadas à comunidade científica através de publicação, a fim de que os interessados possam utilizá-las. E esta corrida em busca da prioridade da descoberta científica implica originalidade, vista como a capacidade de levar a ciência para a frente, de explorar suas potencialidades, de criar alternativas, enfim, de garantir a dinamicidade intrínseca à ciência (TARGINO, 2000, p.15).

De acordo com Hunter e Guy (2004, p. 7, tradução nossa) “a comunicação científica tem sido descrita como uma estrutura contendo quatro componentes essenciais”:

- **Registro** - estabelece a prioridade intelectual de uma idéia, conceito ou pesquisa;

- **Certificação** - certifica a qualidade da pesquisa e/ou a validade da sentença reivindicada;
- **Responsabilidade moral** - assegura a disseminação e acessibilidade da pesquisa, provendo meios pelos quais os pesquisadores possam estar a par de novas pesquisas; e
- **Arquivamento** - preserva a herança intelectual para uso futuro.

No entanto, estas funções devem ser úteis não somente no sistema tradicional mas em qualquer sistema de comunicação científica existente (HUNTER e GUY, 2004).

No modelo tradicional, eram os publicadores que concentravam os componentes de distribuição como, impressão, marketing e outras tarefas especializadas e caras. Devido a evolução da publicação digital e das tecnologias de distribuição e rede essas funções, no novo modelo, puderam ser desagregadas em seus vários agentes distribuídos cooperativamente, funcionalmente e economicamente, responsáveis por vários aspectos do Registro, Certificação, Responsabilidade e Arquivamento, ou seja, essas funções são realizadas por diferentes atores como, instituições acadêmicas, departamentos acadêmicos, bibliotecas, repositórios etc. Quando efetuadas separadamente, cada uma delas pode operar mais eficazmente e competitivamente (CROW, 2002).

Os quadros 1 e 2 exemplificam essas quatro funções dentro do sistema tradicional e do novo modelo desagregado de comunicação científica:

FUNÇÃO	PROCESSO	ATORES	PATROCINADOR DO PROCESSO
Registro	papel (ou eletrônico) submissão ao periódico	Autor acadêmico-Pesquisador	Publicador
Certificação	Revisão pelos pares	Consultores acadêmicos	Publicador
Responsabilidade moral	Seleção do periódico de biblioteca e suporte	Bibliotecários Publicadores	Instituição acadêmica Publicador
Arquivamento	Acesso perpétuo	Bibliotecários	Instituição acadêmica

QUADRO 1: Funções da comunicação científica no **sistema tradicional** de periódicos acadêmicos

FONTE: CROW, 2002, p. 9, tradução nossa.

FUNÇÃO	PROCESSO	ATORES	PATROCINADOR DO PROCESSO
Registro	Divulgar o documento eletrônico no repositório	Autor acadêmico - Pesquisador	Repositório patrocinador
Certificação	Revisão pelos pares Certificação associativa Resposta on-line	Consultores acadêmicos Consultores acadêmicos Respondentes acadêmicos	Periódicos cobertos Departamentos acadêmicos Repositório patrocinador
Responsabilidade moral	Repositórios abertos interoperáveis e serviços de apoio	Bibliotecários	Instituições acadêmicas Sociedades profissionais Provedores de terceiros
Arquivamento	Acesso perpétuo	Bibliotecários	Instituição acadêmica

QUADRO 2: Funções da comunicação científica no **novo modelo desagregado**

FONTE: CROW, 2002, p. 9, tradução nossa.

Os Repositórios Institucionais demonstrarão ser essenciais por possuírem uma estrutura de comunicação científica reconstituída de maneira que os benefícios do novo sistema possam ser totalmente realizados. O fundamental para implementar este modelo desagregado é a separação lógica do conteúdo e dos componentes de serviços defendida por Van de Sompel e outros. Esta separação permite acesso aberto ao conteúdo dos repositórios distribuídos, por estes serem mantidos independentemente de serviços de valor-agregado de múltiplos provedores de serviços (CROW, 2002).

Ou seja, o movimento de arquivos abertos tem acelerado a desconstrução do modelo de publicação científica tradicional. Com o estabelecimento de soluções de interoperabilidade⁴ para facilitar a disseminação de conteúdos, um novo modelo de publicação desagregado e interoperável foi construído, o qual os conteúdos dispostos nos *data providers* (Repositórios institucionais e outros), que fazem parte de um sistema global de repositórios distribuídos e descentralizados, ficam separados, ou seja, desagregados dos

⁴ A OAI estabeleceu o protocolo OAI-PMH para coleta de metadados que favorece a interoperabilidade de repositórios digitais independentemente do tipo (institucional, de disciplina específica, comercial, etc.) ou do conteúdo.

service providers (*havesters* de metadados, busca e recuperação, e outras ferramentas de acesso de valor-agregado) (CROW,2002).

Como parte da solução de interoperabilidade, o protocolo OAI-PMH, resultado da Iniciativa de Arquivos Abertos (OAI) e que partiu de uma necessidade essencialmente técnica, surgiu sem grandes preocupações filosóficas. Mas ao fornecer uma base estável para a interoperabilidade entre arquivos abertos, e devido ao crescente número de servidores que o implementam, a iniciativa OAI e o protocolo OAI-PMH contribuem para dar maior visibilidade e encorajamento ao movimento do Acesso Livre à literatura científica (RODRIGUES, 2004).

2.2 Protocolo da Iniciativa de Arquivos Abertos para coleta de metadados – OAI-PMH

A idéia de interoperabilidade foi apresentada com o advento da World Wide Web e das redes de computadores. Apesar da definição exata do termo ainda ser discutida, seu conceito é muito utilizado na literatura corrente e em prática atual (MAI, 2003).

De acordo com Sompel e Lagoze (2000), interoperabilidade é um termo abrangente, relacionado a diversos aspectos da Iniciativa de Arquivos Abertos, incluindo os formatos de metadados, arquitetura básica, abertura para a criação de serviços de biblioteca digital para terceiros, integração com o mecanismo de comunicação científica estabelecido, usabilidade em um contexto interdisciplinar, capacidade para contribuir com um sistema métrico coletivo para uso e citação etc.

Para Arms et al (2002) a meta de interoperabilidade é construir serviços coerentes para usuários, de componentes que são tecnicamente diferentes e administrados por diferentes organizações. Isto requer acordos de cooperação em três níveis: técnico, de conteúdo e organizacional.

A Iniciativa de Arquivos Abertos estabeleceu um conjunto mínimo do que é requerido para interoperabilidade:

- Um protocolo para o intercâmbio de metadados;
- A XML deve ser a sintaxe para representar e transportar os metadados;
- Os metadados devem ser expostos para serviços de usuário final;
- Os metadados devem ser coletados para facilitar a descoberta de conteúdos armazenados em arquivos de eprints distribuídos (HUNTER; GUY, 2004).

O *Open Archives Initiative Protocol for Metadata Harvesting* - OAI-PMH, em português, Protocolo da Iniciativa de Arquivos Abertos para Coleta de Metadados, foi desenvolvido pela Iniciativa de Arquivos Abertos, que é um esforço colaborativo para desenvolver e promover padrões de interoperabilidade com o encargo de facilitar a disseminação eficiente de conteúdo digital na comunidade científica (SHEARER, 2003).

Este protocolo foi projetado para facilitar o compartilhamento e descoberta de recursos de informação científica através de uma estrutura simples, contudo eficiente, e que por meio desta os repositórios tornam seus metadados de recursos informacionais disponíveis via um protocolo bem definido. Ele foi autorizado para uso pela OAI em janeiro de 2001 (LAGOZE; SOMPEL, 2002).

Embora a sua origem esteja motivada pela necessidade de descobrir recursos eletrônicos, o protocolo não especifica nada virtualmente sobre a questão da comunicação científica. E nada também sobre a especificação de formatos de metadados específicos (HUNTER ; GUY, 2004, p. 4, tradução nossa).

No entanto, dentro do sistema de comunicação científica este protocolo tem sido muito bem aceito como um meio efetivo para tornar metadados disponíveis para coleta e transmissão de conteúdos na Internet. Ele tem sido tema de muitos debates e estudos, vários

artigos já foram publicados em revistas especializadas e muitos projetos de pesquisa foram aprovados por muitas instituições estrangeiras. E agora com quase 7 anos de existência, “os implementadores e pesquisadores estão começando a manifestar seus feedbacks, provendo avaliações do OAI-PMH e avançando em novas aplicações para o protocolo” (SHEARER, 2003, p.1, tradução nossa).

O percurso do OAI-PMH começou em 1999 em Santa Fé – Novo México, numa reunião convocada para discutir uma série de problemas no mundo dos fornecedores de e-prints. O problema principal discutido foi que, como os servidores de e-prints de conteúdos específicos tanto de instituições quanto disciplinares se proliferaram, havia a necessidade de desenvolver serviços que permitissem o cruzamento de busca de documentos hospedados em múltiplos repositórios (SHEARER, 2003).

Os repositórios também necessitariam ser capazes de automaticamente identificar e copiar trabalhos que tivessem sido depositados neles. Era inconveniente solicitar aos cientistas o depósito de seus trabalhos em múltiplos repositórios, e conseqüentemente seria necessário que os arquivos disciplinares fossem capazes de identificar e duplicar material dos arquivos institucionais e os serviços de busca de cruzar buscas em ambos os arquivos, disciplinares e institucionais (SHEARER, 2003).

Muitos aspectos dos assuntos envolvidos na discussão destes problemas foram explorados pelos participantes da reunião de Santa Fé, incluindo como identificar arquivos de e-prints de interesse e como fixar políticas para encontrar material em cada um dos arquivos (SOMPEL; LAGOZE, 2000 citado por SHEARER, 2003). Contudo, a parte fundamental do trabalho deles era a definição de uma interface que permitisse aos servidores de e-print expor os metadados dos documentos neles armazenados, para que os serviços de busca ou outros repositórios com padrões de metadados similares pudessem então coletar esses metadados. Estes arquivos procederiam então como uma federação de repositórios, indexando

documentos de modo padronizado de forma que múltiplas coleções pudessem ser buscadas como se elas formassem uma única coleção (SOMPEL; LAGOZE, 2003, citado por SHEARER, 2003).

Depois da reunião de Santa Fé em 1999, vários workshops foram realizados para compartilhar a filosofia da OAI com a ampla comunidade dos profissionais da informação. Compareceram nesses workshops representantes da John Wiley & Sons, NASA (*National Aeronautic and Space Administration*), ARL, OCLC (*Online Computer Library Center*), além de diversos profissionais de institutos de pesquisa e bibliotecas (SHEARER, 2003).

Fora desses simpósios emergiu um consenso muito surpreendente. Muitos outros grupos tiveram problemas muito parecidos com os enfrentados pela comunidade de e-print. Os metadados que cada comunidade queria tornar disponíveis possuíam características únicas, porém um mecanismo básico para tornar os metadados disponíveis para coleta era extremamente necessário (LYNCH, 2001, citado por SHEARER, 2003, p. 2, tradução nossa).

A partir da constatação destas necessidades em comum, a Associação dos profissionais da informação (*Coalition for Networked Information*) e a Federação de Bibliotecas Digitais (*Digital Library Federation*), forneceram recursos para estabelecer uma secretaria, na Universidade de Cornell para a Iniciativa de Arquivos Abertos (OAI), inicialmente sob a administração de Herbert Van de Sompel (na época, professor visitante da Cornell) e Carl Lagoze (professor pesquisador da Cornell). Em seguida, um comitê internacional foi estabelecido para conduzir os esforços e um programa foi lançado para a divulgação da Convenção de Santa Fé, por esta apoiar a coleta de todos os tipos de metadados, assim como explorar outras questões de infra-estrutura relacionadas à coleta de metadados como, o registro de sites disponíveis para coleta, além de outras aplicações potencialmente valiosas como as ferramentas que fazem busca cruzada em arquivos e os

serviços de revisão pelos pares (LYNCH, 2001; SOMPEL e LAGOZE, 2000, citado por SHEARER, 2003).

No ano de 2000 a Iniciativa de Arquivos Abertos convocou uma reunião técnica para reorganizar as especificações da convenção de Santa Fé e, posteriormente aperfeiçoá-las através de revisão por e-mail. Em setembro do mesmo ano, a OAI estendeu a sua estrutura de interoperabilidade para além dos e-prints, através do desenvolvimento de um conjunto de padrões de interoperabilidade com o objetivo de facilitar a disseminação eficiente de conteúdo. Esta estrutura foi nomeada de *OAI Protocol for Metadata Harvesting - OAI-PMH* (LYNCH, 2001, citado por SHEARER, 2003).

A primeira versão do protocolo tornou-se pública em 2001 a partir de dois workshops, um em Washington, D.C. e outro em Berlin, para apresentá-lo a potenciais implementadores. O OAI-PMH pode ser utilizado por diversas comunidades empenhadas em publicar conteúdos na Web. Qualquer servidor de trabalho em rede (não necessariamente um servidor de e-print) pode fazer uso dele para criar metadados que descrevam os objetos hospedados nos servidores e torná-los disponíveis para os provedores de serviços ou outros repositórios que desejam colecioná-los (LYNCH, 2001 citado por SHEARER, 2003).

Embora os seus desenvolvedores pretendessem que as especificações da versão 1.0 do protocolo permanecessem estáveis por pelo menos um ano, enquanto a comunidade ganhasse experiência usando-o, já houve uma revisão (SHEARER, 2003). E em 2002, depois de uma revisão técnica de oito meses pelo comitê técnico da OAI, a versão 2.0 do protocolo foi liberada, e está sendo adotada por diversas comunidades além de ser assunto de pesquisa e experimentação (SOMPEL, 2002 citado por SHEARER, 2003).

O protocolo utiliza uma estrutura de transação pedido/resposta muito simples, baseada em **HTTP** (*Hiper Text Transfer Protocol*) para a comunicação entre um *harvester* (coletador) que no caso seria o cliente e um repositório de dados, no caso, o servidor. Esta

estrutura cliente/servidor funciona da seguinte forma: O cliente pede ao servidor que lhe envie metadados segundo determinados critérios como por exemplo a data de criação dos dados, título, autor etc. Em resposta o servidor devolve um conjunto de registros em formato XML, incluindo identificadores dos objetos descritos em cada registro como as **URLs** (*Universal Resource Locator*) por exemplo (ver figura 1) (SHEARER, 2003; BARRUECO, COLL, 2003).

De acordo com Barrueco e Coll (2003, p. 7, tradução nossa) “Os pedidos são feitos utilizando os métodos GET ou POST do protocolo HTTP e constam de uma lista de opções em forma de pares do tipo: Chave = valor. Existem seis tipos de pedido que um cliente pode fazer a um servidor”:

- **GetRecord** – Utilizado para recuperar um registro concreto, real. Necessita dos argumentos: identificador do registro pedido e especificação do formato bibliográfico em que se deve devolver.
- **Identify** – Utilizado para recuperar informações sobre o servidor: nome, versão do protocolo que utiliza, endereço do administrador, etc.
- **ListIdentifiers**- Recupera os títulos dos registros, em lugar dos registros completos. Permite um conjunto de termos como a ordem das datas entre os dados que queremos recuperar.
- **ListRecords** – igual ao anterior, no entanto recupera os registros completos.
- **ListSets** – Recupera um conjunto de registros. Estes conjuntos são criados opcionalmente pelo servidor para facilitar uma recuperação seletiva dos registros. Seria uma classificação dos conteúdos segundo diferentes entradas. Um cliente pode pedir que se recuperem somente os registros pertencentes a uma determinada classe. Os conjuntos podem ser simples listas ou estruturas hierárquicas.
- **ListMetadataFormats** – Devolve a lista de formatos bibliográficos que o servidor utiliza.

O OAI-PMH suporta múltiplos formatos para representar os metadados, no entanto, de acordo com a OAI, é requerido que todos os servidores ofereçam os seus registros

utilizando no mínimo Dublin Core codificado em XML. Contudo, cada servidor tem a liberdade de oferecer os registros em outros formatos como MARC por exemplo (BARRUECO; COLL, 2003).

Um cliente pode pedir que os registros lhe sejam servidos em qualquer um dos formatos suportados pelo servidor. A idéia subjacente aqui é que no futuro as diferentes comunidades que utilizem o protocolo definam seus próprios formatos de metadados que serão mais ricos e mais precisos que o Dublin Core. Já estão sendo feitas aplicações de *harvesting* usando **MODS** (*Metadata Object Description Schema*), que provê uma capacidade descritiva mais rica que o Dublin Core, embora não tão complexa quanto as etiquetas disponíveis no padrão MARC21 completo (NEEDLEMAN, 2005).

Como exemplo , segue um exemplo de pedido/resposta entre um cliente/servidor ou *harvester*/repositório, segundo Barrueco e Coll (2003, p. 8):

Petição:

```
http://an.oa.org/OAI-script?
verb=GetRecord&identifier=oai:arXiv:hep-th/9901001&metadataPrefix=oai_dc
```

Resposta:

```
<?xml version="1.0" encoding="UTF-8" ?>
<OAI-PMH xmlns="http://www.openarchives.org/OAI/2.0/"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/
    http://www.openarchives.org/OAI/2.0/OAI-PMH.xsd">
  <responseDate>2002-05-01T19:20:30Z</responseDate>
  <request verb="GetRecord" identifier="oai:arXiv:hep-th/9901001"
    metadataPrefix="oai_dc">http://an.oa.org/OAI-script</request>
  <GetRecord>
  <record>
  <header>
  <header>
  <identifier>oai:arXiv:cs/0112017</identifier>
  <timestamp>2001-12-14</timestamp>
  <setSpec>cs</setSpec>
  <setSpec>math</setSpec>
  </header>
  <metadata>
  <oai_dc:dc
    xmlns:oai_dc="http://www.openarchives.org/OAI/2.0/oai_dc/"
    xmlns:dc="http://purl.org/dc/elements/1.1/"
    xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
    xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/oai_dc/
      http://www.openarchives.org/OAI/2.0/oai_dc.xsd">
    <dc:title>Using Structural Metadata to Localize Experience of Digital Content
  </dc:title>
  <dc:creator>Dushay, Naomi</dc:creator>
  <dc:subject>Digital Libraries</dc:subject>
  <dc:description>With the increasing technical sophistication of
    both information consumers and providers, there is
    increasing demand for more meaningful experiences of digital
    information. We present a framework that separates digital
    object experience, or rendering, from digital object storage
    and manipulation, so the rendering can be tailored to
    particular communities of users.
  </dc:description>
  <dc:description>Comment: 23 pages including 2 appendices,
    8 figures</dc:description>
  <dc:date>2001-12-14</dc:date>
  </oai_dc:dc>
  </metadata>
  </record>
  </GetRecord>
  </OAI-PMH>
```

FIGURA 1: Exemplo pedido/resposta**FONTE:** BARRUECO e COLL, 2003, p.8

Existem alguns aspectos os quais não são de competência do protocolo, por exemplo, as questões relacionadas à gestão ou autorização para o acesso dos clientes (coletadores); a utilização que os clientes fazem dos dados; como os clientes podem localizar

os servidores que contenham os dados os quais necessitam. Nestes casos, os servidores deverão recorrer a métodos externos (BARRUECO e COLL, 2003).

Devido ao desenvolvimento de protocolos específicos pela iniciativa de Arquivos Abertos e não a utilização ou aperfeiçoamento de protocolos já existentes e muito utilizados como o Z39.50⁵ por exemplo, freqüentemente questões surgem relativo à relação entre a OAI e o Protocolo de recuperação de informação Z39.50 e a razão porque a OAI decidiu desenvolver seu próprio protocolo de metadados do que usar o Z39.50. Uma das metas principais da OAI foi desenvolver mecanismos que fossem fáceis para implementar e que apresentasse poucas barreiras para implementadores potenciais. Acredita-se que o Z39.50 foi examinado como um possível protocolo, mas a OAI não precisava de toda a funcionalidade avançada construída em Z39.50. Assim, a complexidade adicional em se construir um protocolo de *harvesting* usando o Z39.50 poderia não ser justificada. Isto não quer dizer, porém, que os *service providers*, que constroem serviços valor agregado sobre os metadados colhidos da OAI, não podem escolher desdobrar estes serviços através de aplicações baseadas no Z39.50 (NEEDLEMAN, 2002).

Portanto, é importante destacar, não se trata de substituir outras iniciativas, senão desenvolver uma alternativa que seja fácil de implementar e de desenvolver para propósitos diferentes dos que já tratam os sistemas de interoperabilidade existentes (BARRUECO; COLL, 2003).

O OAI-PMH permite aos servidores e-print expor seus metadados e aos *harvesters* coletá-los. Sendo assim, as origens da interoperabilidade técnica estão apoiadas em uma distinção entre duas classes de participantes: os *Data Providers* e os *Service Providers*. Os provedores de dados (*Data Providers*) ou repositórios, adotam a estrutura técnica da OAI como um meio para expor os metadados sobre o seu conteúdo, e os provedores de serviços

⁵ ver seção 4.1.1

(*Service Providers*) ou *harvesters* (coletadores), coletam metadados para os provedores de dados usando o protocolo da OAI e utilizam as réplicas dos metadados como base para construir serviços de valor agregado (SHEARER, 2003).

Conforme as figuras 2 e 3 a seguir, podem ser percebidas as funções dessas duas classes de participantes do processo de interoperabilidade:

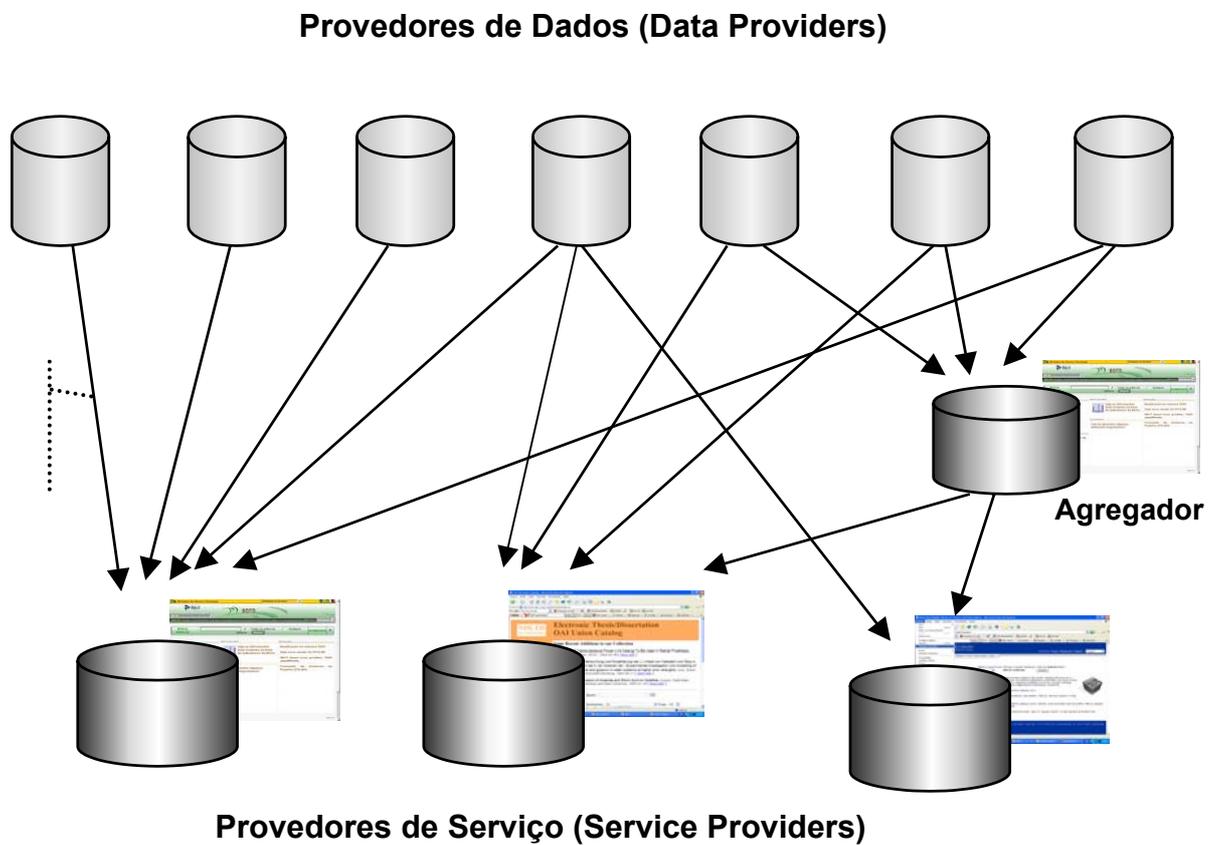


FIGURA 2: Esquema Funcional do OA mais atual
FONTE: KURAMOTO, 2005, slide 8



FIGURA 3: Modelo dos Arquivos Abertos: Características
FONTE: FONSECA, R. M. S., 2005, slide 19

Os repositórios de dados constituem-se de uma rede de servidores acessíveis que expõem seus metadados para *harvesting*, atendendo os pedidos do OAI-PMH. Já um *harvester* é uma aplicação cliente que emite pedidos do OAI-PMH. Um agregador pode fazer o *harvesting* de diferentes *data providers* e disponibilizar serviços aos usuários e/ou metadados para outros *service providers* (figs. 2 e 3) (NELSON, 2005).

Assim, os Repositórios Institucionais como parte integrante da *Open Archives Initiative*, inserem-se no crescente movimento de *Open Access Initiative* que objetiva prover o acesso livre à literatura científica e acadêmica.

2.3 Repositórios Institucionais Digitais

As instituições acadêmicas como produtoras de pesquisas primárias tem como algumas de suas principais funções a responsabilidade por assegurar a preservação intelectual dos seus membros e a disseminação deste capital intelectual. São “coleções digitais que asseguram e preservam a produção intelectual de uma ou várias comunidades universitárias” (CROW, 2002, p. 4, tradução nossa). Para Café et al (2003, p. 4),

Um repositório institucional agrega um conjunto avançado de serviços relativos a organização, tratamento, acesso e disseminação do conteúdo digital produzido por

uma instituição e sua comunidade acadêmica e de pesquisa. Dentre estes serviços, inclui-se principalmente o de preservação digital, uma vez que o gerenciamento da migração do conteúdo digital de uma tecnologia em vias de ser desativada para um sistema de ponta deve ocupar um espaço primordial nas preocupações das organizações que detêm repositórios institucionais.

De acordo com Crow (2002), os repositórios institucionais enquanto centralizam, preservam, e tornam acessível o capital intelectual de uma instituição, ao mesmo tempo em que fazem parte de um sistema global de distribuição, são repositórios interoperáveis que fornecerão a base para um novo modelo de publicação científica desagregado.

As razões pelas quais universidades e faculdades implementam repositórios institucionais baseiam-se em duas proposições inter-relacionadas, de acordo com Crow (2002, p. 6, tradução nossa):

- um maior suporte para um esforço pan-institucional (todas as instituições) e,
- oferecer benefícios imediatos e sem intermediários para cada instituição que implementa um repositório.

E ainda que, os Repositórios institucionais “por assegurar, preservar e disseminar o capital intelectual coletivo de uma universidade, é visto como um dos indicadores significantes da qualidade acadêmica de uma instituição” (CROW, 2002, p. 6, tradução nossa).

No sistema atual de comunicação científica, grande parte da produção intelectual e da importância da propriedade intelectual de uma instituição são difundidas por milhares de periódicos científicos. Mas enquanto as publicações de docentes nestes periódicos refletem positivamente para a universidade, um repositório institucional concentra a produção intelectual criada pelos pesquisadores da universidade, e com isso fica mais fácil de demonstrar seu valor científico, social e financeiro (CROW, 2002, p. 6, tradução nossa).

Desta forma, os repositórios institucionais complementam o sistema existente para medir a produtividade e o prestígio institucional, com isso aumenta a visibilidade da instituição e reflete a qualidade de bolsas de estudos, e esta demonstração de valor pode converter em benefícios tangíveis, inclusive em recursos financeiros – de fontes públicas e privadas – que deriva em parte do status e da reputação de uma instituição (CROW, 2002).

O atual sistema de comunicação científica, como diz Crow (2002, p. 7, tradução nossa) “limita, ao invés de expandir, a leitura e a disponibilidade de mais pesquisas científicas (enquanto também obscurece suas origens institucionais)”, e os aumentos crescentes dos preços dos periódicos e os cancelamentos das assinaturas, são fatores que contribuirão fortemente para reduzir a audiência deste tradicional sistema de comunicação científica futuramente.

Para Harnad (2001) em citação de Hunter e Guy (2004), os custos de publicar trabalhos no formato tradicional são agora tão grandes que são cobradas das bibliotecas taxas de assinaturas enormes por acesso aos periódicos, o que significa que cada vez menos podem arcar com estas subscrições. E no final das contas, a comunidade a qual o estudante/cientista deseja se dirigir em muitos casos já não tem acesso. Em casos extremos isto poderia significar que o autor, ele mesmo, já não pode entrar na biblioteca de seu departamento, e poder ver o seu próprio trabalho nas estantes

Os periódicos científicos existem por mais de trezentos anos porque eles suprem com sucesso uma grande gama das necessidades dos autores:

- comunicar as suas descobertas aos colegas,
- estabelecer precedência de seus trabalhos,
- adquirir validação através de revisão pelos pares,
- estabelecer sua reputação,
- ter a certeza de que a versão final de seu trabalho está segura, e
- saber que seus trabalhos serão acessíveis por outros cientistas no futuro (BUCK, FLAGAN, COLES, 1999, p. 1, tradução nossa).

Segundo os autores supracitados, da mesma forma como o serviço postal tornou possível o compartilhamento de trabalhos científicos regularmente e entre um grande número de leitores, a Internet agora proporciona um canal de distribuição com o poder de reduzir o tempo de publicação e ampliar os formatos de impressão tradicionais suportando opções de multimídia e discurso em linha.

No entanto, apesar da aceitação difundida da Web pela comunidade acadêmica e de pesquisa, com a incorporação de tecnologia de rede avançada dentro de um novo paradigma para a comunicação científica, os publicadores

não aplicaram um custo mais baixo de distribuição na Web para fazer versões on-lines de periódicos disponíveis a um preço mais acessível que as versões impressas. Está ficando cada vez mais claro para a comunidade científica que devemos prever e desenvolver para nós mesmos um modelo novo e de recursos o suficiente para disseminar e preservar resultados, que sintetizem a tecnologia digital e as necessidades contínuas dos cientistas. (BUCK, FLAGAN, COLES, 1999, p. 1, tradução nossa).

Neste contexto, os repositórios institucionais, um dos modelos alternativos de publicação científica, quebram os monopólios dos publicadores e aumentam a consciência da produção intelectual universitária transparente, se forem implementados em campus individuais ou em projetos de consórcios em colaboração (CROW, 2002, p.7).

Nos últimos anos, devido aos avanços das tecnologias de informação e comunicação, não somente a estrutura e o fluxo da comunicação científica vem sofrendo consideráveis alterações, como também o próprio conceito e suporte dos documentos científicos.

2.4 O documento científico e as estruturas textuais no ambiente digital

Embora, esta pesquisa não tenha como propósito definir o termo documento por ser um vocábulo polissêmico e que apresenta um conceito aberto e polêmico em diversas áreas do conhecimento, alguns conceitos serão apresentados. Para Pinto Molina (1992, p. 16, tradução nossa), documento é a “memória materializada da humanidade em que se registram os fatos, idéias, ações, sentimentos..., que tem impressionado o espírito humano. Trata-se de um ente dual: contém o físico (forma), por um lado, e difusor científico (conteúdo) por outro”. Ribeiro e Santos (2003, p. 87), enumeram várias definições para documento entre elas:

- É qualquer base de conhecimento fixada materialmente e disposta de maneira que se possa utilizar para consulta, estudo, prova ou pesquisa;
- Reunião de informações e dados em geral gravados de forma permanente e legível por máquina ou por pessoa. Diz-se também de todo meio impresso (livro, revista, tese, monografia, etc. relacionado em uma base de dados bibliográfica);
- Termo genérico utilizado para referência a arquivos que contenham objetos que farão parte de compartilhamentos de dados entre aplicativos.

“Paul Otlet, considerado por muitos como o precursor da Documentação, concebia o documento como um suporte que pelos signos nele fixados, registra e conserva determinados dados da consciência referentes ao mundo” (SAGREDO E IZQUIERDO, 1983 citado por PINTO MOLINA, 1992, p. 36, tradução nossa). É “uma entidade material, artificial e de utilidade intelectual, em que Otlet distingue duas unidades: a ‘física’, ou matéria; e a ‘intelectual’, ou conteúdo, com os seguintes grupos de elementos: materiais, gráficos, lingüísticos e intelectuais” (PINTO MOLINA, 1992, p. 36, tradução nossa).

Para López Yepes (1981, p. 121) citado por Pinto Molina (1992, p. 36, tradução nossa), “a objetivação do conhecimento a um suporte e a possibilidade de comunicação ou

acessibilidade ao mesmo tempo, são as duas linhas que configuram o conceito de documento”.

Para Garcia Gutiérrez (1984, p. 32) citado por Pinto Molina (1992, p. 37, tradução nossa), “a história do documento transcorre paralela à história da humanidade, já que nenhum animal transmite documentos a seus descendentes, mas sim somente tradições geradas (e não acumuláveis) pelo instinto de sobrevivência”. Ainda de acordo com o autor, “sem o homem, a informação não pode ser fixada num suporte nem se transmitir”.

Desantes (1987, p. 230) citado por Pinto Molina (1992, p. 37), “nos fala da mensagem como centro objetivo do processo informativo, núcleo igual ao do processo documental”. E entende o documento como,

o produto da incorporação, e posterior tratamento documental, de uma “mensagem” a um “suporte”, um “conteúdo” a um “continente”, uma “substância” a um “acidente”, dando-nos a entender deste modo a supremacia da “mensagem” no contexto da unidade documental, apesar de que tradicionalmente os estudiosos, por mais paradójico que isto pareça, têm se preocupado mais com o continente sem prestar demasiada atenção ao conteúdo.

No entanto, graças ao suporte, a informação se consolida, ou seja, se transforma em documento ou “objeto informativo”. Porém, “a polissemia do vocábulo gera sérias dificuldades na hora de estabelecer sua conceitualização” (PINTO MOLINA, 1992, p. 34, tradução nossa).

Mas, com toda a transição tecnológica que estamos presenciando, e o rápido crescimento de publicações eletrônicas, vê-se que o desafio, no momento, é o de redefinir o conceito fundamental de “documento” que está por trás das apreensões dos profissionais da informação sobre mudanças na geração de informação, controle e acesso, juntamente com as suas preocupações sobre o futuro das profissões de informação (Schamber, 1996).

Segundo Schamber (1996, p. 669, tradução nossa),

Um conceito fundamental para os profissionais da informação é o de documento, um conceito arraigado em centenas de anos de tradição, firmemente fixado em sistemas enormes e complexos para publicação, organização e acesso. Contudo, tornou-se mais evidente que o conceito arquetipo de "documento" como "livro", que está por de trás destes sistemas, é insuficiente para lidar com uma multidão de formatos de mídias, particularmente os diversos formatos eletrônicos como periódicos, índices, catálogos, grupos de discussão na Internet, softwares de negócios, educacionais e de entretenimento.

O conceito tradicional de documento está baseado em suposições que ignoram muitas características dos formatos eletrônicos e como as pessoas os usam. Os documentos eletrônicos compartilham certas características que tendem a distingui-los dos de documentos impressos (SCHAMBER, 1996, p. 669, tradução nossa). Alguns exemplos são:

- **Facilmente manipulável** - o conteúdo pode ser produzido, revisado, copiado, colado e adaptado.
- **Linkado internamente e externamente** - múltiplos formatos de mídia (texto, imagem, áudio, vídeo) podem ser conectados e re-conectados dentro e entre documentos.
- **Prontamente transformável** - os formatos de armazenamento (disco, fita, impresso) e de exibições são mutáveis e trocáveis. O armazenamento requer espaço em disco, não armazéns.
- **Inerentemente recuperável** - as estruturas lógicas do documento, o processamento em linguagem natural e os sistemas especialistas, facilitam o acesso para a interpretação do conteúdo.
- **Imediatamente transportável** - as informações podem ser imediatamente transmitidas para qualquer local geográfico. A distribuição requer redes eletrônicas, não caminhões.
- **Infinidamente replicável** - a distribuição realmente é replicação; os originais permanecem na fonte. Quantidades são tecnicamente ilimitadas. A produção em massa ou cópia não degrada a qualidade das cópias (ou de originais).

De acordo com Schamber (1996), estas características também descrevem conceitos tradicionais de documento, e elas são a extensão do que velhas tecnologias

permitiram. E ainda, as características dos documentos estão ligadas as tecnologias: “Nossas percepções dos documentos tradicionais foram moldadas através das tecnologias de impressão da mesma forma que nossas percepções de documentos eletrônicos estão sendo moldadas através das tecnologias eletrônicas” (SCHAMBER, 1996, p. 670, tradução nossa).

Sendo assim, Schamber faz um questionamento: como o conceito de documento pode ser razoavelmente redefinido para abarcar os documentos eletrônicos com as suas características distintas? É preciso “começar assumindo que qualquer documento é uma unidade que contém alguma representação de informação com significado potencial e de uso”. Então, a autora, considerando uma definição de documento como unidade, define-o da seguinte forma: “Consiste de um conteúdo dinâmico, flexível e não linear, representado como um conjunto de itens informacionais *linkados*, armazenados em uma ou mais mídias físicas ou *sites* em rede; criado e usado por um ou mais indivíduos na facilitação de algum processo ou projeto” (SCHAMBER, 1996, p. 671, tradução nossa).

“Contudo esta definição não é limitada a um ambiente eletrônico/digital. Essencialmente, descreve modos os quais as pessoas sempre usaram o conteúdo do documento cognitivamente para organizar, interpretar, e sintetizar idéias” (SCHAMBER, 1996, p. 671, tradução nossa) . A autora ainda complementa: “Documentos e coleções sempre mudaram em valor e utilidade da informação com o passar do tempo”.

No entanto, é interessante notar as dicotomias que caracterizam e distinguem o documento, de acordo com Pinto Molina (1992, p. 41, tradução nossa),

Funcionalmente, tem duas missões básicas a satisfazer, que vem a ser como as duas caras de uma mesma moeda: o documento suporta ou contém informação (aspecto estático); e ao mesmo tempo a difunde (aspecto dinâmico). [...] Dupla personalidade funcional (“*contenedora*” - difusora).

Além disso o documento é substância, conteúdo científico, “continuum ideológico”, por um lado; e forma, aspecto físico, “limitações físicas” por outro.

Numa cultura eminentemente científica, o documento científico é de fundamental importância, visto que é considerado a autêntica seiva da árvore da ciência. Estes documentos possuem uma personalidade característica que nos permite distingui-los dos demais, não somente pelo seu conteúdo singular, evidentemente científico, mas sobretudo pela maneira de estruturar a apresentação desse conteúdo (PINTO MOLINA, 1992).

Convencionalmente os documentos científicos são classificados, no que se refere ao conteúdo e a singular configuração da árvore da ciência, da seguinte maneira: os produzidos no âmbito das Ciências Naturais; e os derivados das Ciências Sociais e Humanidades. Além do mais, e sobretudo,

os documentos científicos possuem uma estrutura formal peculiar e característica. Constitui esta estrutura formal do discurso ou tratado científico uma variante especial das estruturas formais argumentativas, e nela encontraremos quatro grandes categorias que, por ordem cronológica no desenvolvimento do discurso, são as seguintes: OBJETIVOS, MÉTODOS, RESULTADOS E CONCLUSÕES (O M R C) (PINTO MOLINA, 1992, p. 65).

No entanto, não necessariamente todos os textos científicos precisam conter todas essas categorias estruturais. Estudos comprovam que há diferenças entre as superestruturas dos textos de Ciências Naturais e seus equivalentes de Ciências Sociais e Humanidades.

“A produção textual, como canal de expressão, tem uma importância diferente nos domínios das Ciências Naturais e das Ciências Sociais e Humanidades. No primeiro caso, o texto é primordialmente o meio.” (PINTO MOLINA, 1992, p. 66, tradução nossa). Neste domínio o autor, segundo Pinto Molina (1992), está mais interessado no que quer dizer e menos em como dizê-lo.

Diferentemente ocorre nas Ciências Sociais e Humanidades, onde os autores estabelecem com frequência amplas e complexas questões que são analisadas por diferentes pontos de vista. Portanto, na maioria das ocasiões se dedica a mesma atenção ao “que” e ao

“como” da mensagem. Tornando o discurso não somente o meio mas também o fim (PINTO MOLINA, 1992).

Outra diferença significativa surge quando nos fixamos na área de operações de cada um dos domínios considerados, visto que em (CN), tanto o início como o final da investigação tem lugar no laboratório, em (CS), e particularmente em (H), este papel é julgado pelo texto. Daí a predileção dos cientistas sociais e humanistas por documentos volumosos (PINTO MOLINA, 1992, p.67).

É natural portanto, que os textos de CS e H (Ciências Sociais e História) tenham uma estrutura formal muito mais versátil, dependendo em grande medida da personalidade do autor. A análise baseada na divisão (OMRC), da superestrutura dos documentos científicos, tem confirmado estas diferenças entre um estilo de escrita normalizado e um altamente estruturado como acontece nas CN (Ciências Naturais), que utiliza uma estrutura formal explícita OMRC, e o estilo mais idiossincrático dos autores em CS e H (PINTO MOLINA, 1992).

O termo texto é definido por Petofi e García Berrio (1978, p. 56) citado por Pinto Molina (1992, p. 48, tradução nossa), como o “conjunto de unidades lingüísticas vinculadas em um conglomerado total de intenção comunicativa”. De acordo com Pinto Molina (1992, p. 48, tradução nossa), “a lingüística do texto se fundamenta não por frases que falamos, mas sim por textos: ou seja, todas as unidades lingüísticas regulam sua interação operativa conforme a idéia textual em que aparecem inseridas”.

Assim, a estruturação geral da unidade textual implica a superposição e inter-relação de três estruturas fundamentais: Estrutura Superficial, Estrutura Profunda e Superestrutura. Portanto, a essência da entidade textual se baseia na singular integração desta trilogia de estruturas complementares (PINTO MOLINA, 1992).

A autora faz uma analogia da estrutura geral do texto com um esquema arbóreo, criando uma imagem esclarecedora de alguns conceitos básicos na lingüística textual:

se as partes fundamentais da árvore são raiz, tronco, galho e folhas, também no texto podemos encontrar seus equivalentes: as folhas corresponderiam à estrutura de superfície (ES) (microestrutura); os galhos representariam a superestrutura; o tronco seria a estrutura profunda (EP) (macroestrutura); e a raiz se equipararia ao tópico do discurso (PINTO MOLINA, p. 49, tradução nossa).

Portanto, de acordo com Pinto Molina, a estrutura geral do texto obedece a um esquema arbóreo, tal como pode ser analisado na figura a seguir.

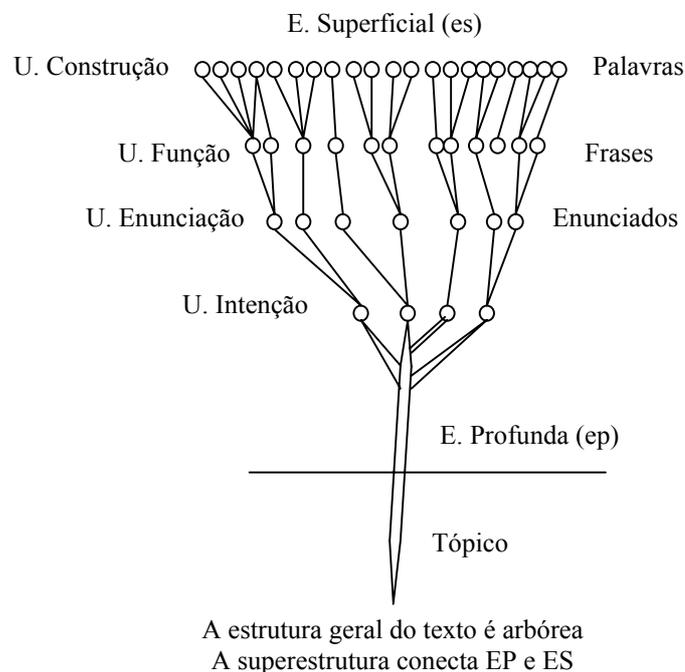


FIGURA 4: Tipologia do documento científico.

FONTE: PINTO-SEGURA (1991) citado por PINTO MOLINA (1992, p. 50, tradução nossa)

A estrutura superficial (ES), ou microestrutura, é a única estrutura concreta, não abstrata, visto que ela se corresponde com a realidade física do texto; as palavras, que são os signos que moldam esta ES, tem a capacidade de projetar-se sobre nossos sentidos, permitindo o processo de percepção que resulta na compreensão integral do texto (PINTO MOLINA, 1992).

Já a estrutura profunda (EP), ou macroestrutura, é uma “estrutura abstrata e teórica, ainda quando se fundamente sobre categorias e regras de tipo geral e convencional” (PINTO MOLINA, 1992, p. 53, tradução nossa). De acordo com Van Dijk (1992, p. 55, tradução nossa),

a macroestrutura de um texto [...] é uma representação abstrata da estrutura global de significado de um texto. Enquanto que as seqüências [de orações] devem cumprir as condições de coerência linear, os textos não somente devem cumprir estas condições (porque se ‘apresentam’ como seqüências de orações), mas também as de coerência global.

O termo macroestrutura é considerado por Van Dijk muito relativo: “o que em um texto pode considerar-se uma microestrutura, em outro seria uma macroestrutura” (VAN DIJK, 1992, p. 56, tradução nossa).

Quanto à superestrutura, denominação criada por Van Dijk para conceituar uma série de estruturas globais especiais, “é um tipo de esquema abstrato que estabelece a ordem global de um texto e que se compõem de uma série de categorias, cujas possibilidades de combinação se baseiam em regras convencionais” (VAN DIJK, 1992, p. 144, tradução nossa).

Como disse o autor:

Denominaremos superestruturas as estruturas globais que caracterizam o tipo de um texto. Portanto, uma estrutura narrativa é uma superestrutura, independente do *conteúdo* (ou seja, da macroestrutura) da narração, mesmo que as superestruturas imponham certas limitações ao conteúdo de um texto. Para dizer metaforicamente: uma superestrutura é um tipo de *forma do texto*, cujo objeto, o tema, ou seja: a macroestrutura, é o *conteúdo do texto* (VAN DIJK, 1992, p. 142, tradução nossa).

A superestrutura é uma espécie de esquema ao qual o texto se adapta. Ela existe independentemente do conteúdo. No entanto, não se pode dizer que todos os textos possuem uma superestrutura (VAN DIJK, 1992).

Os textos científicos apresentam uma superestrutura argumentativa baseada não somente em uma conclusão e sua justificação, mas também em um esboço do problema e uma solução (VAN DIJK, 1992). Um exemplo de uma estrutura hierárquica de um artigo científico poderia ser, segundo Van Dijk (1992, p. 164), o apresentado na figura 5:

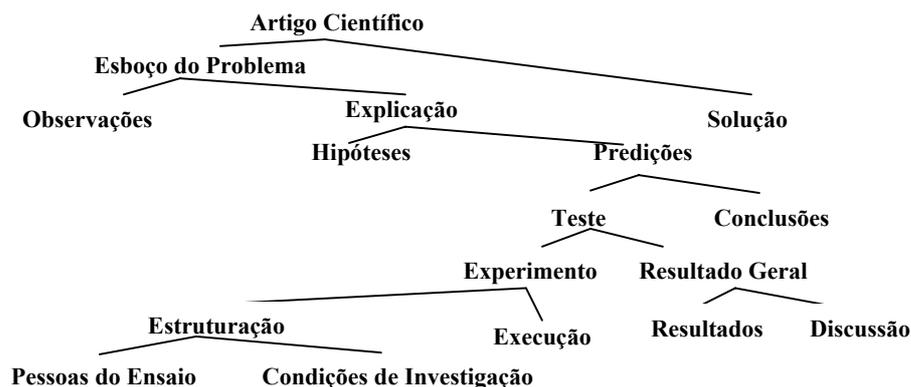


FIGURA 5: Estrutura hierárquica de um artigo científico

FONTE: VAN DIJK (1992, p. 164, tradução nossa)

Evidentemente, nem todos os discursos científicos baseiam-se exclusivamente nesta estrutura, “[...], sobretudo nas disciplinas científicas não experimentais” (VAN DIJK, 1992, p. 165, tradução nossa).

Recentemente a estrutura do texto vem sendo amplamente considerada em virtude de inovações tecnológicas aplicadas nos atuais sistemas de informações. E uma dessas inovações é a linguagem XML⁶, utilizada dentre outras coisas, na marcação de documentos.

Para Arellano (2004, p. 15), “a natureza dos documentos digitais está permitindo ampla produção e disseminação de informação no mundo atual”. A preservação desses documentos é um dos grandes desafios deste século, devido a fatores como obsolescência e, principalmente, a fácil alteração de seu conteúdo, suporte e estrutura. De acordo com Arellano

⁶ A XML será abordada com maiores detalhes na Seção 4.

(2004, p. 18), com essas alterações, “a perda do conteúdo intelectual original pode ser declarada inaceitável pelo autor”.

Desta forma, a preservação da propriedade intelectual (*Intellectual Property Rights* – IPR), é considerada por Bullock, (1999) citado por (ARELLANO, 2004, p. 18) “uma das principais barreiras que interferem na preservação dos objetos digitais”, porque de acordo com Arellano (2004, p. 18),

o IPR deve considerar não apenas o conteúdo, mas também qualquer ação relacionada ao software (cópias, encapsulamento de conteúdo, emulação de software, migração de conteúdo) que envolva atividades que podem infringir permissões específicas daqueles que mantêm os direitos.

O autor ainda enumera os requisitos de preservação dos documentos digitais:

Fixar os limites do objeto a ser preservado; Preservar a presença física; Preservar o conteúdo; Preservar a apresentação; Preservar a funcionalidade; Preservar a autenticidade; Localizar e rastrear o objeto digital; Preservar a proveniência; Preservar o contexto (BULLOCK, 1999 citado por ARELLANO, 2004, p. 18, grifo do autor).

Neste sentido, os profissionais da informação que trabalham com documentos em formatos digitais “estão elaborando normas necessárias para armazenar e compartilhar de maneira adequada esses materiais, assim como buscam a formulação de políticas institucionais de preservação” (ARELLANO, 2004, p. 16). E, “na preservação de documentos digitais, assim como na dos documentos em papel, é necessária a adoção de ferramentas que protejam e garantam a sua manutenção” (ARELLANO, 2004, p. 17). É imprescindível que essas ferramentas sejam padrões abertos, possibilitando a criação de documentos nos formatos livres, “para que eles sejam acessados após a obsolescência dos equipamentos e programas informáticos em que foram criados. Usar padrões abertos permite seu estudo e sua conversão para novos padrões” (ARELLANO, 2004, p. 16). Os documentos digitais “não podem ser

deixados em formatos obsoletos para serem transferidos depois de longos períodos de negligência para repositórios digitais” (ARELLANO, 2004, p. 16).

A Linguagem de marcação XML é um padrão aberto, não impõe restrições à sua utilização nem a criação e incorporação de outras ferramentas que a manipulem. Além de ser um dos componentes do conjunto mínimo exigido para interoperabilidade estabelecido pela Iniciativa de Arquivos Abertos, sendo a sintaxe obrigatória para representar e transportar metadados.

Uma linguagem de marcação é um conjunto de símbolos que pode ser colocado no texto de um documento para demarcar e rotular as partes desse documento, levando em conta a semântica da informação, ou seja, cada pessoa ou instituição pode criar um conjunto de etiquetas que melhor represente os elementos que compõem os seus documentos. Ela pode armazenar e organizar praticamente qualquer tipo de informação em um formato adequado às necessidades de determinada instituição (RAY, 2001; MARTÍNEZ GONZALES, 2000).

Portanto, o uso da XML na marcação dos documentos científicos nos repositórios institucionais digitais possibilita que a estrutura textual do texto científico se torne evidente de modo que sua superestrutura possa ser visualizada e com isso dar subsídios para o desenvolvimento de normas para tornar as marcações ou namespaces universalmente únicos e assim evitar conflitos entre sistemas no processo de troca de informações e dinamizar a busca e a recuperação de documentos em ambientes digitais. Um exemplo de marcação de conteúdo de um texto científico de acordo com a estrutura hierárquica de Van Dijk (1992) pode ser visto na figura 6:

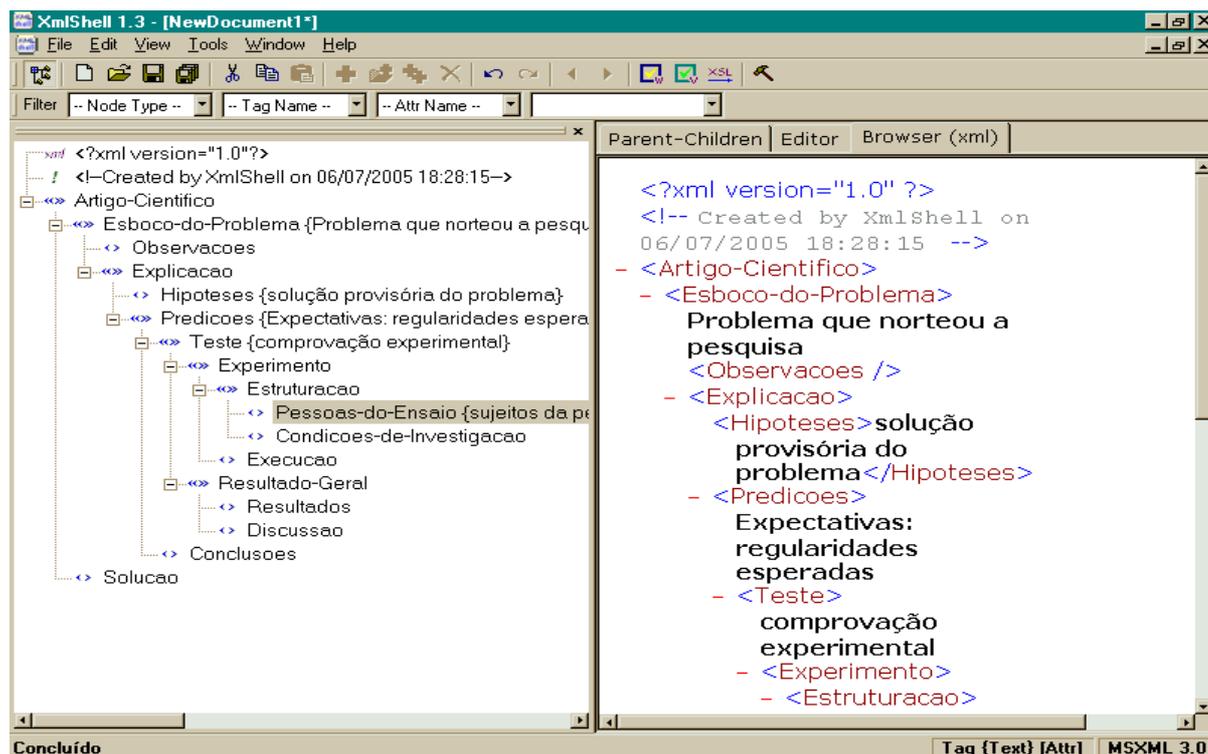


FIGURA 6: Exemplo de marcação de conteúdo de um texto científico de acordo com a estrutura hierárquica de Van Dijk (1992).

Percebe-se claramente, em comparação com a **figura 5**, que a XML é capaz de explicitar a estrutura hierárquica de um documento.

No entanto, a XML sozinha não faz nada. São as tecnologias associadas a ela que a torna eficiente. Essas tecnologias são: DTD, XML Schema, XSL, Xlink, Xpointer, Xpath, XML namespace entre outras.

A DTD é o esquema da linguagem, ou seja, são as regras pré-estabelecidas que os documentos precisarão obedecer para ser um documento válido e bem formado, é chamada por alguns autores de vocabulário da linguagem; O XML Schema define a estrutura, conteúdo e semântica dos documentos XML, ou seja, são conjuntos de regras que especificam quais elementos e atributos são permitidos ou necessários em um documento, é como se fosse uma DTD melhorada; a XSL é a ferramenta que possibilita diferentes apresentações de um mesmo documento, são as folhas de estilos; Xlink permite criar e descrever ligações ou links entre os recursos; Xpointer é um padrão associado ao Xlink e que permite o direcionamento de

fragmentos internos dos documentos, por exemplo: um capítulo de livro ou um parágrafo ou somente as referências etc.; Xpath direciona o caminho que uma parte do documento deve seguir para encontrar outro documento ou recurso; XML namespace são os nomes dos elementos relacionados no interior das tags.

Como a XML permite que cada pessoa ou instituição crie suas próprias tags, isso pode gerar alguns problemas quando uma aplicação XML desenvolvida por várias pessoas ou instituições gera nomes de elementos ou vocabulários de marcação diversos para o mesmo tipo de informação. A função principal dos namespaces é identificar de forma única os elementos e atributos de um documento XML. Mas para isso é preciso a criação e adoção de normas, visto que, as formas de representação (tanto temáticas quanto descritivas) dos recursos informacionais têm grande importância uma vez que buscam satisfazer diversas necessidades em diversos contextos, incluindo a preservação dos recursos informacionais nos Arquivos Abertos e Repositórios Institucionais Digitais e a otimização dos processos de busca e recuperação da informação.

A próxima seção aborda brevemente o polêmico conceito de formas de representação, informação e a Ciência da Informação.

3 INFORMAÇÃO, CIÊNCIA DA INFORMAÇÃO E FORMAS DE REPRESENTAÇÃO

“E a informação existe sempre, até mesmo na ausência de produtos de informação”
Michel J. Menou

Nesta seção faz-se uma breve abordagem a respeito da polêmica questão sobre os conceitos de informação, formas de representação e a Ciência da Informação. Destacando-se a importância da adoção de uma estrutura de representação complexa, rica e já consagrada na área da Ciência da Informação pelas Instituições provedoras de metadados.

As mais diversas áreas do conhecimento (entre elas a Física, a Biologia, a Filosofia, a Ciência da Informação, entre outras) estudam os conceitos que envolvem a informação, porém, não se tem ainda um consenso sobre a sua natureza última. Para Milidoni, Gonzalez e Broens (2004, p. 290) a informação é

entendida ora como artefato (Dretske, 1981, 1983, 1992), ora como um processo de transmissão de forma (Pereira Jr. & Gonzalez, 1995), ora como um elemento básico constitutivo do universo físico (Stonier, 1990), ora como uma diferença que faz diferença (Bateson, 1979), a informação entra na agenda dos estudos contemporâneos sem, contudo, alcançar uma definição que seja consensual entre seus pesquisadores.

A polêmica a respeito do estatuto ontológico da informação se torna mais complexa à medida que tentamos procurar explicações em correntes distintas (GARCIA et al, 2005). Por exemplo na Cibernética, ciência que estuda todo o campo de controle e comunicação de um sistema seja nos organismos vivos ou nas máquinas, Wiener (1968, p. 19) diz que

Informação é o termo que designa o conteúdo daquilo que permutamos com o mundo exterior ao ajustar-nos a ele, e o que faz com que nosso ajustamento seja nele percebido. E, o processo de receber e utilizar informação é o processo de nosso ajuste às contingências do meio ambiente e de nosso efetivo viver nesse meio ambiente.

Já o campo da Biologia Molecular, Lwoff (1970, p. 110) considera a informação como matéria,

O que podemos denominar informação para um ser vivo é, pois, uma série de estruturas, de seqüências, uma ordem bem determinada. É esta ordem que representa a informação biológica. O conceito de informação corresponde a este conjunto de dados bastante complexos. Como vêm, para o biólogo, o termo informação, o termo mensagem, representam algo bem material, totalmente desprovido de poesia: é uma seqüência de pequenas moléculas e o conjunto de funções por elas realizadas.

Gonzalez (2004, p. 7, tradução nossa) define

[...] informação como um processo de auto-organização que permite a expansão de padrões de ação para organismos situados. Nós reivindicamos que este processo de auto-organização evolui em consequência da sistêmica, relações significantes estabelecidas entre os organismos e seus ambientes.

Já Belkin e Robertson (1975) citados por Oliveira (2005) definem informação como “aquilo que é capaz de mudar uma estrutura”. Oliveira (2005, grifo do autor) reconhece a utilidade desse conceito para a Ciência da Informação, entretanto, sugere “que é mais conveniente dizer que: informação é aquilo que é capaz de **desencadear** uma **possível** mudança numa estrutura”. De acordo com a autora – baseada em Bateson (1986), “só posso dizer que algo tem informação, se contraste isso com um referencial que já tenho e que me diz que ali tem informação. Pois, não é possível perceber uma diferença sem uma estrutura através da qual se possa medir se, o que se vê, é ou não uma diferença”.

Dretske (1981) propõe que a informação é um artefato, um modo de descrever o significado, para algum agente de eventos intrinsecamente significativos.

Dretske propõe uma definição nuclear de informação segundo a qual: um sinal carrega informação sobre o que ocorre em uma fonte (o seu conteúdo) se ele for capaz de reproduzir factualmente as relações que se estabelecem na fonte, tornando-as acessíveis para qualquer observador que se encontre em condições recebê-las. Ainda que ele não exija uma correspondência biunívoca entre o sinal e o seu conteúdo (assim, por exemplo, o sinal que carrega a informação sobre um objeto grande não precisa ele próprio ser grande), a hipótese da informação nuclear pressupõe que os eventos que ocorrem na fonte impõem certos constraints [limitações] ao sinal que carrega informação sobre eles, tornando-os nomologicamente dependentes (MILIDONI; GONZALEZ; BROENS, 2004, p. 291).

Nesta tendência de pensamento, pode-se perceber também que Buckland (1991) citado por Capurro (2003) propôs a informação como sendo “[...] algo tangível como documentos e livros, ou, mais genericamente, qualquer tipo de objeto que possa ter valor informativo, o qual pode ser, em princípio, literalmente qualquer coisa”. Entretanto,

[...] é claro que o valor informativo a que alude Buckland não é uma coisa nem a propriedade de uma coisa, mas um predicado de segunda ordem, isto é, algo que o usuário ou o sujeito cognoscente adjudica a “qualquer coisa” num processo interpretativo demarcado por limites sociais de pré-compreensão que o sustentam (CAPURRO, 2003).

Neste sentido, Hjørland (2003) aponta que o foco principal da Ciência da Informação está no conhecimento documentado (conhecimento explícito) produzido pelo ser humano em alguns tipos de documentos de **uso potencial** para outros seres humanos. A luz das estrelas não é informação para a Ciência da Informação, mas a informação astronômica produzida e utilizada pelos astrônomos é.

Assim, no que concerne esta pesquisa, será adotada a definição sobre informação de Smit e Barreto (2002, p. 21), que definem:

Informação – estruturas simbolicamente significantes, codificadas de forma socialmente decodificável e registradas (para garantir permanência no tempo e portabilidade no espaço) e que apresentam a competência de gerar conhecimento para o indivíduo e para o seu meio. Estas estruturas significantes são estocadas em função de um uso futuro, causando a institucionalização da informação.

Ou seja, deve-se considerar que a Ciência da informação, traçada dentro do paradigma da Pós-Modernidade,

[...] instaura seu objeto – a informação – no seu universo simbólico original – distinguindo objeto material – conteúdo documentado – e objeto formal – processos de elaboração de estruturas significantes (formas sintéticas), passíveis de integrarem fluxos sociais e de serem apropriadas subjetivamente. A abordagem do seu objeto é interdisciplinar, já que esta é uma exigência das temáticas que trata, mas a Ciência da Informação enquanto tal constitui campo específico – uma disciplina (KOBASHI; TÁLAMO, 2003, p. 15).

Malheiro da Silva e Ribeiro (2002) representam em um esquema ilustrativo estes enfoques da Ciência da Informação:

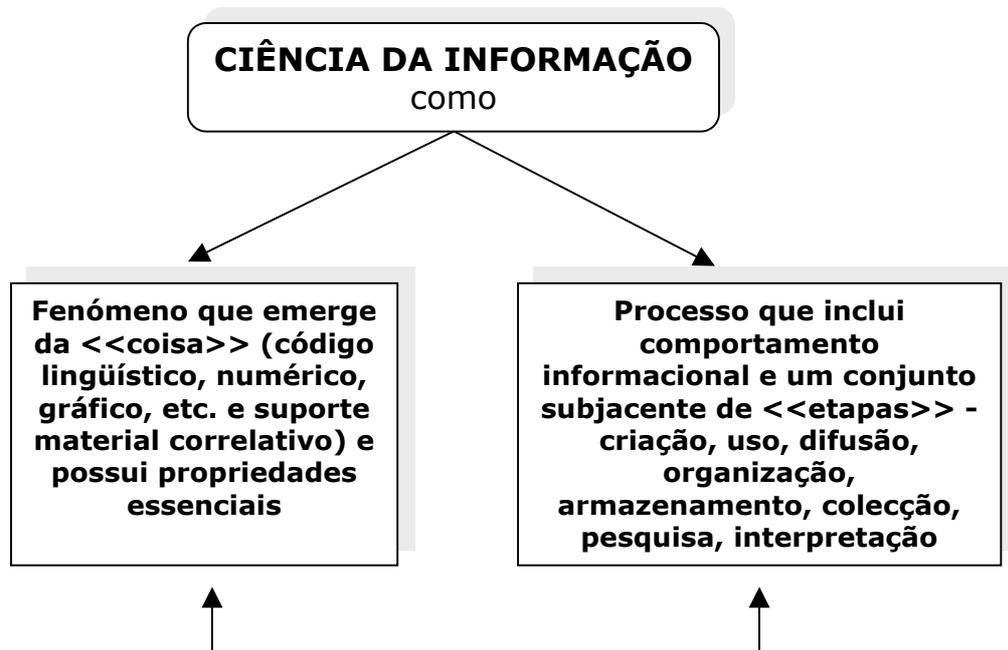


FIGURA 7: Representação sugestiva da Ciência da Informação
FONTE: MALHEIRO DA SILVA; RIBEIRO (2002, p. 84)

Desta forma, a Ciência da Informação esta preocupada com a geração, coleção, organização, interpretação, armazenamento, recuperação, disseminação, transformação e uso da informação, ou seja, tanto a organização do conteúdo documentado, quanto o processo e fluxo da informação com, conforme Hjørland (2003), particular ênfase nas aplicações das modernas tecnologias de informação e comunicação nestas áreas.

Vê-se então que a Ciência da Informação, disciplina do Campo das Ciências Sociais Aplicadas, requer em princípio, a admissão de paradigmas opostos, entretanto não excludentes entre si e até, complementares entre si ou tendo pontos de convergência. A exemplo, segundo Kobashi e Tálamo (2003, p. 13),

A operação com a informação no campo da Ciência da Informação parece, portanto, requerer a adoção de dois paradigmas em princípio opostos: a Teoria da Informação para lidar com sua dimensão formal (como estrutura e como sinal) e a Teoria da Comunicação para dar conta da mensagem e sua recepção (como significado).

No Entanto, as discussões e diálogos que outras disciplinas como a Ciência Cognitiva e Semiótica, a Linguística, a Ciência da Computação, entre outras, realizam sobre a informação, são fundamentais para se construir conceitos, modelos e teorias que possam contribuir para o desenvolvimento da Ciência da Informação.

3.1 REPRESENTAÇÃO DA INFORMAÇÃO

As atividades inerentes à Ciência da Informação (geração, coleção, organização, interpretação, armazenamento, recuperação, disseminação, transformação e uso da informação) envolvem principalmente o processo de produção de **representações** de recursos informacionais, uma vez que, Conforme Marcondes (2001, p. 61, grifo nosso),

O trabalho informacional de desenvolvimento e sistematização de representações dos recursos informacionais hoje disponíveis na **Internet** joga um papel essencial como facilitador da identificação, avaliação e acesso a estes recursos, intermediando um usuário com suas necessidades e informações potencialmente relevantes.

Os conceitos de representação encontram-se na teoria da Ciência Cognitiva que trata de temas como representação analógica, digital, proposicional, cognitiva ou, de maneira geral, representação mental. Entretanto é

[...] na semiótica geral que se encontram definições muito variadas do conceito de representação. O âmbito da sua significação situa-se entre apresentação e imaginação e estende-se, assim, a conceitos semióticos chave como signo, veículo do signo, imagem (representação imagética), assim como significação e referência. As tentativas de delimitação do conceito são variadas, mas, freqüentemente, imprecisas (NÖTH, 1997, p. 63).

Embora a conceituação sobre representação tenha seus problemas (tal como a informação), algumas definições serão aqui apresentadas para o entendimento do termo para as atividades da Ciência da Informação.

No âmbito da representação como relação de objeto, para Peirce (CP 1.540, citado por NÖTH, 1997, p. 65, grifo do autor) a representação

é o processo da apresentação de um objeto a um intérprete de um signo, mais exatamente, a relação entre o signo e o objeto: “Eu restrinjo a palavra representação à operação do signo ou sua relação com o objeto para o interprete da representação”. A fim de delimitar os conceitos de representação e signo, ele introduz o termo *representamen* para o veículo do signo: “Quando é desejável distinguir entre aquilo que representa e o ato ou relação de representar, o primeiro pode ser chamado de ‘representamen’, o último de ‘representação’” (Peirce, CP 2.273).

Neste contexto, Peirce (CP 2.273, citado por NÖTH, 1997, p. 65), define representar como: “estar para, quer dizer, algo está numa relação tal com um outro que, para certos propósitos, ele é tratado por uma mente como se fosse aquele outro”. E como exemplo, Pierce (CP 1.554, citado por NÖTH, 1997, p. 66) esclarece:

Uma palavra representa algo para a concepção na mente do ouvinte, um retrato representa a pessoa para quem ele dirige a concepção de reconhecimento, um catavento representa a direção do vento para a concepção daquele que o entende, um advogado representa seu cliente para o juiz e júri que ele influencia.

Neste sentido, representação é um processo acontecendo na mente de uma pessoa, originando nesta mente algo que não é o objeto a que se refere. “A representação então relaciona o objeto que ela representa com a mente que o percebe” (MARCONDES, 2001, p. 64). Esta noção é importante quando pensamos no processo de busca de informações dos usuários.

Já para Kaczmarek (1986, p. 88, citado por NÖTH, 1997, p. 67), na perspectiva de representação como referência, representar significa,

[...] apresentar algo por meio de algo materialmente distinto de acordo com regras exatas, nas quais certas características ou estruturas daquilo representado devem ser expressas, acentuadas e tornadas compreensíveis pelo tipo de apresentação, enquanto outras devem ser conscientemente suprimidas.

Neste sentido, Alvarenga (2003) diz que, partindo-se da definição de que representar significa o “*ato de colocar algo no lugar de*” classifica-se em nível **primário** a representação feita pelos autores no instante da expressão dos resultados de seus pensamentos, estes derivados de observações metódicas da natureza e dos fatos sociais, utilizando-se das linguagens disponíveis no contexto da produção e comunicação de conhecimentos. Nesse nível de representação, as línguas dos diversos povos e das especialidades (campos de conhecimento) desempenham papel primordial, incluindo-se também nesse instrumental outros símbolos icônicos e sonoros.

Depois de produzidos, os **registros de conhecimentos** constantes de documentos passam a integrar acervos de arquivos, bibliotecas, serviços ou centros de documentação/informações, sendo então novamente representados (representação **secundária**), visando-se à sua inclusão em sistemas documentais referenciais (ALVARENGA, 2003).

Isto é, no âmbito da Ciência da Informação as representações¹ são as descrições temáticas e descritivas do item documentário (recurso informacional).

“Na representação de um documento por meio de uma referência bibliográfica [registro de metadados], o objeto seria o documento original, o signo seria a referência

¹ No caso “representações que podem ser produzidas com lápis e papel. Em contraste com representações mentais internas” (CASA, 1997, p. 209).

bibliográfica e o interpretante seria a idéia do documento criada pela referência na mente do usuário” (MARCONDES, 2001, p. 66).

Assim, este signo (registro de metadados que aparece ao usuário) seria um ícone², ou seja, se apresentaria como o documento original, um registro de metadados contendo o mesmo autor, título, palavras-chave associadas ao assunto (como os termos de indexação), resumo descrevendo o conteúdo, a fonte do documento original etc.

3.1.1 FORMAS DE REPRESENTAÇÃO

Tanto o processo da representação descritiva (catalogação) quanto o processo da representação temática (indexação, elaboração de resumos, etc), utilizam-se de ferramentas e técnicas (de forma cada vez mais integrada) para a geração de produtos que possibilitarão que um documento específico seja recuperado. Estes produtos podem ser definidos como Metadados³, que podem ser mais comumente definidos como dados sobre outros dados (SOUZA; VENDRUSCULO; MELO, 2000). Para Milstead e Feldman (1999, p. 2, tradução nossa) “[...] a informação bibliográfica padrão, os índices, os termos indexadores, e os ‘abstracts’ são todos representantes do material original, portanto metadados”.

Assim, o uso de vocabulários controlados, a limitação de buscas aos descritores, identificadores, autores, títulos, ou campos de fontes, tem sido capaz de melhorar a precisão das buscas por muitos anos. Com a utilização do modelo booleano, os metadados tornaram possível encontrar a informação sem o excessivo aparecimento de disparates estranhos (MILSTEAD; FELDMAN, 1999). Entretanto, a utilização dos

² Um ícone é um tipo de signo que pode representar um objeto, porque seus atributos têm relações análogas com os atributos do objeto sendo representado (MARCONDES, 2001, p. 66).

³ Os conceitos e definições sobre metadado serão abordados na **Seção 4: MARCXML** para a OAI

[...] metadados, procedimento que no campo da Ciência da Informação tem-se utilizado mais para a atividade descritiva do que em qualquer outra da gestão de informação, mas que podem ter distintos usos potenciais, está direcionada [todos eles] ao descobrimento e recuperação de objetos de informação da forma mais simples e eficaz possível. Neste sentido se tem desenvolvido alguns formatos para metadados específicos no campo da Ciência da Informação, que parecem estar dando bons resultados, embora não estejam totalmente consolidados, devido a fase de experimentação em que se encontram [...] (GARCÍA MARTÍNEZ, 2001, p. 46, tradução nossa).

Pode-se ver na literatura especializada, muitos exemplos da aplicação dos metadados para a representação da informação, principalmente com as iniciativas que a XML (*eXtensible Markup Language*) vem trazendo para a **potencialização** da representação do item documentário para a otimização dos processos de recuperação da informação.

As formas de representação do item documentário (informação potencial registrada) têm grande importância uma vez que buscam satisfazer diversas necessidades em diversos contextos.

Neste sentido, Casa (1997, p. 209) diz que as “formas diferentes de representação podem também mostrar um objeto ou conceito a partir de diferentes pontos de vista. Na engenharia elétrica, por exemplo, diagramas de tipos diferentes são empregados para representar e raciocinar sobre as características de um circuito eletrônico”.

As diferentes formas de representação utilizadas na descrição de recursos informacionais no contexto dos Repositórios de dados ou das Iniciativas de Arquivos Abertos, por exemplo, podem facilitar de maneira efetiva e eficiente a recuperação e intercâmbio de informações. Conforme CASA (1997, p.210),

o uso de diversas formas de representação, cada uma delas específica para uma determinada tarefa, pode aumentar o número de oportunidades para se empregar elementos secundários de representação, os quais, por sua vez, contribuem para facilitar a interpretação da representação em questão.

Assim, uma instituição provedora de metadados ou provedora de dados (*data provider*) precisa oferecer várias formas de representação como DC (Dublin Core), MODS (*Metadata Object Description Schema*), MARC, MARCXML, METS (*Metadata Encoding and Transfere Sintaxe*), ou outros formatos baseados na XML, para que as diversas instituições beneficiárias ou *service providers* adquiram as representações que melhor atendam as suas exigências e necessidades informacionais, pois, “o fato de que diferentes formas de representação são úteis para propósitos diversos não é surpreendente ou mesmo um conceito inteiramente novo!” (CASA, 1997, p. 211). Mas, para isso, a adoção de um formato de metadados mais completo possível (como o MARCXML) pela instituição provedora torna-se necessária.

Entretanto, devido à falta de conhecimento da estrutura MARC pelos analistas de sistemas (computacionais) e também pelos próprios bibliotecários, essas instituições provedoras de metadados tem utilizado outros formatos mais simples para a representação de informações nos desenvolvimentos de softwares e na Iniciativa de Arquivos Abertos (OAI). Desta forma, as instituições que querem fazer parte dessas iniciativas acabam por adotar estes formatos de descrição de recursos informacionais mais simples. Exemplo disso, dentre vários outros parecidos, é a experiência do IMPA (Instituto de Matemática Pura e Aplicada) que para fazer parte da iniciativa internacional do projeto MATH-NET⁴ teve exigida a adoção do padrão de metadados Dublin Core e não a estrutura MARC já adotada para a conversão retrospectiva de seus registros bibliográficos.

⁴ O projeto MATH-NET é uma iniciativa de instituições alemãs, com o apoio do Deutsches Forschungsnetz (DFN) e do Deutsche Telekom, de criar uma rede de informações sobre matemática na Internet. O objetivo do projeto é estabelecer uma infra-estrutura técnica e organizacional para que informações eletrônicas de matemática possam ser disponibilizadas a baixo custo, com eficiência e de fácil uso pelos usuários em geral. A ênfase são informações sobre publicações, coleções de softwares e dados, atividades de ensino e pesquisa, bem como informações organizacionais e bibliográficas (CHATAIGNIER; SILVA, 2001, p. 9).

Deve-se assim, levar estas instituições à conscientização da importância de mudança de postura quanto à adoção da estrutura de representação MARCXML e o seu poder de influência para mudar as exigências da iniciativa de arquivos abertos, pois ao adotar uma estrutura completa, rica e já consagrada na área os benefícios e a qualidade dos metadados serão bem maiores levando-se em conta que, são estas representações que os usuários manipulam no processo de busca e recuperação da informação, e um ambiente computacional provedor de múltiplas formas de representação torna-se necessário para atender a diferentes tipos e necessidades dos usuários, pois conforme Marcondes (2001, p. 66),

O valor da representação consiste no fato de que ela pode economizar energia do usuário, tornando o processo de busca de informações mais econômico de um ponto de vista do seu dispêndio de energia; ler um resumo despende muito menos energia que a leitura do documento completo.

O atual ambiente informacional caracterizado pelos mais diferentes tipos de recursos informacionais como textos, imagens, sons, audiovisuais etc, demonstra a importância de representações de registros de metadados adequados para as atividades de busca e recuperação da informação por diferentes tipos e necessidades dos usuários.

A próxima seção apresenta o MARCXML como um padrão de descrição que possui uma estrutura completa, rica e já consagrada na área da Ciência da Informação, para recursos informacionais nos *open archives*.

4 MARC e XML

“A maior parte das idéias fundamentais da ciência são essencialmente simples e, por regra geral, podem ser expressas em uma linguagem compreensível para todos”.

Albert Einstein (1879-1955)

Aborda-se o conceito de metadados, o formato MARC, a linguagem de marcação XML e as ferramentas tecnológicas associadas.

A representação e a organização da informação na era digital, assim como os negócios e a indústria, sofreram grandes mudanças em quase todas as fases de seus processos. As mudanças incluem não somente os métodos e a tecnologia usada na criação de formas de representação de documentos e informações, mas também nos padrões, que são essenciais para o sucesso e a efetividade das ferramentas de busca e recuperação da informação.

Diante desse cenário, fundamentalmente enraizado em ambientes e infraestruturas para a informação digital, o cerne deste novo espaço informacional está na existência de padrões, ferramentas que possibilitam a diferentes tipos de dados serem intercambiados e entendidos por plataformas e sistemas heterogêneos (QIN, 2000).

Pode-se definir ferramenta como qualquer implemento ou objeto destinado a realizar uma operação ou executar um trabalho de qualquer natureza. Por exemplo, o escritor cria com a mente, mas necessita de uma ferramenta para anotar seus pensamentos. Um homem que dirige um caminhão, assim como uma pessoa que esteja utilizando um computador está fazendo uso de uma ferramenta, do mesmo modo que aqueles que os fabricaram. O emprego de ferramentas, o desenvolvimento da mecânica, a aplicação da

eletricidade, foram os indispensáveis pioneiros de nossos recentes aperfeiçoamentos da automação, da aplicação de controles eletrônicos e das tecnologias de informática (ARNOLD; WHITE, 1965).

Nesse sentido, as ferramentas, fruto das inovações tecnológicas, são as portas de entrada para o mundo globalizado que as unidades de informação de qualidade necessitam para estarem conectadas com o novo ambiente digital. A qualidade de uma Unidade de Informação é representada pela sua capacidade em tratar as informações por meio dos processos de armazenamento, recuperação e disseminação com qualidade e padronização e a isso podemos chamar de otimização na gestão da informação.

É com essa preocupação que “Bibliotecários e indexadores têm produzido e padronizado metadados por séculos” (MILSTEAD; FELDMAN, 1999, p. 1, tradução nossa), e o Formato de Intercâmbio *Machine Readable Cataloging* - **MARC21**, é um exemplo disso, ou seja, permite a construção de um registro catalográfico legível por computador e é uma estrutura e um padrão para a construção de metadados bibliográficos e catalográficos na atualidade.

A definição mais comum de metadados é dados sobre dados. Grácio (2002, p. 114) define metadados como “um conjunto de elementos que descrevem as informações contidas em um recurso, com o objetivo de possibilitar sua busca e recuperação”. Há também definições de outros autores. Para Souza, Vendrusculo e Melo (2000, p.93) metadado “significa dado sobre o dado. É a catalogação do dado ou descrição do recurso eletrônico”. Segundo Souza, Catarino e Santos (1997, p.2) metadados “são descrições de dados armazenados em banco de dados, ou como é comumente definido dados sobre dados a partir de um dicionário digital de dados”.

Nessa pesquisa entende-se por metadados um conjunto de dados usados para descrever e representar objetos informacionais, assim como, descrever o seu conteúdo e sua

localização e desta forma facilitar o acesso e recuperação de informações seja na Internet ou não, pois o termo metadado se refere a dado em sentido amplo e não somente para fontes eletrônicas. O resumo de um texto ou as suas palavras-chave, por exemplo, representam o texto original, são dados sobre outros dados, portanto considerados metadados.

De acordo com Milstead; Feldman (1999, p. 3, tradução nossa), “todos os motivos porque a indexação e a catalogação são necessárias para fontes impressas se aplicam ainda mais enfaticamente aos metadados para documentos eletrônicos”. Ainda segundo as autoras, metadado não é nada de novo para os pesquisadores profissionais, pois eles têm sido capazes de melhorar a precisão de suas buscas através do uso de vocabulários controlados ou limitando as buscas aos descritores, identificadores, autores, títulos, assuntos e outros por muitos anos. Embora metadado tenha se tornado um “*buzzword*”¹ no mercado de informação, o conceito é importante para autores e pesquisadores de informação eletrônica. Usado eficientemente, torna a informação acessível através de rótulos consistentes de seus conteúdos.

Para Milstead e Feldman, (1999, p. 1, tradução nossa),

Como o homem que esteve escrevendo prosa toda sua vida sem saber, os bibliotecários e indexadores têm produzido e padronizado metadados há séculos. Ignorando este legado, uma imensa variedade de outros profissionais tem recentemente entrado neste campo, e muitos deles não têm idéia de que alguém mais já “esteve ali e feito isto” antes. Vários sistemas estão sendo desenvolvidos para diferentes e algumas vezes até para o mesmo tipo de informação, resultando numa atmosfera caótica de padrões.

Não importa o nome que se use, catalogação, indexação ou metadados, os profissionais da informação estão familiarizados com os seus conceitos. “Agora o mundo eletrônico finalmente o descobriu. Até poucos anos atrás, somente alguns filósofos tinham

¹ Entende-se por “*buzzword*” um termo técnico que soa importante, geralmente usado para impressionar pessoas leigas no assunto.

ouvido a palavra ‘metadado’. Hoje, é difícil encontrar uma publicação que a ignore” (MILSTEAD; FELDMAN, 1999, p. 1, tradução nossa).

Metadados como um novo conceito surgiu da necessidade de se ter representantes para o conteúdo dos bancos de dados e que pudessem fornecer mais informações sobre eles, devido o crescente número e tipos de objetos digitais disponíveis.

A idéia de metadados como um novo conceito surgiu fora da arena tradicional dos textos e da bibliografia. Com o desenvolvimento dos arquivos de dados, especialmente dos dados geoespaciais, tornou-se evidente que seriam necessários representantes para fornecer mais informações sobre o conteúdo dos bancos de dados. Um crescente número e tipos de objetos foram disponibilizados digitalmente, mas foi reconhecido que os dados crus tinham pouco valor sem a informação sobre como foi coletado, seu propósito, os formatos, as plataformas para exibição e manipulação dos dados, suas restrições para reprodução e reutilização, assim como as informações de identificação mais convencionais como, autor, título, assunto e resumo (MILSTEAD; FELDMAN, 1999, p.3, tradução nossa).

Diante de tudo isso, os metadados em si não são algo novo, o que é novo hoje em dia são as variedades de padrões que estão sendo desenvolvidos e as formas de utilização. Os registros bibliográficos que se tem criado a muitos anos no mundo bibliotecário são essencialmente metadados. Proporcionam informação descritiva e analítica² sobre um objeto de informação. Os registros MARC nos Catálogos de Acesso Público Online - OPACs, assim como as fichas catalográficas nos catálogos manuais não são outra coisa senão um conjunto de metadados (ORTIZ-REPISO JIMÉNEZ, 1999).

Para que os metadados alcancem os seus objetivos com sucesso é imprescindível o uso de listas ou bases de dados com assuntos e autoridades controladas, como as linguagens controladas ou os tesouros.

² o autor utiliza o termo ‘analítica’ para se referir ao resultado da **análise documentária**, ou seja, a ‘informação temática’.

Por exemplo, se em um sistema de informação não for adotado um padrão para a descrição do campo de responsabilidade intelectual, por exemplo, para a descrição de **autor**, seriam permitidas várias formas de descrição, como por exemplo, **Silva, Maria Ap.**; **Silva, M. Ap.**; ou **Maria Aparecida Silva**, entre outras seriam possíveis, o que impediria a reunião eficiente e confiável das obras da autora, pois, diversas formas de descrição para a indicação de um mesmo autor foram efetuadas no sistema, ou seja, não foi usada uma forma autorizada ou padronizada disponível em bases de dados para a indicação de autoria, de padronização oficial como é o caso, por exemplo das bases **BIBLIODATA** e *Libray of Congress – LC*, e isso provocaria uma inconsistência nas informações obtidas, como resultado da busca.

Magda El-Sherbini (2000, p. 188, tradução nossa) em seu artigo sobre metadados e o futuro da catalogação faz a seguinte pergunta: **Por que bibliotecas precisam de metadados na era de recursos eletrônicos digitais?** Sua pertinente resposta esclarece que, para lançar esta pergunta, é necessário se lembrar de que os bibliotecários ainda estão lidando com informação que é criada para o mesmo propósito como o livro e o periódico, e então, sujeita à mesma classificação e leis de descrição como todo o resto. Ressalta ainda que as bibliotecas usam ferramentas tradicionais para organizar informação que facilita o acesso para materiais impressos e não impressos. AACR2, classificação LC, Classificação Decimal Dewey (CDD), CDU, LSCH, e outras ferramentas foram usadas e continuarão sendo usadas para descrever e classificar informação. Provavelmente as bibliotecas continuarão desenvolvendo e melhorando estes antigos padrões assim como elas se adaptaram às demandas das publicações e da indústria de informação.

A padronização é fundamental tanto para a representação (temática e descritiva) quanto para a recuperação da informação, sendo necessária à preocupação não somente com os campos para recuperação, mas com a padronização do seu conteúdo. Na *Web* os metadados podem ser elaborados no momento da criação de um objeto, como os *sites* por exemplo ou,

então, posteriormente, no caso da tradicional catalogação e indexação dos objetos informacionais, podendo ser confeccionados pelos profissionais bibliotecários no momento do processo da representação descritiva e temática.

No entanto, “na cultura da Web, o autor do documento é a pessoa que aplica o metadado, e como fazer com que milhões de pessoas não profissionais da informação compreendam a importância da catalogação até um certo nível e padrão quando nem sempre os profissionais concordam entre si?” (MILSTEAD e FELDMAN, 1999, p. 5, tradução nossa). No entanto, numerosos esforços vêm sendo feitos para catalogar os recursos da Internet.

“Desde que a informação eletrônica se comporte diferentemente que a mídia de impressão, ela requer tratamento especial para representá-la e classificá-la, e, é aí onde entram os metadados” (EL-SHERBINI, 2000, p. 188, tradução nossa). A autora ainda discorre sobre os usos dos metadados e cita algumas sugestões de outras autoras como Milstead e Feldman. Segundo El-Sherbini (2000 p.188, tradução nossa) os metadados são freqüentemente usados para:

- Organizar e manter dados de investimentos das organizações;
- Prover informação para catálogos de dados;
- Prover informação para auxiliar na transferência de dados;
- Auxiliar na descoberta efetiva e na recuperação da informação, permitir aos usuários determinar se o dados encontrados foram realmente os desejados;
- Controlar a informação de acesso restrito e prevenir alguns usuários (como crianças) de ter acesso a esses dados;
- Prover acordo comum como, qual elemento ou qual conteúdo deve ser usado;
- Fornecer informação sobre o uso dos dados como, condições legais de uso, seu tamanho, data e outros;
- Fornecer o histórico dos dados como, a fonte original e qualquer transformação posterior destes dados;

- Fornecer informação sobre o dono ou o criador do texto e
- Fornecer relacionamentos com outros recursos como links para versões atuais e futuras.

De acordo com El-Sherbini (2000 p. 188, tradução nossa), Milstead e Feldman (1999) outros usos para os metadados podem ser apontados:

- Garantir que todo o material sobre o mesmo assunto seja encontrado reunido, seja na prateleira ou em um banco de dados online;
- Separar conceitos importantes daqueles que são meramente incidentais no trabalho;
- Assegurar que a mesma informação seja encontrada para cada trabalho, e que ela seja colocada no mesmo lugar, de forma que alguém que esteja procurando os trabalhos por um autor de nome Fields não os encontre misturados com tratados agrícolas sobre fertilização de campos de trigo;
- O uso apropriado de vocabulários de indexação e de estruturas de campos, tanto na pesquisa quanto na catalogação, aumenta a precisão e minimiza a chance de resultados falsos.

Segundo El-Sherbini (2000, p. 188, tradução nossa), os Padrões de Metadados “têm um importante papel no suporte ao uso de serviços e recursos eletrônicos. No entanto, para maximizar o uso destes e assegurar consistência, são necessários padrões universalmente aceitos que orientem a criação de metadados”.

Os padrões de metadados da área da Ciência da Informação como, MARC, MARCXML entre outros são considerados ferramentas importantes que têm como objetivo conseguir um grau ótimo de normalização, padronização da representação descritiva automatizada dos dados informacionais, compatibilidade e transferência de dados legíveis por computador, permitindo a cooperação e compartilhamento de recursos e serviços entre diferentes unidades de informação.

É importante ressaltar a diferença existente entre metadados e formatos de metadados. Para Rosetto (2003) **metadados** são um conjunto de dados – atributos – referenciais, metodologicamente estruturados e codificados, conforme padrões internacionais, para localizar, identificar e recuperar pontos informacionais de textos, documentos e imagens disponíveis em meios digitais ou em outros meios convencionais. Já os **formatos de metadados**, para Rosetto (2003), referem-se a padrões que estabelecem regras para a definição de atributos (metadados) de recursos de informacionais, para obter coerência interna entre os elementos por meio de semântica e sintaxe; promover necessária facilidade para esses recursos serem recuperados pelos usuários e permitir a interoperabilidade dos recursos de informação.

4.1 MARC

O padrão de metadados *Machine Readable Cataloging* – **MARC**, ou seja, um registro catalográfico legível por computador, foi criado na década 1960 pela *Library of Congress* (LC) de acordo com o desenvolvimento dos recursos computacionais da época, com o objetivo de minimizar esforços, reduzir custos dos processos técnicos e compartilhar informações.

O formato MARC tornou-se um metapadrão sendo base para outros formatos e na década de 1970 contava-se com mais de 50 derivações do formato: **USMARC** – Estados Unidos, **UKMARC** - Inglaterra, **MONOCLE** - França, **CANMARC** - Canadá, **MARC/BR** - Bélgica, **IBERMARC** - Espanha, **MARC/MÉXICO** – México, **SAMARC** – África do Sul, **ANB MARC** – Austrália, **MAB1** – Alemanha, **DN/MARC** – Dinamarca, **PICA** – Holanda,

NISSAT – Índia, **MARCIS** – Israel, **ANNA MARC** – Itália, **JAPAN MARC** – Japão, **SWEMARC** – Suécia, **NORMARC** – Noruega, **CALCO** – Brasil, e outros.

Com tantos formatos, inevitavelmente surgiram diferenças e dificuldades no intercâmbio de informações. Depois de alguns estudos e tentativas para solucionar o problema de incompatibilidade, iniciou-se então em 1994 a harmonização dos formatos USMARC (desenvolvido pela LC) e o CANMARC (desenvolvido pela Biblioteca Nacional do Canadá). O nome aceito para os formatos harmonizados foi MARC21 que significa MARC para o século XXI.

A estrutura de um registro MARC21 é composta de três elementos básicos: **Líder**, **Diretório** e **Campos**. Os campos são subdivididos em um ou mais “subcampos” e definidos por “indicadores”. O padrão possui um formato de entrada, um formato de armazenamento interno e um formato de intercâmbio de dados catalográficos e bibliográficos. O **formato de entrada** é usado pelo catalogador na entrada dos dados. Veja a figura 8 como exemplo:

Líder	<u>LDR</u>	—	-----22-----4500
CPD[USE CTRL+F]	<u>001</u>	—	-----
Dados fixos	<u>008</u>	—	-----
ISBN	<u>020</u>	—	a
Fonte catal.	<u>040</u>	—	a
			b
Idioma [F8]	<u>041</u>	?	a
			b
Area geog. [F8]	<u>043</u>	—	a
Per. cron.[F2]	<u>045</u>	—	a
N. CDU	<u>080</u>	—	a
N. CDD	<u>082</u>	<u>04</u>	a
			z
Nome pessoal	<u>100</u>	?	a
			b
			c
			d
			e
			f
			g
			h
			i
			j
			k
			l
			m
			n
			o
			p
			q
			r
			s
			t
			u
			v
			w
			x
			y
			z
Nome corpor.	<u>110</u>	?	a
			b
			c
			d
			e
			f
			g
			h
			i
			j
			k
			l
			m
			n
			o
			p
			q
			r
			s
			t
			u
			v
			w
			x
			y
			z
Nome evento	<u>111</u>	?	a
			b
			c
			d
			e
			f
			g
			h
			i
			j
			k
			l
			m
			n
			o
			p
			q
			r
			s
			t
			u
			v
			w
			x
			y
			z
Título	<u>245</u>	<u>1?</u>	a
			b
			c
			d
			e
			f
			g
			h
			i
			j
			k
			l
			m
			n
			o
			p
			q
			r
			s
			t
			u
			v
			w
			x
			y
			z
Edição	<u>250</u>	—	a
			b
			c
			d
			e
			f
			g
			h
			i
			j
			k
			l
			m
			n
			o
			p
			q
			r
			s
			t
			u
			v
			w
			x
			y
			z
Imprensa	<u>260</u>	—	a
			b
			c
			d
			e
			f
			g
			h
			i
			j
			k
			l
			m
			n
			o
			p
			q
			r
			s
			t
			u
			v
			w
			x
			y
			z
Desc.fisica	<u>300</u>	—	a
			b
			c
			d
			e
			f
			g
			h
			i
			j
			k
			l
			m
			n
			o
			p
			q
			r
			s
			t
			u
			v
			w
			x
			y
			z
Serie - título	<u>440</u>	—	a
			b
			c
			d
			e
			f
			g
			h
			i
			j
			k
			l
			m
			n
			o
			p
			q
			r
			s
			t
			u
			v
			w
			x
			y
			z
Notas gerais	<u>500</u>	—	a
Nota bibliog.	<u>504</u>	—	a
Nota conteúdo	<u>505</u>	<u>0</u>	a
Não existe	<u>590</u>	—	a
Assunto-nome	<u>600</u>	<u>?4</u>	a
			b
			c
			d
			e
			f
			g
			h
			i
			j
			k
			l
			m
			n
			o
			p
			q
			r
			s
			t
			u
			v
			w
			x
			y
			z
Assunto-corp.	<u>610</u>	<u>?4</u>	a
			b
			c
			d
			e
			f
			g
			h
			i
			j
			k
			l
			m
			n
			o
			p
			q
			r
			s
			t
			u
			v
			w
			x
			y
			z
Assunto-evento	<u>611</u>	<u>?4</u>	a
			b
			c
			d
			e
			f
			g
			h
			i
			j
			k
			l
			m
			n
			o
			p
			q
			r
			s
			t
			u
			v
			w
			x
			y
			z

Assunto t3pico	<u>650</u>	<u>4</u>	a	_____
			x	_____
			y	_____
			z	_____
Assunto-geogr.	<u>651</u>	<u>4</u>	a	_____
			x	_____
			y	_____
			z	_____
C3d. As.CAPES	<u>697</u>		a	_____
Ent.sec.-nome	<u>700</u>	<u>?</u>	a	_____
			b	_____
			c	_____
			d	_____
			e	_____
Ent.sec.-corp.	<u>710</u>	<u>1</u>	a	_____
			b	_____
			c	_____
			d	_____
Ent.sec.-event.	<u>711</u>	<u>2</u>	a	_____
			b	_____
			c	_____
Ent.sec.-tit.a.	<u>740</u>	<u>?</u>	a	_____

FIGURA 8: Formato de entrada

FONTE: Banco de Dados ATHENA da UNESP, 2003.

Para Balby (1995, p. 31, grifo nosso) “um bom **formato de entrada** deve ser f3cil de entender, ser pr3ximo da forma de trabalho de quem coleta dados para o sistema, e possibilitar o preenchimento imediato, sem recurso a manuais separados de instru33es”.

O **Formato de Armazenamento Interno** diz respeito 3 aloca33o dos dados na mem3ria do computador. Suas caracter3sticas mais desej3veis s3o efici3ncia e economia no armazenamento e recupera33o dos dados (BALBY, 1995).

O **Formato de Interc3mbio** ou comunica33o 3 gerado por programa, a partir do formato de armazenamento interno, com o fim de possibilitar a incorpora33o e interpreta33o de dados bibliogr3ficos por sistemas e computadores diferentes. 3 voltado para a comunica33o de dados bibliogr3ficos entre computadores, e suas caracter3sticas mais desej3veis s3o a efici3ncia e economia na troca de dados. Um formato de interc3mbio deve permitir a comunica33o de dados entre computadores diferentes com o m3nimo disp3ndio em reformata33o dos dados (BALBY, 1995). Veja as figuras 9 e 10 a seguir :

```

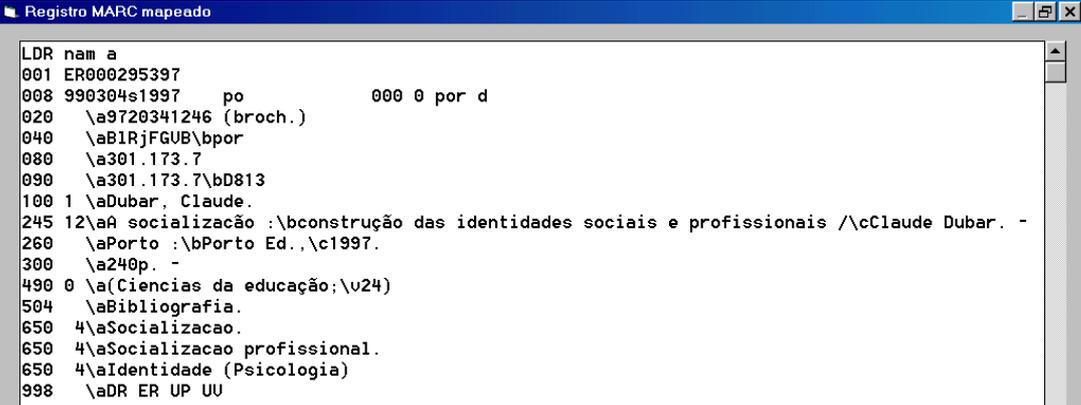
00644nam 2200217 a
4500001001200000000004100012020002400053040001800077080001400095090002000109100001900129245009200
1482600031002403000012002714900032002835040018003156500018003336500031003516500028003829980016004
100ER000295397990304s1997 po 000 0 por d 9720341246 (broch.)
aB1RjFGUBbpor a301.173.7 a301.173.7bD8131 aDubar, Claude.12aA socialização
:bconstrução das identidades sociais e profissionais /cClaude Dubar. - aPorto :bPorto
Ed.,c1997. a240p. -0 a(Ciências da educação;\v24) aBibliografia. 4aSocializacao.
4aSocializacao profissional. 4aIdentidade (Psicologia) aDR ER UP UV

```

FIGURA 9: Formato de intercâmbio

FONTE: Base de dados BIBLIODATA em CD-ROM, 2003.

Para bibliotecários e especialistas, a exibição para a utilização é feita a partir de etiquetas numéricas, conforme a figura abaixo. Esta forma de exibição é proporcionada pelo software do sistema.



```

Registro MARC mapeado
LDR nam a
001 ER000295397
008 990304s1997 po 000 0 por d
020 \a9720341246 (broch.)
040 \aB1RjFGUB\bpor
080 \a301.173.7
090 \a301.173.7\bD813
100 1 \aDubar, Claude.
245 12\aa socialização :\bconstrução das identidades sociais e profissionais /\cClaude Dubar. -
260 \aPorto : \bPorto Ed.,\c1997.
300 \a240p. -
490 0 \a(Ciências da educação;\v24)
504 \aBibliografia.
650 4\aaSocializacao.
650 4\aaSocializacao profissional.
650 4\aaIdentidade (Psicologia)
998 \aDR ER UP UV

```

FIGURA 10 : Exibição com etiqueta

FONTE: Base de dados BIBLIODATA em CD-ROM, 2003.

As telas que são exibidas para os usuários finais dos sistemas de informação, são telas de exibição formatadas visto que, as etiquetas MARC21 são incompreensíveis para o público em geral. Conforme a figura 11 a seguir:



FIGURA 11 : Exibição formatada

FONTE: Base de dados BIBLIODATA Online, 2003.

O padrão MARC comporta qualquer esquema de descrição bibliográfica, entretanto as regras da *Anglo-American Cataloguing Rules* - 2ª edição (AACR2) é que têm sido as mais utilizadas para a descrição de recursos informacionais. Para a normalização no armazenamento dos dados utiliza a norma internacional de padronização ISO 2709 e para a recuperação e intercâmbio de dados catalográficos e bibliográficos em ambiente automatizado utiliza o Protocolo Z39.50 que serão descritos a seguir:

4.1.1 PROTOCOLO Z39.50

No contexto da Internet, protocolo é um conjunto de regras que os computadores utilizam para se comunicarem entre si através das redes. A Internet por exemplo, não seria hoje uma realidade sem o protocolo TCP-IP (*Transmission Control Protocol/Internet Protocol*) que possibilita a comunicação entre os computadores. Da mesma forma, o

protocolo **Z39.50** possibilita o acesso simultâneo aos catálogos *online* de várias bibliotecas via Internet que utilizam esse protocolo para a comunicação e recuperação de dados.

Esse conjunto de regras que administra a comunicação entre os diferentes componentes de um sistema computacional são denominados ‘PROTÓCOLOS’ em alusão aos protocolos usados na sociedade para administrar as relações humanas. Por meio dos protocolos da rede é que se definem os detalhes de cada atividade, incluindo o modo como são enviadas as mensagens, a maneira como a autorização para transmitir mensagens é delegada às máquinas, e a forma como são manipuladas as tarefas de empacotar e desempacotar mensagens para a transmissão (BROOKSHEAR, 2000).

Ou seja, são comandos preestabelecidos entre os componentes de uma rede para que os computadores envolvidos “falem” entre si. Sua finalidade é permitir que os computadores se interliguem e troquem informações com o menor número de erros possível. Os protocolos são, geralmente, documentos em páginas e páginas de material impresso, porém, como ocorre com tantas outras tecnologias complexas, o usuário não precisa conhecê-los em detalhes para que possa deles se utilizar (KROL citado por FERREIRA, 1994; ROSETTO, 1997).

O protocolo Z39.50 foi desenvolvido num projeto conhecido como *Linked Systems Project – LSP*, no início de 1980 . Este projeto envolveu a *Library of Congress - LC* (Biblioteca do Congresso Norte Americano), *Online Computer Library Center - OCLC*, *RLIN* e a *Western Library Network - WLN*. O projeto LSP construiu uma rede experimental para transferir registros bibliográficos entre os participantes e desenvolveu um precursor do protocolo Z39.50.

Em 1984 o protocolo desenvolvido pelo **LSP** foi enviado para a *National Information Standards Organization - NISO*, *American National Standards Institute –*

ANSI, publicadores e provedores de serviços de informação para adicionais desenvolvimento e implementações e assim ser considerado um padrão nacional dos Estados Unidos.

A primeira versão do protocolo Z39.50 foi adotada como um padrão nacional dos Estados Unidos em 1988. Depois desta primeira versão surgiram outras com várias modificações e implementações. Surgiu então a necessidade de se adotar somente uma versão do protocolo. As instituições envolvidas entraram em um consenso e em 1998 foi aprovada oficialmente a versão Z39.50 de 1995 normalizada pela **ISO 23950**, como padrão internacional.

O protocolo Z39.50 embora originalmente tenha surgido no contexto das bibliotecas e do mundo bibliográfico, não há nada no protocolo Z39.50 que restrinja seu uso somente a esses tipos de dados. Ele foi especificamente projetado para ser um mecanismo capaz de manipular qualquer tipo de dados como, imagens, documentos de texto completo, dados bibliográficos e catalográficos e outros, permitindo a pesquisa e recuperação dessas informações em redes de computadores distribuídos (NEEDLEMAN et al, 2001).

O protocolo Z39.50 utiliza-se de um modelo cliente/servidor para possibilitar a busca e recuperação de informações na Internet. Permite a um usuário de um sistema pesquisar e recuperar informações de outros sistemas. Para Abiteboul, Buneman e Suciú (2000) uma arquitetura cliente/servidor funciona da seguinte forma: O cliente (uma pessoa ou um programa) emite uma consulta que é processada, compilada em um código otimizado e executada, e dados de resposta são devolvidos pelo servidor.

Para Cörte et al (1999, p. 247) “o padrão Z39.50 é um padrão de “*middleware*” cliente-servidor”. Segundo Abiteboul, Buneman e Suciú (2000, p. 5) “*middleware* é o software que transforma, integra ou senão adiciona valor aos dados”. Ficando entre a comunicação do cliente com o servidor.

Com o uso deste padrão é possível a interoperabilidade, ou seja, a compatibilidade entre vários sistemas diferentes. Permite-se que determinado servidor possa prestar serviços para diversos clientes ou fornecedores, e um determinado servidor possa usar os serviços de diversos servidores ou fornecedores diferentes. Côrte et al (1999, p. 247) esclarece melhor com o seguinte exemplo:

[...] se este protocolo não existisse, a biblioteca que quisesse recuperar registros bibliográficos de vários bancos de dados deveria dispor de tantos clientes quantos fossem os catálogos ou bancos de dados que se quisesse consultar.

Portanto, o protocolo Z39.50 é considerado uma ferramenta tecnológica utilizada para facilitar o processo de busca, recuperação e intercâmbio de informações em ambiente digital, e sendo melhor utilizado pelos analistas de sistemas do que pelos bibliotecários. Atualmente esse protocolo (amplamente divulgado entre os sistemas de bibliotecas) vem sendo adaptado ao âmbito da Web sob o nome de *Z39.50 International: Next Generation – ZING*, com a promoção da LC e foi,

Desenvolvido por um grupo editorial internacional, examinado através de um *listserv* aberto, a versão 1.1 foi lançada no início de 2004 (uma versão 1.0 experimental já havia sido lançada em 2002) e já tem sido incorporada em alguns produtos, inclusive, pode-se acessar a *Library of Congress*, por meio de um portal SRW/SRU (McCALLUM, 2004, p.4).

O ZING comporta protocolos de recuperação de informação denominados **SRW**³ e **SRU**⁴. Este último, SRU – *Service URL of Search and Retrieval* (Serviço URL de busca e Recuperação) – permite aos usuários fazer uma busca usando título, nome, identificador e outros parâmetros via uma URL e obter registros como resposta. O SRW – *Service Web of Search and Retrieval* (Serviço Web de Busca e Recuperação) – mantém um serviço que opera

³ <http://www.loc.gov/z3950/agency/zing/srw>

⁴ <http://www.loc.gov/standards/sru/cql/>

na Web baseado no protocolo SOAP, no lugar de fazê-lo diretamente através de HTTP. É mais robusto que o SRU e pode ir diretamente sobre o HTTP (McCALLUM, 2004).

Esses protocolos de busca e recuperação, baseados em XML, são semanticamente compatíveis com o protocolo Z39.50, adaptando as partes mais úteis deste último ao âmbito Web. Isto permite aos usuários prover portais SRW/SRU para aplicações Z39.50 já existentes (McCALLUM, 2004).

Segundo McCallum (2004, p.4),

Este empreendimento resultou também no desenvolvimento de uma Linguagem Comum de Consulta – Common Query Language – CQL⁴. A CQL tenta resolver a brecha existente entre os distintos métodos de linguagens de busca: aqueles que são potentes mas complexos e difíceis de usar, como o Xquery ou SQL; e aqueles como o Google que são simples e fáceis de usar mas de limitada funcionalidade.

Sem dúvida, conforme Agenjo (2005), é necessário dizer que os novos desenvolvimentos do Z39.50, até chegar ao atual ZING, estão se aproximando ao Protocolo da OAI justamente através dos SRW/SRU.

Autores como Sanderson, Young e Le Van (2005) apontam para a sinergia surgida entre os dois protocolos (ZING e OAI-PMH):

Um exemplo perfeito disto são os arquivos dos repositórios OAI que Thomas Habing criou para a UIUC. Os registros nos repositórios são descritos usando um esquema Dublin Core e o esquema de descrição de serviço ZeeRex. O registro pode ser buscado via SRU, coletado via OAI, monitorado via um RSS baseado em serviço de alerta, e buscado ou manuseado interativamente por um browser via uma interface HTMLWeb. [...] Um arquivo similar de serviços de SRW também está disponível com as mesmas características do arquivo do repositório do OAI. O arquivo SRW também tem vários serviços de valor-agregado, incluindo um *gateway* que pode habilitar para as transformações de registro. Assim, por exemplo, se um repositório SRW somente suportar um esquema de registro MARC, acessando-o através do gateway SRU do arquivo proverá um host de outros formatos de saída,

feitos a partir de um ‘catálogo’ de metadados XSLT disponíveis (SANDERSON; YOUNG; LE VAN, 2005, tradução nossa).

Isto é, SRW/SRU e OAI-PMH complementam um ao outro. Embora os dois protocolos tenham escolhido diferentes respostas para certas questões, isto não os impedem de fazerem parte de configurações muito diferentes e interessantes. A baixa barreira do OAI-PMH para implementação e metas específicas faz dele de fácil recomendação para qualquer um implementar, considerando que o SRW é um pouco mais complexo, mas reproduz as funções essenciais do Z39.50, facilitando, assim, a busca distribuída no lugar do *harvesting* (SANDERSON; YOUNG; LE VAN, 2005).

A sinergia surgida entre o OAI-PMH e os protocolos de recuperação de informação dentro do ZING (SRW/SRU) de acordo com Agenjo (2005) era previsível, pois, em última instância ambos os protocolos não deixam de buscar o mesmo objetivo, ou seja, o acesso à difusão da informação bibliográfica, a partir de determinados padrões.

4.1.2 NORMA ISO 2709

Uma norma é uma fórmula que tem valor de regra, em geral indicativo e algumas vezes imperativo. Ela define as características que deve ter um objeto e as suas características de uso, bem como as características de um procedimento e /ou de um método. É adotada como base ou medida para a realização ou avaliação de algo (FERREIRA, 1999; GUINCHAT e MENO, 1994).

A normalização para os dados catalográficos e bibliográficos é fundamental para que possa haver o compartilhamento, a cooperação e a interoperabilidade entre várias unidades de informação, algo indispensável hoje em dia, possibilita a simplificação e

racionalização dos métodos e técnicas utilizadas padronizando os dados informacionais e desta forma diminuindo custos e tempo para efetuar o intercâmbio de informações.

A *International Standardization Organization* - **ISO**, ou seja, Organização Internacional para Padronização é o organismo principal de normalização mundial e sua atividade se estende a todos os campos. É uma organização não governamental estabelecida em 1947. Tem por objetivo promover o desenvolvimento da padronização e atividades relacionadas no mundo inteiro, para facilitar a troca internacional de bens e serviços, e a cooperação nas esferas intelectual, científica, tecnológica e econômica (ISO, 2002), assim como a *World Wide Web Consortium* – W3C, organização criada em 1994 e que tem por objetivo padronizar e normatizar os padrões para a *Web*, conduzindo-a a seu potencial máximo e garantindo a sua interoperabilidade (W3C, 1994-2002).

O trabalho da ISO resulta em acordos internacionais que são publicados como Padrões Internacionais, ou seja, documentos gerados a partir de acordos e que contêm especificações técnicas ou outros critérios que são precisos para serem usados constantemente como regras, diretrizes, ou definições de características, para assegurarem que materiais, produtos, processos e serviços são ajustados, padronizados para servir a um determinado propósito. Por exemplo, o formato dos cartões de crédito e cartões telefônicos que se tornaram comuns, é derivado de uma norma internacional padrão ISO. Aderindo ao padrão que define tal característica como uma espessura de (0,76 mm), os cartões podem ser usados mundialmente (ISO, 2002).

Portanto, os padrões internacionais contribuem para tornar a vida mais simples, e aumentar a confiança e efetividade dos bens e serviços que todos nós usamos (ISO, 2002). Com este mesmo propósito se aplica a norma **ISO 2709** para a normalização do formato MARC21. A norma internacional de padronização ISO 2709 (*Documentation Format for Bibliographic Interchange on Magnetic tape*), foi desenvolvida pelo Comitê Técnico

ISO/TC46, encarregado da elaboração de normas no campo da documentação, juntamente com subcomitê **SC4**- encarregado da automação em documentação.

A norma ISO 2709 é a versão internacional da norma nacional norte-americana **ANSI Z39.2**⁵ (atual **NISO Z39.2**) que especifica os requisitos para o formato de intercâmbio de registros bibliográficos que descrevam todas as formas e tipos de materiais e dados relacionados como autoridade, cabeçalhos de assunto, circulação e outros sujeitos à descrição bibliográfica. Por exemplo, o MARC21 está baseado em uma **estrutura de formato ISO 2709**. Essa norma não define a extensão do conteúdo de documentos individuais, nem designa significado algum para os parágrafos, indicadores ou identificadores, sendo essas especificações as funções dos formatos de implementação. A ISO 2709 apresenta uma estrutura geral, ou seja um molde ou esqueleto projetado especialmente para a comunicação entre sistemas de processamento de dados e não para uso como formato de processamento dentro dos sistemas (CÔRTE et al., 2000).

A forma como os dados estão estruturados em meio magnético possibilitam o intercâmbio de registros bibliográficos. Mesmo assim, esta característica não elimina a incompatibilidade entre os registros que utilizam diferentes formatos de entrada e, principalmente, diferentes regras de entrada de dados.

Ou seja, a estrutura do formato, mesmo quando semelhante, não garante a compatibilidade do registro. É preciso considerar o **conteúdo dos registros** que irão influenciar fortemente na recuperação da informação. É o que trata a próxima subseção.

⁵ O *American National Standards Institute* - **ANSI**, é uma organização privada sem fins lucrativos que administra e coordena a padronização norte-americana, fundada em 1918. O objetivo do Instituto é aumentar a competitividade global de negócios dos Estados Unidos e a qualidade de vida norte-americana promovendo padrões e seus sistemas de avaliação (NISO, 1994).

4.1.3 AACR, FRBR e RDA

Los usuarios son la razón por la cual catalogamos.
(Joint Steering Committee for Revision of AACR, 2005).

Vale ressaltar que, esse assunto não é o foco desta pesquisa, mas serão feitas algumas considerações a respeito, pois são as regras, os esquemas, os modelos e as práticas de catalogação que provêm o conteúdo para as estruturas de dados definidas pelos formatos de intercâmbio de dados.

Originalmente publicada em 1978 e revisada em 1988 em edições americana e britânica, as regras de catalogação Anglo-Americana (**AACR2**), por exemplo, **padronizam a entrada dos dados** no formato MARC 21, definindo os campos para determinado item documentário e a entrada dos seus respectivos dados.

As bibliotecas têm tradicionalmente empregado a abordagem ‘filosófica’ MARC/AACR2 para metadados. Este modo emprega um único esquema para cobrir todos os tipos de trabalhos e todos os grupos de usuários. Como tipos novos de recursos informacionais surgem, campos novos são acrescentados à estrutura/esquema MARC/AACR2, ou as regras para campos existentes são mudadas para acomodar estes novos recursos. E como as comunidades emergem com novas necessidades de metadados, estes também são incorporados no esquema existente. A ‘filosofia’ MARC/AACR2 mantém que este grande esquema deveria servir a todas as necessidades dos usuários e para todos os tipos de trabalhos. Os críticos desta abordagem mostram que o esquema se tornou complexo demais e que somente especialistas altamente capacitados (os bibliotecários) podem ordenar o uso destes metadados, e que o sistema é muito lento para adaptar a tipos emergentes de recursos eletrônicos e digitais. Eles também reivindicam que grupos de usuários

freqüentemente têm necessidades de metadados que os controladores do MARC/AACR2 estão pouco dispostos a acomodar (BESSER, 2002).

Conforme Moreno e Arellano (2005, p. 22), “parece-nos, portanto, que o foco da representação descritiva sempre se deu a partir das necessidades do usuário-meio, quais sejam, o bibliotecário e, não sob a perspectiva do usuário-fim”. Após quatro décadas dos Princípios de Paris⁶ e do documento **ISBD** – *International Standard Bibliographic Description* – publicado pela **IFLA** – *International Federation Library Associations and Institutions* – em 1971 (os quais regem os códigos nacionais, inclusive o AACR), iniciou-se uma reavaliação teórica e prática da representação descritiva (catalogação), em âmbito internacional. Assim em dezembro de 2003, na Reunião da IFLA de especialistas sobre um Código Internacional de Catalogação, realizada em Frankfurt, na Alemanha, foram estabelecidos os

[...] novos princípios que substituem os Princípios de Paris e os ampliam de somente obras textuais para todo tipo de materiais e de somente a eleição e forma de registro para todos os aspectos dos registros bibliográficos e de autoridade utilizados nos catálogos de bibliotecas (DECLARACIÓN DE PRINCIPIOS INTERNACIONALES DE CATALOGACIÓN, 2003, tradução nossa).

Os seguintes princípios preliminares incluem:

1. Alcance
2. Entidades, Atributos e Relações
3. Funções do Catálogo
4. Descrição Bibliográfica
5. Pontos de Acesso

⁶ A Declaração de Princípios – conhecida comumente como os Princípios de Paris (1961) – foi aprovada pela Conferencia Internacional sobre Princípios de Catalogação em 1961. Tem-se conseguido, sem nenhuma dúvida, seu objetivo de servir como base para a normalização internacional da catalogação: a maioria dos códigos de catalogação que se desenvolveram no mundo desde então seguiram os Princípios estritamente ou, ao menos, em um alto grau (DECLARACIÓN DE PRINCIPIOS INTERNACIONALES DE CATALOGACIÓN, 2003).

6. Registros de Autoridade
7. Fundamentos para as Capacidades de Busca

Estes novos princípios (**Princípios Internacionais de Catalogação**) têm a intenção de orientar o desenvolvimento dos códigos de catalogação dando um enfoque consistente à catalogação descritiva e à catalogação de recursos bibliográficos de todo tipo, além da conveniência dos usuários do catálogo.

Segundo o documento preliminar da *Declaración de Principios Internacionales de Catalogación* (2003), estes novos princípios se constroem sobre a base das grandes tradições catalográficas do mundo⁷ e também sobre os modelos conceituais dos documentos da IFLA *Functional Requirements for Bibliographic Records* – Requisitos Funcionais para Registros Bibliográficos (**FRBR**) e *Functional Requirements and Numbering for Authority Records* (**FRANAR**), que estendem os Princípios de Paris ao campo da catalogação por assunto (representação temática).

Entre os anos de 1992 a 1997, o grupo de estudos proveniente da seção de catalogação e da seção de indexação da IFLA desenvolveu um modelo de entidade-relacionamento como uma visão generalizada do universo bibliográfico, que pretendia ser independente de qualquer código de catalogação e que inclui uma descrição do modelo conceitual (as entidades, relações e atributos ou *metadados*). O modelo, publicado em 1998, é denominado como Requisitos Funcionais para Registros Bibliográficos (**FRBR**) (TILLET, 2004; MORENO; ARELLANO, 2005).

⁷ CUTTER, Charles A. **Rules for a dictionary catalog**. 4. ed. rev. Washington D. C.: Government Printing office, 1904.

RANGANATHAN, S. R. **Heading and canons**. Madras [India]: S. Viswanathan, 1955.

LUBETZKY, Seymour. **Principles of cataloging**. Final Report. Phase I: Descriptive Cataloging. Los Angeles, Calif: University of California, Institute of Library Research, 1969.

Desde a 56ª IFLA Council General Conference (Estocolmo, 1990), a IFLA havia apostado por uma catalogação mais simples e em consonância com as possibilidades técnicas atuais. Os FRBR são a culminância das múltiplas propostas feitas durante a segunda metade do século XX para que as regulamentações catalográficas fossem breves e essencialmente práticas (LÓPES GUILLAMÓN, p. 133-134).

Segundo Moreno e Arellano (2005, p. 23, grifo dos autores) os FRBR configuram “[...] uma recomendação para **reestruturar** os registros bibliográficos de maneira a refletir a estrutura conceitual de **buscas** de informação, levando em conta a diversidade de”:

- Usuários – usuários de bibliotecas, pesquisadores, bibliotecários, da seção de aquisição, publicadores, editores, vendedores;
- Materiais – textuais, musicais, cartográficos, audiovisuais, gráficos e tridimensionais;
- Suporte físico – papel, filme, fita magnética, meios óticos de armazenagem, etc. e;
- Formatos – livros, folhas, discos, cassetes, cartuchos, etc. que o registro possa conter (MORENO; ARELLANO, 2005, p. 24).

Os FRBR apresentam conceitos e definições de **entidades**, **relacionamentos** e **atributos**, sob uma perspectiva inovadora do objeto bibliográfico, centrado no usuário e suas ações com as diferentes entidades, ou seja, tarefas⁸ como: **encontrar** (localizar e reunir), **identificar**, **selecionar**, **obter**; além de: **relacionar/navegar**, atribuir direitos de autor, preservar.

Segundo Chen (1990), citado por Moreno & Arellano (2005, p. 26, grifo dos autores),

Entidade é aqui entendida como uma “*coisa*” ou um “*objeto*” no mundo real que pode ser identificada de forma unívoca em relação a todos os outros objetos. Uma entidade pode ser concreta ou abstrata. Por sua vez, **atributos** são as diversas

⁸ Estas tarefas do usuário reforçam os objetivos tradicionais do catálogo, descritos por Cutter em 1876 de permitir ao usuário encontrar e colocar juntas as obras (TILLET, 2004).

características que um tipo de entidade possui, ou propriedades descritivas de cada membro de um conjunto de entidades. Um **relacionamento** ‘é uma associação entre uma ou várias entidades’.

No modelo conceitual dos FRBR, o universo bibliográfico consiste em diversas entidades que se relacionam entre si e podem ser descritas mediante elementos de informação (ou atributos). Estas entidades, segundo Tillett (2003), se classificam em três grupos:

O **Grupo 1** são os produtos do esforço intelectual e artístico que são mencionados ou descritos nos registos bibliográficos: **obra**, **expressão**, **manifestação** e **item** (Figura 12).

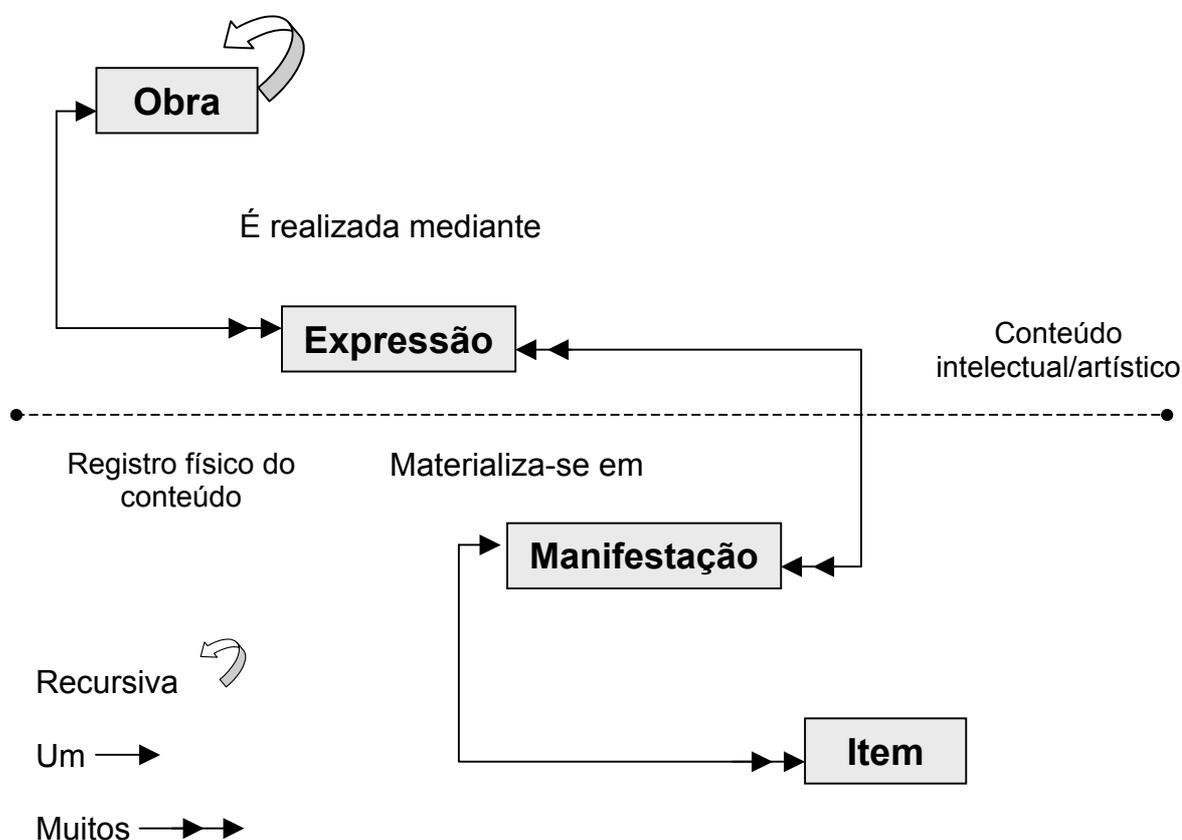


FIGURA12: Entidades do “Grupo 1” dos FRBR.
FONTE: TILLET (2003)

Obra é uma entidade abstrata, *uma criação intelectual ou artística distinta*. A entidade **Expressão** de uma obra é a *realização intelectual ou artística específica que assume uma obra ao ser realizada*, excluindo-se aí aspectos da alteração da forma física. Uma **Manifestação** é a *materialização de uma expressão de uma obra*, ou seja, seus suportes físicos, que podem ser livros, periódicos, kits multimídia, filmes, etc, que é representada pelo **Item**, um *único exemplar de uma manifestação*.

As duas últimas entidades *refletem a forma física*, são entidades concretas, enquanto as duas primeiras refletem o conteúdo intelectual ou artístico (MORENO; ARELLANO, 2005, p. 27, grifo dos autores).

Os FRBR concentram sua atenção nas entidades do Grupo 1. O FRANAR (*Functional Requirements for Authority Numbers and Records*), se ocupa das entidades do **Grupo 2 (Pessoa e Entidade corporativa)**, ou seja, as entidades responsáveis pelo conteúdo intelectual ou artístico, a produção e disseminação física ou a custódia de tais produtos (TILLET, 2003).

As entidades do **Grupo 3** dos FRBR são as que servem como assuntos de obras. Estas podem ser: **Conceitos, objetos, eventos, lugares** e qualquer das entidades do Grupo 1 ou 2. Por exemplo: pode-se ter uma obra acerca de outra obra ou uma obra acerca de uma pessoa ou entidade corporativa, etc (TILLET, 2003; MORENO; ARELLANO, 2005).

A terminologia FRBR, suas relações e as tarefas do usuário estão auxiliando a revisão das regras e princípios da catalogação tradicional, ‘a luz do atual ambiente digital’ (TILLET, 2003, 2004).

Assim, com o patrocínio da *Library of Congress* (LC) aconteceu em Washington em novembro de 2000 a *Bicentennial Conference of Networked Resources and the Web*. Nesta conferência se discutiram as perspectivas de futuro do AACR2 diante das novas propostas de metadados, o estado das autoridades bibliográficas ou a tecnologia web como meio de apresentação dos catálogos, entre outros.

Desde 2002 o *Joint Steering Committee for Revision of AACR* – Comitê de Direção para a Revisão do AACR – (**JSCAACR**) vem, neste sentido, trabalhando no desenvolvimento de um “[...] plano estratégico para as regras de catalogação Anglo-americanas [...]” (JOINT STEERING COMMITTEE FOR REVISION OF AACR, p. 16). ou seja, um novo conjunto de regras, o que seria o **AACR3: Resource Description and Access**

(**RDA**)⁹. O comitê, surgido em 1966 como criação das associações profissionais, respectivas, dos Estados Unidos e do Reino Unido, juntamente com o Comitê em Catalogação, Descrição e Acesso – *Committee on Cataloguing, Description & Access* – são os responsáveis da atualização periódica das *Anglo-American Cataloguing Rules* (AACR), as quais exercem, pela extensão de sua aplicação, pelo grau de desenvolvimento e atualidade de suas normas, a liderança entre os códigos de catalogação no panorama internacional (LÓPEZ GUILLAMÓN, 2004).

Com a dinamização dos recursos eletrônicos e digitais, devido aos desenvolvimentos tecnológicos e ao intenso uso da Internet, “estamos agora vendo emergir outras estruturas no mundo digital e novas formas de marcar a informação que também descrevem recursos e provêm acesso à informação” (JOINT STEERING COMMITTEE FOR REVISION OF AACR, 2005, p. 14, tradução nossa). Neste sentido,

As regras de catalogação necessitam permanecer independentes de qualquer formato de comunicação. Também [devem] prover um conteúdo normalizado para os elementos da descrição e o acesso bibliográfico que pode ser usado por qualquer uma das normas de metadados emergentes como o Dublin Core. As normas para metadados nos dão as categorias dos elementos de dados a serem incluídos no registro, mas normalmente não nos dizem como estruturar o conteúdo destes elementos o que deveríamos usar como fonte para encontrar o conteúdo desses elementos – como, por exemplo, “título” ou “data” (JOINT STEERING COMMITTEE FOR REVISION OF AACR, 2005, p. 14, tradução nossa).

São Propósitos do **RDA**, segundo o *Joint Steering Committee for Revision of AACR* (2005):

- Permitir aos usuários de catálogos de bibliotecas, encontrar e usar recursos apropriados a suas necessidades de informação;

⁹ O *Joint Steering Committee for Revision of AACR* (2005, p. 1, tradução nossa) sinaliza para que o nome seja abreviado para **RDA** pois preferem “[...] que a sigla RDA se use em todos os idiomas para maior reconhecimento”.

- Apoiar os conceitos FRBR relativos às tarefas dos usuários (encontrar, identificar, selecionar, obter);
- E ir mais além dos conceitos FRBR relativos às tarefas dos usuários (uso de recursos, administração de coleções, sistemas de navegação).

O novo código (com data prevista para a 1ª edição em 2008) incluirá regras de conteúdo e exemplos atualizados quando necessário. Incorporará o conceito de controle de autoridade e os conceitos dos **FRBR** ao mesmo tempo em que reduzirá as atuais redundâncias das regras e aumentará a consistência através de diferentes tipos de conteúdos. Uma versão do novo código será baseada na Web, que terá maior funcionalidade (JOINT STEERING COMMITTEE FOR REVISION OF AACR, 2005).

Assim como os padrões de conteúdos (AACR, e outros) vêm sofrendo consideráveis revisões para acomodar a catalogação de recursos, sobretudo digitais, o padrão de dados estruturados MARC também vem passando por consideráveis revisões/adaptações para o ambiente da web.

4.2 XML

A XML, uma linguagem de marcação estendida, baseada na **SGML**¹⁰ (*Standard Generalized Markup Language*), é o resultado dos esforços de um grupo de especialistas estabelecido em 1996 pelo **W3C**¹¹ (*World Wide Web Consortium*) e liberada oficialmente em 1998, foi concebida especialmente para gerenciar, armazenar e transmitir dados via Internet.

¹⁰ (Meta)liguagem desenvolvida há aproximadamente 30 anos (EDWARDS, 1997 citado por BAX, 2001). Foi base para um novo paradigma numa tentativa de padronização dos diversos formatos empregados para representar a informação. Reconhecida como padrão ISO em 1986.

¹¹ W3C - www.w3c.org - é a organização que se encarrega do desenvolvimento e manutenção dos padrões da Web.

Segundo Ray (2001, p.2) a XML “[...] é um conjunto de símbolos que pode ser colocado no texto de um documento para demarcar e rotular as partes desse documento”. Por exemplo: **<titulo>** *Mente e Natureza* **</tilulo>**, **<autor>** Gregory Bateson **</autor>**, as palavras entre “*Tags*”, marcas ou marcação correspondem, respectivamente, ao título e ao autor do documento. Ou seja, a XML “fornece pistas semânticas na forma de *tags* (marcas) sobre o significado dos dados” (LIBERTY; KRALEY, 2001, p. 5).

Segundo Furgeri (1999, p. 26) “Um documento XML é composto, basicamente, de três elementos distintos”:

- Conteúdo dos dados: são as informações entre as *Tags*;
- Estrutura: a organização dos elementos dentro do documento (pode possuir a estrutura de um memorando, contrato, receita etc.);
- Apresentação: é a forma com que as informações são apresentadas ao leitor do documento.

Com inúmeras vantagens a XML possibilita, entre outras (SANTOS; FLAMINO, 2004; ORTIZ-REPISO JIMÉNES, 1999; FURGERI, 1999):

- Pode ser usada para criar registros catalográficos e para criar os textos eletrônicos do catálogo;
- O armazenamento da informação e sua descrição num mesmo local, ou seja, os documentos XML são essencialmente banco de dados de informações, podendo ser consultados e processados como qualquer banco de dados;
- Não há limitações para os tamanhos dos documentos;
- Sua estrutura permite a indexação e a recuperação do documento na íntegra, ou as partes do documento, mediante busca booleana, adjacência, proximidade, ranking de relevância e outros; O documento de mesmo

conteúdo pode ser visualizado de diferentes formas, para diferentes usuários, através da utilização de folhas de estilo;

- Permite alto grau de automação para agentes de software (agentes inteligentes);
- A XML possibilita a entrega de qualquer tipo de dados estruturado por toda a WEB. Aplicações XML padronizadas possibilitarão que diferentes aplicativos trabalhem em conjunto, significando uma maior **Interoperabilidade**, ou seja, a compatibilidade entre sistemas e plataformas heterogêneos.

Para Furgeri (1999, p. 26),

A XML torna o documento mais “inteligente” criando uma estrutura que mostra claramente qual é o significado das informações, pois elas estão agrupadas, e não dispersas pelo documento. Essas novas etiquetas são criadas a critério do desenvolvedor, [...] Sendo assim as etiquetas são criadas conforme as necessidades dos autores do documento.

A XML permite marcar os documentos levando em conta a semântica da informação, ou seja, cada pessoa ou instituição pode criar um conjunto de etiquetas que melhor represente os elementos que compõem os seus documentos (MARTÍNEZ GONZALEZ, 2000).

A esta propriedade que possibilita a criação de novas etiquetas é denominada extensibilidade. Desta propriedade se deriva uma outra, a legibilidade. Visto que, um documento XML etiquetado com marcas suficientemente representativas, qualquer usuário, com ou sem conhecimentos em XML é capaz de reconhecer os distintos elementos de um documento XML [...] (MARTÍNEZ GONZALEZ, 2000 p, 125, tradução nossa).

Dois exemplos de documentos XML podem ser observados logo abaixo:

```
<?xml version="1.0" ?>
<!-- Exemplo de um documento XML -->
- <Cadastro>
  <nome>Sofia</nome>
  <idade>25</idade>
  <email>sofiaet@tvnet.com</email>
</Cadastro>
```

doc. 1

```
<?xml version="1.0" ?>
<!-- Exemplo de um documento XML -->
- <Catalogo>
  <Autor>Flamino, Adriana Nascimento</Autor>
  - <Titulo>
    MARC21 e XML como ferramentas para a
    consolidação da Catalogação Cooperativa
    Automatizada
    <subtitulo>uma revisão de literatura</subtitulo>
  </Titulo>
  <Ano>2003</Ano>
</Catalogo>
```

doc 2

FIGURA 13: Exemplos de documentos em XML
FONTE: FLAMINO, 2003, p. 84

A expressão `<Cadastro>` é chamada de marca de início e a `</Cadastro>` de marca de fim, assim como as outras *tags* como `catalogo`, `nome`, `autor` e outras. Pode-se observar que todas elas têm as suas marca de início e fim. As marcas de início e de fim são também chamadas de marcações (ABITEBOUL; BUNEMAN; SUCIU, 2000).

As marcas ou marcações em XML de acordo com Abiteboul, Buneman e Suciú (2000, p. 28),

são definidas pelos próprios usuários, não há marcas predefinidas como em HTML. O texto entre uma marca de início e a correspondente marca de fim, incluindo as marcas embutidas, é chamado um elemento e as estruturas entre as marcas são o conteúdo. O termo subelemento também é utilizado para descrever a relação entre um elemento e seus elementos componentes.

Assim, `<Titulo> ... </Titulo>` é um sub-elemento de `<Catalogo> ... </Catalogo>` no exemplo citado (FIG. 13).

É importante ressaltar que “são as tecnologias associadas a XML que ampliam seu potencial” (SANTOS; FLAMINO, 2004, p. 128), entre algumas delas, destacam-se (SOUSA; ALVARENGA, 2004; FLAMINO, 2003):

- **XML Schema** – expressa vocabulários compartilhados e permite que máquinas executem regras elaboradas por pessoas. Provê meios para definir a estrutura, conteúdo e semântica de documentos XML;
- **XSL e CSS** – usadas para converter um documento XML em diferentes formatos de exibição e para diferentes dispositivos;
- **Xlink** – permite inserir elementos em documentos XML para criar e descrever ligações entre recursos (*hiperlinks*), mais potentes com características de extensibilidade, inclusão de informação acerca da semântica e da relação, links bidirecionais, links múltiplos, links fora dos documentos, ligar fragmentos de documentos, manipulação como qualquer outro dado XML.
- **XML Namespaces** – o desenvolvedor pode criar os seus próprios elementos com os nomes que quiser. Os nomes dos elementos relacionados no interior das etiquetas são, um dos trunfos da XML.

A característica principal da XML é a facilidade que ela proporciona para modelar os dados. Por meio da modelagem dos dados, pode-se especificar quais elementos o documento terá, assim como a ligação dos elementos entre si. A modelagem dos dados pode ser feita através da *Document Type Definition* – **DTD**, ou seja, Definição de Tipo de Documento ou por intermédio de esquemas XML – **XML Schemas**¹² (RAMALHO, 2002).

¹² A linguagem XML *Schema*, apesar de ter a mesma função da DTD, possui muitas características que a torna mais poderosa (e mais completa) que a DTD. A DTD (*Document Type Definition*) é um arquivo do tipo texto onde estão definidas as *tags* [etiquetas], a ordem em que elas devem aparecer no documento XML e sua obrigatoriedade (FERNEDA, 2003).

“A estrutura dos dados pode ser chamada de esquema ou até mesmo de vocabulário. Ao definir um esquema, pode-se criar regras para que os dados de um documento obedecem a determinadas regras pré-estabelecidas” (RAMALHO, 2002, p. 31).

A DTD (*Document Type Definition*) define a estrutura do documento, assim como especifica uma lista dos elementos válidos e seu possível conteúdo. Quando se trabalha sozinho, a pessoa ou instituição tem a liberdade de definir seus elementos do jeito que quiser. No entanto, como a XML é uma linguagem cuja principal finalidade é compartilhar dados, deve-se pensar nas outras pessoas ou outras instituições que estejam envolvidas no processo de criação e intercâmbio dos dados (RAMALHO, 2002).

Ao utilizar a DTD pode-se criar um modelo de dados que pode ser usado por outros grupos de pessoas ou instituições envolvidas no processo de intercâmbio de dados e por meio dela verificar se os dados recebidos de terceiros são válidos e se estão seguindo as normas definidas (RAMALHO, 2002).

Segundo Ramalho (2002, p. 32) “a definição do tipo de documento pode ser interna ou externa ao documento XML.”

- Uma DTD externa pode ser usada por diversos documentos.
- Um documento que use uma DTD interna é completamente independente de fontes externas.
- Nos dois casos, o formato da DTD é o mesmo (RAMALHO, 2002, p. 32).

< DTD INTERNA

Uma DTD interna é criada dentro do próprio arquivo XML e não depende de arquivos externos. A *tag* **DOCTYPE** é usada para declarar uma DTD.

Veja um exemplo de sua sintaxe (FIG. 14):

```

<?xml version= "1.0" encoding= "ISO8859-1" ?>
<!DOCTYPE relatório [
  <!ELEMENT relatório (para, de, título, mensagem)>
  <!ELEMENT para (#PCDATA)*>
  <!ELEMENT de (#PCDATA)*>
  <!ELEMENT título (#PCDATA)*>
  <!ELEMENT mensagem (#PCDATA)*>
]>
<relatorio>
<para>Programa de Pós-Graduação em Ciência da Informação</para>
<de>Flamino, Adriana Nascimento</de>
<título>MARC21 em XML</título>
<mensagem>Relat&#243;rio cient&#237;fico para disserta&#231;&#227;o </mensagem>
</relatorio>

```

FIGURA 14: Exemplo de uma DTD interna

Analisando o exemplo:

!ELEMENT relatório define que o elemento relatório possui quatro elementos (para, de, título e mensagem).

!ELEMENT para define que o elemento para é do tipo “CDATA”, assim como as demais linhas. “CDATA refere-se a dados do tipo caractere, porém com uma diferença: o texto contido em CDATA não será processado pelo parser, ou seja, referências a entidades e comandos de marcação não serão considerados” (RAMALHO, 2002, p. 33).

Na *tag* <mensagem> estão inseridos alguns caracteres *unicode* que estão representando as letras **ó**, **í** e **çã**. “Unicode é um padrão para a representação de caracteres pertencentes a qualquer idioma. Mais de 65 mil caracteres podem ser associados com um número hexadecimal equivalente ao valor de 16 bits do caractere” (RAMALHO, 2002, p.27).

> DTD EXTERNA

Uma DTD externa é criada independentemente do arquivo XML, ou seja, além do arquivo XML é criado um arquivo de extensão **.dtd**. Enquanto a DTD interna é criada dentro do arquivo XML, a DTD externa é criada fora do arquivo XML e pode ser usada por diversos

documentos, o que não ocorre com a DTD interna. Mas o formato de uma DTD interna e externa é o mesmo. Veja o exemplo:

```
<!DOCTYPE relatório SYSTEM "relatório.dtd">
  <para>Programa de Pós-Graduação em Ciência da Informação </para>
  <de>Flamino, Adriana</de>
  <título>MARC21 em XML</título>
  <mensagem>Dissertação de Mestrado</mensagem>
```

FIGURA 15: Exemplo de uso de uma DTD externa

O conteúdo do arquivo **relatório.dtd** referenciado pela instrução DOCTYPE é o mesmo da DTD interna. Conforme a figura 16:

```
<?xml version="1.0" encoding="ISO8859-1" ?>
<!ELEMENT relatório (para, de, título, mensagem)>
<!ELEMENT para (#PCDATA)*>
<!ELEMENT de (#PCDATA)*>
<!ELEMENT título (#PCDATA)*>
<!ELEMENT mensagem (#PCDATA)*>
```

FIGURA 16: Exemplo de um arquivo DTD.

De acordo com Ramalho (2002, p. 34) “o uso da DTD interna ou externa não afeta a aparência dos dados”.

Segundo Ray (2001) a *Document Type Definition* – DTD, é o tipo mais popular de modelo de documento. “As DTDs, na realidade são mais antigas que a XML, sendo um prosseguimento da SGML, com a sintaxe quase completamente intacta” (RAY, 2001, p. 153).

Para Ray (2001, p. 53) “uma DTD define um tipo de documento da seguinte maneira”:

- Declara um conjunto de elementos permitidos. Não se pode usar qualquer nome de elemento além daqueles nesse conjunto. Funciona como o “vocabulário” da linguagem.
- Define um modelo de conteúdo para cada elemento. O modelo de conteúdo é um padrão que diz quais elementos ou dados podem entrar em um elemento, em que ordem, em que quantidade e se eles são obrigatórios ou opcionais. Funciona como a “gramática” da linguagem.
- Declara um conjunto de atributos permitidos para cada elemento. Cada declaração de atributo define o nome, o tipo de dados, valores default (se houver) e comportamento (por exemplo, se é obrigatório ou opcional) do atributo.
- Oferece uma série de mecanismos para facilitar o gerenciador do modelo, por exemplo, o uso de entidades de parâmetros e a capacidade de importar partes do modelo a partir de um arquivo externo.

Segundo Martínez Gonzalez (2000), cada usuário ou comunidade de usuários pode definir suas próprias DTD adaptadas ao tipo de informação que manipula.

Importante destacar também as Folhas de estilo **XSL**. “As folhas de estilo são conjuntos de regras que se adicionam aos documentos XML e [HTML] e que se aplicam a estes no momento de sua apresentação ao usuário” (MARTÍNEZ GONZALEZ, 2000, p. 132, tradução nossa).

A XSL é composta por duas partes: **XSL Transformation** – XSLT responsável por transformar o documento original e a **XSL Formatting Objects** – XSLFO responsável pela aplicação da formatação ao arquivo transformado (RAMALHO, 2002). Conforme a figura 17:

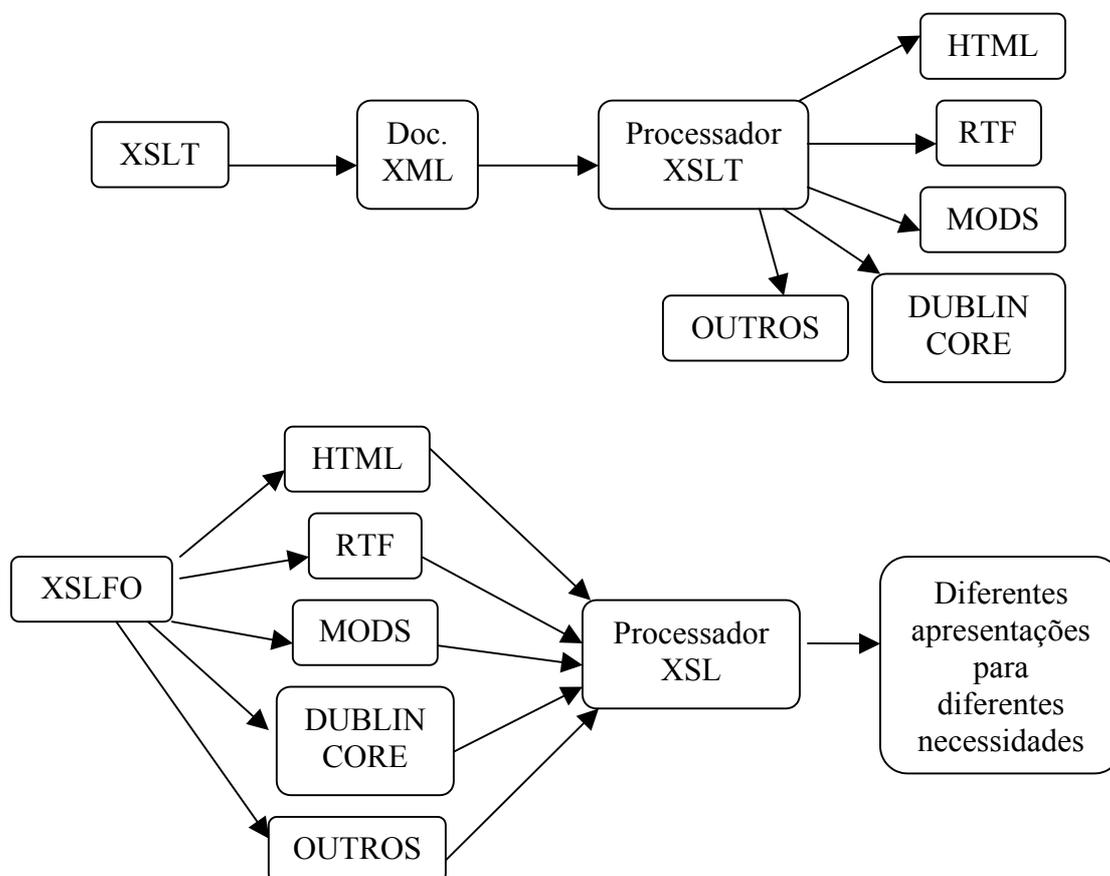


FIGURA 17: Vários estilos de apresentação a partir de um mesmo documento XML
FONTE: FLAMINO, 2003, p. 99

Segundo Hough, Bull e Young (2000, p. 5, tradução nossa),

A tecnologia eXtensible Style Language Transformations (XSLT) – é um modo de converter um arquivo XML em outro arquivo XML, um arquivo HTML ou qualquer outro tipo de arquivo, adicionando, emendando e mudando o conteúdo apropriadamente, como o uso na representação e conversão do MARC. O interesse maior é a conversão de HTML para XML para exibição do Web browser.

A XSLT é usada para converter um documento XML em diferentes formatos, como HTML, RTF (*Rich Text Format*), MODS, DUBLIN CORE, ou qualquer outro estilo de formato existente. Para isso é necessário adicionar no início do documento XML a sintaxe

exigida pela folha de estilo, assim como na DTD (FURGERI, 1999). Veja um exemplo de sua sintaxe abaixo:

```
<?xml-stylesheet href="relatório.xsl" type="text/xsl"?>.
```

A XSLT foi uma importante ferramenta para a representação e conversão de registros MARC para outros formatos de representação como o MODS e outros formatos baseados na XML.

Como se pode perceber a XML é a promessa de tornar a Internet mais “inteligente”, transformando o modo de armazenamento, recuperação e intercâmbio de informações e, principalmente, a economia da informação na forma de redução de custos operacionais (FURGERI, 2001) para as instituições, e desta forma causando grandes mudanças em diversas áreas, inclusive no mundo bibliotecário.

Conforme Miller citado por Johnson (2001, p. 84, tradução nossa),

- Caracteriza a XML como o “formato de dados universal para a Web”, uma “SGML simplificada, usando a sintaxe HTML, com facilidades da Web acrescentadas”.
- Emprega um conjunto único de caracteres, Unicode para representar todos os dados.

Miller, de acordo com Johnson (2001, p. 84, tradução nossa) também observou que, “um aspecto significativo da XML é a sua separação do conteúdo, apresentação e ligação, de forma que cada um possa ser controlado otimamente”.

Segundo Bax (2002) esta é a ‘filosofia’ da XML (a separação da apresentação, do conteúdo e da ligação). O formatador na figura seguinte está representando a XSL, que possibilita diferentes apresentações de um mesmo documento XML, conforme explicado na seção anterior. Veja a figura 18:

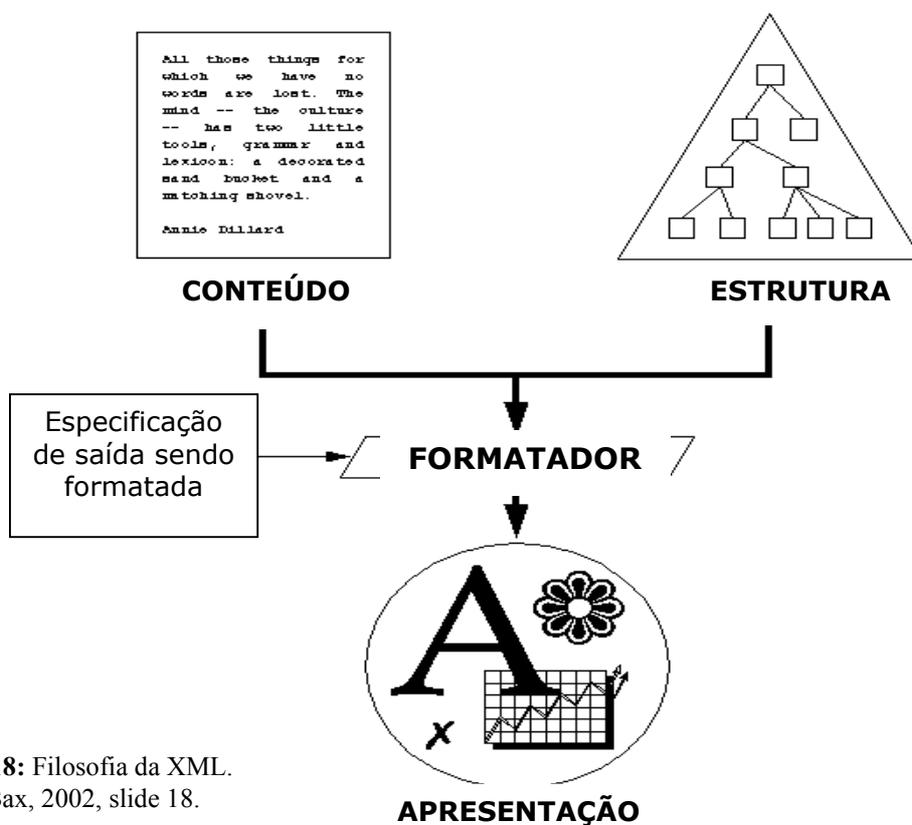


FIGURA 18: Filosofia da XML.

FONTE: Bax, 2002, slide 18.

Para Bax (2002, slide 13) a XML é:

- Padrão para se criar padrões
 - Formato universal para troca de dados e documentos
 - Espécie de “língua franca” da era da informação
 - Poder, precisão, controle e flexibilidade, atraindo a atenção dos bibliotecários
- Padrões abertos (*open source movement*)
 - Não proprietário – qualquer um pode usar
- Extensibilidade (relativamente fácil de expandir)
- Separação entre conteúdo e apresentação
 - Mantém a flexibilidade da informação
- Técnicas enlaces (links) mais sofisticadas
 - Linkar uma parte específica do documento
- Neutralidade de plataforma e software / Interoperabilidade

- Unicode: longevidade, internacionalização
Conjunto de caracteres únicos para todas as línguas humanas
- Interface com banco de dados

A linguagem XML permite ao desenvolvedor criar seus próprios elementos, com nomes e atributos, aliás, uma das grandes vantagens da XML. Porém, isso pode gerar alguns problemas quando uma aplicação XML, desenvolvida por várias pessoas ou instituições, gera nomes de elementos ou atributos duplicados (RAMALHO, 2002). Qin (2000, p. 86, tradução nossa) salienta que “para complicar ainda mais o assunto, futuros documentos XML conterão múltiplos vocabulários de marcação e isso causará problemas para reconhecimento, gerando conflitos”.

Por exemplo, um elemento <titulo> pode se referir a coisas diferentes como, o título de um livro, o título de um artigo de jornal ou o título de uma revista entre outras coisas. Para resolver esse problema usa-se um prefixo para cada elemento <titulo> a fim de identificá-lo corretamente. Por exemplo: <livro: titulo>; <artigodejornal: titulo>; <revista: titulo>. Os nomes dos elementos relacionados no interior das *tags* são denominados *namespaces* (RAMALHO, 2002), “um conjunto de nomes de elementos relacionados” (CASTRO, 2001, p. 113), cuja “função principal [...] é identificar de forma única os elementos e atributos de um documento XML” (RAMALHO 2002, p. 140).

No entanto, de acordo com Qin (2000, p. 86, tradução nossa),

A solução para os problemas de namespaces de XML estará em grande parte nas mãos da comunidade bibliotecária e da Ciência da Informação que, durante anos de pesquisas em organização e representação da informação e do conhecimento, tem desenvolvido uma estrutura inteira de metodologias e sistemas. Um exemplo imediato é que as técnicas usadas na criação e controle de tesouros podem ser aplicadas para padronizar as nomeações ou marcações dos elementos de dados em várias aplicações XML e traçar a semântica dos mesmos nos repositórios de namespaces.

Para não ter problemas com a duplicação de *namespaces*, no processo de troca de informações, é preciso criar normas com o objetivo de torná-los universalmente únicos e assim evitar conflitos, pois, cada vez mais, “surgem aplicações XML, e a demanda para controle e gerenciamento de namespaces também aumentará” (QIN, 2000, p. 86, tradução nossa), visto que uma das exigências das organizações de hoje é ter sistemas de informação efetivos que possam responder rapidamente as necessidades de informação para gerar tomadas de decisões mais seguras e eficientes.

Veja na figura 19 um trecho do *namespace* proposto por Lam (2002, p. 91), Chefe de Sistemas da Biblioteca da Universidade de Ciência e Tecnologia de Hong Kong:

```
<?xml version= "1.0" encoding= "UTF-8"?>
<marc>
  <fd name= "245" ind1= "1" ind2= "4" ... other attributes... >
    <sf name= "a"> The ABCs of XML : </sf>
    <sf name= "b"> the librarian's guide to the extensible markup language / </sf>
    <sf name= "c"> by Norman Desmarais </sf>
  </fd>
  ... more fd elements follow ....
</marc>
```

FIGURA 19: Proposed namespace

FONTE: LAM, 2002, p. 91

No exemplo da FIG. 19, **<marc>** e **</marc>** marcam o início e o término de um registro MARC. **<fd>** e **</fd>** marcam o início e o término das etiquetas MARC. **<sf>** e **</sf>** marcam o início e o término de um subcampo. Os elementos **fd** e **sf** contêm vários atributos. O atributo **name** define o nome da etiqueta MARC ou o código de subcampo. **ind1** e **ind2** são os dois indicadores da etiqueta MARC (LAM, 2002, p. 91, tradução nossa).

Uma das grandes vantagens da XML é a de ser uma estrutura de marcação aberta, ou seja, não impõe restrições à sua utilização nem a criação de ferramentas que a manipulem.

Segundo Shapiro e Varian (1999) com um padrão aberto as Instituições não se preocuparão com o aprisionamento por soluções proprietárias. Veja na Figura 20 o ciclo do aprisionamento segundo os autores.

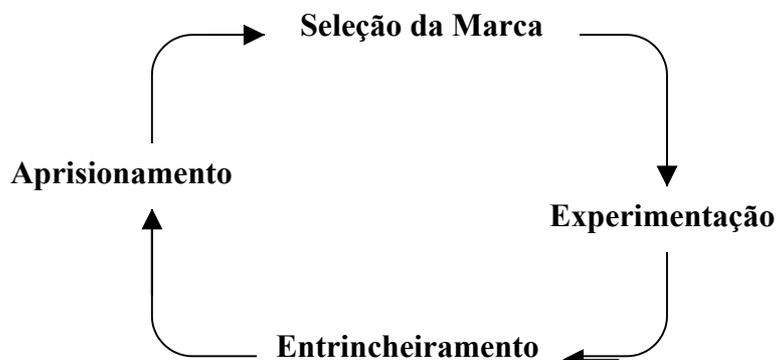


FIGURA 20 – O ciclo do aprisionamento
FONTE: Shapiro e Varian (1999, p. 158)

De acordo com Shapiro e Varian (1999, p. 158) “o lugar mais fácil para lançar-se no ciclo do aprisionamento é no ponto de seleção da marca”, ou seja, a escolha de um novo produto, por exemplo, um novo programa de software ou uma nova base de dados. De acordo com os autores não nascemos “aprisionados”, só ficamos em virtude das escolhas que fazemos (SHAPIRO ; VARIAN, 1999, p.158).

Depois da seleção da marca “segue-se a fase de experimentação durante a qual o cliente usa de maneira ativa a nova marca e usufrui da vantagem de todos os incentivos que teve para dar uma oportunidade a essa marca” (SHAPIRO; VARIAN, 1999, p.158). Por exemplo, as ofertas grátis de produtos como softwares livres para uso 30, 60 ou 90 dias.

Os clientes que fazem mais do que experimentar, entram para a fase de entrincheiramento. Isto é, quando o consumidor [ou instituição] realmente acostuma-se à nova marca [ou novo produto], desenvolve uma preferência por ela em detrimento das outras e talvez fique retido a essa marca ao fazer os investimentos complementares. Normalmente o fornecedor tenta prolongar essa fase e adiar a consideração ativa de outras marcas [ou produtos], desejando que os custos de troca do cliente subam. A fase de entrincheiramento culmina com o

aprisionamento quando os custos de troca tornam-se altos demais. (SHAPIRO; VARIAN, 1999, p.159).

“A essência do aprisionamento é que suas escolhas no futuro serão limitadas pelos seus investimentos de hoje” (SHAPIRO; VARIAN, 1999, p. 161).

Segundo os autores existem 7 (sete) padrões econômicos primários que levam ao aprisionamento: “Compromissos contratuais, equipamentos duráveis e atendimento pós-venda, treinamento específico de marca, informação e banco de dados, fornecedores especializados, custos de busca e programas de lealdade” (SHAPIRO ; VARIAN, 1999, p. 161).

Agora, mesmo os padrões abertos têm algumas desvantagens, por exemplo, segundo Shapiro e Varian (1999, p. 294), eles “são propensos a “estilhaçamento” ou “fragmentação” e podem também serem “seqüestrados”. De acordo com Shapiro e Varian (1999, p. 294) “o estilhaçamento de um padrão refere-se ao surgimento de versões múltiplas e incompatíveis de uma tecnologia padronizada”.

O exemplo clássico dos riscos de administrar padrões abertos e dos riscos do estilhaçamento é a história do sistema operacional Unix. O UNIX foi desenvolvido originalmente nos Laboratórios Bell como uma ferramenta de pesquisa. A AT&T cedeu o código fonte a pesquisadores acadêmicos por muitos anos, e ele tornou-se um padrão na comunidade de pesquisas. Quando o mercado de minicomputadores decolou na década de 1970, o UNIX foi modificado e vendido por diferentes empresas; a disseminação acelerada das estações de trabalho na década de 1980 conduziu a mais versões do UNIX, e nenhum padrão setorial foi estabelecido (SHAPIRO; VARIAN, 1999, p.294).

De acordo com Shapiro e Varian (1999, p. 296),

Os padrões abertos também podem ser “seqüestrados” por empresas que buscam estendê-los em direções proprietárias e, portanto, ganhar controle sobre a base instalada com o tempo. A Microsoft tem sido acusada de tentar estender tanto a Java como a HTML em direções proprietárias.

São por estas razões que os padrões abertos necessitam de responsáveis que possam administrá-los, e que sejam encarregados de estabelecer a direção na qual o padrão evoluirá e mantê-los livre da estagnação, fazendo investimentos em recursos para o seu desenvolvimento e manutenção.

Segundo Shapiro e Varian (1999, p. 298) “Sem um responsável, os padrões abertos podem estagnar-se ou estilhaçar-se em pedaços incompatíveis”. No caso da XML, o órgão responsável pelo seu desenvolvimento e manutenção é o W3C, como já foi explicado anteriormente.

A aplicação da XML (*eXtensible Markup Language*) em bibliotecas vem aumentando em considerável interesse, desde que se tornou um potencial padrão para a Web de impacto universal. Vários projetos exploratórios estão sendo acometidos e algumas bibliotecas importantes (sobretudo nos EUA) estão comprometidas a prover versões em XML de seus registros.

Para alguns autores como Ortiz-Repiso Jiménez (1999, p. 222, tradução nossa) “tanto a [...] [XML] como o MARC são linguagens de marcação descritiva que criam textos estruturados”. E a XML permite flexibilidades que não podem ser encontradas no MARC. No entanto, de acordo com a literatura estudada, MARC é uma estrutura de descrição documentária legível por máquina e a XML é uma linguagem computacional de marcação de dados, portanto, são ferramentas diferentes, mas que, no entanto, se complementam.

De acordo com Carvalho e Cordeiro (2002), o tópico mais discutido atualmente é a relação entre o formato MARC e a XML, variando da atual suficiência do MARC, até a que vantagens se ganhariam provendo registros em XML no lugar do MARC e, a pontos mais específicos de interesse sobre a forma correta de representação da XML para registros bibliográficos (CARVALHO; CORDEIRO, 2002). Tal discussão (substituir MARC por XML), não se fundamenta uma vez que, um registro MARC pode ou não ser em XML.

MARC é estrutura e não linguagem, e a XML aceita qualquer tipo de estrutura, como por exemplo, MARC, Dublin Core, e outras.

Deve-se, ainda, destacar a diferença entre um formato de conteúdo e um formato de transporte. Para Andresen, (2004, tradução nossa),

[...] é importante diferenciar entre formatos de conteúdo e formatos de transporte. Um formato de conteúdo é um formato que define as descrições atuais e ordens, por exemplo, que um título deve estar no campo 245 sub-campo *a, e que um número de campo tem que ter três dígitos. Um formato de transporte é um recipiente no qual podem ser colocados vários formatos de conteúdo. MARC21, UNIMARC e danMARC2 são todos formatos de conteúdos. O único formato de transporte internacionalmente unificado é a ISO 2709 que pode conter todos os diferentes formatos MARC.

Ou seja, o formato MARC serve como um recipiente no qual a informação é armazenada, enquanto que a XML proporciona, com vantagens, o transporte desta informação em um ambiente digital. Segundo Eito Brun (2002, slide 3, tradução nossa) a “[...] XML tem se apresentado como uma alternativa para representar e transferir registros MARC através da Internet e facilitar seu tratamento automatizado”.

Pode-se dizer que, com os avanços das tecnologias o MARC vem novamente se adaptando as novas tendências mostrando que, a união entre a estrutura de representações bibliográficas e catalográficas MARC e a linguagem XML será extremamente valiosa. O formato MARCXML será abordado no próximo capítulo.

5 MARCXML para a OAI

“As *novidades* precisam ser compreendidas; as *potencialidades* precisam ser valorizadas; as *ambivalências* precisam ser superadas.”

Plácida L.V. A. C. Santos

Aborda-se o MARCXML como uma estrutura para representação descritiva de recursos informacionais legível por máquina, sobretudo em arquivos abertos e não como um mero mecanismo de conversão.

Ao longo dos últimos anos (meados da última década do século XX) surgiram algumas iniciativas interessadas no uso da XML em relação com os registros MARC, sobretudo para a adaptação deste último à **Internet**. A biblioteca do Congresso Americano (*Library of Congress – LC*) desenvolveu em 1995 um projeto chamado ‘**MARCDTD**’¹ para definir os formatos USMARC como DTDs e desenvolver utilitários para a conversão entre os formatos DTDs e a ISO 2709. O projeto usou a **SGML** (*Standard Generalized Markup Language*, um padrão popular na indústria de publicação) em lugar de XML DTDs (HOUGH; BULL; YOUNG, 2000, p. 13), ou seja, “o objetivo era criar uma DTD que permitisse a conversão de registros MARC para SGML – e vice-versa – sem perda de informação” (ÉITO BRUN, 2002, slide 16).

Foram desenvolvidas 2 DTDs. Uma para dados bibliográficos e autoridades e uma aplicação ‘PERL’² para a conversão de MARC 2709 para MARC SGML e vice-versa. As DTDs foram publicadas em maio de 1996. A utilidade de conversão

¹ A página Web <http://www.loc.gov/marc/marcdtd/marcdtdback.html> apresenta informações relevantes sobre o projeto.

² <http://www.perl.com>

começou em julho de 1997. Em janeiro de 1998 se tornou pública (EÍTO BRUN, 2002, slide 16).

Porém como a tecnologia evoluiu, esta DTD SGML foi convertida para uma DTD XML, para adaptá-la ao novo cenário tecnológico. A diferença entre as DTDs SGML e XML está em uma variação sintática na própria estrutura do documento.

Segundo Keith (2004, p.122, tradução nossa) em relação a DTD XML,

Uma característica importante desta DTD é o seu uso para a validação do MARC21. Como uma consequência de codificar toda a regra de validação no schema, ela tornou-se muito grande, tão grande que há duas DTDs; uma com aproximadamente 500k em tamanho que valida registros bibliográficos, holdings, e informação da comunidade, e a outra com aproximadamente 240k em tamanho que valida registros de autoridade e de classificação. Para validar a estrutura MARC21, para toda tag, foram enumerados nos schemas os códigos de subcampo válidos e valores de indicadores, deste modo resultando em um tamanho muito extenso. O tamanho dessas DTDs gera um desconforto no trabalho com elas, especialmente para usuários de aplicações baseadas em XML desktop.

Uma amostra de registro em DTD XML para MARC pode ser vista a seguir:

```

<!DOCTYPE mrcbfile SYSTEM "ftp://
ftp.loc.gov/pub/marc/dtd/mrcbxml
file.dtd">
<mrcbfile>
  <mrcb format-type="bd">
    <mrcbldr-bd>
      <mrcbldr-bd-05 value="c"/>
      <mrcbldr-bd-06 value="a"/>
      <mrcbldr-bd-07 value="m"/>
      <mrcbldr-bd-08 value="blank"/>
      <mrcbldr-bd-09 value="blank"/>
      <mrcbldr-bd-17 value="blank"/>
      <mrcbldr-bd-18 value="a"/>
      <mrcbldr-bd-19 value="blank"/>
    </mrcbldr-bd>
    <mrcb-control-fields>
      <mrcb001> 92005291 </mrcb001>
      <mrcb003>DLC</mrcb003>
      <mrcb005>19930521155141.9
      </mrcb005>
      <mrcb008-bk>
      <mrcb008-bk-00-05 value="920219"/>
      <mrcb008-bk-06 value="s"/>
      <mrcb008-bk-07-10 value="1993"/>
      <mrcb008-bk-11-14 value="blank"/>
      <mrcb008-bk-15-17 value="cau"/>
      <mrcb008-bk-18-21 value="a "/>
      <mrcb008-bk-22 value="j"/>
      <mrcb008-bk-23 value="blank"/>
      <mrcb008-bk-24-27 value="blank"/>
      <mrcb008-bk-28 value="blank"/>
      <mrcb008-bk-29 value="0"/>
      <mrcb008-bk-30 value="0"/>
      <mrcb008-bk-31 value="0"/>
      <mrcb008-bk-32 value="blank"/>
      <mrcb008-bk-33 value="0"/>
      <mrcb008-bk-34 value="blank"/>
      <mrcb008-bk-35-37 value="eng"/>
      <mrcb008-bk-38 value="blank"/>
      <mrcb008-bk-39 value="blank"/>
    </mrcb008-bk>
    </mrcb-control-fields>
    <mrcb-main-entry>
      <mrcb100 i1="1-1" i2="12-blank">
      <mrcb100-a>Sandburg, Carl,</mrcb100-a>
      <mrcb100-d>1878-1967.</mrcb100-d>
      </mrcb100>
    </mrcb-main-entry>
    <mrcb-title-and-title-related>
      <mrcb245 i1="1-1" i2="12-0">
      <mrcb245-a>Arithmetic/</mrcb245-a>
      <mrcb245-c>Carl Sandburg ; illustrated as
      an anamorphic adventure by Ted Rand.</
      mrcb245-c>
      </mrcb245>
    </mrcb-title-and-title-related>
    <mrcb-edition-imprint-etc>
      <mrcb260 i1="11-blank" i2="12-blank">
      <mrcb260-a>San Diego :</mrcb260-a>
      <mrcb260-b>Harcourt Brace Jovanovich,
      </mrcb260-b>
      <mrcb260-c>c1993.</mrcb260-c>
      </mrcb260>
    </mrcb-edition-imprint-etc>
    <mrcb-physical-description>
      <mrcb300 i1="11-blank" i2="12-blank">
      <mrcb300-a>1 v. (unpaged) :</mrcb300
      <mrcb300-b>ill. (some col.) ;</mrcb300
      <mrcb300-c>26 cm.</mrcb300-c>
      </mrcb300>
    </mrcb-physical-description>
    <mrcb-notes>
      <mrcb500 i1="11-blank" i2="12-blank">
      <mrcb500-a>One Mylar sheet included i
      pocket.</mrcb500-a>
      </mrcb500>
    </mrcb-notes>
    <mrcb-subject-access>
      <mrcb650 i1="11-blank" i2="12-0">
      <mrcb650-a>Arithmetic</mrcb650-a>
      <mrcb650-x>Juvenile poetry.
      </mrcb650-x>
      </mrcb650>
      <mrcb650 i1="11-blank" i2="12-0">
      <mrcb650-a>Children's poetry,
      American.</mrcb650-a>
      </mrcb650>
      <mrcb650 i1="11-blank" i2="12-1">
      <mrcb650-a>Arithmetic</mrcb650-a>
      <mrcb650-x>Poetry.</mrcb650-x>
      </mrcb650>
      <mrcb650 i1="11-blank" i2="12-1">
      <mrcb650-a>American poetry.
      </mrcb650-a>
      </mrcb650>
      <mrcb650 i1="11-blank" i2="12-1">
      <mrcb650-a>Visual perception.
      </mrcb650-a>
      </mrcb650>
    </mrcb-subject-access>
    <mrcb-added-entry>
      <mrcb700 i1="11-1" i2="12-blank">
      <mrcb700-a>Rand, Ted,</mrcb700-a>
      <mrcb700-e>ill.</mrcb700-e>
      </mrcb700>
    </mrcb-added-entry>
    </mrcb>
  </mrcbfile>

```

FIGURA 21: Exemplo de um registro DTD XML para MARC.
Fonte: KEITH, 2004, p. 122-123.

Como visto, as DTDs XML contém um elemento XML para cada campo e subcampo do MARC. Por exemplo: para o elemento 245 do MARC, na DTD se inclui um elemento mrcb245, um mrcb245-b, mrcb245-c etc. Estes campos se agrupam em outros

elementos, como, por exemplo, *mrcb-title-and-title-related*, *mrcb-subject-access* etc. Já os campos de controle variáveis – 008 – eles têm um elemento para cada "posição" ou "item" dos quais contêm informação etc. Isto fez com que a DTD e os registros se tornassem muito grandes (EÍTO BRUN, 2002, slide 18; KEITH, 2004).

Para resolver este problema, conforme Keith (2004, p. 122, tradução nossa),

Atualmente a LC desenvolveu um esquema XML (www.w3.org/XML/Schema) que incorpora lições aprendidas ao longo do desenvolvimento de DTDs, e tira proveito dos padrões tecnológicos atuais para prover uma aproximação melhorada para os dados MARC21 em XML.

Preparado pelo escritório de desenvolvimento de redes e padrões MARC da *Library of Congress*, em colaboração com a **OCLC** e a **RLG** e revisado pela Biblioteca Nacional do Canadá e pela Biblioteca Nacional de Medicina (NLM), o ‘**XML Schema**’³ foi desenvolvido para uso na comunicação com os registros MARC (KEITH, 2002). Isto é, “o novo esquema pretende ir mais além de um mero mecanismo de conversão e assim facilitar” (EÍTO BRUN, 2002, slide 20, grifo nosso):

- A representação de registros MARC existentes em formato XML;
- A representação de metadados para facilitar o harvesting do projeto **OAI** (Open Archives Initiative);
- A descrição original de recursos utilizando MARC;
- A codificação de metadados MARC que podem ser “incluídos” como parte do conjunto de recursos de informação disponíveis em formato eletrônico;
- Além de ser usado potencialmente como um esquema de extensão para **METS** (Metadata Encoding and Transmission Standard).

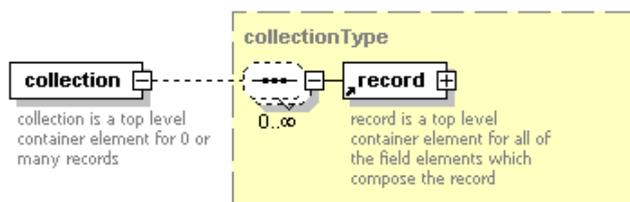
³ Também a Universidade de Stanford tem produzido um pacote chamado ‘**XMLMARC**’ o qual usa um DTD **USMARC** junto com um ‘arquivo mapa especial’ para a geração automática de arquivo XML (HOUGH; BULL; YOUNG, 2000, p. 14).

O atual esquema MARCXML foi desenvolvido como uma alternativa “menos carregada” (mais simples) ao DTD XML para MARC. Este esquema é composto de apenas seis elementos: (1) **coleção**; (2) **registro**; (3) **líder**; (4) **campo de controle**; (5) **campo de dados**; e (6) **subcampo**. O esquema provê a preservação de espaços em branco e a validação da integridade básica dos elementos de registro apropriados da ISO 2709. Devido a característica baseada na natureza do padrão MARC21, especialmente nos campos de controle e líder, o espaço em branco precisa ser preservado em ordem para manter o equilíbrio dos caracteres e, deste modo, assegurar a própria interpretação dos dados (KEITH, 2004).

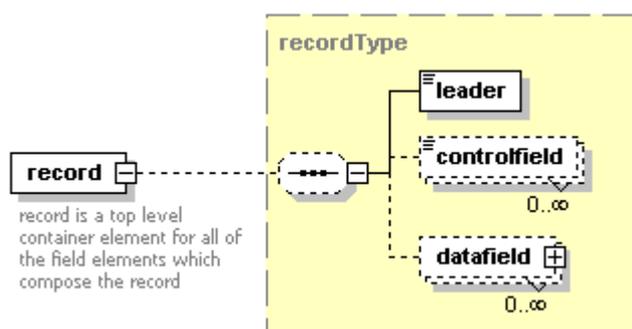
Enquanto os elementos líder, *controlfield*, *datafield*, e *subfield* devem ser familiares para aqueles da comunidade bibliotecária, o uso dos elementos **coleção** e **registro** no contexto do esquema MARCXML devem ser explicados. O fundamental para o uso do esquema mencionado é um único registro MARC21 em um único documento XML. Neste caso, o elemento registro do *namespace* <http://www.loc.gov/MARC21/slim> (ver figura 23) é o elemento raiz. No entanto, uma flexibilidade é necessária para casos quando múltiplos registros MARC21 existirem no mesmo arquivo (ver figura 24). Isto é útil principalmente para o intercâmbio de múltiplos registros em um arquivo, para conjuntos de resultados que podem retornar zero ou muitos registros, ou transformações que operam em múltiplos registros tal como uma transformação de requerimento funcional para registros bibliográficos (FRBR). Nestes casos, o elemento **coleção** do mesmo *namespace* deve ser o elemento raiz do documento XML com múltiplos sub elementos **registro** contendo cada registro.

Como um adendo, a próxima versão do esquema MARCXML definirá os elementos líder, campo de controle, campo de dado, e subcampo como elementos globais para que outros esquemas XML possam referenciá-los (KEITH, 2004). Um esquema e dois exemplos de registros podem ser vistos logo em seguida:

Elemento Coleção (e Registro)



Elemento Registro (e Líder; Campo de Controle, e Campo de Dados)



Elemento: Campo de Dados (e subcampo)

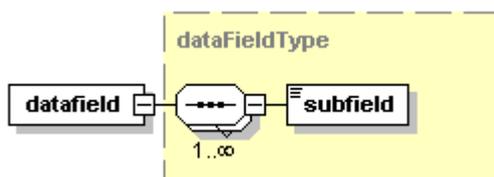


FIGURA 22: Esquema MARCXML (simplificado)

FONTE: <http://www.loc.gov/standards/marcxml/schema/MARC21slim.xsd>. Acesso em: 02 abr. 2004.

```
<record xmlns="http://www.loc.gov/MARC21/slim">
  <leader>01142cam 2200301 a 4500
</leader>
  <controlfield tag="001"> 92005291
</controlfield>
  <controlfield tag="003">DLC</controlfield>
  <controlfield tag="005">
19930521155141.9
</controlfield>
  <controlfield tag="008">920219s1993 caua
j 000 0 eng </controlfield>
  <datafield tag="100" ind1="1" ind2=" ">
  <subfield code="a"> Sandburg, Carl,
</subfield>
  <subfield code="d"> 1878-1967. </subfield>
</datafield>
  <datafield tag="245" ind1="1" ind2="0">
  <subfield code="a"> Arithmetic </subfield>
  <subfield code="c"> Carl Sandburg ;
illustrated as an anamorphic adventure by
Ted Rand. </subfield>
</datafield>
  <datafield tag="260" ind1=" " ind2=" ">
  <subfield code="a"> San Diego </subfield>
  <subfield code="b"> Harcourt Brace
Jovanovich, </subfield>
  <subfield code="c"> c1993. </subfield>
</datafield>
  <datafield tag="300" ind1=" " ind2=" ">
  <subfield code="a"> 1 v. (unpaged) :
</subfield>
  <subfield code="b"> ill. (some col.) ;
</subfield>
  <subfield code="c"> 26 cm. </subfield>
</datafield>
  <datafield tag="650" ind1=" " ind2="0">
  <subfield code="a"> Arithmetic</subfield>
  <subfield code="x"> Juvenile poetry.
</subfield>
</datafield>
  <datafield tag="650" ind1=" " ind2="0">
  <subfield code="a"> Children's poetry,
American. </subfield>
</datafield>
  <datafield tag="650" ind1=" " ind2="1">
  <subfield code="a"> Arithmetic</subfield>
  <subfield code="x"> Poetry. </subfield>
</datafield>
  <datafield tag="650" ind1=" " ind2="1">
  <subfield code="a"> American poetry.
</subfield>
</datafield>
  <datafield tag="650" ind1=" " ind2="1">
  <subfield code="a"> Visual perception.
</subfield>
</datafield>
  <datafield tag="700" ind1="1" ind2=" ">
  <subfield code="a"> Rand, Ted, </subfield>
  <subfield code="e"> ill. </subfield>
</datafield>
</record>
```

FIGURA 23: Exemplo de um registro MARCXML
FONTE: KEITH, 2004, p. 124.

```

<?xml version="1.0" encoding="UTF-8" ?>
- <!--
  edited with XML Spy v4.3 U (http://www.xmlspy.com) by Morgan Cundiff (Library of
  Congress)
-->
- <marc:collection xmlns:marc="http://www.loc.gov/MARC21/slim"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-
  instance" xsi:schemaLocation="http://www.loc.gov/MARC21/slim
  http://www.loc.gov/standards/marcxml/schema/MARC21slim.xsd">
  <marc:record>
  <marc:leader>00925njm 22002777a 4500</marc:leader>
  <marc:controlfieldtag="001">5637241</marc:controlfield>
  <marc:controlfieldtag="003">DLC</marc:controlfield>
  <marc:controlfieldtag="005">19920826084036.0</marc:controlfield>
  <marc:controlfieldtag="007">sdubumennmplu</marc:controlfield>
  <marc:controlfieldtag="008">910926s1957 nyuun eng</marc:controlfield>
  <marc:datafieldtag="010"ind1=""ind2="">
    <marc:subfieldcode="a">91758335</marc:subfield>
  </marc:datafield>
  <marc:datafieldtag="028"ind1="0"ind2="0">
    <marc:subfieldcode="a">1259</marc:subfield>
    <marc:subfieldcode="b">Atlantic</marc:subfield>
  </marc:datafield>
  <marc:datafieldtag="040"ind1=""ind2="">
    <marc:subfieldcode="a">DLC</marc:subfield>
    <marc:subfieldcode="c">DLC</marc:subfield>
  </marc:datafield>
  <marc:datafieldtag="050"ind1="0"ind2="0">
    <marc:subfieldcode="a">Atlantic 1259</marc:subfield>
  </marc:datafield>
  <marc:datafieldtag="245"ind1="0"ind2="4">
    <marc:subfieldcode="a">The Great Ray Charles</marc:subfield>
    <marc:subfieldcode="h">[sound recording].</marc:subfield>
  </marc:datafield>
  <marc:datafieldtag="260"ind1=""ind2="">
    <marc:subfieldcode="a">New York, N.Y. :</marc:subfield>
    <marc:subfieldcode="b">Atlantic,</marc:subfield>
    <marc:subfieldcode="c">[1957?]</marc:subfield>
  </marc:datafield>
  <marc:datafieldtag="300"ind1=""ind2="">
    <marc:subfieldcode="a">1 sound disc :</marc:subfield>
    <marc:subfieldcode="b">analog, 33 1/3 rpm ;</marc:subfield>
    <marc:subfieldcode="c">12 in.</marc:subfield>
  </marc:datafield>
  <marc:datafieldtag="511"ind1="0"ind2="">
    <marc:subfieldcode="a">Ray Charles, piano & celeste.</marc:subfield>
  </marc:datafield>
  <marc:datafieldtag="505"ind1="0"ind2="">
    <marc:subfieldcode="a">The Ray -- My melancholy baby -- Black coffee --
    There's no you -- Doodlin' -- Sweet sixteen bars -- I surrender dear --
    Undecided.</marc:subfield>
  </marc:datafield>
  <marc:datafieldtag="500"ind1=""ind2="">
    <marc:subfieldcode="a">Brief record.</marc:subfield>
  </marc:datafield>
  <marc:datafieldtag="650"ind1=""ind2="0">
    <marc:subfieldcode="a">Jazz</marc:subfield>
    <marc:subfieldcode="y">1951-1960.</marc:subfield>
  </marc:datafield>
  <marc:datafieldtag="650"ind1=""ind2="0">
    <marc:subfieldcode="a">Piano with jazz ensemble.</marc:subfield>
  </marc:datafield>
  <marc:datafieldtag="700"ind1="1"ind2="">
    <marc:subfieldcode="a">Charles, Ray,</marc:subfield>
    <marc:subfieldcode="d">1930-</marc:subfield>
    <marc:subfieldcode="4">prf</marc:subfield>
  </marc:datafield>
  </marc:record>
  </marc:collection>

```

```

</marc:record>
= <marc:record>
  <marc:leader>01832cmma 2200349 a 4500</marc:leader>
  <marc:controlfieldtag="001">12149120</marc:controlfield>
  <marc:controlfieldtag="005">20001005175443.0</marc:controlfield>
  <marc:controlfieldtag="007">cr |||</marc:controlfield>
  <marc:controlfieldtag="008">000407m19949999dcu g m eng d</marc:controlfield>
  = <marc:datafieldtag="906"ind1=""ind2="">
    <marc:subfieldcode="a">0</marc:subfield>
    <marc:subfieldcode="b">ibc</marc:subfield>
    <marc:subfieldcode="c">copycat</marc:subfield>
    <marc:subfieldcode="d">1</marc:subfield>
    <marc:subfieldcode="e">ncip</marc:subfield>
    <marc:subfieldcode="f">20</marc:subfield>
    <marc:subfieldcode="g">y-gencompf</marc:subfield>
  </marc:datafield>
  = <marc:datafieldtag="925"ind1="0"ind2="">
    <marc:subfieldcode="a">undetermined</marc:subfield>
    <marc:subfieldcode="x">web preservation project (wpp)</marc:subfield>
  </marc:datafield>
  = <marc:datafieldtag="955"ind1=""ind2="">
    <marc:subfieldcode="a">vb07 (stars done) 08-19-00 to HLCD lk00; AA3s lk29
      received for subject Aug 25, 2000; to DEWEY 08-25-00; aa11 08-28-
      00</marc:subfield>
  </marc:datafield>
  = <marc:datafieldtag="010"ind1=""ind2="">
    <marc:subfieldcode="a">00530046</marc:subfield>
  </marc:datafield>
  = <marc:datafieldtag="035"ind1=""ind2="">
    <marc:subfieldcode="a">(OCoLC)ocm44279786</marc:subfield>
  </marc:datafield>
  = <marc:datafieldtag="040"ind1=""ind2="">
    <marc:subfieldcode="a">IEU</marc:subfield>
    <marc:subfieldcode="c">IEU</marc:subfield>
    <marc:subfieldcode="d">N@F</marc:subfield>
    <marc:subfieldcode="d">DLC</marc:subfield>
  </marc:datafield>
  = <marc:datafieldtag="042"ind1=""ind2="">
    <marc:subfieldcode="a">lccopycat</marc:subfield>
  </marc:datafield>
  = <marc:datafieldtag="043"ind1=""ind2="">
    <marc:subfieldcode="a">n-us-dc</marc:subfield>
    <marc:subfieldcode="a">n-us---</marc:subfield>
  </marc:datafield>
  = <marc:datafieldtag="050"ind1="0"ind2="0">
    <marc:subfieldcode="a">F204.W5</marc:subfield>
  </marc:datafield>
  = <marc:datafieldtag="082"ind1="1"ind2="0">
    <marc:subfieldcode="a">975.3</marc:subfield>
    <marc:subfieldcode="2">13</marc:subfield>
  </marc:datafield>
  = <marc:datafieldtag="245"ind1="0"ind2="4">
    <marc:subfieldcode="a">The White House</marc:subfield>
    <marc:subfieldcode="h">[computer file].</marc:subfield>
  </marc:datafield>
  = <marc:datafieldtag="256"ind1=""ind2="">
    <marc:subfieldcode="a">Computer data.</marc:subfield>
  </marc:datafield>
  = <marc:datafieldtag="260"ind1=""ind2="">
    <marc:subfieldcode="a">Washington, D.C. :</marc:subfield>
    <marc:subfieldcode="b">White House Web Team,</marc:subfield>
    <marc:subfieldcode="c">1994-</marc:subfield>
  </marc:datafield>
  = <marc:datafieldtag="538"ind1=""ind2="">
    <marc:subfieldcode="a">Mode of access: Internet.</marc:subfield>
  </marc:datafield>
  = <marc:datafieldtag="500"ind1=""ind2="">

```

```

    <marc:subfieldcode="a">Title from home page as viewed on Aug. 19,
      2000.</marc:subfield>
  </marc:datafield>
- <marc:datafieldtag="520"ind1="8"ind2="">
  <marc:subfieldcode="a">Features the White House. Highlights the Executive
    Office of the President, which includes senior policy advisors and offices
    responsible for the President's correspondence and communications,
    the Office of the Vice President, and the Office of the First Lady. Posts
    contact information via mailing address, telephone and fax numbers,
    and e-mail. Contains the Interactive Citizens' Handbook with
    information on health, travel and tourism, education and training, and
    housing. Provides a tour and the history of the White House. Links to
    White House for Kids.</marc:subfield>
  </marc:datafield>
- <marc:datafieldtag="610"ind1="2"ind2="0">
  <marc:subfieldcode="a">White House (Washington, D.C.)</marc:subfield>
  </marc:datafield>
- <marc:datafieldtag="610"ind1="1"ind2="0">
  <marc:subfieldcode="a">United States.</marc:subfield>
  <marc:subfieldcode="b">Executive Office of the President.</marc:subfield>
  </marc:datafield>
- <marc:datafieldtag="610"ind1="1"ind2="0">
  <marc:subfieldcode="a">United States.</marc:subfield>
  <marc:subfieldcode="b">Office of the Vice President.</marc:subfield>
  </marc:datafield>
- <marc:datafieldtag="610"ind1="1"ind2="0">
  <marc:subfieldcode="a">United States.</marc:subfield>
  <marc:subfieldcode="b">Office of the First Lady.</marc:subfield>
  </marc:datafield>
- <marc:datafieldtag="710"ind1="2"ind2="">
  <marc:subfieldcode="a">White House Web Team.</marc:subfield>
  </marc:datafield>
- <marc:datafieldtag="856"ind1="4"ind2="0">
  <marc:subfieldcode="u">http://www.whitehouse.gov</marc:subfield>
  </marc:datafield>
- <marc:datafieldtag="856"ind1="4"ind2="0">
  <marc:subfieldcode="u">http://lcweb.loc.gov/staff/wpp/whitehouse.ht
    ml</marc:subfield>
  <marc:subfieldcode="z">Web site archive</marc:subfield>
  </marc:datafield>
</marc:record>
</marc:collection>

```

FIGURA 24: Exemplo de uma coleção (registros múltiplos)

FONTE: <http://www.loc.gov/standards/marcxml/xml/collection.xml>. Acesso em: 23 jun. 2003.

Segundo Eíto Brun (2002, slide 21, tradução nossa), o esquema MARCXML (como no exemplo mostrado acima) contém:

O elemento raiz **Collection** [primeiro elemento que pode conter 0 ou muitos registros], e que tem um elemento **Record** para cada registro;

O elemento **Record** contém:

- Um elemento não vazio **Leader** para o início,
- Elementos não vazios **Controlfield** para cada campo de controle e;
- Elementos não vazios **Datafield** para cada campo de dados;

O início é indicado por um elemento **Leader**;

Os campos de controle são indicados com elementos **Controlfield**;

Cada campo de controle tem um atributo – **Tag** – que abriga seu código (este indicado mediante 3 dígitos). Entre suas etiquetas de início e de fim é escrito seu valor;

Os campos são representados mediante elementos **Datafield**;

A etiqueta (nome do campo) e os indicadores são escritos com os valores dos atributos **Tag, ind 1 e ind 2** do elemento **Datafield**;

Os subcampos são como elementos filhos – **Subfield** – do elemento **Datafield** correspondente ao campo;

Os códigos dos subcampos são escritos com o valor dos atributos **Code**;

O valor do subcampo é escrito entre as etiquetas de início e fim do elemento **Subfield**. (ver figura 24).

O esquema MARCXML é uma parte do núcleo da arquitetura do MARC DTD XML principal. Devido à variedade de etiquetas no formato MARC21, o esquema serve como um formato ideal que outros formatos de metadados possam projetar-se para outros formatos e vice-versa. Por exemplo, pode-se criar registros MODS a partir de metadados MARC21 e também se pode criar registros MARC21 a partir de metadados MODS. A discrepância dos conjuntos de elementos de metadados para os meios normalmente traçados para MARC21 e traçados de MARC21 podem não ser espelhos exatos um do outro, requerendo uma especificação e transformação para cada direção (KEITH, 2004). Isto quer dizer, em outras palavras, que um registro MARC21 poderia ser convertido para MODS, mas um registro MODS pode não ser capaz de ser convertido de volta para o registro MARC21 original por causa das múltiplas *tags* do MARC21 que foram traçadas para uma única *tag* do MODS (NEEDLEMAN, 2005).

Segundo Keith (2004, p. 125, tradução nossa, grifo nosso),

A representação MARCXML dos registros MARC21 serve como um ônibus no qual a informação é levada junta. Quando um formato de metadados específico é traçado para MARCXML, idealmente por XSLT, a informação é livre para fluir sobre o ‘ônibus’, expondo assim os metadados para toda as características ligadas

ao ônibus. Mas também traçando na direção oposta, de MARCXML de volta para um formato de metadados específico, a informação pode ser **derrubada para fora do ‘ônibus’**, criando registros nestes formatos específicos de outros formatos que são habilitados no ‘ônibus’ MARCXML, por exemplo, registros Dublin Core podem ser criados de registros MODS porque os traçados existem de MODS para MARCXML e de MARCXML para Dublin Core.

Isto é ilustrado na figura que se segue:

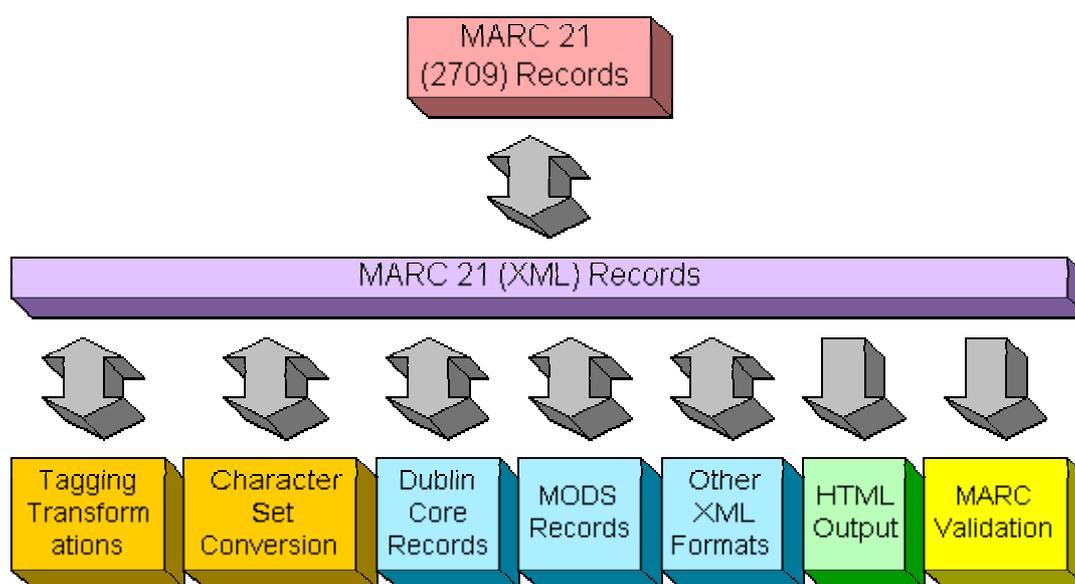


FIGURA 25: Conversão MARC 21 (2709) de/para MARC 21 (XML)
FONTE: CHAVES, 2002, p. 33

No entanto torna-se necessário explicar que, na conversão de registros MARC (um padrão mais complexo) para registros Dublin Core (mais simples que o primeiro), por exemplo, são desconsiderados alguns dados (muitos dos dados contemplados no formato MARC não o são no Dublin Core), o que pode inviabilizar ou tornar pouco representativa a conversão de volta (de Dublin Core para MARC), pois esta conversão resultará em um registro MARC incompleto, ou seja, com menos elementos de representação.

Porém, através da conversão entre MARC 21 (2709) e MARCXML, conforme ilustrado na Figura 25, atividades como a transformação de etiquetas, geração e conversão

para outros formatos de dados e validação de registros MARC, entre outras, podem ser realizadas. De acordo com Keith (2004, p. 124, tradução nossa),

a meta em desenvolver o esquema foi permitir acesso fácil a trechos discretos de dados. Com este esquema é muito fácil acessar a informação até o nível do sub campo com um XPATH simples (www.w3.org/TR/xslt) para manipular e transformar os dados. Possíveis tipos de transformação de uma folha de estilo usada para criar uma página de HTML sem os registros MARC como exibição, até a folhas de estilo complexas para transformação para outros formatos de metadados como o MODS. Embora a XSLT não ofereça as características completas e o controle de uma linguagem de programação como Java, é bastante surpreendente o que pode ser realizado com a XSLT.

As transformações mantidas dos dados XML em folhas de estilo resultam em muitas conseqüências positivas. As folhas de estilo XSLT são facilmente modificáveis. Não há nenhuma necessidade de recompilar ou ter ferramentas de software especiais para mudar as folhas de estilo. Profissionais de biblioteca que não são desenvolvedores de softwares podem fazer mudanças para transformações com a mínima assistência. Documentos XSLT são arquivos de texto simples parecidos aos documentos XML que são editáveis com processadores de textos básicos. Muitas ferramentas de software existem para tornar mais fácil desenvolver e manter folhas de estilo.

Como conseqüência de fazer do MARCXML um esquema mais simples do que o DTD XML, a habilidade para validar a estrutura MARC21 do registro foi perdida. Embora o esquema não valide a estrutura MARC21, 'ferramentas' externas ao esquema estão sendo desenvolvidas para realizar a validação. A LC começou a desenvolver folhas de estilo que podem ser usadas para validar os registros MARC. Escrevendo por uma folha de estilo para executar a validação, implementadores são capazes de modificar as regras de validação para as suas próprias práticas locais (KEITH, 2004).

A LC está com uma abordagem de multi-níveis para a validação. O primeiro nível é uma simples checagem se a XML está bem-formatada. O segundo nível de validação é a comparação com o esquema MARCXML (como declarado anteriormente, esta é a validação ISO2709 básica de estrutura de registro). O terceiro nível de validação examina a estrutura do próprio registro MARC21, conferindo as *tags* válidas, indicadores válidos, e códigos de subcampos com a informação extraída da documentação do MARC. O quarto nível está para ser desenvolvido, ele faria a validação do conteúdo atual dos registros MARC com as regras de catalogação utilizadas, tal como o AACR2 (KEITH, 2004).

A LC tem desenvolvido também um ‘Kit de Ferramentas’ – *toolkit* – que nada mais é do que um conjunto de aplicações/utilidades escritas em Java para facilitar a conversão de registros MARC21 para ‘XML’ e vice-versa e também permite o usuário aplicar suas próprias XSLT para transformar registros MARC21. O Kit de ferramentas MARCXML usa o MARC4J⁴, software *open-source* para manipulação de registros MARC (KEITH, 2004).

O esquema **MARCXML** foi desenvolvido pela LC para a padronização do intercâmbio de dados estruturados do MARC em XML. É um exemplo recente de adaptação para as novas tecnologias na história do MARC. Segundo Keith (2004), o desafio para a ‘comunidade dos metadados’ é trazer o **valor semântico** dos padrões existentes, tal como o MARC, e implementá-los com as novas tecnologias e ferramentas para prover esta contínua evolução. Desta forma, estimular o esquema MARCXML como um padrão de intercâmbio sólido permite a comunidade bibliotecária criar uma base ampla de ferramentas de softwares reutilizáveis (de fontes livres e de baixo custo) e assim, facilitar o fluxo da informação independentemente do formato.

⁴ <http://marc4j.tigris.org/>

5.1 MODS

O escritório de desenvolvimento de redes e padrões MARC da Biblioteca do Congresso Americano desenvolveu também o já citado anteriormente **MODS**⁵ – *Metadata Object Description Schema* – em consulta com especialistas interessados em satisfazer a necessidade expressada por uma versão XML abreviada do MARC 21. Basicamente o MODS

[...] é um esquema XML com etiquetas (tags) baseadas em linguagem que incluem um subconjunto de elementos de dados derivados do MARC21. Ele é planejado para carregar dados selecionados dos registros MARC21 existentes como também habilitar a criação de registros originais de descrição de recurso (GUENTHER; McCALLUM, 2003, p. 12, tradução nossa).

Ou seja, o MODS é um esquema de linguagem de marcação extensível para descrever uma entidade bibliográfica da mesma maneira como o MARC21. É planejado para uma variedade de usos e aplicações como um formato para transmissão que usa o Z39.50. Atualmente esta sendo usado nos experimentos do Z39.50 *International Next Generation* (ZING), na descrição bibliográfica básica, em aplicações de *harvesting* da *Open Archives Initiative* (OAI), e outras (NEEDLEMAN, 2004).

O MODS foi desenvolvido para complementar outros formatos de metadados e proporcionar uma alternativa entre um formato de metadados simples, com um mínimo de campos e pouca ou nenhuma sub-estrutura como o Dublin Core, e um formato mais detalhado, com muitos elementos de dados e tendo várias complexidades estruturais tal como o MARC21. O MODS tem um alto nível de compatibilidade com os registros MARC porque herda a semântica dos elementos de dados equivalentes no MARC21, ou seja, a maioria dos elementos que estão definidos em MODS tem equivalentes no formato bibliográfico

⁵ <http://www.loc.gov/standards/mods/>

MARC21, assim, um elemento em MODS <titleInfo> tem o seu significado detalhado na *tag* equivalente no MARC21 ‘245’ (GUENTHER; McCALLUM, 2003; NEEDLEMAN, 2004). Desta forma, a perda de informação nas transformações⁶ (de MARC para MODS) é menor.

Assim, é mais rico que o Dublin Core e mais compatível com dados de biblioteca, porém, é também mais simples que o formato MARC completo (ou como a ISO2709 ou o MARCXML completo), ou seja, “ele provê uma capacidade de descrição mais rica que o Dublin Core (DC) contudo não é tão complexo quanto as *tags* disponíveis no padrão MARC21 completo, assim eles está posicionado em algum lugar entre o DC e o MARC21” (NEEDLEMAN, 2004, p. 70). O MODS contém um subconjunto de *tags* disponíveis em MARC e é mais “amigável” porque usa etiquetas baseadas em linguagem (alfanuméricas) que podem ser entendidas facilmente por qualquer um que lida com o registro “cru”, ao invés das tradicionais etiquetas numéricas para MARC.

A figura 26 em seguida é um exemplo de um registro bibliográfico em MODS:

```
<?xml version="1.0" encoding="UTF-8" ?>
<mods xmlns:xlink="http://www.w3.org/1999/xlink" version="3.0"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xmlns="http://www.loc.gov/mods/v3"
  xsi:schemaLocation="http://www.loc.gov/mods/v3
  http://www.loc.gov/standards/mods/v3/mods-3-0.xsd">
  <titleInfo>
    <title>Sound and fury :</title>
    <subTitle>the making of the punditocracy /</subTitle>
  </titleInfo>
  <name type="personal">
    <namePart>Alterman, Eric</namePart>
    <role>
      <roleTerm type="text">creator</roleTerm>
    </role>
  </name>
  <typeOfResource>text</typeOfResource>
  <genre authority="marc">bibliography</genre>
  <originInfo>
    <place>
      <placeTerm authority="marccountry" type="code">nyu</placeTerm>
    </place>
    <place>
      <placeTerm type="text">Ithaca, N.Y</placeTerm>
    </place>
  </originInfo>
</mods>
```

⁶ A conversão entre MARC 2709 e MODS sempre será feita mediante MARCXML; nunca diretamente, conforme a figura 20.

```

    <publisher>Cornell University Press</publisher>
    <dateIssued>c1999</dateIssued>
    <dateIssued encoding="marc">1999</dateIssued>
    <issuance>monographic</issuance>
  </originInfo>
  = <language>
    <languageTerm authority="iso639-2b" type="code">eng</languageTerm>
  </language>
  = <physicalDescription>
    <form authority="marcform">print</form>
    <extent>vii, 322 p. ; 23 cm.</extent>
  </physicalDescription>
  <note type="statement of responsibility">Eric Alterman.</note>
  <note>Includes bibliographical references (p. 291-312) and index.</note>
  = <subject authority="lcsht">
    <topic>Journalism</topic>
    <topic>Political aspects</topic>
    <geographic>United States.</geographic>
  </subject>
  = <subject authority="lcsht">
    <geographic>United States</geographic>
    <topic>Politics and government</topic>
    <temporal>20th century.</temporal>
  </subject>
  = <subject authority="lcsht">
    <topic>Mass media</topic>
    <topic>Political aspects</topic>
    <geographic>United States.</geographic>
  </subject>
  = <subject authority="lcsht">
    <topic>Television and politics</topic>
    <geographic>United States.</geographic>
  </subject>
  = <subject authority="lcsht">
    <topic>Press and politics</topic>
    <geographic>United States.</geographic>
  </subject>
  = <subject authority="lcsht">
    <topic>Talk shows</topic>
    <geographic>United States.</geographic>
  </subject>
  <classification authority="lcc">PN4888.P6 A48 1999</classification>
  <classification edition="21" authority="ddc">071/.3</classification>
  = <recordInfo>
    <recordContentSource>DLC</recordContentSource>
    <recordCreationDate encoding="marc">990730</recordCreationDate>
    <recordChangeDate encoding="iso8601">20000406144503.0</recordChangeDate>
    <recordIdentifier>11761548</recordIdentifier>
  </recordInfo>
</mods>

```

FIGURA 26: Exemplo de um registro de livro em MODS

FONTE: www.loc.gov/standards/mods/. Acesso em: 29 ago. 2005.

Os principais elementos em MODS, segundo Eíto Brun (2003, slides 38-41):

- **titleInfo** – é obrigatório. Contém ao menos um sub-elemento chamado **title**
- **name** – corresponderia ao “cabeçalho”. Pode ser de pessoa, organização ou conferência. O tipo de cabeçalho se indica mediante um atributo **type** que aceitará os valores **personal**, **corporate** ou **conference** o **name** pode incluir os sub-elementos : **namePart**, **role**, **affiliation** e **displayForm**

- **typeOfResource** – tipo de recurso
- **genre** – gênero. Não tem sub-elementos. Só tem um atributo **authority** que levará o valor **marc**. Seu valor levará da lista disponível na URL: <http://www.loc.gov/marc/sourcecode/genre/>
- **publicationInfo** – agrupa os dados sobre o “editor”, mediante os sub-elementos **placeCode**, **place**, **publisher**, **dateIssued**, **dateCreated**, **edition**, etc.
- **language** – idioma do recurso
- **physicalDescription** – descrição física do recurso. Inclui os sub-elementos: **form**, **extent**, **digitalOrigin**, etc.
- **abstract** – resumo; pode apontar para um recurso remoto
- **tableOfContents** – incluirá uma URL que aponta para um recurso remoto com esta informação
- **targetAudience** – público para o qual é dirigido conteúdo
- **note** – para codificar notas. Pode apontar para uma nota disponível como um recurso externo mediante um hipervínculo
- **cartographics** – para informação cartográfica, contém os sub-elementos coordenadas, escala e projeção
- **subject** – temas de que trata o recurso. Inclui os sub-elementos **topic**, **geographic** e **temporal**
- **Classification** – código de classificação segundo um esquema de classificação que se indicará o atributo **authority** deste elemento
- **relatedItem** – permite relacionar elementos
- **Identifier** – ID do recurso. Pode ser um ISBN, ISSN, etc. Seu tipo se indicará mediante um atributo **type**
- **Location** – Sub-elementos: **physicalLocation**, **URL**
- **accessConditions** – (MARC 21 fields 506 and 540)
- **Extension** – (MARC 21 campo 9XX)
- **recordInfo** – contém os seguintes sub-elementos: **recordCreationDate**, **recordChangeDate**, **recordIdentifier**, etc.

Certos elementos do MODS definem conceitos que ocorrem periodicamente em mais de um elemento como sub-elementos. A XML facilita usando a mesma definição para elementos múltiplos. Por exemplo, “*name*” pode ser o nome primário associado com o recurso ou um nome associado com um item relacionado. Em outras palavras, “[...] em certos casos se agrupam elementos de MARC em um mesmo elemento” (EÍTO BRUN, 2003, slide 35); em MODS, ambos usam a mesma definição. Este conceito é certamente presente em MARC 21, mas não tão constantemente quanto em MODS.

“Aparentemente, o MODS é a opção que deveria ser imposta como sistema para a descrição de recursos na Internet” (EÍTO BRUN, 2002, slide 36). Neste sentido, segundo Guenther e Mccallum (2003, p. 13, tradução nossa),

A necessidade por um padrão de metadados rico, tal como o MODS tem sido expressada por membros de bibliotecas digitais e comunidades relacionadas que tentam implementar projetos que envolvem a busca e recuperação, a administração de objetos digitais complexos, a integração de metadados de bases de dados de biblioteca com outras fontes não MARC e outras funções.

O OAI- PMH permite a coleta de registros MARC de múltiplos sistemas tornando-os amplamente disponíveis. Geralmente, os registros têm estado disponíveis em MARC (usando etiquetas MARC e sintaxe em MARCXML) ou Dublin Core simples em XML. A LC está planejando incorporar o MODS como um formato alternativo para seus mais de 100.000 registros de metadados que descrevem várias formas de materiais. Isto permitirá a exportação de metadados mais ricos que o registro Dublin Core, que não contempla muito dos metadados, porém provendo dados mais simples que o MARCXML completo.

5.1.1 MADS e METS: Metadados associados

A LC tem desenvolvido também o MADS⁷ um esquema XML para um conjunto de elementos de autoridade que pode ser usado para prover metadados sobre agentes (pessoas, organizações), eventos, e termos (tópicos, geográficos, gêneros, etc.). O MADS foi criado para servir como um companheiro ao *Metadata Object Description Schema* (MODS). Como tal, o MADS tem uma relação com o formato de Autoridade do MARC 21, como o MODS tem com o Bibliográfico MARC 21 – ambos levam dados selecionados do MARC 21. O MADS é expresso usando a linguagem XML *schema* do *World Wide Web Consortium* (W3C). O padrão está sendo desenvolvido e será mantido pelo escritório de desenvolvimento de redes e padrões MARC da Biblioteca do Congresso Americano com a contribuição dos usuários.

Um registro MADS inclui um ou mais itens. (O termo "item" é usado para conveniência de vocabulário e não está presente no XML *schema* ou no exemplo.) Um item é um elemento <authority>, <related>, ou <variant>. Deve haver exatamente um <authority> e pode haver qualquer quantidade de itens <related> and/or <variant>. Os itens são seguidos através de elementos adicionais misturados. Cada item tem um ou mais descritores; cada descritor é um dos elementos seguintes: <name>, <titleInfo>, <tópico>, <temporal>, <gêner>, <geographic>, <hierarchicalGeographic> ou <occupation>. Um exemplo de registro MADS segue logo abaixo:

```
<?xml version="1.0" encoding="UTF-8" ?>
=<mads xmlns="http://www.loc.gov/mads/" xmlns:mods="http://www.loc.gov/mods/v3"
  xmlns:xlink="http://www.w3.org/1999/xlink"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xsi:schemaLocation="http://www.loc.gov/mads/ mads.xsd">
=<authority>
  =<name authority="naf">
    <namePart>Bach, P D Q</namePart>
  </name>
  =<titleInfo>
```

⁷ <http://www.loc.gov/standards/mads/>

```

      <mods:title>Symphony in C</mods:title>
      <mods:partName>Adagio</mods:partName>
    </titleInfo>
  </authority>
  = <variant type="other">
    = <name>
      <namePart>Bach, Pretty D Q</namePart>
    </name>
    = <titleInfo>
      <mods:title>Symphony in Sea</mods:title>
      <mods:subTitle>the drowning symphony</mods:subTitle>
      <mods:partName>Rapidissimo</mods:partName>
    </titleInfo>
  </variant>
  <note>This guy is nuts.</note>
</mads>

```

FIGURA 27: Exemplo de registro MADS

FONTE: <http://www.loc.gov/standards/mads/>. Acesso em: 29 ago. 2005.

O METS⁸ é um esquema baseado em XML que provê os meios para “empacotar” (ou encapsular), ou evidenciar a descrição, administração, estruturação, correção e condução de metadados para recursos digitais, além de outros dados necessários para a recuperação, preservação e o fornecimento em recursos digitais. Ou seja, o METS permite que a biblioteca tenha acesso a metadados técnicos apropriados para lhe permitir atualizar e migrar os dados, garantindo a durabilidade dos recursos.

O METS apóia o fluxo sem costura de metadados e recursos eletrônicos entre sistemas em rede. Segundo Guenther e McCallum (2003, p. 13, tradução nossa), o “METS: Codificador de metadados e Padrão de Transmissão é um esquema XML altamente flexível para empacotar os metadados descritivos e vários outros tipos importantes de metadados, necessários para assegurar o uso e preservação dos recursos digitais”. Pois sem metadados estruturais, os arquivos com imagens ou texto que compõem uma obra digital serão de pouca utilidade, e sem metadados técnicos sobre o processo de digitalização, os pesquisadores poderão ter dúvidas sobre a exatidão da reflexão do original que a versão digital oferece (METS - *Metadata Encoding & Transmission Standard*, 2003).

⁸ <http://www.loc.gov/standards/mets/>

Os metadados METS são essenciais para um repositório de material digital, onde os recursos digitais são armazenados juntos com a informação sobre os recursos. Um repositório que pode levar muitas configurações é o instrumento para acesso e preservação dos objetos. Os dados METS também são importantes para o intercâmbio de objetos digitais para visualização e uso através de outros sistemas. Se o recurso digital estiver com a descrição do METS, o arquivo poderá ser utilizável em muitas atividades no sistema receptor.

Um documento METS consiste em sete seções principais (METS - *Metadata Encoding & Transmission Standard*, 2003):

1. **Cabeçalho METS** - O cabeçalho METS contém metadados descrevendo o documento METS em si, incluindo informação como o criador, editor, a data e a hora em que o registro METS foi criado, o estado do processamento do registro, etc.

Exemplo:

```
<metsHdr CREATEDATE="2003-07-04T15:00:00" RECORDSTATUS="Complete">
  <agent ROLE="CREATOR" TYPE="INDIVIDUAL">
    <name>Jerome McDonough</name>
  </agent>
  <agent ROLE="ARCHIVIST" TYPE="INDIVIDUAL">
    <name>Ann Butler</name>
  </agent>
</metsHdr>
```

2. **Metadados Descritivos** - A seção de metadados descritivos pode apontar para metadados descritivos externos ao documento METS (ex., um registro MARC num OPAC ou um registro EAD mantido num servidor Web), ou conter metadados descritivos embebidos (internos), ou ambos. Múltiplas instâncias de metadados descritivos, tanto internas como externas, podem ser incluídos na seção de metadados descritivos.

Ex: metadados descritivos **Externos**:

```
<dmdSec ID="dmd001">
  <mdRef LOCTYPE="URN" MIMETYPE="application/xml" MDTYPE="EAD"
    LABEL="Berol Collection Finding Aid">urn:x-nyu:fales1735</mdRef>
</dmdSec>
```

Ex: metadados descritivos **Internos**:

```
<dmdSec ID="dmd002">
  <mdWrap MIMETYPE="text/xml" MDTYPE="DC" LABEL="Dublin Core Metadata">
    <xmlData>
      <dc:title>Alice's Adventures in Wonderland</dc:title>
      <dc:creator>Lewis Carroll</dc:creator>
      <dc:date>between 1872 and 1890</dc:date>

      <dc:publisher>McCloughlin Brothers</dc:publisher>
      <dc:type>text</dc:type>
    </xmlData>
  </mdWrap>
</dmdSec>

<dmdSec ID="dmd003">
  <mdWrap MIMETYPE="application/marc" MDTYPE="MARC" LABEL="OPAC Record">
    <binData>MDI0ODdjam0gIDIyMDA1ODkgYSA0NU0wMDAxMDA...(etc.)
    </binData>
  </mdWrap>
</dmdSec>
```

3. **Metadados Administrativos** - A seção de metadados administrativos oferece informação sobre como os arquivos foram criados e armazenados, direitos de propriedade intelectual, metadados sobre o objeto original a partir do qual o objeto digital foi derivado, e informação sobre a proveniência dos arquivos que compõem o objeto digital (i.e., relações de arquivos originais/derivados, e informação de migração/transformação). Tal como os metadados descritivos, os metadados administrativos podem ser tanto externos ao documento METS, ou codificados internamente.

Exemplo:

```
<techMD ID="AMD001">
  <mdWrap MIMETYPE="text/xml" MDTYPE="NISOIMG" LABEL="NISO Img. Data">
    <xmlData>
      <niso:MIMETYPE>image/tiff</niso:MIMETYPE>
      <niso:Compression>LZW</niso:Compression>
      <niso:PhotometricInterpretation>8</niso:PhotometricInterpretation>

      <niso:Orientation>1</niso:Orientation>
      <niso:ScanningAgency>NYU Press</niso:ScanningAgency>
    </xmlData>
  </mdWrap>
</techMD>
```

4. **Seção de Arquivos** - A seção de arquivos lista todos os arquivos que contêm as versões eletrônicas do objeto digital. Elementos <file> podem ser agrupados em elementos <fileGrp>, para permitir a subdivisão de arquivos por versão do objeto.

Exemplo:

```
<fileSec>
  <fileGrp ID="VERS1">
    <file ID="FILE001" MIMETYPE="application/xml" SIZE="257537"
      CREATED="2001-06-10">
      <FLocat LOCTYPE="URL">http://dlib.nyu.edu/tamwag/beame.xml</FLocat>
    </file>
  </fileGrp>
  <fileGrp ID="VERS2">
    <file ID="FILE002" MIMETYPE="audio/wav" SIZE="64232836"
      CREATED="2001-05-17" GROUPID="AUDIO1">
      <FLocat LOCTYPE="URL">http://dlib.nyu.edu/tamwag/beame.wav</FLocat>
    </file>
  </fileGrp>
  <fileGrp ID="VERS3" VERSDATE="2001-05-18">
    <file ID="FILE003" MIMETYPE="audio/mpeg" SIZE="8238866"
      CREATED="2001-05-18" GROUPID="AUDIO1">
      <FLocat LOCTYPE="URL">http://dlib.nyu.edu/tamwag/beame.mp3</FLocat>
    </file>
  </fileGrp>
</fileSec>
```

5. **Mapa Estrutural** - O Mapa Estrutural é o coração do documento METS. Ele esboça uma estrutura hierárquica para o objeto da biblioteca digital, e liga os elementos dessa estrutura a arquivos com conteúdos e metadados referentes a cada elemento.

Exemplo:

```
<structMap TYPE="logical">
  <div ID="div1" LABEL="Oral History: Mayor Abraham Beame"
    TYPE="história falada">
    <div ID="div1.1" LABEL="Interviewer Introduction"
      ORDER="1">
    <fptr FILEID="FILE001">
    <area FILEID="FILE001" BEGIN="INTVWBG" END="INTVWND"
      BETYPE="IDREF" />
    </fptr>
    <fptr FILEID="FILE002">
    <area FILEID="FILE002" BEGIN="00:00:00" END="00:01:47"
      BETYPE="TIME" />
    </fptr>
    <fptr FILEID="FILE003">
    <area FILEID="FILE003" BEGIN="00:00:00" END="00:01:47"
      BETYPE="TIME" />
    </fptr>
  </div>
  <div ID="div1.2" LABEL="História da Família" ORDER="2">
  <fptr FILEID="FILE001">
  <area FILEID="FILE001" BEGIN="FHBG" END="FHND"
    BETYPE="IDREF" />
  </fptr>
```

```

    </fptr>

    <fptr FILEID="FILE002">
      <area FILEID="FILE002" BEGIN="00:01:48"END="00:06:17"
        BETYPE="TIME" />
    </fptr>
    <fptr FILEID="FILE003">
      <area FILEID="FILE003" BEGIN="00:01:48" END="00:06:17"
        BETYPE="TIME" />
    </fptr>
  </div>
  <div ID="div1.3" LABEL="Introdução ao Sindicato dos Professores"
    ORDER="3">
    <fptr FILEID="FILE001">
      <area FILEID="FILE001" BEGIN="TUBG" END="TUND"
        BETYPE="IDREF" />
    </fptr>
    <fptr FILEID="FILE002">

      <area FILEID="FILE002" BEGIN="00:06:18" END="00:10:03"
        BETYPE="TIME" />
    </fptr>
    <fptr FILEID="FILE003">
      <area FILEID="FILE003" BEGIN="00:06:18" END="00:10:03"
        BETYPE="TIME" />
    </fptr>
  </div>
</div>
</structMap>

```

6. **Ligações Estruturais** - A seção de Ligações Estruturais do METS permite aos criadores METS registrar a existência de hiperligações entre nós na hierarquia esboçada no Mapa Estrutural. Esta seção tem um valor particular na utilização do METS para arquivar *sites*.

Exemplo:

```

<div ID="P1" TYPE="page" LABEL="Page 1">
  <fptr FILEID="HTMLF1"/>
  <div ID="IMG1" TYPE="image" LABEL="Imagem com hiperligação
para a página 2">
    <fptr FILEID="JPGF1"/>
  </div>
</div>
<div ID="P2" TYPE="page" LABEL="Page 2">
  <fptr FILEID="HTMLF2"/>
</div>

```

7. **Comportamento** - Uma seção de comportamento pode ser usada para associar comportamentos executáveis com o conteúdo no objeto METS. Cada comportamento numa seção de comportamento tem um elemento de definição de interface que representa uma definição abstrata do conjunto de comportamentos representado por uma seção de comportamento particular. Cada comportamento também tem um

elemento de mecanismo que identifica um módulo de código executável que implementa e executa os comportamentos definidos de forma abstrata pela definição de interface.

Exemplo:

```
<METS:behavior ID="DISS1.1" STRUCTID="S1.1" BTYPE="uva-bdef:stdImage"
  CREATED="2002-05-25T08:32:00" LABEL="UVA Std Image Disseminator"
  GROUPID="DISS1" ADMID="AUDREC1">
  <METS:interfaceDef LABEL="UVA Standard Image Behavior Definition"
    LOCTYPE="URN" xlink:href="uva-bdef:stdImage"/>
  <METS:mechanism LABEL="A NEW AND IMPROVED Image Mechanism"
    LOCTYPE="URN" xlink:href="uva-bmech:BETTER-imageMech"/>
</METS:behavior>
```

O esquema METS oferece um mecanismo flexível para codificar metadados descritivos, administrativos e estruturais para um objeto de uma biblioteca digital, e para exprimir as ligações complexas entre estas várias formas de metadados. Assim o METS oferece uma norma útil para a troca de objetos digitais entre repositórios. Adicionalmente, o METS oferece a possibilidade de associar um objeto digital com comportamentos ou serviços. [...] uma examinação mais detalhada do esquema e da sua documentação é necessária para compreender todo o alcance das suas capacidades (METS - Metadata Encoding & Transmission Standard, 2003).

As comunidades de biblioteca têm desenvolvido bem as descrições bibliográficas tradicionais que com alguns ajustes para os recursos digitais, o desenvolvimento do MODS, servirá o futuro digital. No quadro dos metadados extensos, o desenvolvimento do METS é um grande passo para trazer aos metadados não descritivos a estabilidade necessária para um ambiente de Internet funcional onde os recursos eletrônicos fluem perfeitamente entre sistemas. Estes desenvolvimentos relacionam-se bem ao OAIS (*Open Archival Information System* – Sistema de Informação de Arquivos Abertos) Modelo de Referência (www.ccsds.org/RP9905/RP9905.html) que ajuda a definir os processos e limites na criação, administração, sustentação e serviços de recursos digitais. O pacote METS pode ser usado para colecionar metadados de recurso digital para submissão ao repositório, serve como lugar para os metadados dentro do repositório e é o provedor de informação às ferramentas que provêm os recursos aos clientes (GUENTHER; McCALLUM, 2003). Conforme Keith

(2004), colocar os metadados MARC em ambiente XML expõe a informação do MARC21 para um mundo novo de recursos de softwares para as pessoas. A arquitetura MARCXML, e seus componentes incluindo o MARCXML *Schema*, criam uma base e padronizam uma abordagem para os metadados MARC em XML. As aplicações apresentadas pretendem demonstrar apenas alguns dos muitos modos possíveis de manipular metadados na forma XML e provocar a reflexão para usos adicionais da arquitetura MARCXML sobretudo, para a descrição, preservação e recuperação dos recursos informacionais nos *open archives*.

6 CONSIDERAÇÕES FINAIS

“A descoberta consiste em ver o que todos vêem e pensar no que ninguém pensou.”

Albert Von Szent Gvorgvi (1893-1986)

As inovações tecnológicas, principalmente aquelas ligadas à informação e a comunicação, a todo instante promovem mudanças que atingem o fazer das pessoas e os processos das organizações, além de dinamizar, cada vez mais, o fluxo de informações.

Nesse sentido, a comunicação científica vem sofrendo alterações consideráveis tanto no processo quanto na estrutura e filosofia de comunicação dos resultados de pesquisa. As iniciativas de arquivos abertos e de acesso livre estão contribuindo significativamente para a desconstrução do modelo tradicional de comunicação científica e para a construção de um novo modelo desagregado e interoperável, mais justo e eficiente para disseminar o conhecimento gerado pelas comunidades científicas. Acredita-se que com este novo modelo e filosofia, a questão da visibilidade e acessibilidade à literatura científica se tornará mais eficiente. De acordo com Rodrigues (2004, p. 33), tudo indica que o acesso livre fará parte “do futuro da publicação científica e que as bibliotecas das instituições de ciência e tecnologia, e os seus profissionais, terão a oportunidade e a obrigação de reequacionar o seu posicionamento e demonstrar o seu valor na nova paisagem informativa”.

Assim, na medida em que este novo modelo for consolidado, e com ele os hábitos e a confiança em seus usos por parte dos pesquisadores, cientistas e profissionais da informação, é possível que se estabeleça uma significativa mudança cultural no sentido da gestão, socialização e acesso às informações resultantes das investigações científicas.

Atualmente, toda essa revolução digital faz com que o modo como é tratada a informação no aspecto documental adquira uma posição central, pois determina, não somente

um novo modo de transmissão da informação, como também uma proposta de novos objetivos, novas funções e novos modelos organizacionais (AUGUSTÍN LACRUZ, 1998). Até alguns anos atrás não se tinha o dinamismo de múltiplas formas de representação de um mesmo registro de metadados em ambiente digital. Hoje com as facilidades proporcionadas pela XML e tecnologias associadas como a XSL, essas questões vêm se tornando possíveis no ambiente computacional/digital.

Sabe-se que nos últimos anos, devido aos avanços das tecnologias de informação e comunicação, não somente a estrutura e o fluxo da comunicação científica vem sofrendo consideráveis alterações, como também o próprio conceito e suporte dos documentos científicos. Recentemente a estrutura do texto vem sendo amplamente considerada em virtude de inovações tecnológicas aplicadas nos atuais sistemas de informações e um dos resultados identificados nessa pesquisa diz respeito à linguagem de marcação XML como uma dessas inovações. Utilizada dentre outras formas, na marcação de documentos, foi possível realizar uma analogia entre a estrutura textual do texto científico e a linguagem XML a partir dos subsídios teóricos de Van Dijk, e concluir que a XML pode “marcar” e explicitar a estrutura textual do texto científico, de modo que sua superestrutura torne-se “visível”.

Isso significa que, a XML pode utilizar os subsídios teóricos da lingüística textual para criar normas com o objetivo de tornar as marcações ou *namespaces* de documentos científicos universalmente únicos e assim evitar conflitos entre sistemas no processo de intercâmbio de informações e dinamizar a busca e a recuperação de documentos em ambientes digitais, e uma das formas de facilitar todo esse processo é levar em consideração uma descrição bibliográfica completa e de alta qualidade dos recursos informacionais, que somente uma estrutura como a do padrão MARC pode proporcionar.

Conforme se pôde constatar e como aponta McCallum (2004), o MARC21 (como apontado anteriormente¹) está baseado em uma estrutura de formato ISO 2709, que é uma norma estabelecida mundialmente. Esta estrutura, que também é usada para UNIMARC e muitos outros formatos similares a MARC, têm funcionado bem e seguirá sendo importante por muitos mais anos. Porém, atualmente muitos protocolos novos estão mais adaptados a uma estrutura de registro bibliográfico baseado em XML. Em resposta a essa necessidade, nos últimos anos têm sido desenvolvidas estruturas alternativas para os elementos de dados em MARC21. A mais importante é o MARCXML, que provê um caminho de ida e volta ininterrupto de MARC21 a MARC em XML.

Outras estruturas têm sido desenvolvidas para descrever entidades bibliográficas – sobretudo digitais – da mesma maneira como o MARC21, tal como o MODS que herda a semântica dos elementos de dados equivalentes no formato bibliográfico MARC21. Sendo mais rico que o Dublin Core, porém, mais simples que o formato MARC completo, o MODS é a opção para a necessidade expressada pelos membros de bibliotecas digitais e comunidades relacionadas, por um padrão de metadados mais completo para a descrição de recursos na Internet (GUENTHER, McCALLUM, 2003).

Assim, estes desenvolvimentos têm sido importantes para permitir que os vastos ‘depósitos’ de dados em MARC21 sejam mais facilmente intercambiáveis com ferramentas e protocolos em XML. Exemplos destes protocolos e ferramentas são o protocolo da Iniciativa de Arquivos Abertos (OAI-PMH), o novo protocolo de busca e recuperação (SRW/SRU) baseado no Z39.50 e o padrão de codificação e recuperação de metadados (METS). A LC, usando estas ferramentas, vem permitindo que as respostas a buscas em Z39.50 e SRW sejam enviadas em vários formatos, incluindo MARC21, MARCXML, MODS e Dublin Core (MCCALLUM, 2004).

¹ ver seção 4.1.2 NORMA ISO 2709.

Assim, a estrutura MARC (*Machine Readable Cataloging*), consolidada no âmbito da Ciência da Informação e que tem sido utilizada há décadas, foi grandemente beneficiada pela linguagem de marcação XML, sendo potencializada na forma de extrair dessa estrutura diversas formas de representação da informação.

Nos últimos anos surgiram várias iniciativas de adaptação do MARC em XML para o ambiente da Internet, pois somente a estrutura MARC não efetiva a interoperabilidade no atual cenário digital. É preciso outras tecnologias acopladas a esta estrutura para garantir e efetivar a interoperabilidade entre sistemas de informação heterogêneos. Visto que esse formato possui uma estrutura completa que atende as necessidades de padronização de metadados e facilita a interoperabilidade entre sistemas de informação, devido a sua arquitetura organizacional, buscou-se adequar o MARC a um novo paradigma de comunicação científica baseado em um novo modelo de comunicação e no livre acesso às informações.

O formato MARCXML adequado pela *Library of Congress* a essa nova realidade, foi identificado nessa pesquisa como um padrão adequado para a representação descritiva dos recursos informacionais nos *open archives* visto que, de acordo com a literatura estudada, o MARC é uma **estrutura de descrição documentária** legível por máquina e a XML é uma linguagem computacional de **marcação de dados**, portanto, ferramentas diferentes mas que, no entanto, se complementam no sentido de que, o formato MARC serve como um recipiente no qual a informação é armazenada, enquanto que a XML proporciona, com inúmeras vantagens, o transporte desta informação em ambiente digital.

Sendo assim, a hipótese levantada no início dessa pesquisa foi confirmada, ou seja, com a implantação efetiva do formato MARCXML como uma estrutura de representação de recursos informacionais da área da Ciência da Informação, as questões do intercâmbio de dados e da interoperabilidade entre sistemas de informação, em especial na implantação dos

arquivos abertos, ficariam mais eficientes considerando-se que o MARC, como um padrão de metadados completo e flexível, garante a qualidade do registro enquanto que a XML garante a interoperabilidade entre os sistemas de informação heterogêneos, possibilitando, sobretudo, a otimização dos processos de busca e recuperação da informação.

A partir desse resultado, acredita-se que com o amadurecimento e consolidação da iniciativa de arquivos abertos (OAI), algumas de suas exigências passe a ser a utilização não do mínimo requerido para descrição bibliográfica, mas de uma estrutura descritiva mais completa na perspectiva da descrição de recursos informacionais como o MARCXML, por exemplo (ao menos para aquelas instituições que desejam atuar como provedoras de informações).

A medida que se aumenta a complexidade dos documentos nos repositórios e *open archives*, mais justifica-se uma estrutura de metadados, como a do formato MARCXML, que suporte a descrição das especificidades dos recursos informacionais, uma vez que esta iniciativa não está e nem estará se restringindo a documentos científicos, mas se expandindo a outros tipos de recursos informacionais cada vez mais complexos e específicos, demandando também uma descrição apropriada para a especificidade das entidades bibliográficas.

Ressalta-se ainda que, os formatos e estruturas de metadados, as linguagens de marcação de dados, as regras de catalogação, os modelos conceituais, os protocolos, as normas etc; estão, cada vez mais, indissociáveis, ou seja, fazem parte de uma intrincada relação e não podem mais ser tratados ou estudados separados uns dos outros (ou ao menos deixarem de ser levados em consideração uns dos outros), sobretudo nas pesquisas que abordem o tema da representação no campo da Biblioteconomia e Ciência da Informação. Recomenda-se que na sistematização de um novo modelo para o ensino da Representação tanto descritiva quanto temática (Catalogação e Indexação) nos cursos de Biblioteconomia, esta visão holística seja levada em consideração.

Com esta pesquisa, objetivou-se divulgar à comunidade de profissionais da informação a adequação do formato MARCXML para a descrição de recursos informacionais nos arquivos abertos como um padrão de metadados completo, flexível e adequado que possibilitará a interoperabilidade entre sistemas de informação heterogêneos.

Espera-se que esta pesquisa venha trazer subsídios para a área de Biblioteconomia e Ciência da Informação, contribuindo no que se refere à representação descritiva de recursos informacionais digitais em *open archives* suscitando outras discussões sobre este tema para que, desta forma, se de continuidade a este assunto atual e de extrema relevância para a área.

REFERÊNCIAS

- ABITEBOUL, S.; BUNEMAN, P.; SUCIU, D. **Gerenciando dados na Web**. Tradução de: Mônica Córdia. Rio de Janeiro: Campus, 2000, 251 p.
- AGENJO, Xavier. **Las bibliotecas nacionales del siglo XXI**. 2005.
- ALVARENGA, L. Representação do conhecimento na perspectiva da ciência da informação em tempo e espaço digitais. **Enc. Bibli: R. Eletr. Bibliotecon. Ci. Inf.**, Florianópolis, n. 15, 2003.
- ARELLANO, M. A. Preservação de documentos digitais. **Ciência da Informação**, v. 33, n.2, p. 15-27, maio/ago. 2004.
- ARMS et al. A spectrum of interoperability: the site for science prototype for the NSDL. **D-Lib Magazine**, v. 8, n. 1, 2002.
- ARNOLD, P. ; WHITE, P. **A era da automação**. Rio de Janeiro: Lidador, 1965, 206 p. Tradução de: Hersílio Soares.
- AUGUSTÍN LACRUZ, M. C. Bibliotecas digitales y sociedad de la información. **Scire**, v. 4, n. 2, jul./dic. p. 47-62, 1998.
- ANDRESEN, Leif. After MARC: what then? **Library Hi tech**, v. 22, n. 1, p. 40-51, 2004.
- ARAUJO, C. A. A.; SOUZA, R. R. As potencialidades das bibliotecas digitais de teses e dissertações ante a explosão informacional da pesquisa em comunicação. 2004. In: ENDOCOM – Encontro Nacional de Informação em Ciências da Comunicação, 14., 2004. **Anais...** Disponível em: <<http://www.portcom.intercom.org.br/portcom2/endocom2004/Araujo.PDF>>. Acesso em: 20 abr. 2005.
- BALBY, C. N. Formatos de intercâmbio de registros bibliográficos: conceitos básicos. **Cadernos da F.F.C.** Marília, v.4, n.1, 1995. p. 29-35.
- BARRUECO, J. M. ; COLL, I. S. 2003. **OAI-PMH**: protocolo para la transmisión de contenidos en Internet. Disponível em: <<http://www.uv.es/=barrueco/cardedu.doc>> Acesso em: 27 mar. 2004.
- BATESON, Gregory. **Mente e natureza**. Trad. Claudia Gerpe. Rio de Janeiro: Francisco Alves, 1986. (tradução do original; *Mind and nature: a necessary unity*, 1979).
- BAX, M. P. Introdução às linguagens de marcas. **Ciência da Informação**, Brasília, v. 30, n. 1, p. 32- 38, jan./abr. 2001.

BAX, M. P. **XML – Introdução e impactos nas bibliotecas**. 2002. Slides. Disponível em: <www.bax.com.br>. Acesso em: 20 nov. 2002.

BESSER, Howard. **Moving from Isolated Digital Collections to Interoperable Digital Libraries**. 2002. Disponível em: <<http://www.gseis.ucla.edu/~howard/Papers/vala01.html>>. Acesso em: 21 fev. 2005.

BRASIL. Ministério da Educação. Gabinete do Ministro. Fundação Coordenação de Aperfeiçoamento de Pessoal de Nível Superior. Portaria nº13, de 15 de fevereiro de 2006. Institui a divulgação digital das teses e dissertações produzidas pelos programas de doutorado e mestrado reconhecidos. **Diário Oficial [da] República Federativa do Brasil**, Brasília, DF, 17 fev. 2006. Disponível em: <<http://www.in.gov.br>>. Acesso em: 28 fev. 2006.

BROOKSHEAR, J. G. **Ciência da Computação: uma visão abrangente**. 5. ed. Porto Alegre: Bookman, 2000. 499p.

BUCK, A. M.; FLAGAN, R. C. ; COLES, B. 1999. **Scholar's Forum: A New Model For Scholarly Communication**. Disponível em: <<http://library.caltech.edu/publications/scholarsforum/scholarsforum.pdf>>. Acesso em: 17 jul. 2004.

BUSH, Vannevar. As We May Think. **The Atlantic Monthly**, v.176, n.1 p. 101-108, July. 1945. Disponível em: <<http://www.theatlantic.com/unbound/flashbks/computer/bushf.htm>>. Acesso em 09 jan. 2005.

CAFE, L. ; LAGE, M. B. Auto-arquivamento: Uma opção inovadora para a produção científica. **DataGramZero – Revista de Ciência da Informação**, 3, 3 (Junho 2002). Disponível em: <http://www.dgz.org.br/jun02/Art_04.htm>. Acesso em: 20 mai. 2004.

CAFÉ, L. et. al. Repositórios institucionais: nova estratégia para publicação científica na Rede. In: **ENDOCOM – Encontro Nacional de Informação em Ciências da Comunicação**, 13., 2003. *Anais...*. Disponível em: <http://www.intercom.org.br/papers/2003/endocom/ENDOCOM_CAFE.pdf> . Acesso em: 10 jul. 2004.

CAMPELLO, B. S.; CÉNDON, B. V.; KREMER, J. M. (Org.). **Fontes de informação para pesquisadores e profissionais**. Belo Horizonte: Ed. UFMG, 2000. 319p. (Aprender).

CAPURRO, R. **Epistemologia e Ciência da Informação**. 2003. Disponível em: <www.capurro.de/enancib_p.htm>. Acesso em: 27 maio 2005.

CARVALHO, Joaquim de; CORDEIRO, Maria Inês. XML and bibliographic data: the TVS (Transport, Validation and Services) model. In: **IFLA Council and General Conference**, 68th. 2002.

CARVALHO, J.O.F. de. O papel da interação humano-computador na inclusão digital. **Transinformação**, Campinas, 15 (Edição especial) set./dez. p. 75-89, 2003.

CASA, M. Ambientes computacionais com múltiplas formas de representação. In: GONZALES, M. E. Q. et al. (Org.). **Encontro com as ciências cognitivas**. 2. ed. rev. e ampl. Marília: FFC, 1997. p 209-220.

CASTRO, E. **XML para a World Wide Web**. Rio de Janeiro: Campus, 2001. 269 p. (Visual quickstart guide). Tradução de Hugo de Souza Melo.

CHATAIGNIER, M. C. P.; SILVA, M. P. Biblioteca digital: a experiência do Impa. **Ciência da Informação**, v. 30, n. 3, p. 7-12, 2001.

CHAVES, Marcirio Silveira. **Padrões em biblioteca digitais**. Porto Alegre, Pontificia Universidade Católica do Rio Grande do Sul – Faculdade de Informática, Programa de Pós-Graduação em Ciência da Computação – Curso de Mestrado. Ago., 2002, 54f. Trabalho individual I.

CORREA DA CRUZ, Angelo Antonio Alves et al. Impacto dos periódicos eletrônicos em bibliotecas universitárias. **Ciência da Informação**, Brasília, v. 32, n. 2, p. 47-53, maio/ago. 2003.

CÔRTE, A. R. et al. Automação de bibliotecas e centros de documentação: o processo de avaliação e seleção de softwares. **Ciência da Informação**. Brasília, v.28, n.3, p. 241-256, set./dez. 1999.

CROW, R. **The case for institutional repositories**: A SPARC position paper. SPARC: Washington. 2002. Disponível em: <http://www.arl.org/sparc/IR/IR_Final_Release_102.pdf> Acesso em: 12 jul. 2004.

DECLARACIÓN DE PRINCIPIOS INTERNACIONALES DE CATALOGACIÓN.

Tradução de Elena Escolano Rodríguez. 2003. Disponível em:< http://www.loc.gov/loc/ifla/imeicc/source/declaracion_2003-sp.pdf>. Acesso em: 30 de set. de 2005.

DRETSKE, F. I. **Knowledge and flow of the information**. Oxford: basil Blackwell, 1981.

DZIEKANIAK, G. V. Participação do bibliotecário na criação e planejamento de projetos de softwares: o envolvimento com a tecnologia da informação. **Revista Digital de Biblioteconomia e Ciência da Informação**. Campinas, v. 2, n.1, p. 37-56, jul./dez. 2004. Disponível em: <<http://server01.bc.unicamp.br/revbib/sumario.php?vol=2&num=1&mes=jul./dez.&edit=3&ano=2004>>. Acesso em: 23 set. 2004.

EÍTO BRUN, Ricardo. **Tema 5;XML en la descripción de recursos**. 2002. 42 slides. Disponível em:

<<http://www.forpas.us.es/aula/xml/doc/09.XML%20en%20la%20descripci%C3%B3n%20de%20recursos%20MARC.ppt>>. Acesso em: 20 nov. 2002.

EL-SHERBINI, M. Metadata and the future of cataloging. **Library Computing**, [S.l.], v.19, n.3/4, p.180-191, 2000.

FERNEDA, Edberto. **Recuperação de Informação**: análise sobre a contribuição da Ciência da Computação para a Ciência da Informação. Tese (Doutorado) – Escola de Comunicação e Artes. USP, São Paulo, 2003.

FERREIRA, A. B. H. **Novo dicionário Aurélio eletrônico: século XXI**. 3. ed. rev. ampl. Rio de Janeiro: Nova Fronteira, 1999.

FERREIRA, S. M. S. P. Introdução às redes eletrônicas de comunicação. **Ciência da Informação**, Brasília, v.23, n.2, p. 258-263, maio./ago. 1994.

FLAMINO, Adriana Nascimento. **MARC21 e XML como ferramentas para a consolidação da catalogação cooperativa automatizada** : uma revisão de literatura. 2003. 142f. TCC (trabalho de conclusão de curso) - Faculdade de Filosofia e Ciências, Universidade Estadual Paulista, Marília, 2003.

FONSECA, Ramón M. S. **Acesso Público à Informação Científica, Modelo dos Arquivos Abertos e Código Livre**. 2005. 38 slides. Disponível em: <http://www.ibict.br/openaccess/arquivos/openaccess_ramon.ppt >. Acesso em: 17 fev. 2006.

FURGERI, S. **Utilização de catálogos XML para o desenvolvimento do comércio eletrônico**. 1999. 92 f. Dissertação (Mestrado em Sistemas de Informação) - Pontifícia Universidade Católica de Campinas, Instituto de Informática, Campinas, 1999.

GARCIA, Rodrigo Moreira et al. Reflexões filosóficas sobre a informação no contexto cibernético e biológico. In: PAGNI, P. (Org.). **Universidade e contemporaneidade: produção do conhecimento e formação profissional** – coletânea de textos do VI Simpósio em Filosofia e Ciência. Marília: FFC/Marília/UNESP – Comissão Permanente de publicações, 2005. CD-ROM. ISBN 858673828-X.

GARCÍA MARTÍNEZ, Ana M^a. Definición y estilo de los objetos de información digitales y metadatos para la descripción. **Boletín de la Asociación Andaluza de Bibliotecarios**. [S. l.], n. 63, p. 23-47, Junio. 2001.

GONZALEZ, Maria Eunice Quilici. Information, co-evolutionary robotics and action: a systemic approach. In: VI Encontro brasileiro internacional de ciência cognitiva: auto-organização e informação: uma abordagem sistêmica do conhecimento, 2004, Maranhão, Marília. **Caderno de Resumos...** Marília: UNESP, 2004, p.7.

GRÁCIO, J. C. A. **Metadados para descrição de recursos da Internet**: o padrão Dublin Core, aplicações e a questão da interoperabilidade. 2002. 127 f. Dissertação (Mestrado em Ciência da Informação) – Faculdade de Filosofia e Ciências, Universidade Estadual Paulista, Marília, 2002.

GUENTHER, Rebecca; McCALLUM, Sally. New metadata standards for digital resources: MODS and METS. **Bulletin of the American Society for Information Science and Technology**. v. 29, n. 2, p. 12-15, Dec./Jan., 2003.

GUINCHAT, C.; MENOU, M. **Introdução geral às ciências e técnicas da informação e documentação**. 2. ed. corr. aum. Por Marie France Blanquet. Brasília: IBICT, 1994. 540 p. Tradução de: Miriam Vieira da Cunha.

HJØRLAND, B. Fundaments of knowledge organization. **Know. Org.**, v. 30, n. 2, p. 87-111, 2003.

HOUGH, John; BULL, Robert; YOUNG, Brenda. **Using XSLT For XML MARC Record Conversion**. 2000. 49 p. Disponível em: <http://www.crxnet.com/one2/xslt_marc_report.pdf>. Acesso em: 25 mar. 2003.

HUNTER, P.; GUY, M. Metadata for Harvesting: the open archives initiative, and how to find things on the Web. **Electronic Library**. v. 22, n. 2, p. 167-174, 2004.

IBICT. **Manifesto Brasileiro de apoio ao Acesso Livre à Informação Científica**. 2006a. Disponível em: http://www.ibict.br/openaccess/index.php?option=com_content&task=blogcategory&id=27&Itemid=63 >. Acesso em: 1º fev. 2006.

IBICT. **Declarações sobre o acesso livre (open access)**. 2006b. Disponível em: http://www.ibict.br/openaccess/index.php?option=com_content&task=blogcategory&id=27&Itemid=63 >. Acesso em: 1º fev. 2006.

ISO - INTERNATIONAL ORGANIZATION STANDARDIZATION. 2002. Disponível em: <<http://www.iso.ch/iso/en/aboutiso/introduction/whatisISO.html>>. Acesso em: 15 abr. 2003.

JOHNSON, B. C. XML and MARC: which is “right”? **Cataloging & Classification Quartely**, New York, v. 32, n.1, p. 81-89, 2001.

JOINT STEERING COMMITTEE FOR REVISION OF AACR. **RDA: Descripción y Acceso de Recursos**. 2005. Disponível em: <www.collectionscanada.ca/jsc/docs/rdapptjuly2005_spa.pdf>. Acesso em: 27 set. 2005.

KEITH, Corey. **Library of Congress Publishes MARC 21 XML Schema and Transformation Tools**. 2002. Disponível em: <<http://xml.coverpages.org/ni2002-06-05-a.html>>. Acesso em: 25 mar. 2004.

KEITH, Corey. Using XSLT to manipulate MARC Metadata. **Library Hi Tech**, v. 22, n. 2, p. 122-130, 2004.

KOBASHI N. Y.; TÁLAMO, M. F. G. M. Informação; fenômeno e objeto de estudo da sociedade contemporânea. **Transinformação**, Campinas, n. 15 p. 7-21, 2003.

KURAMOTO, Hélio. **Tecnologias para a Comunicação Científica: Open Archives**. 2005. 15 slides. Disponível em: <http://www.ibict.br/openaccess/arquivos/Palestra_I_CVALC.ppt>. Acesso em: 17 fev. 2006.

LAGOZE, Carl; SOMPEL, Herbert van de. The Making of the Open Archives Initiative Protocol for Metadata Harvesting. **Open Archives Initiative**. 2002. Disponível em: <<http://www.cs.cornell.edu/lagoze/papers/The%20Making%20of%20the%20Open%20Archives%20Initiative.pdf>>. Acesso em: 03 mar. 2006.

LAM, K. XML and global name access control. **OCLC Systems & Services**, [S. l.], v.18, n.2, p. 88-96, 2002.

LIBERTY, J.; KRALEY, M. **Aprendendo a desenvolver documentos XML para a Web**. Tradução de: Flávia Cruz. São Paulo: Makron Books, 2001. 274 p.

LÓPEZ GUILLAMÓN, Ignacio. Apuntes para una historia de la catalogación internacional en los siglos XIX y XX. **Scire: Representación y Organización del Conocimiento**, Zaragoza, v.10, n.1, p.121-144, en./jun. 2004.

LWOFF, André. **O conceito de informação na Biologia Molecular**. IN: CUNHA, Fausto; FELIX, Moacyr (dir.). **O conceito de informação na Ciência contemporânea**. Trad. Maria Helena Kühner. v.2. Rio de Janeiro: Paz e Terra, 1970.

MAI, Jens-Erik. The future of general Classification. **Cataloging & Classification Quarterly**, New York, v.37, n.1/2, p.3-12, 2003.

MALHEIRO DA SILVA, A.; RIBEIRO, F. **Das Ciências documentais à ciência da informação**: ensaio epistemológico para um novo modelo curricular. Porto: Afrontamento, 2002.

MARCONDES, C. H. Representação e economia da informação. **Ciência da Informação**, Brasília, v. 30, n. 1, p. 61-70, 2001.

MARTÍNEZ GONZÁLEZ, M. M. Extended Markup Language (XML) : Una solución para modelar documentos y sus interrelaciones basada en la semántica de la información. **Scire: Representación y Organización del Conocimiento**, Zaragoza, v.6, n.2, p.121-151, jul./dic. 2000.

MEADOWS, A. J. **A comunicação científica**. Tradução: Antonio Agenor Briquet de Lemos. Brasília: Briquet de Lemos/ Livros, 1999.

McCALLUM, Sally H. Metadatos, Protocolo y Actividades de Identificación: Alianza Biblioteca del Congreso IFLA/CDNL para el Reporte de Estándares Bibliográficos. 2004. In: **World Library and Information Congress: 70th IFLA General Conference and Council**. 2004. Disponível em: <<http://www.ifla.org/IV/ifla70/prog04.htm>>. Acesso em: 20 abr. 2005.

McGARRY, K. **O contexto dinâmico da informação**: uma análise introdutória. Tradução de Helena Vilar de Lemos. Brasília: Brinquet de Lemos/Livros, 1999. 206 p.

METS - Metadata Encoding & Transmission Standard, 2003. Disponível em: <<http://www.loc.gov/standards/mets/METSOverview.v2.html>>. Acesso em: 24 maio 2005.

MILIDONI, Carmen Beatriz; GONZALEZ, Maria Eunice Quilici; BROENS, Mariana C. Informação e percepção no modelo de mente do “projeto” Freudiano. In: SOUZA, Gustavo M.; D’OTTAVIANO, Itala M. Loffredo; GONZALEZ, Maria Eunice Q. (Orgs.). **Auto-organização**: estudos interdisciplinares. Campinas: UNICAMP, Centro de Lógica, Epistemologia e História da Ciência, 2004, p. 279-299. (Coleção CLE; v. 38).

- MILSTEAD, J. ; FELDMAN, S. Metadata: Cataloging by Any Other Name. **Online**, Jan. 1999. Disponível em: <<http://www.onlineinc.com/online/OL1999/milstead1.html>>. Acesso em: 10 set. 2002.
- MORENO, Fernanda Passini; ARELLANO, Miguel Ángel Márdero. Requisitos Funcionais para Registros Bibliográficos – FRBR: uma apresentação. **Revista Digital de Biblioteconomia e Ciência da Informação**, Campinas, v.3, n. 1, p. 20-38, jul./dez. 5005. ISSN: 1678-765X.
- MUELLER, S.P.M. O crescimento da ciência, o comportamento científico e a comunicação científica: algumas reflexões. **Revista Escola de Biblioteconomia UFMG**, Belo Horizonte, v. 24, n.1, p. 63-84, jan./jun. 1995.
- MUELLER, S. P. M. A ciência, o sistema de comunicação científica e a literatura científica. In: CAMPELLO, Bernadete Santos; CENDÓN, Beatriz Valadares; KREMER, Jeannette Marguerite (Org.). **Fontes de Informação para Pesquisadores e Profissionais**. Belo Horizonte, Ed. UFMG, 2000, p. 21-34.
- NEEDLEMAN, M. et al. The NISO circulation interchange protocol (NCIP) – an XML based Standard. **Library Hi Tech**, [S. l.], v.19, n.3, p. 223-230, 2001.
- NEEDLEMAN, Mark. The Open Archives Initiative. **Serials Reviews**. v.28, n. 2, p. 156-158, 2002.
- NEEDLEMAN, Mark. **Standards Update**: some interesting xml standards. **Serials Reviews**. v.31, p. 70-71, 2005.
- NELSON, M. L. **A New Model for Web Resource Harvesting**. 2005. 51 slides.
- NÖTH, W. Signo, representação e representação mental. In: GONZALES , M. E. Q. et al. (Org.). **Encontro com as ciências cognitivas**. 2. ed. rev. e ampl. Marília: FFC, 1997. p 63-91.
- OLIVEIRA, Vitória Peres de. Uma informação tácita. **Datagramazero**: Revista de Ciência da Informação, v.6, n.3, jun. 2005.
- ORTIZ-REPISO JIMÉNEZ, V. Qué enseñamos después Del MARC?. **Organización del Conocimiento en Sistemas de Información y Documentación**. Zaragoza, v.3, p. 217-225, 1999. Actas del III Encuentro de ISKO-España Getafe, 19 al 21 de noviembre de 1997
- PINTO MOLINA, M. **El resumem documental**: princípios y métodos. Madrid: Pirâmide, 1992. 584 p.
- QIN, J. Representation and Organization of Information in the Web Space: From MARC to XML. **Informing Science**, [S. l.], v.3, n.2, p. 83-87, 2000.
- RAMALHO, J. A. **XML teoria e prática**. São Paulo: Berkeley Brasil, 2002, 146 p.
- RAY, E. T. **Aprendendo XML**. Rio de Janeiro: Campus, 2001. 372 p. Tradução de: Daniel Vieira.

RIBEIRO, C. M.; SANTOS, G C. **Acrônimos, siglas e termos técnicos**: arquivística, biblioteconomia, documentação, informática. Campinas: Editora Átomo, 2003. 277 p.

ROBREDO, Jaime, **Da ciência da informação revisitada aos sistemas humanos de informação**. Brasília: Thesaurus; SSRR Informações, 2003. 262 p.

RODRIGUES, Eloy. **Acesso Livre ao Conhecimento**: a mudança do sistema de comunicação da ciência e os profissionais de informação. 2004. Disponível em: <<http://www.apbad.pt/CadernosBAD/Caderno2004/Rodrigues.pdf>>. Acesso em: 26 jan. 2005.

RODRIGUES M. E. F.; LIMA M H.T. F.; GARCIA, M. J. O. A normalização no contexto da comunicação científica. **Perspectivas em Ciência da Informação**, Belo Horizonte, v.3, n.1, p. 147-156, jul./dez. 1998.

ROSETTO, M. **Uso do protocolo Z39.50 para recuperação de informação em redes eletrônicas**, 1997. Disponível em:<<http://www.internic.net/z3950/z3950.html>>. Acesso em 10 abr. 2002.

ROSETTO, Márcia. **Metadados e formatos de metadados em sistemas de informação**: caracterização e definição. 2003. 95 p. Dissertação (Mestrado) – Escola de Comunicação e Artes, Universidade de São Paulo.

SANTOS, Plácida L. V. A. Costa; FLAMINO, Adriana Nascimento. MARC21 e XML como ferramentas para a consolidação da catalogação cooperativa automatizada: uma revisão de literatura. In: VIDOTTI, Ap. B. Gregório (coord). **Tecnologia e conteúdos informacionais**: abordagens teóricas e práticas. São Paulo: Polis, 2004. p. 114-138.

SANDERSON, Robert; YOUNG, Jeffrey; LE VAN, Ralph. SRW/U with OAI: expected and unexpected synergies. **D-Lib Magazine**, v. 11, n.2, feb. 2005.

SENA, N. K. Open archives: caminho alternativo para a comunicação científica. **Ciência da Informação**, Brasília, v.29, n.3, p. 71-78, set./dez. 2000.

SCHAMBER, L. What is a document? Rethinking the concept in uneasy times. **Journal of the American Society for information Science**, v. 47, n. 9 p. 669-671, 1996.

SHAPIRO, C.; VARIAN, H. R. **A economia da informação**: como os princípios econômicos se aplicam a era da Internet. 5ª edição, Rio de Janeiro: Campus, 1999. Tradução de: Information rules por Ricardo Inofosa. 397 p.

SHEARER, K. **The Open Archives Initiative Protocol for Metadata Harvesting**: Three Years Running. 2003.

SMIT, J. W. **O que é documentação**. São Paulo: Brasiliense, 1987 (Coleção primeiros passos; 174).

SMIT, J. W.; BARRETO, A. A. Ciência da informação: base conceitual para a formação profissional. IN: VALENTIM, M. L. **Formação do profissional da informação**. São Paulo, Polis, 2002, p. 9-23.

SOMPEL, Herbert van de; LAGOZE, Carl. The Santa Fe convention of the open archives initiative. **D-Lib Magazine**, v. 6, n. 2, Feb. 2000. Disponível em : <<http://www.dlib.org/dlib/february00/vandesompel-oai/vandesompel-oai.html>>. Acesso em: 20 mai. 2004.

SOUZA, Lídia; ALVARENGA, Renato Rocha. A Web semântica e suas contribuições para a Ciência da Informação. **Ciência da Informação**, Brasília, v. 33, n.1, p. 132-141, jan./ abr. 2004.

SOUZA, M. I. F.; VENDRUSCULO, L. G.; MELO, G. C. Metadados para a descrição de recursos de informação eletrônica : utilização do padrão Dublin Core. **Ciência da Informação**, Brasília, v. 29, n.1, p. 93-102, jan./ abr. 2000.

SOUZA, T. B.; CATARINO, M. E.; SANTOS, P. C. Metadados: catalogando dados na Internet. **Transinformação**, v.9, n.2, mai/ago, 1997.

TARGINO, M. G. Comunicação Científica: uma revisão de seus elementos básicos. **Informação & Sociedade: estudos**, v. 10, n.2, 2000. Disponível em: <<http://www.informacaoesociedade.ufpb.br/1020002.pdf>>. Acesso em: 16 jul. 2004.

TILLET, Barbara. **El modelo FRBR (Requerimientos Funcionales para Registros Bibliográficos)**. 2003. Disponível em: <<http://www.loc.gov/catdir/cpso/frbrspan.pdf>>. Acesso em: 29 set. 2005.

TILLET, Barbara. **¿Qué es FRBR?: un modelo conceptual del universo bibliográfico**. 2004. Disponível em: <<http://www.loc.gov/catdir/cpso/Que-es-FRBR.pdf>>. Acesso em: 29 set. 2005.

VAN DIJK, T. A. **La ciencia del texto: un enfoque interdisciplinario**. Tradução Sibila Hunzinger. Barcelona: Paidós, 1992. 309 p. Tradução de: Tekstwetenschap. Een interdisciplinaire inleiding.

W3C WORLD WIDE WEB CONSORTIUM. 1994 – 2002. Disponível em: <<http://www.w3c.org/consortium>>. Acesso em: 20 maio 2004.

WIENER, Norbert. **Cibernética e sociedade: o uso humano de seres humanos**. 2 ed. Trad. José Paula Paes. São Paulo: Cultrix, 1968.

BIBLIOGRAFIA CONSULTADA

ALVES, R. C. V. **Web Semântica**: uma análise focada no uso de metadados. 2005. 180 f. Dissertação (Mestrado em Ciência da Informação) – Faculdade de Filosofia Ciências, Universidade Estadual Paulista, Marília, 2005.

BARRETO, CÁSSIA MARIA. **Modelo de metadados para a descrição de documentos eletrônicos na web**. 1999. 190f. Dissertação (mestrado em Ciências em Sistemas e Computação) - Instituto Militar de Engenharia, Rio de Janeiro, 1999.

BERTO, R. M. V. de S. Novas práticas de comunicação e produção de publicações científicas. In: **ENDOCOM** – Encontro Nacional de Informação em Ciências da Comunicação, 13., 2003. *Anais...* Disponível em: <http://www.intercom.org.br/papers/congresso2003/pdf/2003_ENDOCOM_TRABALHO_berto.pdf> acesso em: 10 jul. 2004.

BOMFÁ, Cláudia Regina Ziliotto; CASTRO, João Ernesto E.. Desenvolvimento de revistas científicas em mídia digital – o caso da Revista Produção Online. **Ciência da Informação**, Brasília, v. 33, n. 2, p. 39-48, maio/ago. 2004.

BONIFÁCIO, Ailton Sergio; HEUSER, Carlos Alberto. **Metadados Semânticos para Buscas em Bibliotecas Digitais**. disponível em: <<http://www.uel.br/pessoal/ailton/Trabalhos/SemanaAcad-Ailton.html>>. Acesso em: 10 mar. 2004.

CAFÉ, L. et. al. **Arquivos abertos** : inovação para a comunicação científica na rede. 2002. In: INTERCOM – Sociedade Brasileira de Estudos Interdisciplinares da Comunicação, 25., 2003. *Anais...* Disponível em: <http://www.portcom.intercom.org.br/papers/2002/endocom/ENDOCOM_CAFE.pdf>. Acesso em: 15 jul. 2004.

CALANAG, Maria Luisa; TABATA, Koichi; SUGIMOTO, Shigeo. Linking preservation metadata and collection management policies. **Collection Building**, v. 23, n. 2, p. 56-63, 2004.

COSTA, S.M.S. O novo papel das tecnologias digitais na comunicação científica. In: MARCONDES et al (Orgs). **Bibliotecas digitais**: saberes e práticas. Salvador: EDUFBA; Brasília: IBICT, 2005. p. 167-183.

FIANDER, D. Applying XML to the Bibliographic Description. **Cataloging & Classification Quartely**, New York, v.33, n.2, p.17-28, 2001.

FURRIE, B. **O MARC bibliográfico**: um guia introdutório; catalogação legível por computador. Brasília : Thesaurus, 2000. 95 p. Tradução de: Beatriz Valadares Cendón, Sonia Burnier, Maria Helena Santos e Natália Guiné de Mello Carvalho.

GALINDO, Marcos; PEREIRA, Marcos S.; LIMA, Cleiton M. V. **Bibliotecas digitais e metadados: uma abordagem integradora**. Disponível em: <<http://libdigi.unicamp.br/document/?view=8283>>. Acesso em: 18 abr. 2005.

GALVÃO, Maria Cristiane Barbosa; BORGES, Paulo César Rodrigues. Ciência da informação: ciência recursiva no contexto da sociedade da informação. **Ciência da Informação**, Brasília, v. 29, n. 3, p. 40-49, set./dez. 2000.

HAMMOND, Tony; HANNAY, Timo; LUND, Ben. The Role of RSS in Science Publishing. **D-Lib Magazine**. v. 10, n. 12, 2004. Disponível em: <<http://www.dlib.org/dlib/december04/hammond/12hammond.html>>. Acesso em: 23 dez. 2004.

HASELAGER, W. F. G. **A teoria dos sistemas dinâmicos**. Disponível em: <<http://www.nici.ru.nl/~haselag/port/talks/01dst.html>>. Acesso em: 1º jun. 2005.

HERWIJNEN, Eric van. **The impact of XML on library procedures and services**. 2000. Disponível em: <<http://lhcb.web.cern.ch/lhcb/~evh/xmlandlibrary.htm>>. Acesso em: 07 jul. 2004.

MARCONDES, Carlos Henrique; SAYÃO, Luis Fernando. Documentos digitais e novas formas de cooperação entre sistemas de informação em C&T. **Ciência da Informação**, Brasília, v. 31, n. 3, p. 42-54, set./dez. 2002.

MARTINS, Eduardo Vieira. O contexto político e o discurso da ciência da informação no Brasil: uma análise a partir do IbiCT. **Ciência da Informação**, Brasília, v. 33, n. 1, p. 91-100, jan./abr. 2004.

MARTINS, Ives Gandra da Silva; MARTINS, Rogério Vidal Gandra da Silva. Privacidade na comunicação eletrônica. **Ciência da Informação**, Brasília, v. 30, n. 1, p. 13-18, jan./abr. 2001.

NISO - NATIONAL INFORMATION STANDARDS ORGANIZATION. ANSI/NISO Z39.2-1994: information interchange format. United States of America, 1994. 6 p. Disponível em: <<http://www.niso.org/standards/resources/Z39.2.PDF>>. Acesso em: 15 abr. 2003.

RODRIGUES, Eloy et al. **RepositóriUM**: implementação do DSpace em português: Lições para o futuro e linhas de investigação. 2004. Disponível em: <<https://repositorium.sdum.uminho.pt/bitstream/1822/679/1/RepositoriUM+-+implementa%C3%A7%C3%A3o+do+Dspace+em+portugu%C3%AAs.+Li%C3%A7%C3%B5es+para+o+futuro+e+linhas+de+investiga%C3%A7%C3%A3o.pdf>>. Acesso em: 05 dez. 2004.

SANTOS, Plácida L. V. A. Costa. **Metadados e bibliotecas digitais**. 2005. 27 slides.

TARGINO, M. G. **Novas tecnologias e produção científica**: uma relação de causa e efeito ou uma relação de muitos efeitos? 2003. In: INTERCOM – Sociedade Brasileira de Estudos Interdisciplinares da Comunicação, 26., 2003. **Anais...** Disponível em: <http://intercom.locaweb.com.br/papers/congresso2003/pdf/2003_NP09_targino.pdf>. Acesso em: 16 jul. 2004.

TRIGUEIRO, Michelangelo Giotto Santoro. A comunidade científica, o Estado e as universidades, no atual estágio de desenvolvimento científico tecnológico. **Sociologias**, Porto Alegre, ano 3, nº 6, p. 30-50, jul/dez 2001.

LISTA DE LINKS

<http://bibliotecas-cruesp.usp.br/3sibd/docs/flamino194.pdf>

<http://portal.cid.unb.br/CIPECCbr/papers.php>

<http://www.ibict.br>

<http://www.arxiv.org/>

<http://www.loc.gov/z3950/agency/zing/srw>

<http://www.loc.gov/standards/sru/cql/>

www.w3c.org

www.w3.org/XML/Schema

<http://www.loc.gov/MARC21/slim>

<http://www.loc.gov/standards/marcxml/schema/MARC21slim.xsd>

<http://www.loc.gov/standards/marcxml/xml/collection.xml>

www.w3.org/TR/xslt

www.loc.gov/standards/mods/

<http://www.loc.gov/marc/sourcecode/genre/>

<http://www.loc.gov/standards/mads/>

www.ccsds.org/RP9905/RP9905.html

<http://www.loc.gov/marc/marcdtd/marcdtdback.html>

<http://www.perl.com>

<http://www.loc.gov/standards/mets/>

<http://marc4j.tigris.org/>