

anais do



3<sup>o</sup> encontro de bases de dados  
sobre informações  
arquivísticas

16 e 17  
de março  
de 2009

O acesso aos  
arquivos e  
as tecnologias  
de informação

**ANAIS DO**  
**III ENCONTRO DE BASES DE DADOS**  
**SOBRE INFORMAÇÕES ARQUIVÍSTICAS**  
O Acesso aos Arquivos e as Tecnologias de Informação

**ANAIS DO  
III ENCONTRO DE BASES DE DADOS  
SOBRE INFORMAÇÕES ARQUIVÍSTICAS**  
O Acesso aos Arquivos e as Tecnologias de Informação

16 e 17 de março  
de 2009

Rio de Janeiro, RJ  
Brasil

## **DIRETORIA (BIÊNIO 2007-2009)**

Lucia Maria Velloso de Oliveira	Presidente
Eliana Balbina Flora Sales	Vice-presidência
Renata Silva Borges	1° Tesoureiro
Alcides Sodré Peçanha Filho	2° Tesoureiro
Isabel Cristina Borges de Oliveira	1° Secretário
Laura Regina Xavier	2° Secretário

## **CONSELHO DELIBERATIVO E DE ÉTICA**

### **Membros efetivos**

Paulino Lemes de Sousa Cardoso  
Maria Rosângela da Cunha (licenciada desde julho de 2008)  
Leila Estephanio de Moura  
Mariza Bottino  
Rosângela Florido Rangel

### **Suplentes**

Ivonete Pereira Tavares  
Margareth da Silva  
Suzanna Alves Martins

## **CONSELHO FISCAL**

### **Membros efetivos**

Maria Celina Soares de Mello e Silva  
Sérgio Conde de Albite Silva  
Jacilene Alves Brejo (membro efetivo desde de 31 de setembro de 2008)

# III ENCONTRO DE BASES DE DADOS SOBRE INFORMAÇÕES ARQUIVÍSTICAS

## COMISSÃO ORGANIZADORA

Lucia Maria Velloso de Oliveira                      Presidente  
Leila Estephanio de Moura  
Isabel Cristina Borges de Oliveira  
Bianca Therezinha Carvalho Panisset  
Fabiana da Silva Fonseca                              (Assistente)

## COMISSÃO CIENTÍFICA

Ana Maria de Almeida Camargo,  
José Maria Jardim  
Maria Celina Soares de Mello e Silva  
Paulo Roberto Elian dos Santos,  
Renato Tarciso Barbosa de Sousa  
Sérgio Conde de Albite Silva

## CONFERENCISTAS E PRESIDENTES DE MESA

Ana Maria de Almeida Camargo  
Jackie Dooley  
José Maria Jardim  
Peter Van Garderen  
Wendy Duff

## PALESTRANTES

Alan Vasconcelos  
Claudiane emanuele Nazario  
Elisangela Cristina Agentte  
Fábio Rodrigo Pinheiro da Silva  
Flávia Carneiro Leão  
Galba Di Mambro  
Isabel Cristina Borges de Oliveira  
José Guilherme Ribeiro  
Leandro Ribeiro Negreiros  
Lucia Maria Velloso de Oliveira  
Maria Aparecida Forti  
Neire do Rossio Martins  
Telma Maria Murari

## SUMÁRIO

APRESENTAÇÃO	5
QUEM É QUEM	6
THE ICA-ATOM PROJECT AND TECHNOLOGY Peter Van Garderen	10
THE EVOLUTION OF ARCHIVAL DESCRIPTIVE STANDARDS IN THE UNITED STATES Jackie Dooley	32
SISTEMA DE ARQUIVOS HISTÓRICOS DA UNICAMP (PESQUISARQH): MÓDULO DE REGISTRO, ARMAZENAMENTO E ACESSO Fábio Rodrigo Pinheiro da Silva; Telma Maria Murari; Maria Aparecida Forti; Neire do Rossio Martins	42
DESCRIÇÃO ARQUIVÍSTICA: CONTEXTO ARQUIVÍSTICO, CONTROLE DE VOCABULÁRIO E O USUÁRIO Lucia Maria Velloso de Oliveira; Isabel Cristina Borges de Oliveira	46
A NORMALIZAÇÃO DA DESCRIÇÃO ARQUIVÍSTICA: A ISAD(G) E OS ARQUIVOS PERMANENTES Flávia Carneiro Leão	58
INTERACTING WITH ARCHIVAL USERS: A STATUS REPORT Dr. Wendy M. Duff	71
“UM HOMEM BATEU EM MINHA PORTA E EU ABRI (?)”: UMA AVALIAÇÃO DOS SERVIÇOS DE REFERÊNCIA E DA ACESSIBILIDADE DOS SÍTES DE INSTITUIÇÕES PÚBLICAS ARQUIVÍSTICAS Leandro Ribeiro Negreiros e Alan Vasconcelos	80
COMO INTEGRAR SÍTIOS DE INSTITUIÇÕES ARQUIVÍSTICAS NUM CONTEXTO DE BAIXA PADRONIZAÇÃO DE SUAS DESCRIÇÕES? Galba Di Mambro; José Guilherme Ribeiro	92

O III Encontro de Bases de Dados sobre informações arquivísticas, com a temática central O Acesso aos Arquivos e as Tecnologias de Informação, foi realizado nos dias 16 e 17 de março de 2009. O evento, que já se firmou na agenda de eventos científicos da área arquivística, é mais uma das promoções da Associação dos Arquivistas Brasileiros – AAB, com o intuito de propiciar aos profissionais da Arquivologia espaços para a discussão científica e para a difusão da produção de conhecimento da área.

O evento, realizado na Fundação Casa de Rui Barbosa – FCRB, contou também com o apoio do CNPQ, da CAPES, do Arquivo Nacional – Casa Civil, bem como das empresas Molducenter, Femade e Spool Systems.

Durante o III Encontro, professores e pesquisadores, como Ana Maria de Almeida Camargo (Universidade de São Paulo – São Paulo), Wendy Duff (Universidade de Toronto – Canadá), José Maria Jardim (Universidade Federal Fluminense – Rio de Janeiro) e os profissionais Peter Van Garderen (Artefactual Systems – Canadá) e Jackie Dooley (RLG e OCLC Research – Estados Unidos), abordaram questões de máxima relevância para a área, tais como o uso de normas e padrões na Arquivologia voltados para a descrição arquivística, o uso da informação da arquivística e os usuários em face das Tecnologias de Informação e Comunicação, bem como o uso de software livre para a disponibilização de informações sobre acervos arquivísticos na Web.

Além disso, foram apresentados trabalhos avaliados pela Comissão Científica do evento. Os trabalhos selecionados representam projetos concluídos ou em desenvolvimento por instituições sediadas nos estados do Rio de Janeiro, São Paulo e Minas Gerais.

O evento reuniu 120 participantes, dentre os quais 8 professores de cursos de graduação e pós-graduação, 12 doutores e pesquisadores e 20 estudantes de graduação.

Apresentamos a seguir as conferências na íntegra. O conteúdo das conferências é de responsabilidade dos autores.

**Lucia Maria Velloso de Oliveira**

*Presidente da Associação dos Arquivistas Brasileiros*

**Alan Vasconcelos,**

mestrando em Ciência da Informação pela Escola de Ciência da Informação da Universidade Federal de Minas Gerais. Colaborador da Synergia – Engenharia de Softwares e Sistemas.

**Ana Maria de Almeida Camargo,**

graduada em História pela Universidade de São Paulo (1966) e doutorado em História Social pela Universidade de São Paulo (1974). Atualmente é Professora doutora da Universidade de São Paulo. Tem experiência na área de História, com ênfase em Teoria e Filosofia da História. Atuando principalmente nos seguintes temas: Arquivos Pessoais, Catálogo, Documentação, Hemeroteca, Instituto Histórico e Geográfico de São Paulo, Imprensa Periódica e Imprensa Brasileira.

**Bianca Therezinha Carvalho Panisset,**

graduada em Arquivologia pela Universidade Federal do Estado do Rio de Janeiro (UNIRIO), atualmente é bolsista FAPERJ do Programa de Inovação Tecnológica atuando na organização e padronização dos Arquivos Correntes do Museu de Astronomia e Ciências Afins (MAST).

**Claudiane Emanuele Nazario,**

consultora Sênior em Documentação e Informação, Profissional da Biblioteconomia, com grande experiência em análise de processos, diagnóstico situacional e implantação de projetos de Gestão Informacional em empresas de grande porte, de segmentos variados.

**Fábio Rodrigo Pinheiro da Silva,**

graduado em Ciência da Informação - habilit. Biblioteconomia pela Pontifícia Universidade Católica de Campinas (2004). Atualmente é profissional em organização de arquivos no Arquivo Central do Sistema de Arquivos da Universidade Estadual de Campinas. Tem experiência na área de Arquivologia e Documentação, com ênfase em Gestão de Documentos e Arquivos, atuando principalmente nas áreas de gestão e preservação de documentos.

**Flávia Carneiro Leão,**

graduada em Ciências Sociais pela Unicamp, possui o curso de “Especialização em Organização de Arquivos pela Unicamp/USP. É mestre em Ciência da Informação pela Escola de Comunicação e Artes da USP (Depto. de Biblioteconomia e Documentação) onde desenvolveu estudo sobre a padronização da descrição arquivística. É supervisora do Centro de Documentação “Alexandre Eulalio”, CEDAE/IEL-Unicamp, desde 1995, onde realiza trabalho voltado, sobretudo à preservação de fundos pessoais, institucionais e de coleções relacionadas à literatura e à lingüística brasileiras.

**Galba Di Mambro,**

graduado em História pela Universidade Federal de Minas Gerais, é professor do Departamento de História da Universidade Federal de Juiz de Fora onde leciona as disciplinas Metodologia do Trabalho

Científico e Patrimônio Histórico III (Patrimônio Documental) e Diretor do Arquivo Histórico da mesma universidade. Tem atuado como coordenador ou consultor em projetos de organização / reformatação de arquivos, tendo a Norma Brasileira de Descrição Arquivística como referencial descritivo.

**Isabel Cristina Borges de Oliveira,**

graduada em Arquivologia e História. Experiência na área de descrição e controle terminológico. Bolsista de Desenvolvimento Tecnológico no Arquivo Histórico e Institucional da Fundação Casa de Rui Barbosa. Mestranda do Mestrado Profissionalizante em Bens Culturais e Projetos Sociais, do Centro de Pesquisas e Documentação em História Contemporânea-CPDOC, da Fundação Getúlio Vargas.

**Jackie Dooley,**

arquivista-consultora dos programas Research Libraries Group (RLG) e OCLC Research, antes foi Diretora dos Arquivos e coleções Especiais da Universidade da Califórnia, Irvine. Possui experiência com coleções especiais de livros raros, arquivos e material visual. Como especialista em descrição arquivística e terminologia, foi membro do grupo de pesquisa designado Encoded Archival Description (EAD). Foi uma das idealizadoras do Thesaurus for Graphic Materials e co-autora do Descriptive Cataloging of Rare Books. É membro participativo da Sociedade dos Arquivistas Americanos e da Seção de Livros raros e manuscritos da Association de Bibliotecas Acadêmicas ou de Pesquisa. É membro do Grupo de Trabalho de Coleções Especiais da Associação de Bibliotecas de Pesquisa.

**José Guilherme Ribeiro,**

graduado em História pela Universidade Federal de Juiz de Fora- UFJF, e cursa especialização em Análise de Sistemas pela Universidade Federal de Minas Gerais. É consultor do Arquivo Histórico da UFJF, desenvolvendo atividades relacionadas à disponibilização de descrições arquivísticas e imagens digitais em ambiente de rede.

**José Maria Jardim,**

graduado em História. Doutor em Ciência da Informação. Professor do Dep. de Ciência da Informação da Universidade Federal Fluminense e do Programa de Pós-Graduação em Ciência da Informação UFF-IBICT. No momento, desenvolve pesquisa intitulada “Políticas públicas de informação: a (não) construção da política nacional de arquivos públicos e privados”. É membro do Laboratório de Investigação Arquivística do Departamento de Biblioteconomia e Documentação da Universidade Carlos III de Madrid.

**Leandro Ribeiro Negreiros,**

mestre em Ciência da Informação pela Escola de Ciência da Informação da Universidade Federal de Minas Gerais. Colaborador Synergia – Engenharia de Softwares e Sistemas.

**Leila Estephania de Moura,**

graduada em História com especialização em Arquivos pela Universidade de São Paulo. Trabalha no Arquivo Histórico e Institucional da Fundação Casa de Rui Barbosa desde de 1995. Mestre do Mestrado Profissionalizante em Bens Culturais e Projetos Sociais, do Centro de Pesquisas e Documentação em História Contemporânea-CPDOC, da Fundação Getúlio Vargas

**Lucia Maria Velloso de Oliveira,**

graduada em História e Arquivologia. MBA em Administração e Gestão de Negócios pela Fundação Getúlio Vargas. Atua na área desde de 1987. Chefe do Serviço de Arquivo Histórico e Institucional da Fundação Casa de Rui Barbosa desde 2002. Mestre em Ciência da Informação do Programa de Pós-Graduação em Ciência da Informação UFF-IBICT. Doutoranda em História Social pela Universidade de São Paulo – USP.

**Maria Celina Soares de Mello e Silva,**

arquivista do Museu de Astronomia e Ciências Afim/MCT, mestre em Memória Social e Documentação pela Universidade Federal do Estado do Rio de Janeiro – UNIRIO. Doutora em História Social pela Universidade de São Paulo – USP.

**Neire do Rossio Martins,**

graduada em Biblioteconomia pela Pontifícia Universidade Católica de Campinas (1982), com especialização em Organização de Arquivos pela Universidade de São Paulo (1982) e especialização em Planejamento e Administração Sistemas Informação pela Pontifícia Universidade Católica de Campinas (1989) . Atualmente é Bibliotecária da Universidade Estadual de Campinas.

**Paulo Roberto Elian dos Santos,**

graduado em História pela Pontifícia Universidade Católica do Rio de Janeiro (PUC/RJ). Mestre e doutor em História Social pela Universidade de São Paulo (USP). É tecnologista senior da Casa de Oswaldo Cruz / Fundação Oswaldo Cruz (Fiocruz), onde coordena o Sistema de Gestão de Documentos e Arquivos. É membro do Conselho Nacional de Arquivos (CONARQ). Integra o Conselho Editorial da Revista Arquivo & Administração da Associação dos Arquivistas Brasileiros (AAB). Tem experiência na área de Arquivologia, com ênfase em Organização de Arquivos, atuando principalmente nos seguintes temas: classificação e descrição em arquivos permanentes; arquivos pessoais; e gestão de arquivos de instituições de ciência & tecnologia e saúde.

**Peter Van Garderen,**

arquiteto do sistema e gerente do projeto ICA-AToM. Mr. Van Garderen é graduado pela University of British Columbia's Master of Archival Studies and Software Engineering programs. Atualmente faz doutorado na University of Amsterdam onde está pesquisando o uso da tecnologia Web 2.0 para sistemas de acesso voltados para arquivos. Mr. Van Garderen iniciou sua carreira como arquivista em 1997 como gerente de produção de softwares voltados para arquivos e como coordenador do projeto InterPARES - research on the long-term preservation of electronic records. Sua empresa, Artefactual Systems, é a principal contratada pelo ICA para desenvolver e dar suporte ao software ICA-AToM. A Artefactual Systems trabalha com projetos incluindo o piloto do ICA-AToM, e com a concepção de um software de fonte aberta para preservação digital, em colaboração com o Arquivo de Vancouver.

**Renato Tarciso Barbosa de Sousa,**

graduado em História pela Universidade de Brasília (1990), mestre em Biblioteconomia e Documentação pela Universidade de Brasília (1995) e doutor em História Social pela Universidade de São Paulo (2005). Atualmente é professor adjunto da Universidade de Brasília. Tem experiência na área de Ciência da Informação, com ênfase em Organização de Arquivos, atuando principalmente nos seguintes temas: arquivologia, organização de arquivos, gestão de documentos, formação profissional e políticas públicas de arquivo.

**Sérgio Conde de Albite Silva,**

arquivista-conservador. Doutor em Ciência da Informação pela Universidade Federal Fluminense em convênio com o Instituto Brasileiro de Informação em Ciência e Tecnologia UFF-IBICT, mestre em Memória Social e Documento. É professor adjunto da Universidade Federal do Estado do Rio de Janeiro - UNIRIO. Publica na área de Arquivologia, Ciência da Informação e Biblioteconomia, com ênfase na preservação da informação arquivística, políticas públicas de preservação e conservação preventiva em arquivos e bibliotecas.

**Wendy Duff**

professora da Faculdade de Estudos da Informação da Universidade de Toronto (Canadá), obteve seu bacharelado pela Universidade Kings College, MLS na Universidade de Dalhousie e seu PHD pela Universidade de Pittsburgh. Foi co-coordenadora do Projeto sobre guarda e preservação de documentos eletrônicos da Universidade de Pittsburgh. Fundadora do AX-NET, grupo internacional de pesquisadores interessados em facilitar o acesso ao material arquivístico. Atuou na Comissão do Conselho Internacional de Arquivos – ICA para elaboração de padrões de Descrição, no Grupo de trabalho sobre arquivamento e preservação digital e dirigiu o Comitê Canadense de Descrição Arquivística. Seus interesses de pesquisa são: estudo de usuário, descrição arquivística e documentos eletrônicos. Autora de inúmeros artigos publicados em Periódicos da área. Atualmente pesquisa o desenvolvimento de ferramentas de avaliação do usuário, comportamento do usuário, serviço de referência arquivístico, educação arquivística e preservação digital. Ano passado, recebeu o Prêmio de Excelência em Pesquisa.

# The ICA-AtoM Project and Technology

Peter Van Garderen

## Introduction

*Bom dia.* I want to thank the Association of Brazilian Archivists for this opportunity to talk about the ICA-AtoM project. I am humbled by the opportunity to address this conference twice this week. I gave another presentation at your annual conference in June so I would like to think that you just love listening to me talk. However, I realize that this invitation is more likely an indication of the growing interest in open-source software within the Brazilian archival community. Before I begin, I would like to pass on greetings and best wishes from the ICA Secretary-General David Leitch. The ICA is very excited that our Brazilian colleagues have taken an early lead role in helping to realize the vision and full potential of the ICA-AtoM software collaboration.

Of course, I myself do not represent the ICA directly. My Canadian company, Artefactual Systems, is the primary contractor that is leading the technical development of the software. My formal role within the project is Software Release Manager and, as such, I work closely with the ICA Secretariat and the ICA-AtoM Steering Committee. However, I should note that I am an archivist myself and that the roots of this project go back to my own personal motivation to provide a free and open-source software application for use by my fellow archivists.

After I graduated from the Master of Archival Studies program at the University of British Columbia in 1997, I worked briefly for a commercial software vendor. There I learned pretty quickly that most archival institutions have very limited resources.

I took a call from a sweet elderly lady one morning who was a volunteer at a small community archives. I had spoken to her previously about using our software to manage her archival description project. She told me excitedly that they had a very successful Bingo night and that the archives now had \$500 dollars to purchase the software. Unfortunately I had to tell her that the software cost several times that amount. "Oh," she said sadly. "I guess we need a few more bingos." Another time I was on-site with an existing client who had some difficulty getting the software to work. After taking a look at their system I realized that their problem would be solved if they simply implemented an additional module. Unfortunately, that module cost several thousands of dollars. "We've already missed the budget for this year," the archivist noted. "Maybe, we can get the money in the next fiscal year." They were forced to continue their manual process of writing and printing out finding aids and eventually migrated their data to another system. These early episodes made quite an impression on me. I had the technical knowledge to help my fellow archivists but that was only if they had the money to pay for access to the required tools.

I soon moved on to another job as the first Project Coordinator for the InterPARES Project back at UBC. Then, in 2001, I launched Artefactual Systems to begin my electronic records and digital preservation consulting practice. Throughout that time I became increasingly interested in the free and open-source software movement, which was characterized by the growing mainstream popularity of the Linux operating system and the Apache web server as well as a number of digital library projects such as Dspace.

## Free and Open Source Software

The free software movement began when Richard Stallman released the GNU operating systems in 1983 as a replacement for the UNIX operating system. GNU stands for “GNU is Not UNIX”. GNU still forms the basis of the current Linux operating system. Stallman was frustrated by the restrictions placed on his ability as a computer scientist to study and share the design of software systems. Therefore, he released GNU under his own GNU Public License (GPL) which gave users of the software four basic freedoms:

1. The freedom to run the program for any purpose
2. The freedom to study how the program works, and adapt it to their own needs, meaning that easy access to the source code must be provided
3. The freedom to redistribute copies to help friends, family, colleagues or society in general
4. The freedom to improve the program, and release their own improvements to the public, so that the whole community benefits. Again, easy access to the source code is a precondition for this.

The GPL is now the most widely implemented free and open-source license in use. The ICA-AtoM software is released under version 2 of the GPL. There are also a number of other open-source licenses available. The Open Source Initiative maintains a full list at <http://osi.org> as well as an open source definition in the form of ten criteria that expand upon Stallman’s four freedoms.

It is important to note that these freedoms or criteria do not restrict the ability to charge money for free and open-source software. That said, very few open-source projects actually charge money for their software. Providing software free of charge is certainly one of the more popular characteristics of most open-source projects, and it is what first attracted me and many others. Therefore, there are no fees at all to download, install or revise the ICA-AtoM software. It is completely free of charge. It is like free beer. If I buy you a beer, you can drink it and that costs you nothing. This makes me feel good about myself and it is pretty exciting for you, assuming you like beer. Using the freedom to redistribute free copies, I am finally able to pass on my technical knowledge in the form of software, even to those who could not otherwise afford my consulting services.

However, as exciting as free beer is, the ability for you to study, improve and re-distribute the software is arguably more exciting and was certainly revolutionary when the idea was first proposed by Stallman. This is freedom as in ‘free as a dove’. If you choose to use ICA-AtoM or any other open-source software you have the ability to make modifications and add modules as you see fit. You don’t have to ask anyone’s permission. If you don’t have the technical ability to make the modifications you need, you can get the help from a colleague or friend who does or hire anyone with the necessary technical skills at the most attractive price. Better yet, you can collaborate with other users of the same software to pool financial and technical resources to improve the software and redistribute those improvements back to all users of the software. This is the approach taken by the ICA-AtoM project.

Of course, free and open-source software comes with its own set of responsibilities. So in addition to being like ‘free beer’ and ‘free as a dove’, open-source software is also like a ‘free kitten’. They are cute, cuddly and exciting but they require feeding, a warm place to sleep and you have to let them out every now and then to go to the bathroom. In the end, you build a two-way relationship with the cat which usually rewards both parties with friendship and trust. Similarly, even if the open-source software is free of charge it is not

without secondary costs or responsibilities. There are always costs associated with installing and maintaining software systems, although I would argue that the total cost of ownership, which factors in all these costs, is often significantly lower for free and open-source software. Commercial vendors whose business models are threatened by the open-source movement are quick to spread fear, uncertainty and doubt about total cost of ownership estimates. I think each organization needs to do its own cost-benefit analysis before making a choice between proprietary or free, open-source software. UNESCO recently prepared an excellent report to assist with such an analysis.<sup>1</sup>

One good example from my home province of British Columbia in Canada is the provincial Public Libraries Association. After encountering questionable upgrade and pricing practices from some major library software vendors, the association decided to support the migration of all the public library information systems to the open-source Evergreen ILS application<sup>2</sup>. They began by hiring one project manager, system administrator and end-user support staff person. The cost of staff salaries and centrally hosting this free software is still significantly less than each library paying for its own license and technical support from the vendors. Opting in to the central Evergreen service is voluntary but about 80% of the public libraries in the province will likely be using it by the end of 2009. The project manager estimates that this switch will save over \$10 million dollars in total costs over the next few years.

Similarly, here in Brazil the federal government has been actively switching to the use of open source for cost reasons. Sergio Amadeu, who runs the government's National Institute for Information Technology, says that "the number one reason for this change is economic." He explains that, for every workstation, the government is currently paying Microsoft fees of around 1200 Brazilian reais (approximately \$500 USD). "If you switch to open source software, you pay less in royalties to foreign companies," says Amadeu. "And that can count for a lot in a country like Brazil, which still has a long way to develop in the IT sector." Overall, Amadeu estimates that the government could save around \$120 million a year by switching from Windows to open-source alternatives.<sup>3</sup>

Amadeu's last point about using open source as a way to stimulate an ICT sector in Brazil is also interesting. This comes back to the responsibility for taking care of a free kitten. I believe that by getting involved in open-source projects, archival institutions will gain more control and knowledge of the technical infrastructure which they require to manage archival functions. Some would argue that this is not a core function for archives and I would agree to some extent. However, archivists are responsible for preserving and providing access to information. Today, over 90% of the information being created in the world is in digital format. These are tomorrow's archives. In fact, many archives are already struggling with transfers and accruals of electronic records. Archivists can no longer ignore digital technologies simply because they don't fully understand them or because they find the topic overwhelming. Mechanics need to understand automobile engines, doctors need to understand human bodies and archivists need to understand information in its digital form as well as the systems that are used to create, manage and provide access to them. I believe that the focus on open sharing of technical knowledge and know-how, which is characteristic of healthy open-source communities, will go a long way toward raising the technical capacity of the archival profession.

I also think that the open-source community's willingness to talk about its own faults and shortcomings is a healthy part of that. Archivists should not be embarrassed to ask the 'wrong question' and developers and

---

<sup>1</sup> "Open source and proprietary software" (Sept. 2007) UNESCO Information for All Programme.

<sup>2</sup> <http://sitka.bclibraries.ca/>

<sup>3</sup> 'Brazil adopts open-source software' (2 June 2005) BBC News. <http://news.bbc.co.uk/2/hi/business/4602325.stm..>

technical staff should not be embarrassed to admit to bugs, feature gaps or technical mistakes. These exist in all software projects. The difference is that open-source projects such as ICA-AtoM make this information freely available by providing access to the source code repository, online bug list, developer's wiki and discussion list. At the same time we try to actively involve users in working towards solutions. Commercial vendors tend to hide issues or shortcomings and usually find themselves trapped in marketing speak where they exaggerate both the capabilities of their own products as well as the flaws of their competitors. That does not mean that there are not decent commercial vendors providing good products and services to their clients; however, their software license costs are typically out of the price range of most archival institutions. Furthermore, their proprietary technology tends to create a relationship of dependence rather than one which builds sustainable knowledge, capacity and technological autonomy within the archival community.

## Software Features

So now that I've explained some of the philosophical foundation for the ICA-AtoM project as well as my own personal motivation to develop the software, I'd like to give a demonstration of the features found in the current 1.0.5 beta version which was released last week on March 11.<sup>4</sup>

Even if you choose not to download and install ICA-AtoM, you are still able to test these features for yourself using the online demo version of the application or by burning a copy of the Demo CD. The online demo is a fully-featured copy of the application which grants you full administrator privileges so that you can experiment with all the application's capabilities. The demo website refreshes every hour with default data so you are free to make any changes you like.

Another option is the Demo CD. ICA-AtoM is web-based software which requires a web server and database server to operate. However, we created the ICA-AtoM demo CD to make it easy for you to try out the ICA-AtoM software on your own local computer. The Demo CD will run on any computer. It temporarily loads the Ubuntu Linux operating system into memory, along with the necessary web server, database server, Firefox browser and ICA-AtoM application. When you are finished the demo, your computer restarts using your regular operating system and configuration. Of course you can download and make as many copies as you like and I've also brought some Demo CD copies with me to Rio de Janeiro.

Researchers are able to search the archival descriptions hosted by ICA-AtoM using a basic search box. Advanced users are able to use the same search box to enter more sophisticated boolean or proximity search criteria. ICA-AtoM uses the Zend Lucene search engine and ranks search results based on where the search term appears in the record. For example, hits in the title, creator and access point fields are ranked higher than the archival history field. These criteria can be edited by developers and we would like to add the ability, possibly by release 1.2, for administrators to configure the algorithm ranking themselves.

The users can navigate from the search results to the full archival descriptions which are shown in context to their multi-level description, as well as links to the creator's authority record. The user has the ability to browse by facets such as subject, place, names and media types. The user can also view any links to digital objects or browse all the digital objects for a particular aggregate level of description using a Coverflow viewer. ICA-AtoM creates access derivatives for uploaded digital objects, e.g. JPG images, Flash video. The application also provides a browser for multi-page image and text documents.

---

<sup>4</sup> See [http://ica-atom.org/docs/index.php?title=User\\_manual](http://ica-atom.org/docs/index.php?title=User_manual) and <http://ica-atom.org/demo.html>

Users with log-in permission are able to add and edit archival descriptions, authority records or repository profiles. These are all compliant with the ICA's descriptive standards, namely:

- *International Standard Archival Description (ISAD(G))* - 2nd edition, 1999.
- *International Standard Archival Authority Record (Corporate bodies, Persons, Families) (ISAAR(CPF))* - 2nd edition, 2003.
- *International Standard For Describing Institutions with Archival Holdings (ISDIAH)* - 1st edition, March 2008.

The 1.1 release will also add support for the International Standard For Describing Functions (ISDF) - 1st edition, May 2007. The current release also contains data-entry templates for Dublin Core and the Canadian Rules for Archival Description

Names of creators and other actors can be linked from the archival description to an authority record via an Event entity that also records the dates and other information related to the event. Archival descriptions can be linked to access points and, if the user has permission, they can add additional terms to the controlled vocabularies.

All the data-entry sections correspond to the areas of description for each standard. As well, we have access point, digital object and physical object areas. The physical object area allows for links to the boxes and containers in which the analogue archival materials are stored.

All the terms that are available as access points and menu options throughout the application are maintained as controlled vocabulary taxonomies. By release 1.1 we will make these fully compliant with the ISO Thesauri standard relationships (e.g. Use, Use for, Broad Term, Narrow Term, See also)

All user interface elements (e.g. field labels) as well as database content (e.g. archival descriptions, authority records, static pages, etc.) can be translated into multiple languages. The current version of ICA-AtoM contains translations for Dutch, English, Farsi, French, Italian, Portuguese, Slovenian, and Spanish.

Users are able to export archival descriptions using the EAD XML format and they can also import EAD documents, including any multi-level description hierarchies and physical container elements. The 1.0.5 release includes the ability for ICA-AtoM to act as an OAI repository, making descriptions available to OAI harvesters. The 1.0.6 release will include the ability for ICA-AtoM to harvest and import OAI records from other repositories. We are co-developing the OAI feature with the Library and Archives of Canada which is interested in seeing this functionality in ICA-AtoM to enhance its capability to act as a multi-repository portal system which can receive data from contributors via direct data-entry, EAD XML import or OAI harvesting.

The ability to configure OAI, multi-repository and other settings such as interface languages or user accounts is provided through a basic Administrative interface. Administrators are also able to customize their site titles, static page content (e.g. homepage, contact page) and application menus. The 1.0.6 release will include a theming feature which will allow administrators to change the look and feel of the application in a single click as well as develop their own institutional themes.

As I will explain later in the technical architecture overview, ICA-AtoM is fully web-based software. This type of application is typically more complicated to install than a stand-alone, desktop application.

Therefore, we have included a web-based installer with the application to simplify this task. It performs a full system check to determine that the web server and support files comply with the application's minimal technical requirements. If not, it provides a report with explanations on how to re-configure the server environment accordingly. Thus far, about 80% of the installations we have performed are fully handled by the web installer with the remaining 20% requiring some manual configuration intervention.

I have just demonstrated the features that are available in the current 1.0.5 release of the software. 'Release early and release often' is one of the credos of the open-source and web application community. This is something we take to heart as well. I think a realistic expectation for any software application is that it should always be in a process of getting fixed, enhanced and improved. In late April we will release version 1.0.6 which is mostly a maintenance release but which will also include MODS templates and a theming/skinning module as two major new features.

The next milestone will then be release 1.1 which we will formally launch at the ICA CITRA meeting in Malta in November 2009. This will be the stable, production-ready version of the application. It will include a number of performance and workflow upgrades for the XML import and export, as well as full compliance for the relationship types required by the ICA ISAAR and ISO Thesauri standards. Release 1.1 will also provide support for the ICA's ISDF standard as well as EAC XML templates for importing and exporting authority records. One of the more important new features will be a full Access Control List (ACL) capability. This will enable system administrators to define complex permission rules for the system users (e.g. the ability to restrict edit privileges per taxonomy, or translation privileges per language, or edit privileges per repository).

The ICA-AtoM steering committee will hold another meeting in Malta and decide on the next steps for the software based on the community feedback and how the project's funding and governance model have been defined by that time. However, it is likely that the 1.2 release will address upgrades to the search module as well as adding an accessioning module.

## Beta Testing

The beta 1.0 release of the software was successfully launched at the ICA Congress in July 2008 with a conference presentation, two end-user workshops and one administrator workshop. Over 1200 Demo CDs were distributed to delegates and 26 institutions were recruited to participate in a formal round of beta testing. This began in November 2008 with the 1.0.3 release of the software. A full list of the participants, including links to their sites, is available on the beta testing wiki page.<sup>5</sup> There is wide international representation within this group. Beta testing sites have been deployed in English, French, Spanish, Portuguese, Arabic, Farsi, Italian, Slovenian, Dutch and German.

The beta testing includes active dialogue between the users and developers which is all recorded on the public ICA-AtoM discussion list.<sup>6</sup> Bug fixes and new feature requests are being incorporated into the daily development schedule as part of this work. This first round of beta-testing will end later this month with a survey of the participants to ask for their feedback, impressions, and advice on the application's features. This survey and its replies will be posted to the ICA-AtoM user discussion list. We will launch a

---

<sup>5</sup> [http://www.ica-atom.org/docs/index.php?title=Beta\\_Testing](http://www.ica-atom.org/docs/index.php?title=Beta_Testing)

<sup>6</sup> <http://groups.google.ca/group/ica-atom-users>

second round of beta-testing from May to August, based on the 1.0.6 release of the software. This will assist with finding and resolving any bugs prior to the 1.1 release and will give us further guidance on developing the application for the 1.2 release and beyond.

Much of the beta tester feedback thus far relates to the underlying data model and the implementation of standards. Examples include requests to add a template for the Canadian Rules for Archival Description and EAD XML import/export functionalities; both of these important features have since been added to the software. The Australian testers have requested that ICA-AtoM be adapted for use with the Australian series system. An early analysis demonstrated that the ICA-AtoM data model is flexible enough to accommodate this, including the flexible multi-level hierarchy and multi-provenance links that this requires.<sup>7</sup> Further analysis is now continuing to determine whether a separate template is required or whether small changes to the existing template will be sufficient to accommodate any extra metadata attributes. I should also note that a much earlier analysis carried out in the summer of 2008 at the National Archives of Brazil revealed that ICA-AtoM was compliant with the Brazilian Nobrade standard. Since it is so closely modeled on the ISAD(G), the only change that was required was to add a 'conservation note'. This was done as early as the 1.0.1 release. Finally, the Direction des Archives de France conducted a thorough review of the 1.0.4 release for ICA-AtoM compliance with the full set of ICA standards. We are currently implementing the majority of the recommendations from the this report in the 1.0.6 release. All of the DAF report recommendations, including the ability to use the new ICA International Standard for Describing Functions, will be available in the 1.1 release in November 2009.

Some of the more technical feedback from our beta tester include requests for automated test scripts and a built-in data migration tool, both of which have now been implemented. However, the majority of feedback has been related to usability. Some of the testers have entered large numbers of descriptions, which allows them to become aware of usability considerations that are not necessarily apparent when the system contains only a small sampling of records. For example, the National Archives of Iran imported several tens of thousands of existing descriptions and authority records when they began testing ICA-AtoM. Linking to existing authority records is cumbersome when you have to browse a very, very large list. One way to deal with this is to add an AJAX auto-complete search feature. The developers at the National Archives of Iran coordinated the development of this feature with us and have released it under a GPL license so that we can integrate their work into the upcoming 1.6 or 1.1 release.

Some of the other usability requests have included improvements to adding new users, expandable fields for data entry, and alphabetical sorting in drop-down menus; these features have all been added to the software. Others that are not yet implemented, but are on the development roadmap, include bulk processing capabilities as well as the ability to work with default record templates to avoid having to add the same data repeatedly for large numbers of similar descriptions.

Many beta-tester questions relate to how to use existing functionalities. The most common of these are how to attach creator information to an archival description, whether to add a creator for all levels of description and whether to enter an archival institution for all levels of description. In ICA-AtoM, creator information such as administrative history and dates of existence are maintained in separate authority records and are "added" to the archival description by linking to an existing ISAAR authority record or creating a new one. Both the creator and archival institution are inherited at the lower levels of description,

---

<sup>7</sup> see [http://groups.google.ca/group/ica-atom-users/browse\\_thread/thread/101426fb8444434f/](http://groups.google.ca/group/ica-atom-users/browse_thread/thread/101426fb8444434f/)

which means that they should not be linked to those descriptions. This practice follows the principles of ‘description of the general to the specific’ and ‘non-repetition of information’, both of which are stated in the ISAD(G) and other archival standards. However, somewhat surprisingly, a number of archivists seem unfamiliar with these principles and therefore experience some confusion about how they are implemented in the software. We hope to add a validation feature in the 1.1 release software which gives the user some more explicit feedback on the screen if they are violating some of the data-entry rules or principles.

It seems pertinent at this point to add some other observations about beta-testing behaviour. In general, we have found that users prefer asking questions over reading the manual and forcing us to re-evaluate how much resources we devote to maintaining user documentation. Certainly, context-sensitive help should help fill in some of this gap. We are hoping to include this feature in the 1.2 release.

Moreover, the beta-testers are often reluctant to post questions to the users’ discussion list, preferring instead to rely on one-to-one e-mail communications with software support staff and developers. Possibly, archivists are afraid of asking the “wrong” question and making themselves and their institutions look bad. However, one-on-one communications are not sustainable in the long-term and result in information not being shared with the user community at large. It may take a little longer than expected to encourage archivists to embrace a culture of open collaboration and communication, in the meanwhile we will continue to try and lead by example. Wherever, possible we simply forward communication that does not include sensitive information to the discussion list and carry on the response thread there.

Other observations include the fact that the testers often have limited time to devote to the testing, although this was expected, given what we know about the resources available to most archival institutions today. Therefore, we have never asked for more than a few hours of beta-testers’ time. Some have been able to provide a lot more than that but we appreciate any contribution that is made and all of them are helping to improve the software.

Artefactual Systems is hosting most of the beta-testing participants sites (about 85%). This is partly due to the fact that those archival institutions that have internal technical support cannot spare it on a beta-testing project, they are too busy with day-to-day production responsibilities. However, it also a reflection of a larger technical support gap in the archival community. We anticipate that there will be a need to provide third-party hosting services for ICA-AtoM. Therefore, Artefactual Systems will be providing that for a minimal fee after the beta-testing period. Our first hosting client will be the Archives Association of British Columbia which is currently beta-testing the application and planning to migrate its provincial union list of archival descriptions to ICA-AtoM in a couple of months. We are able to provide hosting, backup and technical support at a cost of \$125 a month which is a fraction of their current fees and less than the current hosting costs for all other commercial archives management applications. We will only charge a monthly fee per website and will not limit the number of users. This means the AABC can give out passwords to those institutions that do not have their own archival description software or do not have the capability to contribute to the provincial union list using EAD export or OAI harvesting. Each archival institution can thereby enter or update their archival descriptions directly in the portal, eliminating the current, cumbersome data migration tasks which are usually carried out by cutting and pasting from word processing documents.

I would also encourage the approach taken by the British Columbia Public Libraries Association where institutions centralize their technical support amongst themselves and share the costs. Here in Rio de Janeiro, the Casa Rui Barbosa Foundation has partnered with the Catholic University. The University Lab is providing hosting support for their beta-testing site while their technical staff are taking the opportunity to study and enhance ICA-AtoM in relation to their own digital library system. This partnership could form the basis for an ongoing hosting arrangement that could be extended to other Brazilian institutions. Of course, that is only one suggestion. I don't want to put unnecessary expectations on either of these excellent beta-testing partners.

My final observation on the beta-testing period is that some testers have unrealistic expectations about the application. To be clear, the current version is a beta release not a stable production version. We are conducting a beta-testing phase to find bugs and get the feedback and recommendations from archivists and their institutions on how to improve the application in time for the 1.1 release.

Also, many users have trouble seeing past the vendor/client type relationship they are used to and appear to have trouble seeing themselves as being part of a broader community that has control over its own software. Yes, Artefactual Systems are the lead developers and the primary source of technical support right now. However, in this open-source eco-system, under the freedoms granted by the GPL license, we are replaceable and compatible with any number of other sources of technical support and software development.

I should note that there are some encouraging exceptions. The Library and Archives of Canada as well as the National Library and Archives of Iran has been very active in testing the full capabilities of the software and contributing to its development, customizing the software to suit their own needs where necessary. As well, developers here at the Catholic University in Rio are looking closely at how they can leverage some of their experience with designing multi-lingual interfaces for their digital library application to enhance the same features in ICA-AtoM.

At the same time, I'll continue to do the best I can to convince archivists that this is their software. This will require more education and acceptance of the open-source model and culture within our community. Therefore, I really appreciate the time I have been given by the AAB to explain the history and vision of the ICA-AtoM Project. In the meanwhile, the ICA and Artefactual Systems are dedicated to providing leadership and support but, in the end, this is your free kitten.

## **Project History**

So now that you have a better idea of the ICA-AtoM's features and objectives, I'd like to take a step back and tell the story of how we arrived at this point. In 2003 the ICA Committee on Information Technology, under the direction of Peter Horsman from the Netherlands Institute for Archival Education Research (Archiefschool), published the functional requirements for an Open Source Archival Resource Information System (OSARIS). It was the Committee's intention to use these requirements as the basis for developing an open-source archival description application; however, it did not have the funding to carry this out.

Then, in 2005, the UNESCO Information For All Programme (IFAP) granted €45,000 Euros to the ICA's Human Rights Task Force to create an online guide to archival sources related to human rights violations. UNESCO's IFAP actively promotes the use and creation of open-source software and that was

one of the criteria for the online guide. Peter Horsman realized that this was a good opportunity to follow up on the OSARIS project. He pointed out that the software created for the online Human Rights database should be based on ICA descriptive standards. The UNESCO grant could be used to build the online database but also double as seed funding to create an open-source archival description application for use by ICA members and other memory institutions. The ICA Secretariat agreed and asked the Dutch Archiefschool to coordinate this work. At that time Peter Horsman and I were working together on the requirements analysis for an E-Depot at the City of Rotterdam archives. He knew I was keen to develop an open-source application for archives since we had discussed such a project several times in the past. Therefore, I eagerly agreed when he asked whether I would be interested in working on the UNESCO contract for the ICA.

During early 2006 I undertook an extensive technical analysis, which I'll revisit in my presentation tomorrow, and developed a first prototype release of the software. We held a project meeting at the ICA Secretariat in mid-2006 to review the prototype and to discuss the vision and objectives for the project, which centered around the need to get more archival descriptions and materials online for those organizations with archival holdings that did not already have the software to do so. We agreed that the ICA Congress in 2008 would mark the first milestone release for the software. It was also around this time that former ICA Secretary-General Joan van Albada chose the acronym for "Access to Memory" as the software name. Joan insisted that this and all other ICA products should use the "ICA" name as a prefix (e.g. ICA-ISAD) to raise awareness about the ICA and its activities. Therefore the software was officially named *ICA-AtoM*.

A second version of the ICA-AtoM prototype (alpha release 0.2) was completed in September 2006 and the results were presented to UNESCO. UNESCO was enthusiastic about the revised approach to the project, which would involve leveraging the online database work as a catalyst to launch an ICA open-source initiative. However, in spite of some early optimistic designs and timelines, the prototype stage had made it clear that to develop a fully-featured beta version of the software by the time of the ICA Congress in 2008 would require more time and money, and it was not entirely clear where any further funding would come from. I was working on a web archiving project for the World Bank Group Archives at that time and they agreed to contribute some of the time and fees, approximately €8,000 Euros, from that Artefactual Systems contract to continue design and development work on the next ICA-AtoM alpha release. Despite the uncertainty about project funding I was determined at this stage to ensure that the project would achieve its initial goals, and I carried on development between other consulting assignments, staying in close touch with Peter Horsman at the Dutch Archiefschool to brief him on my progress.

The faculty at the Archiefschool was very interested in the development of an open-source archival description application, sanctioned by the ICA and compliant with its standards, for use in classroom instruction. This helped them to avoid any conflicts that would arise if they chose to use one commercial application over another for classroom instruction. The Archiefschool was also interested in applying and extending ICA-AtoM as prototyping software in some of its research projects. Therefore, at a project meeting at the ICA Secretariat office to review alpha release 0.3 in early 2007, the Dutch Archiefschool committed to fund the next phase of software development in the amount of €100,000 Euros. Along with an additional €30,000 Euros made available by the Direction des Archives de France, this funding gave me the opportunity to work full-time on software development for the remainder of 2007 and to hire one additional programmer and one archivist to test the software and draft the first user manual.

This funding helped to deliver alpha release 0.5, which I presented to the ICA membership at the CITRA meeting in Quebec City, Canada, in November 2007.

Prior to the CITRA meeting, Secretary-General Joan van Albada had asked me to prepare a budget to complete beta 1.0 release software development in the remaining six months leading up to the ICA Congress in July 2008. I presented a budget for €167,500 Euros which included hiring two additional staff to complete all the remaining development, testing and documentation tasks. Once again, we carried on with the project work with blind faith that, if this project was in fact filling a real need for the archival community, the funding and resources would arrive in some form. After my CITRA presentation we had a roundtable discussion about next steps for the project and the variety of compromises or work-arounds I might implement to keep the project moving forward. At this point Abdullah Al Reyes, the Director of the United Arab Emirates Centre for Documentation and Research (CDR), and also vice-president for ICA Marketing, saw the potential that ICA-AtoM has to promote the ICA worldwide and committed the CDR to fund the 1.0 development budget, giving us the last push we needed to complete the first public beta release of the software in time for the ICA Congress in July 2008.

## Project Governance

At this stage the project gained greater visibility and the various stakeholders agreed that a more stable governance model had to be put into place to ensure the long-term sustainability of the software and its community of users, developers and sponsors. The ICA-AtoM Steering Committee was formed to address this need. The current list of ICA-AtoM Steering Committee members is posted on the project website.<sup>8</sup> The Committee held its first meeting at the Dutch Archiefschool in Amsterdam in May 2008. There it drafted the following Project Purpose and Values statement.

The purpose of the ICA-AtoM Project is to provide free and open-source software:

1. that enables institutions to make their archival holdings available online, especially those who could not otherwise afford to do so
2. that manages archival descriptions in accord with ICA standards
3. that provides multi-lingual interfaces and content translation features
4. that supports multiple collection types
5. that is fully web-based, user-friendly and follows accessibility best practices
6. that is flexible and customizable
7. that is useful to both small and large institutions alike
8. that supports single or multi-repository implementations

The values of the ICA-AtoM Project are to conduct its affairs in the spirit of open-source collaboration, including:

1. openly sharing technical expertise
2. promoting professional best practices

---

<sup>8</sup> <http://ica-atom.org/about.html#governance>

3. nurturing an active community of users and developers
4. leveraging existing open web technology to deliver best-of-breed archives software
5. providing solutions to organizations with limited financial and technical resources
6. engaging active participation from organizations that have the necessary financial and technical resources
7. providing a common ground for cross-discipline collaboration with related communities
8. generating revenue to support ICA-AtoM and other ICA activities through a business model that benefits from widespread ICA-AtoM adoption

At the Amsterdam meeting in May 2008, the Committee also began discussion about the project's organizational structure and business model but these were not finalized at that time. Another meeting of the ICA-AtoM Steering Committee was held at the ICA Congress in Kuala Lumpur and I made a presentation on the project to the ICA Board. I reported that following extensive beta testing and additional core development, ICA-AtoM 1.2 should be the first production-ready, full release of the software. With the appropriate funding, release 1.2 could be made available by Fall 2009 in time for the next CITRA meeting in Malta.

However, the ICA-AtoM project costs are quite a bit higher than traditional ICA programs or projects and this was raising some concern within the ICA management. To be clear, the project has not taken away funds from existing program budgets. In fact, it has attracted new funding and resources into the ICA network. As well, the ICA-AtoM project expenses are actually quite reasonable compared to the budgets of similar projects, e.g. Dspace \$2 million USD (2002), Fedora \$4 million USD (2007). The main difference is that these other open-source projects have received large, one-time grants from major funding sources like the Andrew Mellon foundation. This is different from the more grassroots and serendipitous approach which characterized the early years of the ICA-AtoM project. That is not to say that the ICA-AtoM project is opposed to receiving grants. This type of funding model (for example, an EU funding request) is something that has been suggested for further action at the previous two steering committee meetings.. However, the way that the project has unfolded thus far is to move forward with voluntary contributions from some key institutions that share the vision and objectives of the ICA-AtoM project.

After the 1.0 release this has included further funding from the United Arab Emirates Center for Documentation and Research, the Direction des Archives de France as well as some related ICA-AtoM implementation projects that Artefactual Systems is carrying out in Canada involving the Archives Association of British Columbia, the Library and Archives of Canada, the City of Vancouver Archives and Canadiana.org. These contributions will be used to deliver a stable, production 1.1 release in time for the next ICA CITRA meeting in November 2009. However, the more ambitious 1.2 release roadmap has been scaled back while the ICA continues to define the governance model of the ICA-AtoM project.

There is ongoing dialogue within the ICA Board, Secretariat and ICA-AtoM Steering Committee about the role of the ICA-AtoM Project. This project has challenged some of the traditional ways of doing things, in large part due to the speed and scope of the project developments. In many ways the governance is catching up to the technical developments: whereas I had defined a lot of the early direction and decision-making for the project, Artefactual Systems is now stepping back into a more formal, third-party contractor role while

the Steering Committee assumes control over the central decision-making, promotion and communication. The Committee will likely be assisted shortly by a project coordinator based out of the ICA Secretariat office in Paris. My formal position as Software Release Manager is limited by the terms and length of any active contract between Artefactual Systems and the ICA. We are currently negotiating a contract for the release 1.1 development which will take place between May and October this year.

These are all positive changes in the growth and evolution of the project. They help to maximize the transparency and sustainability of the ICA-AtoM Project for years to come. I am pleased with Artefactual Systems' role as a third-party service provider and I hope that soon a number of other companies or consultants will emerge worldwide to provide ongoing ICA-AtoM support, training, development and hosting services for those archival institutions that don't have these access to these kinds of resources within their own organization. This will encourage collaboration and competition equally, which is a signature quality of healthy open-source communities. Artefactual Systems will collaborate on the continuing development and enhancement of ICA-AtoM with other companies or institutions. At the same time, we will work hard to stay at the cutting edge of ICA-AtoM innovation and deployment to play a technical leadership role in such a community.

## **ICA-AtoM and the ICA Strategic Objectives**

In the meanwhile, the ICA leadership is pleased to see more formal management structures put into place for the ICA-AtoM project and are encouraged by the potential of this project to meet all six of its recently defined 2008-2018 Strategic Objectives .

These objectives are:

1. Raising Awareness
2. Harnessing New Technologies
3. Building Capacity in the Archival Profession
4. Strengthening the ICA Network
5. Improving the Performance and Accountability of the ICA:
6. Building Partnerships

As free and open-source software that addresses a very practical need for archivists around the world, the ICA-AtoM application is already helping to address the ICA's first strategic objective to 'raise awareness'. The fact that I am standing here speaking to you today is direct evidence of that. Since the beta launch at the ICA Congress in July 2008, the ica-atom.org website has received 4,000 visits to the online demo version of ICA-AtoM. Much of the feedback we are receiving, including that from senior IFLA and UNESCO members, reflects pleasant surprise that ICA is pursuing such 'innovative' and 'cutting-edge' work.

We are helping to achieve the ICA's 'harnessing new technologies' strategic objective. ICA-AtoM is fully web-based software that is positioned to take advantage of new technologies which will allow archivists to provide timely and relevant technical solutions to the institutions and clients they serve. For example, the Archives Association of British Columbia and the Canadian Council on Archives have evaluated a number of options for upgrading their archival description portals and have decided that ICA-AtoM provides the

best architecture and features to meet their requirements. I will speak more tomorrow on the design and evolution of ICA-AtoM's technical architecture.

The ICA-AtoM project is also helping to achieve the ICA's third strategic objective to 'build capacity in the archival profession.' The ICA-AtoM release 1.1 will be fully-compliant with all ICA descriptive standards and will thereby serve as an excellent tool to promote and teach archival best practice. The Dutch Archiefschool and the ICA Section on Education have both beta-tested ICA-AtoM for use in classroom and professional workshop instruction.

The ICA-AtoM project is also helping to achieve the strategic objective of strengthening the ICA Network. By providing a free but high-quality software solution, the ICA-AtoM project is drawing interest throughout the ICA's international network, allowing archivists to work together and organize themselves regionally around a concrete project that is directly relevant to their day-to-day practice. This is currently happening most visibly in Canada, where I have spoken on behalf of the ICA and the ICA-AtoM project at several conferences and meetings. ICA-AtoM is the focus of a provincial and national portal upgrade and it was the predominant agenda item at the Canadian Council on Archives' Annual General Meeting in October 2008. Also, following this AAB conference we will hold a beta-testers meeting here in Rio de Janeiro for all those institutions that have already begun to implement the software.

The ICA-AtoM project is also helping to achieve the ICA's strategic objective to 'improve the performance and accountability of the ICA' by openly discussing the funding and governance issues as the project grows and evolves. As an open-source project, ICA-AtoM introduces a culture of open access and transparency to project management. Information about the day-to-day technical design, development and decisions is all available for monitoring online in the projects discussion lists, documentation wiki, code repository, issues list, and the logs of the online meeting room.

Finally, the ICA-AtoM is helping to achieve the sixth of the ICA's strategic objective, 'building partnerships.' From its outset, ICA-AtoM has been a collaborative project involving partners inside and outside the archival community (e.g. the Canadiana.org project). It has also provided a common ground of collaboration with UNESCO, which is continuing with the Human Rights database project, beta testing ICA-AtoM for its own archives, and possibly working with its Memory of the World Subcommittee on Technology with Artefactual Systems and the City of Vancouver to integrate ICA-AtoM with other open-source tools to create a digital preservation system prototype. The ongoing success of the ICA-AtoM project will continue to depend on the ability to nurture and grow these types of partnerships.

## **Technical Architecture**

So far, I have demonstrated the ICA-AtoM software and talked about the history and vision for the project. I would like to conclude my presentation now by taking a closer look at the analysis that led to software's technical architecture and data model. It is almost stating the obvious to say that most information seekers today turn to the Internet first to find resources related to their research needs. Although erroneous, there is a common belief that if information (or at least metadata about that information) is not on the Web, it does not exist, or it might as well not exist.

One of the primary motivations for the ICA-AtoM project was a concern that not enough archival materials, or at least archival descriptions, were available on the World Wide Web to represent the richness, variety and depth

of the archival holdings that are available to researchers around the world. Therefore, the primary goal of the project is to provide free and open-source software that enables institutions to make their archival holdings available online, especially those which could not otherwise afford to do so. This means that the primary requirement for the ICA-AtoM software is that it be fully web-based and allow archival institutions to get their descriptions online where they will have the widest exposure to an international audience. Hence the name 'Access to Memory'.

So what does it take to get archival information on the web? The World Wide Web is based on the concept of resources, such as html pages, images, video, etc.. Each of these resources has its own unique resource identifier made available by web servers that respond to requests sent by client applications, namely web browsers such as Mozilla Firefox, Safari, Internet Explorer, etc. The simplest way to publish information on the web is to manually create a static HTML page. Many archival institutions still maintain their own websites this way. As well, over the past decade many archival institutions have invested time in creating EAD XML finding aids, and they make these available as stand-alone XML files or convert them to HTML or Acrobat PDF files and post those as static files. However, these all have to be managed as individual resources and it is often difficult or not possible at all to provide integrated search, browse and presentation functionality for these resources.

A more efficient way to publish information to the web is with database-backed web applications where all the resources being made available on a website are being maintained and stored as an integrated system. Blogs, wikis, and web content management systems such as Drupal or Joomla! are all examples of this type of web application.

In early 2006 I conducted a technical analysis of existing open-source software applications to determine whether we could simply extend these types of applications to meet the needs of the ICA-AtoM project. I evaluated Fedora, Dspace and Greenstone. All three are excellent open-source projects originating from the digital library or digital repository communities. However, at that time, each of these projects had fairly rigid internal technical architectures that made them difficult to adapt to archival description requirements. This was especially true of their data models, which would need to be adapted to allow for compliance with the ISAD(G) archival description standards and principles, such as flexible, multi-level description. However, an ever bigger issue is that these tools were Java framework applications which are difficult to package and deploy on cheap, shared hosting services. Part of the ICA-AtoM objective is to make an application that is easy to download, install and maintain on basic shared hosting services. These hosting providers tend to favour simpler LAMP stack applications.

LAMP is an acronym for Linux, Apache, MySQL and Python, Perl or PHP. It represents the basic and most popular open-source web application architecture that uses Linux for an operating system, Apache HTTP server as a web server and MySQL as the database server, in combination with a web application written in the Python, Perl or PHP scripting language.

Consequently, I turned my attention to some of the more popular open-source PHP web publishing platforms, namely Wordpress and Drupal. Both were already in wide use in 2006 and had large communities of users and active developers. However, I ran into the same issue again. It was difficult to adapt these applications to implement the data model required for archival description.<sup>9</sup> More troubling was that both of these applications grew very organically within a loosely organized open-source community. This was part of the reason for their popularity but it also meant that the underlying technical architecture for each

---

<sup>9</sup> The Drupal CCK feature, which allows for the creation of custom content types in Drupal, was only in its design stage in early 2006.

of these applications were not very cleanly or clearly designed. The vision for ICA-AtoM was to implement a well-designed, scalable architecture that can form the basis for a long-standing archives management system which is adaptable to increasingly complex requirements for web services and digital preservation. It required a clearly articulated technical design and this ruled out Wordpress and Drupal at the time.

I should note also that in early 2006 there was no evidence yet of any other open-source archival description applications. I had heard about the Archivists' Toolkit project but it did not make its first version available until early 2007. Likewise, I did not hear about the Archon project until I attended its launch session at the Society of American Archivists conference in August 2006. However, I did stop and take a close look at both of these applications when they were released. At that time both used EAD as their core data model instead of ICA standards and neither offered multi-lingual capabilities or multi-repository capabilities, all of which are critical requirements for the ICA-AtoM application. Furthermore, the Archivists' Toolkit was always intended as an application to support the internal processing of archival materials, rather than providing access to archival holdings. It allows for the export of EAD finding aids but it does not provide a web-interface to its database of archival descriptions.

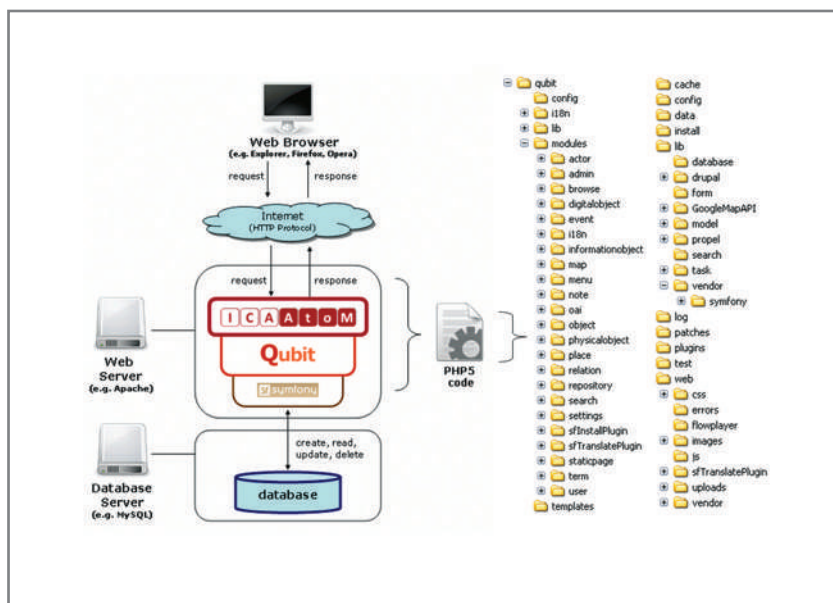
Both of these applications have recently added support for the American DACS standard which is very closely related to the ISAD(G). As well, Archon has added the ability to translate the application interface (although not the database content) and the Archivists' Toolkit can be installed in a multi-repository configuration. However, neither of these two tools address all of the same critical requirements for the ICA-AtoM Project. I also think it is fair to say that both of these projects are targeted largely to the needs of the American archival community and its sponsors. So merging the ICA-AtoM development with one of these projects did not make sense back in 2006. I believe that there are still enough differences in the technical architectures and community priorities to warrant separate projects today. I also believe this is a good thing for the archival community as a whole. Whereas we had been starved far too long for legitimate open-source archival management applications, archivists now have three excellent projects to choose from, allowing them to select the best tool and community to address their particular needs.

So, after completing my analysis of technical options in early 2006, it looked as though the only remaining option was to develop an application from scratch, a daunting and ambitious task given the very limited resources available to the project. However, at that time there was another trend emerging in web development, namely the popularity of Model-View-Controller (MVC) web frameworks for rapid development of web applications. I tested the most popular of these, namely Ruby-on-Rails, Django, and CakePHP before settling on Symfony.

I favoured Symfony because it was PHP-based, making it easier to deploy on a wide variety of shared hosts; it was also designed for developing enterprise applications and was based on fully object-oriented PHP5, allowing for the implementation of a number of best practice design patterns. There are many advantages to developing ICA-AtoM using the Symfony framework. By using the Symfony framework we avoid having to develop and document a whole number of common web application features, such as caching, user sessions, routing, forms templating, etc.. Instead we can concentrate on developing the business logic and templates required for an archival description application. This has saved us a large amount of development time and cost. We are also benefitting from being part of the larger Symfony open-source community so that we are not "going it alone". As well, it is much easier to add new developers to the project, since Symfony provides rich documentation and a consistent structure and logic for organizing the application code.

Symfony is a fairly typical implementation of the Model-View-Controller pattern. MVC is a software engineering architectural pattern which isolates data persistence (i.e. the model) from business logic (i.e. the controller) and from user interface components (i.e. the view), resulting in an application in which it is easier to modify the visual appearance of the application without affecting the underlying

business rules and vice-versa. In a Symfony application, client browsers interact with the application through the front controller which is the *index.php* file located in the publicly accessible */web* directory. Using the Symfony routing system rules, the front controller sends the client requests to the appropriate action code in one of the application's modules.



For example, this request maps to the information object module:  
<http://ica-atom.org/demo/web/index.php?informationobject/edit/id/100>

It will load the edit action and pass 100 as the id parameter value. This will trigger some application logic that reads the informationobject with the id value of 100 from the database. The corresponding edit template is populated. The template is then decorated with the global layout template, which includes the header, footer and other generic layout elements. The output is sent to the user's browser where ICA-Atom waits for the next request. Entering some data and clicking the save button will send an *update* request to the application, and entering a search query will send a new *read* request.

Each module such as informationobject, actor, or repository corresponds closely to an object in the application's data model. Most requests usually include some database create, read, update, or delete (CRUD) activity before passing the response back to a template in the view layer. The database CRUD actions are handled through an object relational mapping (ORM) layer which converts programmable application objects into rows or records in the underlying relational database, which is still the most effective and scalable data store for web-based applications. The MySQL database is used in development but ICA-Atom uses the PEAR PDO database abstraction layer within the ORM and it is therefore also compatible with Postgres, SQLite, SQLServer, Oracle, etc..

We have also integrated Symfony with a number of other open-source libraries that provide additional features, such as JQuery and YUI for AJAX controls and effects. AJAX is a methodology that allows for web application controls to provide dynamic visual effects or get and post data to the web server without requiring the web page to reload. AJAX controls allow web applications to behave more like desktop applications and provide for a better usability experience.

By mid-2006 I had developed the first two alpha releases of the ICA-AtoM application using the Symphony framework. I had demonstrated it to the ICA Secretariat and we received approval from UNESCO to apply the 40,000 Euros they had made available for the Human Rights Archives web portal as seed money for this prototype. However, project financing and therefore development was limited through to mid-2007 when the Dutch Archiefschool committed funds to work towards the first beta release of the application.

Around the same time I had been in contact with developers of the Alouette Toolkit, another early-stage open-source development project based in Canada. They were also looking to develop a web-based application to help memory institutions in Canada that did not have any existing tools to make their digitized collections available online and to produce standards-compliant metadata that could be used in the Alouette website, a search and browse portal for digitized Canadian heritage content made available by libraries, museums, galleries and archives.

The underlying technical vision for the two projects was almost identical, as were the multi-lingual requirements since Canada is officially a bilingual country. The only slight differences were the data standards and the target communities for which these were being developed. The primary sponsors and target community for the Alouette Toolkit was Canadian research and special libraries with a focus on the Dublin Core and MODS standards. After some discussions and evaluation of the initial ICA-AtoM prototypes, the technical teams for both projects decided that, since both projects were starting from scratch, it was to the benefit of both to share the limited time, technical resources and finances available to each and to collaborate on the underlying technical platform. I should add that governance for the Alouette Toolkit has recently been taken over by [Canadiana.org](http://Canadiana.org), a consortium of major Canadian memory institutions promoting digitization and access to Canadian works. [Canadiana.org](http://Canadiana.org) will launch its first public release of the software in May 2009, renaming its version of the application to Digital Collection Builder.

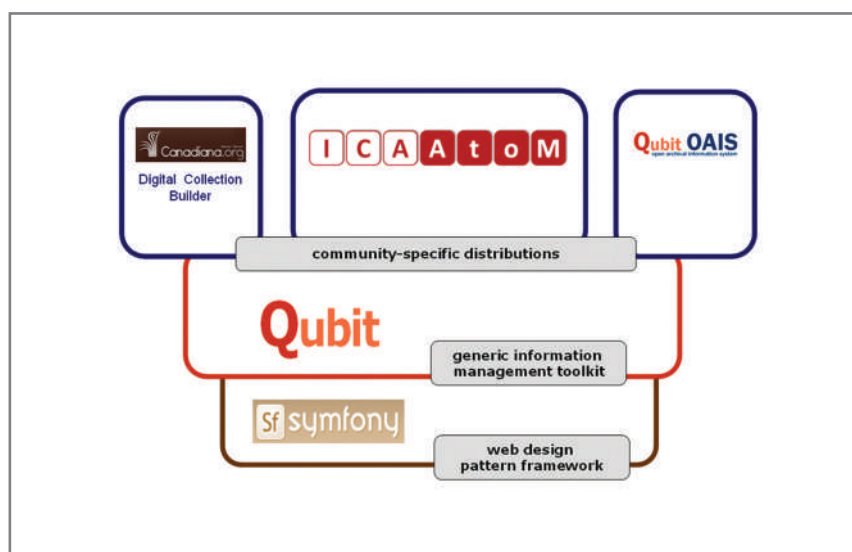
I had already designed the ICA-AtoM object model to be as flexible and generic as possible. Part of the rationale was that archival institutions would also need software to help catalogue their published reference material or any artefact collections for which they may be responsible. I felt that this capability might as well be provided by the same application. From a technical point of view, managing metadata about information objects, whether these are archival materials, published materials or artefact materials, can be made generic enough to accommodate all collection types while still respecting specific descriptive standards and practices. I also designed the data model to ultimately apply archival description to electronic records and digital collections as well as to map the ICA metadata to compatible standards. This flexibility made it possible for the Alouette Toolkit to adopt and adapt it for its own use.

The remaining issue was to appease the sponsors and target community of each respective project with an application that was branded and customized specifically to their audiences. To allow for this we created a generic underlying toolkit, which we eventually codenamed Qubit. A 'Qubit' is a quantum bit of information, the smallest conceptual unit of information known to science. It is an appropriate name for a system that is designed to manage untold Qubits of information and provide access and context for them. Both the ICA-AtoM and Digital Collection Builder applications are managed as distributions of the Qubit toolkit.

The concept of *distributions* is fairly common in the open-source community. For example, there are a large number of Linux distributions (such as Red Hat and Ubuntu) each of which is based on the Linux kernel that is officially maintained by the original Linux developer Linus Torvalds. Torvalds decides what new code is incorporated into the Linux kernel on which the distributions are based. Another example is the Drupal open-

source content management platform which has several distributions that are made available as installation profiles. These include distributions for education sites, civic organizations or the music industry.

Similarly to the Linux and Drupal projects, Qubit now has several distributions. *ICA-AtoM* is a Qubit distribution for the international archival community, *Digital Collection Builder* is a Qubit distribution for the Canadian Heritage Community, and *Qubit OAIS* is a more recent collaboration between Artefactual Systems, the City of Vancouver Archives, the InterPARES Project and possibly the UNESCO Memory of the World to create a Qubit distribution for institutions implementing OAIS-compliant digital preservation systems. Each Qubit distribution and Qubit itself is released under a GPL free and open-source license which gives each project and community the right to revise and re-distribute the software as it sees fit.



Each of these projects is contributing technical resources and financial sponsorship to develop the underlying Qubit architecture to meet the requirements of its particular sponsors and the needs of their community. Each one benefits in turn from contributions made to the underlying platform which become available to them in each subsequent release of the software. For example, much of the application core was developed via the ICA-AtoM project but the recent ICA-AtoM 1.0.5 beta release contains a custom menu module, multi-page digital object browser, and Dublin Core templates, all of which have been funded by the Alouette Toolkit project as contributions to Qubit.

So this is how we arrived at the current technical architecture for ICA-AtoM. As demonstrated yesterday, the application is fully web-based. All user interactions with the system take place through the user's web browser. Users access resources such as archival descriptions, digital objects, authority records, etc. on the web server; clicking a button or link triggers a PHP script that sends a command to the database and returns the output as HTML back to the user's browser.

ICA-AtoM is developed and supported on a fully-open source LAMP stack rather than proprietary software. Consequently, there is no cost to download any of the software required to run the ICA-AtoM application, and users are free to make any enhancements or changes as they see fit.

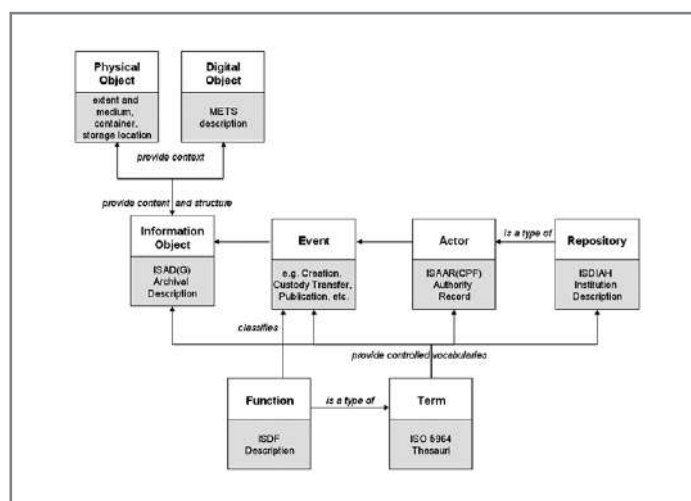
ICA-AtoM can be used by a single institution for its own descriptions or it can be set up as a multi-

repository “union list” accepting descriptions from any number of contributing institutions via direct data-entry, EAD import or OAI harvesting. Artefactual Systems is pilot testing this scenario and architecture right now in a collaboration with the Archives Association of British Columbia, the Canadian Council of Archives and the Library and Archives of Canada. There are many small archives in Canada that still do not have their own archival description software. Also, they don’t really have the technical capacity to install and host a web application like ICA-AtoM. Representatives from these institutions will be given a repository-specific account so that they can enter and update archival descriptions directly in the central union portal. All they need is web access. Other institutions that have software with EAD export capability can use this method to contribute to the union list. Finally, our ideal scenario is to use OAI harvesting for the automatic export and import of archival descriptions throughout a hierarchy of provincial, thematic, or national networks. The 1.0.6 release will have the capability to harvest Dublin Core records, the standard metadata format used for OAI. However, simple Dublin Core is not rich enough to support archival description relationships, therefore, we are looking ahead to use EAD records within the OAI protocol.

## Data Model

The ICA-AtoM data model is designed to be flexible enough for adaptation to other descriptive standards and it currently supports crosswalks from ISAD(G) to EAD, Dublin Core, the Canadian Rules for Archival Description, and MODS. The ICA-AtoM data model consists of a number of core entities, namely, archival descriptions, authority records, archival institutions, terms and events.

Archival materials are a type of information object that we preserve for future access. By thinking of archival materials in a more generic way, as information objects, it makes the underlying data model more flexible and allows it to be re-used for other types of information objects, such as published material or cultural artefacts.



Archival descriptions provide contextual information about archival materials; they are representations of the archival materials. Archival descriptions are arranged into hierarchical levels (e.g., fonds, series, files, items). This is done in ICA-AtoM using a flexible and sophisticated nested set hierarchy model. Each archival description includes the data elements found in the ICA’s ISAD(G) standard. Each archival description may

be linked to a physical object or a digital object describing location, physical characteristics, structure and technical information about the archival material or record itself. Physical objects and digital objects possess content and structure. They need the context provided by representation information to become information objects. The full name for the information object in the ICA-AtoM data model could be 'information object representation' but that would be redundant. In the end, all of the objects in the data model contain representation information for the objects they surrogate, so the 'representation' suffix is implied.

Authority records provide descriptions of the actors that interact with the archival materials represented by archival descriptions. The primary actor in an ISAD(G) description is a creator. ICA-AtoM authority records include the data elements found in the ICA's ISAAR(CPF) standard.

Archival institution entities provide descriptions of repositories that preserve and provide access to archival materials. An archival institution is a type of actor. Like all actors, an archival institution has its own ISAAR authority record. But its characteristics as a repository (e.g. its opening hours, research services, contact information) are described separately in an ISDIAH institution record. In other words, an archival institution record inherits all the applicable fields from the Actor's authority record (e.g. Authorized Form of Name, History, etc.) and extends the repository-specific fields in an ISDIAH record.

Term entities organized into separate taxonomies provide controlled vocabularies used throughout the system (e.g. as subject access points or as drop-down menu options for levels of description or authority types, for example). Release 1.1 will be upgraded to make all the term taxonomies compliant with the structure and behaviour of the ISO Thesauri standard, which calls for the implementation of equivalence (Use, Use for), hierarchical (Broad term, Narrow term) and associative relationships (See also).

Most definitions of records in archival science emphasize the importance of 'acts', 'transactions', or 'business processes'. The Event entity formalizes this concept in the underlying conceptual model rather than being lost in a weak link directly between an archival description and an authority record. Event relationships link ISAAR authority records (descriptions of actors) and ISAD(G) records (descriptions of archival materials).

Through events, a single actor can have zero to many relationships with zero to many archival units and each relationship can be treated as a separate entity, delimited by normalized start and end dates. The ICA-AtoM Events are very similar to the 'Business' entity in the Australian SPIRT metadata model<sup>10</sup> but also the Event object found in the PREMIS digital preservation metadata standard, the former capturing the context of the record's relationships in its business, juridical and procedural context and the latter capturing the physical context of the record's relationships to its content and structure.

The Event entity is designed to capture all events related to the archival materials, not just creation. This could include custodial history changes, conservation treatments, digital object reformatting, access logs, description revision history, etc. The Event entities provide the true context of what has happened to the archival materials over their lifetime. This is especially useful in managing digital archival materials and adapting ISAD to the description of digital archival materials.

Functions described using the new ICA International Standard for Describing Functions (ISDF) can be used to further classify Events. For example,

---

<sup>10</sup> <http://www.sims.monash.edu.au/research/rcrg/research/spirt/deliver/conrelmod.html>

- the Event entity indicates when a record was created,
- a link from the Event entity to the ISAAR authority record indicates which Actor created it,
- a link from the Event entity to the ISAD record describes the document that was created
- and a link from the Event entity to the ISDF record classifies what type of act the Event is in relation to the organization's functions (e.g. the Event is an 'instance' of the function described in the ISDF record).
- Also, a link to the Place term taxonomy can classify the geographical location of the event.

All of these Event relations can be used to build rich search and browse interfaces to the archival materials. I believe this design is a practical and powerful way to implement the otherwise theoretical definition of a record as 'process-bound information' (as the Dutch school calls it) or the dispositive and probative records defined by *diplomatics*.

However, the recent ICA-AtoM beta testing phase did raise some concerns about how the Event concept related to the existing ICA standards and how archivists would normally expect to enter creation dates and links to the authority records for creators. Therefore, based on the feedback from beta-testers and the DAF report, it has been decided to completely hide the Event entity from the end-user in release 1.1 and to limit the Event types to creation events only in the ISAD(G) templates. Unfortunately, this will mean that creation dates will not be entered in direct relation to creators and it limits the potential of the Event entity concept. However, it will remain in the underlying data model and could possibly be re-introduced in some other form as the ICA Committee on Descriptive Standards begins its own work on a conceptual model for archival description. Also, the Event entity is still accessible from the Dublin Core template, which calls for Creator, Publisher and Contributor actors, and the Canadian Rules for Archival Description template which incorporates a number of different types of dates and actor relationships.

## Closing remarks

In the original abstract for this presentation I had also planned to discuss Web 2.0 and Semantic Web technologies and practices. However, I'm afraid that I've exhausted my generous allotment of presentation time on the primary and most important topic I came here to discuss. Nevertheless, Web 2.0 and the Semantic Web are two topics closely related to my doctoral research which, I must admit, has been put on hold for a couple of years due to the heavy demands of the ICA-AtoM project and other consulting contracts. However, I do fully intend to return to this research soon and apply it as much as possible to any future development of the ICA-AtoM software. I believe these areas hold some interesting potential for improving the usability of archival information systems and access to archival materials. However, I will have to save that for any future presentations or papers. Of course, I am fascinated by technology and archival systems and I don't really tire from talking about these topics. However, I also recognize that not every archivist shares this interest as strongly. Therefore, I want to thank the AAB once again for the honour to speak to you at length about the ICA-AtoM project and technology. I hope that some of you will decide to check out and join the ICA-AtoM community or that I've at least discussed some things that help you to evaluate or analyze the technical architectures within your own institutions. *Obrigado, tchau e ate a próxima.*

# The Evolution of Archival Descriptive Standards in the United States

Jackie Dooley

This paper provides an overview of the development of key archival descriptive standards in the United States since the early 1980s when the first relevant standard came into existence, as well as some observations about current issues with regard to access to archives. The evolution of descriptive standards has transformed how we do our work as archivists and has added to our success in making our materials available to researchers of all kinds who wish to use them. The standards-based infrastructure that we now use can provide the necessary infrastructure of consistent, structured data that we need in order to build a global system of networked databases of archival information—but only if archives implement these standards and contribute their data to powerful shared systems.

I will begin by describing some characteristics of the archival landscape in the U.S., since it differs significantly in some respects from that of other countries in terms of both institutional environments and archival terminology.

Next I will describe the development and significance of the following descriptive standards:

- The MARC format for archival and manuscripts control (MARC AMC)
- *Archives, Personal Papers and Manuscripts (APPM)*
- Encoded Archival Description (EAD)
- *Describing Archives: A Content Standard (DACS)*

Third, I will examine several aspects of our current environment relative to the production and use of archival data:

- Archival collections management systems
- Archival discovery systems
- Current challenges for the archives community
- Our archives research agenda in RLG Programs and OCLC Research

## The archival community in the United States

The United States is diverse nation, both geographically and demographically, in which most aspects of domestic governance are highly decentralized. Our archival community is the same: enormous and heterogeneous, with no centralized governmental funding, mandates, or organization. Unlike other nations,

our National Archives plays no national coordinating role, except for Federal government archives. In the realm of standards, the National Archives not only does not take the lead; it has not implemented any of the standards I will be describing.

We use the term “repositories” to refer to all types of institutions and organizations that are the custodians of archives, and those types vary greatly. Public archives--by which I mean those holding the archives of governments--and private archives are equally significant in terms of the importance and extent of the materials they hold. Each of the fifty states has a state archives, the quality of which varies a great deal. Local governmental archives for counties and cities or municipalities often do not exist.

Thousands of colleges and universities throughout the nation also hold archives. Some of these are public universities owned by the state in which they are located, while others are private and receive no governmental funds. Archives usually are located administratively within the university library, which generally leads to close integration with library standards and services. This may be one reason why the community of archivists who work at universities is the most professionalized in the archival sector.

Countless other private archives exist in private companies, religious organizations, museums, labor unions, local historical societies, and many others. In such archives, particularly if they are small, there may be only one staff member who has little or no formal education as an archivist. I understand that the situation is similar in Brazil.

Records management generally is administered separately from archives in most of these institutional environments.

Professional education for archivists also is decentralized, and there is no requirement for certification. The Society of American Archivists (SAA), our national professional organization, plays the major role in articulating the necessary characteristics of archival education, but universities are not required to follow the SAA guidelines. The quality of the programs and the depth of archival coursework provided therefore differ significantly from one program to another. Professional archives degrees are at the post-baccalaureate Master's level, and the programs generally are allied with either history departments or library schools. Many people work as archivists who have no professional training, particularly in smaller institutions. SAA has strong continuing education programs through which archivists may develop their skills in new areas; the courses on descriptive standards are among the most heavily subscribed.

SAA takes the lead role in standards development in addition to educational guidelines and workshops. The Library of Congress, our *de facto* national library, also plays a significant role in managing some standards in coordination with SAA. Many archives still use no formal descriptive standards; they are most used within the highly professionalized university environment, and therefore more with private papers and manuscripts than with government records.

U.S. archivists have little—if any--awareness of international archival practices and standards; this, unfortunately, mirrors the low level of awareness that most Americans have about international affairs.

## Archival terminology

Archivists in the U.S. speak our own archival language, and usage of archival terminology varies even within our country. We do not use the term *fonds* for a body of archival materials. Government and organizational archives speak of “record groups” or “record series,” while archivists in universities and private archives refer to “collections,” “manuscripts” or “papers,” in addition to “archives.” All of these terms can have roughly the same meaning.

In the context of description, we refer to “collection-level” descriptions, meaning the top level of description for any body of archival materials—the *fonds*-level description, for those who use that term. We generally use the term “finding aid” for detailed listings of materials that are presented in a hierarchical structure.

Archival “processing” is the term we use for the work of physical organization and preservation of materials.

## Development of descriptive standards

All descriptive standards have two essential features: a well defined set of data elements and a clear set of relationships among those elements. Without these two essentials, we can have no intelligent archival information systems, and archival information cannot be reliably or intelligibly exchanged across institutions. It is my assumption that all of us at this conference believe in the importance of descriptive standards, since we are here to discuss the effective design and implementation of such systems. It is our duty as archivists to carry this message to those in our profession whose principal attitude about standards is that they are too difficult, too time consuming, and therefore too expensive to implement. It is also our responsibility to try to lessen these inhibiting factors.

U.S. archivists began to think about the possibility of developing standards in the late 1970s. A group of visionary leaders began a project called the National Information Systems Task Force (NISTF),<sup>11</sup> and one of their initial questions was this: Is it possible to define a set of standard data elements to describe archival materials? Archivists generally believed that the answer would be no, since every archival collection is unique. The NISTF report published in 1980 showed, however, that generally the same data categories were used across archives, even though the resulting descriptions were organized and presented very differently.

Another important group, the U.S. Working Group on Standards for Archival Description, did excellent work eight years later to articulate an overarching set of principals for archival descriptive standards and systems environments. Their report, published in 1988 as a special issue of *The American Archivist*,<sup>12</sup> articulated a matrix of three types of standards that is an exceptionally useful construct:

Data *structure*: categories of information, usually called data elements

Data *content*: guidelines for formulating the content of data elements

Data *value*: particular indexing terms or names as found in thesauri, vocabularies, and authority files

I will refer to the first two types—structure and content--as I discuss particular standards.

---

<sup>11</sup> National Information Systems Task Force ...

<sup>12</sup> Working Group on Standards for Archival Description ... [was it one or two AA issues?]

## The MARC format for Archival and Manuscripts Control (MARC AMC)

Led by the U.S. Library of Congress, librarians implemented the MARC format, the first data *structure* standard for library cataloging records, in 1968. MARC was an important milestone in the history of data processing and instantly created the ability for libraries throughout the world to exchange standardized online cataloging records. MARC was, and still is, used in concert with the library *content* standard, the *Anglo-American Cataloging Rules* (AACR2), in the U.S., Canada, the United Kingdom, and Australia. MARC is used in libraries worldwide, and records are exchanged principally via OCLC's WorldCat database of 125,000,000 cataloging records.

Over the forty years since MARC came into use, most libraries have had to upgrade their catalog software multiple times. Without a standardized format such as MARC, successful migration of data from one system to another never would have been possible. This is one of the most important lessons for archivists who are unsure of the value of standards: a data *structure* standard helps assure the permanence of data by enabling its inevitable transfer from one systems environment to another.

The most important outcome of the pioneering National Information Systems Task Force was that archivists decided to develop an archival component of the MARC format to enable exchange of archival data. As a result, the MARC Archival and Manuscripts Control format, or MARC AMC, was launched in 1983. The NISTF members did not see MARC as the ideal environment for archival information, but it nevertheless served as an important vehicle for exchanging information about archival collections in a standardized way. MARC was routinely used in libraries throughout the U.S. by this time, and so archivists could easily add their data into a computerized infrastructure that was readily available at their institution. In addition, library users could locate descriptions of archival collections in the same online catalogs in which they search for books, journals, and other materials. This integration has been very important in bringing the rich resources of archives to the attention of broader scholarly and educational audiences.

MARC AMC was intended for use principally for *collection-level* summary descriptions of archival collections, which was a new practice for U.S. archives, in which detailed finding aids were (and still are) the usual informational tool. In many respects these new collection-level records were similar to the top level of description in a finding aid, but to many archivists it was disconcerting to see them standing alone and in a library-centric environment. Use of MARC AMC was controversial, and it has never been implemented in most governmental archives. Universities and private research institutions eager to make their archival collections known to potential users were the biggest users, and in the past 25 years, archivists have created more than one million records in the MARC format.

While MARC AMC became used over time in many U.S. repositories, it has never been much used in other nations, where the links between libraries and archives are not as prevalent as in the United States.

## Archives, Personal Papers and Manuscripts (APPM)

Because the archivists who developed the MARC AMC format sought to integrate archival descriptions into a library-based environment, they realized the need for a companion *content* standard to parallel the MARC *structure* standard. I mentioned earlier that for U.S. libraries, AACR2 is the companion

content standard for MARC. Steven Hensen of the Library of Congress, a member of the NISTF group, took on the task of formulating such a content standard.

The result was *Archives, Personal Papers and Manuscripts (APPM)*, published in 1983, the same year as MARC AMC. Like MARC AMC, APPM was intended for collection-level descriptions, not detailed finding aids. Because this *content* standard was to be used for modern manuscripts in addition to archives, Hensen also included rules for describing single manuscripts. The data elements detailed in APPM parallel the descriptive elements in a MARC record, such as author, title, date, and various specific notes describing the materials in narrative form.

One brief example will serve to illustrate how MARC AMC and APPM work together. MARC very briefly defines the purpose of field 520 for recording a summary of the contents of a collection, while APPM describes in detail the types of information that comprise a useful summary. These include the types or forms of documents that are present, the major activities or achievements recorded, and the most significant people, organizations, topics, and places reflected in the collection.

Other content standards have been published over the years for particular types of archival material. Two of the most important are *Graphic Materials*<sup>13</sup> for describing two-dimensional pictorial materials such as photographs and prints, and *Archival Moving Image Materials*<sup>14</sup> for films and videorecordings. Other manuals provide detailed methodologies for cataloging oral histories, maps, medieval manuscripts, and other formats. Many of the descriptions created using these content standards are encoded in the MARC format and thereby made available to the global information community.

## Encoded Archival Description (EAD)

Publication of Encoded Archival Description version 1.0 in 1998<sup>15</sup> was a watershed event for archives in the United States. Fifteen years after MARC AMC and APPM were published, archivists finally had a *structure* standard for multi-level descriptions, and thus the ability to encode complete archival finding aids. As many of you know from the extensive professional literature on EAD, it was initially developed as a research project by a group of archivists led by Daniel Pitti of the University of California at Berkeley. I was fortunate to be a member of this group, which began its work in 1995 after being awarded a Bentley Library Fellowship from the University of Michigan. After EAD was published, the Society of American Archivists formed the EAD Working Group, of which Wendy Duff and I have been members at different points in time.

EAD's original design took into account archival practice in the United States only. Nevertheless, interest quickly grew in Western Europe, and important projects quickly emerged from the United Kingdom, France, and Germany. The Society of American Archivists, which manages EAD in collaboration with the Library of Congress, responded to international interest by adding international members to the EAD Working Group. Version 2002<sup>16</sup> is fully compatible with ISAD(G),<sup>17</sup> and many of the changes made in that version were in direct responses to issues raised by international members of the working group.

---

<sup>13</sup> Betz, Elisabeth. *Graphic Materials: ...* Washington, D.C.: Library of Congress, 1982.

<sup>14</sup> White-Hensen, Wendy. *Archival Moving Image Materials: ...* Washington, D.C.: Library of Congress, 1992.

<sup>15</sup> *Encoded Archival Description ...* Chicago, Illinois: Society of American Archivists, 1998.

<sup>16</sup> *Encoded Archival Description ...* Chicago, Illinois: Society of American Archivists, 2002.

<sup>17</sup> *International Standard Archival Description (General) ...* International Council on Archives.

While the basic data elements in EAD are very similar to those in MARC AMC and APPM, EAD has been and remains controversial. Some resistance is tied simply to the fact that learning and implementing a standard is perceived as being more time consuming than is constructing finding aids informally. It is a simple matter to convert a finding aid from any common data-creation tool into HTML, whereas EAD encoding requires a more sophisticated software environment. Publishing EAD and developing an effective searchable interface is the most complex of all, a problem that many repositories have solved by joining a consortium and sharing one discovery system. Those designing archival information systems lament the fact that EAD's flexibility—very few elements are required in any particular context—makes it difficult to design a system that has strong functionality. In addition, some archivists who are working to envision new system functionalities argue that EAD's design renders it too closely tied to the structure of static finding aids printed on paper.

Despite all these concerns, EAD has served the essential purpose that the MARC format did for libraries thirty years earlier: EAD has made it possible for archives throughout the world to exchange archival descriptions in a standard format. These descriptions can then be migrated among technical platforms, flexibly retooled for effective interface design, and disseminated to users via global discovery systems. Whatever EAD's weaknesses may be, archivists will resolve them in time, and many thousands of archival descriptions will smoothly transfer to whatever new environment that they create.

## **Describing Archives: A Content Standard (DACS)**

DACS was published in 2004 as the new U.S. *content* standard to replace APPM.<sup>18</sup> Whereas APPM focuses only on collection-level description, DACS very consciously enables the multi-level description prescribed by the international archival descriptive standard ISAD(G). It is the U.S. implementation of the ISAD(G) and the companion ISAAR(CPF)) for the description of archival materials and their creators. In addition, whereas MARC was the only type of data output envisioned for APPM descriptions, DACS is output-neutral: MARC, EAD, Dublin Core, METS, and other record structures can be exported.

DACS includes examples to illustrate how the rules might be implemented in MARC and EAD; it also includes crosswalks to these and other standards, as EAD does. DACS can be used to describe archival materials at any level of specificity, from the collection to the item level. It includes an overview of archival description (including the requirements for effective multilevel description), outlines the elements that must be included at different levels of description, and describes how those elements should be implemented. DACS also includes guidance in describing creators of archival material, constructing archival authority records, and recording forms of names.

DACS originally was intended to be a joint U.S.-Canadian standard, replacing both APPM and the Canadian *Rules for Archival Description* (RAD).<sup>19</sup> Unfortunately, philosophical and practical differences between the Americans and Canadians led to divorce; the Americans finished their work and published DACS, while the Canadian archival community found the overall nature of the rules in DACS too controversial and maintained RAD in its existing form. While I was not a member of the committee that worked on DACS

---

<sup>18</sup> *Describing Archives: A Content Standard*. Chicago, Illinois: Society of American Archivists, 2004.

<sup>19</sup> *Rules for Archival Description* ...

and therefore do not know the detailed arguments on either side, I wonder to what extent the differences are ones that would be noticeable by end users of archival information systems.

## Archival collections management systems

We are learning a great deal about the AtoM<sup>20</sup> software at this conference. Its promise for providing an open-source software environment that helps archivists create ISAD(G)-compatible descriptions addresses a problem that U.S. archivists share. We have standards, and we know what we want our data to be like, but it is time consuming to prepare, manage, and disseminate that data, particularly for archivists who have access to insufficient expertise in information technology.

At least a dozen existing software tools seek to remedy this problem; some are freely available as open-source software, while others are marketed by commercial firms who sell or license their software, sometimes for rather significant fees. AtoM is an open-source solution, as are two tools under development in the United States: the Archivists' Toolkit and Archon.

Every one of these software packages has very different features and functionality. Rather than detail their characteristics, I will cite the important new report *Archival Management Software*, published in January 2009 by the U.S. Council on Library and Information Science, that compares AtoM and nine other systems.<sup>21</sup>

## Archival discovery systems

We develop archival descriptive standards so that we can make the materials in our archives known and used; our descriptive work is a means to an end. For U.S. archives, the chief systems for global discovery of archival collections are WorldCat, ArchiveGrid, and a variety of consortial finding aid hubs. Some archives disseminate their descriptions only via a local institutional website.

I mentioned earlier that WorldCat, hosted by OCLC and including cataloging records contributed by nearly 70,000 institutions around the world, contains 125,000,000 MARC records describing materials in libraries and archives. While WorldCat has historically been focused on MARC catalog records, in recent years OCLC has expanded it to include many article databases and other information sources. Within the next year WorldCat is going to be greatly expanded once again, this time to include several large databases that include information from archives. An open-access version of WorldCat is freely available on the Internet.<sup>22</sup>

One of these databases is ArchiveGrid, which started in 1998 and combines one million MARC records describing archival collections and 70,000 finding aids, principally from repositories in the United States.<sup>23</sup> Searching for archival collections via ArchiveGrid is efficient because it isolates archives from library materials. ArchiveGrid links to finding aids stored at the host institutions rather than on centralized storage at OCLC. Finding aids contributed to ArchiveGrid can be encoded in either EAD or HTML. We believe strongly in the value of EAD as a data structure standard, but HTML finding aids are permitted as

---

<sup>20</sup> AtoM ...

<sup>21</sup> Spiro, Lisa. *Archival Management Systems* ... Washington, D.C.: Council on Library and Information Resources, 2009. Available at <http://www.clir.org/pubs/reports/spiro2009.html>.

<sup>22</sup> Available at [worldcat.org](http://worldcat.org).

<sup>23</sup> Available at [archivegrid.org](http://archivegrid.org). ???

well in order to make many more collections available to the global audience. Inclusion of unstructured HTML finding aids, however, reduces the design functionality that is possible.

ArchiveGrid has been a stand-alone subscription resource since it began; libraries and archives have had to pay a separate fee in order to obtain access. This very month (March 2009), however, OCLC is adding ArchiveGrid to the basic package of resources that members access via one subscription fee, thus making it available at no extra cost to subscribers worldwide.<sup>24</sup>

ArchiveGrid is one of the projects I work on for RLG and OCLC Research, and we look forward to soon beginning to greatly expand its content and functionality. We know that tens of thousands of additional finding aids are available from U.S. universities alone, not to mention the many other archives in the United States and throughout the world. Our software engineers will develop software connectors to make it very easy for archives to contribute their finding aids to ArchiveGrid. It is our hope that ArchiveGrid can become a unified global discovery system for as many archives as possible. We will be publicizing it widely as plans for expansion develop over the next year.

The third major type of archival discovery system in the U.S. is the category of consortial finding aid hubs. These exist in a variety of states and regions. In my home state of California, for example, the Online Archive of California<sup>25</sup> includes more than 50,000 finding aids, as well as many thousands of digitized images and texts. Other regional systems exist in the Pacific Northwest, Texas, North Carolina, Ohio and elsewhere.<sup>26</sup> Each hub integrates EAD-encoded finding aids from a particular group of institutions, many but not all of which are universities. Each has its own interface design and functionality, and together these systems provide important test beds for archival user studies.

## **Current challenges for the archives community**

The use of standards and the resulting wide dissemination of descriptive information have raised awareness of archives throughout the education and research communities in the United States. This is of course our goal, and use of archives is growing fast. Growth and visibility also create new challenges, however, and we have many.

Whereas in the past archives and other research materials such as rare books principally were used by advanced researchers, in today's society we realize that students of all ages and members of the general public also can learn and benefit from using archives. It is now common for archivists in universities to use archives to teach undergraduate college students, often with exciting results. Archivists in all types of repositories very actively promote use of their collections to the general public through exhibitions, websites, publications, public programs, and visits to schools. The resulting challenges include identifying new ways to interpret materials and finding the time to incorporate these new activities.

Archives also have become far more visible within the larger organizations that support them. In universities, for example, library administrators became aware of the nature of archival data when EAD provided structure to finding aid data, improved its intelligibility, and started to become incorporated

---

<sup>24</sup> How to cite FirstSearch?

<sup>25</sup> Online Archive of California, available at <http://www.cdlib.org/oac>

<sup>26</sup> URLs for all these.

into shared information systems. Realizing the value of archives also has led to new pressure to make all unprocessed collections available so they can be accessed and used. I know of no archive that does not have hundreds or thousands of cubic feet of materials in backlogs, awaiting an archivist who has the time to process and describe them. But few archives receive new staff positions; we must therefore eliminate backlogs by working faster and more efficiently.

You may be aware of an article published in *The American Archivist* in 2005 titled “More product, less process: revamping traditional archival processing,”<sup>27</sup> in which Mark Greene and Dennis Meissner called on archivists to focus on making collections publicly known before organizing and describing them in detail. This article has literally caused a revolution in American archival circles. Letting go of the high-quality work in which we take great professional pride is difficult—but if the rich collections for which we are responsible are unknown, we have failed. Archivists in the U.S. are struggling, as a result, with the many effects, both positive and negative, of making more materials available, thereby further increasing use of our archives, but having less detailed arrangement and description to enable ease of use.

The pressure to digitize every document in every archive is also very strong. For at least the past ten years, archivists have routinely disappointed researchers who discover a collection description online but then learn that the documents themselves have not been digitized. As with backlogs of unprocessed and undescribed archival materials, we now realize that the very slow pace at which we have been digitizing is insufficient. Another strong trend, therefore, is to identify methods for much faster and less expensive digitization of far more material than has been made available online in the past.<sup>28</sup>

At the same time, archivists worldwide are contending with the onslaught of born-digital records. In the U.S., government archivists are far ahead of those in universities and other private environments in learning about electronic records and implementing systems to manage and preserve them. We already have lost years of important information and will lose many more before our governments, universities, and society in general understand the complexity, the enormous cost, and above all the urgency of managing and preserving archival records that have no physical form.

None of us are in danger of running out of challenging work.

## Our research agenda in RLG Programs

My colleagues and I in RLG Programs and OCLC Research work to find ways to help the research library and archives communities work together to address some of these challenges. We currently have seven projects underway,<sup>29</sup> each of which I will briefly summarize.

1. *Collections management*: We are looking at methods for surveying the overall holdings of an archive in order to make it easier for archivists to efficiently capture the information that is most important for evaluating backlogs, setting priorities for processing and preservation, and other management activities. We plan to develop a model methodology for archivists to adopt.
2. *EAD*: We are examining barriers to implementation of EAD. Archivists who do not use EAD cite

---

<sup>27</sup> Greene, Mark and Dennis Meissner. “More product, less process: revamping traditional archival processing.” *The American Archivist* vol. ?? no. ?? (2005), pp.

<sup>28</sup> A report published by RLG Programs proved catalytic in effecting this changed perception. Erway, Ricky and Jennifer Schaffner. *Shifting Gears*: ...

<sup>29</sup> <http://www.oclc.org/programs/ourwork/collectivecoll/archives>

reasons such as lack of training, lack of information technology support, lack of staff, lack of easy-to-use software, and even lack of understanding as to why such a standard is important. A report will summarize the problems and suggest possible solutions.

3. *Data analysis:* We are studying the one million archival MARC records in WorldCat to analyze how archivists have, or have not, implemented the standards that I discussed earlier. We hope to discover patterns, whether good or bad, that will help our colleagues make decisions about the aspects of archival description that are the most valuable for enabling effective discovery. As a result, they may then be able to focus on the most important elements, leaving aside aspects of description that may not be beneficial to users.
4. *Exposing hidden collections:* We will conduct a survey of the archival collections in 400 colleges and universities throughout the U.S. to learn more about the collections that remain “hidden” and unavailable. We will publish the results to establish a new baseline for the state of processed collections, methods used in processing and description, and much else. We also hope that the resulting data will help archives collaborate with each other to identify joint projects and shared problems that warrant high priority for community resolution. We also hope to obtain funds for very inexpensive conversion of “legacy finding aids”—those created before EAD was developed, many thousands of which are not yet online—so that archivists can move ahead to work on new collection and users can readily locate these previously-described collections.
5. *Discovery systems:* We are reviewing user studies that have evaluated discovery (i.e., access and retrieval) environments, particularly the finding aids hubs mentioned earlier, to see how archival descriptive elements are incorporated and used. We hope that there will be interesting connections between this project and the one in which we are analyzing archival MARC records.
6. *Delivery issues:* We are investigating several problems that complicate the final step in the archives process, which is delivering materials to a person for use. Particular issues of concern to U.S. archivists today include efficiently providing digital copies of archival materials to users, clearing intellectual property rights, granting permissions to publish, and loaning materials between archival repositories.
7. *Digitization:* We are investigating methods of increasing the scale and speed of digitization of archives.

## Conclusion

The 21<sup>st</sup> century is a challenging time to be an archivist, as the examples described in this paper illustrate. We must master and promulgate descriptive standards to ensure that our data is clear and consistent. We must then use this data to underpin powerful information systems that are easily and globally available to users of archives everywhere. In the meantime, we must, as we always have, provide high-quality services to those users when they arrive at our archives—whether they arrive physically or via the Internet. And we must do all this with far fewer resources of staff and funding than we know are necessary.

Our passion for archives will get us there.

# Sistema de Arquivos Históricos da Unicamp (PESQUISARQH): módulo de registro, armazenamento e acesso.

Fábio Rodrigo Pinheiro da Silva

Telma Maria Murari

Maria Aparecida Forti

Neire do Rossio Martins

## Introdução

A tecnologia representada pelo uso de computadores isolados ou integrados em redes, com transmissão de dados em alta velocidade, sistemas gerenciadores de banco de dados, armazenamento digital, etc., possibilitaram o acesso instantâneo à informação e a documentação.

Frente a essa realidade, o Arquivo Permanente do SIARQ e os centros de documentação da Unicamp modernizaram seus processos de trabalho, uma vez que a conservação e a preservação da documentação e o suporte ao ensino e à pesquisa, até então proporcionados pelo atendimento convencional, tornaram-se limitados e inadequados.

Com o intuito de estudar e adaptar a realidade das unidades de informação supracitadas e com objetivo de implementar alternativas que otimizassem seus processos de trabalho, foi formada uma equipe técnica multidisciplinar, compreendida por profissionais que atuam no tratamento técnico de documentos e na área de informática, para estudar e implementar alternativas informatizadas. Esta equipe desenvolveu procedimentos técnicos e os implementou em uma ferramenta denominada Sistema de Arquivos Históricos da Unicamp (PESQUISARQH).

Para isso, foram estudados os pré-requisitos necessários para atender as necessidades técnico-operacionais de cada unidade de informação, tendo como premissa principal normalizar a descrição arquivística seguindo os padrões internacionais: ISAD (G) - Norma Internacional de Descrição Arquivística e EAD - Encoded Archival Description, além do preparo para a migração dos documentos dos sistemas existentes até então, cerca de 70 mil registros com 30 elementos de descrição para cada item.

Definido com requisitos de alto nível, o sistema foi projetado nos módulos a seguir: **Módulo 01** – Registrar, acessar e armazenar acervo; **Módulo 02** – Gerenciar atendimento e indicadores de pesquisa; **Módulo 03** – Monitorar preservação; **Módulo 04** – Identificar documentos de acervo; **Módulo 05** – Emitir Relatórios; **Módulo 06** – Integrar com sistemas de informação afins; utilizando a metodologia espiral para o desenvolvimento, com linguagem de programação NetBeans/JAVA/ CVS e banco de dados DB2.

Em 2005 foi colocado em produção o primeiro módulo do sistema, que permite o registro, o armazenamento e o acesso de informações e documentos, compreendendo as seguintes funcionalidades: Manter dados de descrição dos acervos; Conectar documentos digitais armazenados; Identificar e controlar versões de documentos digitais; Manter dados de cadastro de procedência e de documentos recebidos; e Consultar dados pela *Web*.

## O módulo de registro e armazenamento

O módulo de registro e armazenamento é composto da **Área de Manutenção de Dados**, restrita aos responsáveis pela inclusão, alteração, exclusão e manutenção dos dados e documentos através de criação de perfis de acesso. A descrição é feita a partir de uma estruturação multinível, isto é, do geral ao particular, sendo necessária a inserção de fundo e sua respectiva divisão como grupos, sub-grupos, séries, sub-séries, dossiês/processos e itens documentais.

No caso de descrição de itens que permite o acesso aos documentos “nato-digitais” ou capturados digitalmente, foram estabelecidos critérios de segurança e de restrição que garantem a publicação e o acesso somente àqueles cuja autorização esteja expressa no contrato de doação e/ou outra documentação que formalize a sua custódia, conforme previsto na Lei nº 8159/91 e na Lei nº 9610/98.

Além dos 26 elementos de descrição da ISAD (G), foram incluídos pela equipe técnica por demanda local e pautados pela EAD outros 10 elementos. Os novos elementos adicionados são:

- **Natureza do Fundo:** Classifica o fundo e/ou coleção em Privados quando são pessoais e em Institucionais quando são gerados por unidades e órgãos da Unicamp;
- **Gênero:** Identifica o gênero do documento (item documental) por meio de uma tabela constante do sistema. Constitui informação estratégica para planos de pesquisa e gestão do acervo, motivo de sua obrigatoriedade. É possível especificar a designação do gênero documental conforme as opções apresentadas com auxílio de tabela auxiliar;
- **Classe:** Identifica a hierarquia estabelecida, por intermédio de seqüência numérica ou codificada, à série, dossiê e/ou processo ou item documental, a partir das funções e atividades que os geraram;
- **Local:** Identifica, em qualquer nível, o local em que a unidade de descrição foi produzida;
- **Localização Física e ou Catalogação Topográfica:** Identifica a localização topográfica da unidade de descrição;
- **Endereço do Arquivo Eletrônico:** Referencia o endereço eletrônico do nível de descrição em questão quando se tratar de nato-digital ou mesmo de cópia digital;
- **Responsável(is):** Informa o responsável pela descrição e inserção dos dados;
- **Dados Conservação:** Presente somente no nível de item documental, constitui importante fonte de informação para os planos de gestão do acervo e questões que envolvem a conservação pontual dos documentos;
- **Instrução para referência:** Orienta o usuário a referenciar adequadamente o arquivo pesquisado;
- **Descritor(es):** Permite a inserção de palavras-chaves (descritores) que visam facilitar a pesquisa e a recuperação por índices Temático, Onomástico, Geográfico e Cronológico.

Consta ainda do módulo o **Cadastro de aquisições e periódicos**, utilizado para registrar informações do doador e/ou responsável pela remessa do acervo, bem como o processo de negociação

e o conteúdo geral da documentação antes dela ser processada tecnicamente. Além disso, o módulo contempla o **Registro de Entrada de Documentos**, que consiste na descrição preliminar dos documentos recebidos de acordo com sua procedência.

Em se tratando de periódicos é necessário proceder ao **Cadastro de Periódicos**, destinado a catalogação de publicações seriadas, que permite ainda o controle dos vários suportes nos quais tais publicações estão registradas. A catalogação é realizada em duas etapas:

1. Cadastro do título do periódico, quando são registradas as informações relativas a cada publicação: título, editor, ISSN etc.
2. Cadastro dos fascículos (exemplares) do periódico, quando são registrados os dados relativos a cada exemplar dos títulos anteriormente registrados.

## O módulo de acesso

O módulo de acesso dispõe de mecanismos de busca e possui as seguintes variáveis de consultas: **Consulta Avançada** - através da combinação de campos; **Consulta Acervos** - lista os fundos/ coleções; **Consulta Simples** - através de qualquer palavra; e **Lista de Fundos** - relaciona fundos/ coleções por unidade de informação. São consultas públicas disponíveis para qualquer cidadão que tenha acesso à Internet.

## Considerações finais

Atualmente o sistema informatizado integra as bases de dados do Arquivo Edgard Leuenroth (AEL), do Arquivo Central do Sistema de Arquivos (AC/SIARQ), do Centro de Documentação Cultural Alexandre Eulálio (CEDAE), dos Arquivos Históricos do Centro de Lógica, Epistemologia e História da Ciência (CLEHC) e do Centro de Memória da Unicamp (CMU).

Até então, o primeiro módulo desenvolvido tem atendido as expectativas levantadas pelos profissionais envolvidos, no que se refere aos requisitos arquivísticos: registro de dados, conservação de originais a partir da disposição pública, otimização de pesquisa e atendimento ao usuário; e também ao que se refere aos requisitos computacionais: baixo custo, facilidade de manutenção e segurança.

As etapas seguintes serão a de desenvolvimento dos demais módulos que deverão ser incluídos no planejamento do Centro de Computação da Unicamp, que está aberto a estudos de parcerias de desenvolvimento com os órgãos envolvidos; integração com Sistema Informatizado de Gestão Arquivística de Documentos (SIGAD) da Unicamp; disposição e acesso de documentos digitais e; integração de bases de dados de unidades de informação arquivística e biblioteconômica a partir de pontos de acessos padronizados.

## Bibliografia básica

ARQUIVO CENTRAL DO SISTEMA DE ARQUIVOS DA UNIVERSIDADE ESTADUAL DE CAMPINAS. **Desenvolvimento de sistema de instrumentos de pesquisa do Arquivo Central/ SIARQ**: manual de inserção de dados. Campinas: Unicamp, 2004 (obra não publicada por estar em fase de revisão).

BARBIERI, Cristina Correia Dias; INNARELLI, Humberto Celeste; MARTINS, Neire do Rossio. Gerenciamento eletrônico de documentos: criação de um banco de informações e imagens no Arquivo Permanente da UNICAMP. In: CONGRESSO INTERNACIONAL DE ARQUIVOS BIBLIOTECAS CENTROS DE DOCUMENTAÇÃO E MUSEUS, 1., 2002. **Textos...** São Paulo: Imprensa Oficial, 2002. p.53-66.

CAMARGO, Ana Maria de Almeida; BELLOTTO, Heloisa Liberalli (coord.). **Dicionário de terminologia arquivística**. São Paulo: Associação dos Arquivistas Brasileiros/Núcleo Regional de São Paulo: Secretaria de Estado da Cultura, 1996.

CONSELHO INTERNACIONAL DE ARQUIVOS. **ISAD(G)**: norma geral de descrição arquivística. 2. ed. Rio de Janeiro: Arquivo Nacional, 2001. (Adotada pelo Comitê de Normas de Descrição, Estocolmo, Suécia, 19-22 de setembro de 2002).

FUNDACIÓN HISTÓRICA TAVERA. **EAD Descripción Archivística Codificada**: repertorio de etiquetas. España: Fundación Histórica Tavera, 2000. (Versión 1.0).

FUNDACIÓN HISTÓRICA TAVERA. **EAD Descripción Archivística Codificada**: directrices de aplicación. España: Fundación Histórica Tavera, 2000. (Versión 1.0).

NATANYA, Pitts-Moultis; KIRK, Cheryl. **XML Black Book**. São Paulo: Makron Books, 2000.

RIBEIRO, Antonia Motta de Castro Memória. **Descrição e pontos de acesso**. 2. ed. Brasília: Ed. do Autor, 2001. (Edição revista e atualizada pelo AACR2 1998 Revision).

UNIVERSIDADE ESTADUAL DE CAMPINAS. Arquivo Central do Sistema de Arquivos. **SIARQ Pesquisa**. Disponível em: <<http://siarq02.siarq.unicamp.br/acervo/>> Acesso em: 29 jun. 2005.

UNIVERSIDADE ESTADUAL DE CAMPINAS. Sistema de Arquivos Históricos da Unicamp. **Projeto SAH – PESQUISARQH da UNICAMP**. Disponível em: <<http://www.unicamp.br/sahu/projeto.pdf>> Acesso em: 25 jun. 2005.

## Descrição Arquivística: contexto arquivístico, controle de vocabulário e o usuário

Lucia Maria Velloso de Oliveira  
Isabel Cristina Borges de Oliveira

Este trabalho pretende analisar as reflexões metodológicas adotadas após a implantação das bases de dados descritivas sobre os acervos arquivísticos sob a custódia do Serviço de Arquivo Histórico e Institucional da Fundação Casa de Rui Barbosa – FCRB.

A orientação metodológica de implantação levou em consideração fatores internos e externos ao processo, a saber: a política informacional da instituição; a existência, no mercado, de ferramentas para disponibilização das informações em ambiente Web; as normas internacionais de descrição e de catalogação; a abordagem arquivística no decorrer do processo; o usuário da informação arquivística *on-line*.

No ano de 1994, a FCRB, inserida no processo do Ministério da Cultura de desenvolvimento de sistemas de informações culturais, realiza o primeiro diagnóstico sobre as necessidades e os objetivos de suas unidades administrativas, no que se refere a *hardware*, *software*, demandas de informação e de capacitação.

A demanda enviada pelo Serviço de Arquivo em 14 de dezembro de 1995, por meio de Comunicação Interna, refere-se a: informatização do arquivo de Rui Barbosa e de seus instrumentos de pesquisa: inventários e fichários; adoção de solução tecnológica que oferecesse ao usuário a imagem digital da série iconografia e sua descrição em uma mesma tela de computador; disponibilização dos índices (assunto, onomástico e dicionário) e possibilidade de utilização de ferramenta que produzisse um inventário ilustrado como um de seus formatos de saída.

Em 1996, é elaborado um Plano de Informatização, resultado de uma série de reuniões entre chefes de alguns setores da Fundação e especialistas externos da área tecnológica. O grupo recomenda: a capacitação de pessoal; a adesão de padrões internacionais adotados nas áreas de Biblioteconomia (UNIMARC/CALCO), Museologia (recomendações do ICOM) e Arquivologia (não há referência a normas); a criação de bases de dados locais com qualidade, visando à disponibilização, ao público, dos conteúdos dos documentos; e o uso do Micro-Isis (Integrated Set of Information System), *software* desenvolvido pela Unesco, cuja versão em português era então desenvolvida pelo Instituto Brasileiro de Informação em Ciência e Tecnologia – IBICT. Vale mencionar que o IBICT foi responsável pela distribuição do Micro-Isis até julho de 1998.

A equipe interdisciplinar da FCRB, formada por arquivistas e bibliotecários, inicia uma série de visitas a instituições que já disponibilizavam informações sobre seus acervos na Web. Durante o processo de visita às instituições e na fase de pesquisa de mercado, foi possível identificar que não havia disponível um formato de processamento de informações amplamente utilizado que oferecesse uma solução única para todos os acervos, nem tampouco uma ferramenta tecnológica que abarcasse todas as questões específicas das áreas, conforme estabelecia a diretriz mestra da política de informações da Fundação.

No ano de 1998, uma equipe constituída por arquivistas, bibliotecários e museólogos é capacitada para desenvolver o projeto de desenvolvimento das bases de dados: recebe treinamento para trabalhar com o

Micro-Isis, para avaliar e definir critérios metodológicos de organização de conteúdo de bases, de recuperação da informação e de tratamento terminológico.

Uma vez definido o *software*, a equipe começou a analisar qual seria o formato mais adequado para o processo de entrada de dados. Este formato deveria oferecer ampla capacidade de intercâmbio de dados (uma demanda clara da área da Biblioteconomia) e já comprovado uso no mercado no Brasil (uma demanda institucional).

A opção institucional pela adoção do formato MARC (Machine Readable Cataloging) deveu-se a sua ampla utilização nas bibliotecas e por apresentar-se como aplicável aos materiais arquivísticos. A concepção do MARC é composta por um grupo de normas que identificam, armazenam e comunicam as informações, e sua estrutura baseia-se em campos numéricos relacionados ao conteúdo dos registros.

O processo de concepção das bases, sob o ponto de vista de quais conteúdos deveriam ser disponibilizados, e a adaptação da narrativa descritiva para o formato MARC não foram processos simples. A necessidade de uma padronização entre as bases de dados também suscitou ambivalências e críticas. A descrição arquivística de uma prática narrativa necessitava ser modulada por campos e normas. Retomaremos a discussão quando entrarmos na problemática da descrição em si.

No tocante ao levantamento de referências, normas e diretrizes, foram analisadas as normas internacionais arquivísticas: Norma Internacional de Descrição Arquivística (ISAD-G) e Norma Internacional de Registro de Autoridade Arquivística para Entidades Coletivas, Pessoas e Famílias (ISAAR-CPF), elaboradas pelo Conselho Internacional de Arquivos; e a Norma de Catalogação Anglo-American Cataloguing Rules – AACR2, que dispõe sobre a descrição bibliográfica e foi preparada pela American Library Association, British Library, Canadian Committee on Cataloguing, Library Association e Library of Congress. Vale ressaltar que a equipe do Serviço de Arquivo também conhecia a norma canadense Rules for Archival Description (RAD) e a proposta britânica Manual of Archival Description (MAD).

Apesar da capacitação da equipe para utilização do Micro-Isis, a instituição optou, ao final desse processo, por outra ferramenta (orientada para acervo bibliográfico), devido a sua utilização em outras instituições públicas e a sua boa aplicabilidade em ambiente Web.

No ano de 1999, as bases de dados são implementadas e dá-se início ao processo contínuo de disponibilização de informações sobre os acervos.

#### A metodologia e o programa descritivo

O quadro apresentava dois grandes desafios metodológicos:

1– Como oferecer ao usuário o acesso às informações por intermédio de bases de dados *on-line*, usando uma ferramenta tecnológica estruturada nos conceitos biblioteconômicos?

Essa é uma questão central, pois está diretamente relacionada à própria diferença entre a natureza dos acervos. A lógica de que um documento se encerra em si mesmo e que as informações que o traduzem se encontram no próprio documento não se aplica aos documentos de arquivo.

2 – Como representar nas bases de dados as relações contextuais de produção dos arquivos e dos documentos utilizando uma ferramenta que fragmenta as informações sobre os arquivos?

No escopo dessas questões, estava claro que não caberia mais uma visão tradicional da descrição arquivística, até então adotada no Serviço de Arquivo: como uma representação do arranjo e confundida com um de seus inúmeros produtos: os instrumentos de pesquisa.

Em um primeiro momento, de fato, ocorreu uma migração dos inventários e índices em suporte papel para as bases de dados por meio de digitação dos instrumentos de pesquisa existentes, adotando o novo formato, e assim criando-se inventários virtuais. Entretanto, à medida que mais dados foram inseridos, foi tornando-se evidente como as conexões entre os documentos e dossiês, subséries e séries se fragmentavam naquela estrutura de dados.

Algumas soluções foram mais óbvias, tais como a elaboração da base Guia (com informações gerais de cada acervo: sua proveniência, organização lógica, política de acesso e de reprodução etc.); e a elaboração de bases de dados para o controle de termos e nomes. Outras soluções não se mostraram tão imediatas.

A descrição do arranjo na base Guia não fornecia ao usuário a perspectiva do conjunto. Mesmo descrevendo as subordinações e a representação do arranjo, não poderíamos nos assegurar da produção de sentido para o usuário final. A reconstrução do contexto de produção que municiasse o usuário com o máximo de mecanismos para que pudesse realizar sua pesquisa remotamente passou a ser considerada prioritária.

Foi fundamental para a mudança do projeto o entendimento de que, a partir do lançamento das bases de dados em ambiente Web, o primeiro contato do usuário com o arquivista seria posterior ao contato com as informações disponíveis *on-line*.

Desta forma, uma de nossas funções seria, também, fornecer ao usuário instrumentos que lhe assegurassem autonomia. Assim, para suprir a negociação arquivista-usuário típica no atendimento presencial, entendemos que o resultado da pesquisa realizada para a compreensão dos arquivos, de seus titulares, de seus relacionamentos, suas interlocuções sociais e políticas deveria ficar disponível para o usuário, uma vez que se transformam em dispositivos de consulta para que ele possa atingir seus objetivos.

O programa descritivo sofre, então, uma primeira modificação, no sentido de destacar as funções de recomposição das conexões entre os documentos, e também entre os arquivos, e identificar, a partir dos indicadores da sociedade, os valores secundários dos documentos. Mas somente incorporar ao programa descritivo todos os elementos do contexto arquivístico não seria o suficiente para assegurar a comunicação com o usuário.

A alternativa metodológica adotada buscou, em seguida, ampliar a possibilidade de diálogo com o usuário através do controle de vocabulário.

A utilização do recurso da definição de pontos de acesso como resultado da indexação já fazia parte da prática descritiva do Serviço de Arquivo. Contudo, o objetivo do processo de indexação até então seguia a perspectiva biblioteconômica de representação do conteúdo do material e de seus elementos documentários.

As bases de dados de autoridade, inicialmente elaboradas para a realidade biblioteconômica, são definidas para a validação de termos e nomes, mas, em 2003, com a inserção da problemática da contextualização arquivística, são adaptadas.

Os relacionamentos entre os termos e respectivos conceitos propiciam maior compreensão de um contexto no qual os documentos estão inseridos, e essa compreensão só enriquece a metodologia de processamento da informação. Ao tratarmos de uma realidade em que as informações são veiculadas por meio de bases de dados, devemos considerar como pontos de acesso os assuntos referenciados nos conteúdos dos documentos com a mesma relevância com a qual analisamos pontos de acesso tradicionalmente mais utilizados no cenário dos arquivos. Para tal, a padronização da linguagem e a definição dos conceitos funcionam como facilitadores nesse processo de comunicação entre as bases e o usuário.

Dentre ainda o escopo das decisões metodológicas centradas em questões fundamentais da Arquivologia, destacamos casos de adequação e de total incompatibilização. É importante frisar que o processo de decisões desde a fase de elaboração das bases de dados é registrado em instrumentos específicos por área e bases. Os metadados foram organizados de acordo com o MARC e com as especificidades das áreas e de seus documentos.

A primeira questão que se coloca refere-se ao conceito de entrada principal, proposto pelo formato adotado: MARC. Os campos para preenchimento representam nome pessoal, nome corporativo, nome de evento e título uniforme. No âmbito dos arquivos, esse conceito de entrada principal ganha outra dimensão. De acordo com a abordagem multinível em uma etapa da descrição, estaremos analisando o fundo arquivístico propriamente dito e, nesse caso, a entrada principal. Conseqüentemente, o principal ponto de acesso é o nome do fundo e em sua forma direta.

No entanto, o processo de entrada de dados desses campos no caso de pessoa jurídica, ou entidade corporativa com subordinação, por exemplo, indica que o registro deve se iniciar com a identificação do país, o que é incompatível com o conceito arquivístico de produtor, assim como é incompatível com a forma de busca do usuário.

Essa discussão fica clara ao se observar a forma como a norma AACR2 orienta a padronização das entidades coletivas hierarquicamente subordinadas aos ministérios. A orientação fornecida pela norma, por exemplo, é a de dar entrada pela jurisdição, ou seja, *Brasil. Ministério da Fazenda*, e não *Ministério da Fazenda*. Caso seja pessoa física, ou seja, um presidente, governador ou prefeito, a orientação é para que a entrada seja feita também pela jurisdição, ou seja, *Brasil. Presidente (2003- : Lula)*, e não *Luiz Inácio Lula da Silva*.

O programa descritivo, como já mencionamos, está centrado na problemática da pesquisa do usuário e de como ele realiza sua busca, e, definitivamente nesse caso, o usuário não insere Brasil como possibilidade de ponto de acesso.

Assim, considerando a perspectiva do usuário e o conceito arquivístico de produtor, apesar desses indicativos da norma, mantivemos que, no caso da descrição de um fundo arquivístico, a sua forma principal de entrada seria o nome do produtor, em sua forma direta.

Essa orientação predomina nos níveis hierárquicos mais altos de descrição. Contudo, quando descemos ao nível de dossiê, mais uma vez, foi necessário estabelecer diretrizes. Duas situações predominavam:

- 1 – a reunião de documentos sobre uma mesma temática;
- 2 – a reunião de documentos de uma mesma pessoa.

Em ambos os casos poderíamos encontrar um único autor ou inúmeros autores. A opção adotada baseou-se no princípio da ordem original, no caso dos arquivos pessoais, e no princípio da proveniência, no caso dos dossiês institucionais. Se o titular de um arquivo pessoal reuniu dossiês e denominou-os com o nome do missivista, de uma instituição etc., e essa organização original foi mantida, então a mesma denominação foi adotada como ponto de acesso principal. Logo, no caso do acervo institucional, considerou-se a identificação da instituição como entrada principal.

Outra questão nos parece interessante, visto que é decorrente da própria comunidade de arquivistas. A norma internacional de descrição arquivística afirma que devemos identificar um título a cada unidade de descrição. Mesmo que o usuário não realize sua pesquisa dessa forma, o nome das séries, subséries e dossiês são indicados como título. Mas a grande questão se coloca quando tratamos de itens documentais. Na maioria das vezes, o documento arquivístico não possui título, logo, para atendermos a essa demanda da área, atribuímos um título ao documento, na maioria das vezes formulado a partir do conteúdo do documento ou do seu tipo documental. Essa obrigatoriedade pode significar uma repetição de informação, uma vez que tanto o tipo do documento quanto seu conteúdo são identificados em campos distintos do campo específico para título do material: o campo 245, ou também pode levar a uma total liberdade de criação do arquivista. Assim, optamos pela repetição.

No contexto da descrição arquivística do acervo institucional, foi necessário fazer uma diferenciação entre assunto do documento e classificação arquivística. Apesar de a classificação arquivística ser orientada por um quadro de classificação que relaciona as atividades e funções da instituição e as representa atribuindo um termo, a classificação arquivística no processo descritivo do Serviço de Arquivo recebeu um campo distinto do campo do MARC 650: assunto tópico.

Como foi dito anteriormente, a classificação arquivística está vinculada às atividades e funções da instituição, enquanto o assunto refere-se estritamente às temáticas da unidade de descrição objeto de análise. Consideramos que, para o usuário final, essa não se constitui num primeiro momento como uma questão, contudo, no processo de recuperação da informação, alguma inconsistência poderia ocorrer se essa distinção não fosse feita.

No decorrer do processo descritivo da Coleção Família Barbosa de Oliveira, muitos são os documentos que retratam a vida nas grandes fazendas de café e seus negócios. Esse conjunto de documentos é particularmente interessante para os que estudam a história econômica, social, escravidão e primeiros movimentos imigratórios. Dar visibilidade às fazendas no processo de descrição reconhece esse contexto e facilita o pesquisador. Entretanto, segundo a orientação do MARC, era necessário classificar essas fazendas em algum de seus elementos. A escolha de assunto tópico não representaria a realidade que se apresenta. As fazendas são assunto de documentos e podem ser apenas citadas.

As opções do MARC são restritas para esse caso. Apesar de não se configurarem como instituições jurídicas, essa era a melhor aproximação, visto que é necessário seguir o padrão. Como não encontramos parâmetro no Brasil, balizamos o trabalho no exemplo dos “castelos europeus”, também propriedades familiares e que são considerados na *Library of Congress* como entidades jurídicas.

No escopo do programa descritivo desenvolvido priorizando o contexto arquivístico, orientado para o usuário e para o processo de comunicação, torna-se essencial uma especial atenção ao primeiro usuário dos arquivos: o arquivista. O encarregado do trabalho de descrição deve estar comprometido com a abordagem e se qualificar no que tange à definição dos pontos de acesso, pois são as chaves para a pesquisa do usuário final. O arquivista que participa de todo o processo de descrição estabelece quais serão os pontos de acesso em todos os níveis de descrição.

No entanto, por uma questão metodológica, a padronização e normalização, cujo objetivo é a construção de vocabulário controlado, é realizada por um arquivista que não participa do processo de análise dos documentos. Esse arquivista tem como ferramenta de trabalho a descrição arquivística elaborada pelos outros membros da equipe.

A decisão da orientação do programa de descrição de inserir o arquivista-normalizador como segundo usuário das informações sobre um determinado arquivo não é deliberada. O primeiro usuário é o arquivista que organiza e elabora a descrição a partir das diretrizes do programa descritivo. O segundo usuário, com a tarefa de normalizar os pontos de acesso estabelecidos, traz consigo o estranhamento necessário para a consistência do processo. Inúmeras vezes a equipe trabalha em conjunto porque os pontos de acesso definidos podem não ser evidentes para o arquivista-normalizador, apesar de evidentes para o arquivista que os descreve.

Em um certo sentido, reproduzimos o quadro de comunicação, com suas incertezas, que o usuário final dos acervos arquivísticos experimenta diante dos instrumentos de pesquisa que elaboramos. Em um ambiente controlado, os arquivistas, então, analisam essas incertezas, muitas vezes retornam aos documentos, consultam especialistas até encontrarem a melhor forma de representação e de comunicação.

Devemos considerar que o usuário do ambiente Web chega aos arquivos não mais ansiando encontrar inúmeras caixas repletas de documentos pertencentes a um fundo arquivístico. Treinados pela realidade das bibliotecas *on-line*, já formulam suas perguntas com base em temáticas e autoria. Como poderíamos, então, considerar que o título de um fundo, ou de suas séries e subséries, representaria a complexidade de um arquivo?

No processo de comunicação das informações arquivísticas, em geral, elaboramos instrumentos de pesquisa assumindo que o usuário está familiarizado com termos como proveniência, fundos, série, etc. Há casos em que nos atemos às considerações sobre a forma do documento em detrimento às outras possibilidades de pontos de acesso que podem ser oferecidas. Enquanto existe a perspectiva de um atendimento presencial, essas questões podem ser supridas com a negociação entre o arquivista e o usuário. Contudo, ao configurarmos a pesquisa remota por meio de ferramenta para WEB, faz-se necessário olhar de maneira distinta esse processo (OLIVEIRA, 2006, p. 60).

Como a ferramenta tecnológica adotada para disponibilização das informações sobre os acervos apresenta uma informação fragmentada, sem a possibilidade de visualização das relações entre item, dossiê, subsérie, série e fundo, é necessário um investimento maior na pesquisa dos diversos contextos e, em seguida, a disponibilização de mais esse material para o usuário.

Essa foi, sem dúvida, a principal inovação metodológica no trabalho, que provocou mudanças tanto na estrutura do programa descritivo quanto nas bases de dados: ***o programa descritivo passa a incluir os resultados de pesquisa sobre o contexto arquivístico em todos os seus aspectos e, conseqüentemente, os processos de indexação e de elaboração do vocabulário controlado incluem em seus processos os resultados da pesquisa.*** Ao final, o usuário tem acesso *on-line* a todo esse material.

Mas, para tal, outra inovação foi considerada premissa metodológica: os processos de indexação e de elaboração de vocabulário controlado são inseridos no programa descritivo.

## Indexação

Nas últimas duas décadas do século XX, já podemos encontrar na literatura arquivística internacional a discussão sobre indexação. As autoras francesas Christine Pétilat e Hélène Prax (1993) afirmam que o termo “indexação” não é novo na área.

Tradicionalmente, a prática da indexação nos arquivos esteve sempre centrada na produção de índices (onomástico, geográfico, temático, cronológico) dos instrumentos de pesquisa, prevalecendo uma postura em que os índices são vistos como um instrumento secundário e, assim, não o reconhecendo como um produto resultante de uma série de discussões teóricas e definições conceituais.

No Brasil, a discussão a esse respeito é ainda incipiente, porém, nos últimos anos, o quadro vem se alterando, principalmente em função das novas tecnologias.

Durante o VI Congresso Brasileiro de Arquivologia, em 1986, foi apresentado um trabalho intitulado *O desafio da indexação nos arquivos permanentes textuais*, o qual discutia a questão da indexação. Helena Ferrez, do Museu Histórico Nacional, Jerusa Araújo e Rosely Rondineili, da Fundação Casa de Rui Barbosa, afirmaram que a motivação para a apresentação do trabalho era seu ineditismo. As autoras analisam a questão da indexação afirmando que o tema está tradicionalmente sob o domínio dos bibliotecários – realidade que, de certa forma, permanece até hoje.

Quase 20 anos depois, temos a publicação do livro de Johanna W. Smit e Nair Yumiko Kobashi, em 2003, intitulado *Como Elaborar Vocabulário Controlado para Aplicação em Arquivos*. A obra pode ser considerada a primeira publicação específica para área de arquivo cujo tema é vocabulário controlado.

Durante o VI Congresso de Arquivologia do Mercosul, ocorrido no Brasil em 2005, a questão da indexação e da recuperação da informação foram abordadas por Rose Tenório de Oliveira, da Fundação Oswaldo Cruz, e Lucia Maria Velloso de Oliveira, da Fundação Casa de Rui Barbosa.

Em 2006, o XIV Congresso Brasileiro de Arquivologia, organizado pela Associação dos Arquivistas Brasileiros – AAB, apresenta uma mesa-redonda intitulada *Normas e Padrões para a Gestão e Descrição Arquivística*, em que a recuperação da informação esteve presente em duas comunicações. A primeira sobre a NOBRADE e a segunda sobre a indexação e a descrição em arquivos.

Com um teor mais voltado para o ambiente Web, ocorreram o I e II Encontro de Bases de Dados sobre Informações Arquivísticas, respectivamente em 2005 e 2007, organizados pela Associação dos Arquivistas Brasileiros – AAB. Em ambos os encontros, a questão da recuperação da informação esteve presente.

Essa tendência também começa a aparecer nos periódicos da área, como a revista *Arquivo & Administração* e a revista *Acervo*.

A revista *Arquivo & Administração*, nas suas últimas edições, aborda a questão da indexação. Em 2006, é publicado o texto de Maria Luiza Campos, intitulado *Indexação e descrição em arquivos: a questão da representação e recuperação de informações*; em 2007, o artigo intitulado *Análise documentária de crônicas jornalísticas*, de Dilza Ramos Bastos e Maria Luiza Campos; e, em 2008, o artigo *A pesquisa em crônicas jornalísticas: a análise da representação da informação*, de Dilza Ramos Bastos, Maria Luiza Campos e Eliane Vasconcellos.

A revista *Acervo*, do Arquivo Nacional, em 2007, lança uma publicação voltada para a questão das normas e aborda a questão da indexação. O artigo de Maria José Santos merece destaque, pois dedica um item à questão da padronização de entidades coletivas. Segundo a autora (2007, p. 64), a questão é objeto de discussão na Biblioteconomia desde 1976, quando foi criado um grupo de trabalho<sup>30</sup> “que vem discutindo a uniformização, em nível internacional, da forma e da estrutura dessas entradas”.

No entanto, o estudo da temática envolvendo a indexação, construção de vocabulário controlado e a descrição arquivística, no âmbito do Serviço de Arquivo Histórico e Institucional da FCRB, se inicia no ano de 2003. No período de 2003 a 2004, o estudo concentrou-se no cenário do arquivo institucional e na elaboração de um vocabulário controlado para utilização no processo de indexação desse acervo. Já em 2004 inicia-se o estudo no nível de descrição dos outros fundos ou coleções sob a custódia do Serviço de Arquivo e, em 2005, a metodologia é aplicada na descrição dos arquivos pessoais, iniciando o processo com o arquivo Ubaldino do Amaral. No ano de 2006, conclui-se a primeira versão do Vocabulário controlado para uso no Arquivo, com 1.918 itens.

Em termos teóricos, o fundamento básico está na Teoria do Conceito, elaborada por Ingtraut Dahlberg, discípula de Wüester, que introduz no campo das Ciências Sociais os princípios da terminologia (CAMPOS, 2006). Em termos biblioteconômicos, o termo “indexação” passa a ser utilizado, por alguns autores, também para representar o termo “análise documentária”.

Silva e Fujita consideram que dentro de uma perspectiva histórica:

a Indexação tinha uma finalidade específica de construção de índices e o termo “Indexação” se ajustava perfeitamente à atividade, porém com o uso de tecnologias de recuperação da informação a necessidade de elaboração de índices foi sendo substituída pela necessidade de representação do conteúdo documentário por termo de indexação em decorrência da análise de assuntos (SILVA; FUJITA, 2004, p. 137).

Bastos (2007, p. 32), ao analisar Silva e Fujita, identifica três concepções de análise documentária. Na primeira, a análise documentária é a própria indexação, ou seja, a elaboração de índices. Na segunda, a análise documentária “é um processo maior, a indexação corresponde à fase de representação dos conteúdos identificados numa primeira fase de análise”. Na terceira, a análise documentária “é vista de forma ampla, referente à descrição física e temática, isto é, ela trata dos aspectos de forma e conteúdo do documento. Numa primeira fase, a análise descritiva ou bibliográfica trata do aspecto físico ligado ao suporte e, na segunda fase, é dado tratamento temático com fins de representação”.

Nesse sentido, a indexação pode ser compreendida como uma série de etapas visando à elaboração de índices; um processo posterior à análise do documento; ou uma atividade inserida na operação de descrição física e temática do documento.

O Serviço de Arquivo da FCRB se enquadra na terceira abordagem anteriormente mencionada.

Nas bases de autoridades “Nomes” e “Assuntos” constam os termos normalizados a partir da definição dos pontos de acesso durante o processo descritivo dos diferentes setores. Mas vamos nos ater apenas as questões de fundamento arquivístico que nortearam mudanças no desenho e nos objetivos iniciais do projeto.

---

<sup>30</sup> Criado pelo International Federation of Library Association and Institutions (IFLA).

Destacamos:

1 – A primeira alteração foi a criação das notas histórica ou biográfica. Esse item é particularmente interessante, uma vez que o preceito de se elaborar notas históricas (administrativas ou biográficas) não cabe apenas ao produtor do arquivo. Devido à orientação do programa descritivo do Serviço de Arquivo que é voltado para o contexto arquivístico, essas notas se expandem para outras pessoas jurídicas ou físicas que façam parte da rede de relacionamentos e de processos explicitados nos arquivos e que se constituem pontos de acesso para o usuário. Cabe ressaltar que, de acordo com o programa descritivo, os autores identificados durante o processo de análise documentária encontram-se inseridos nesse contexto.

2 – A segunda mudança ocorreu em função da publicação da ISAAR (CPF). Refere-se à criação de um elemento destinado a informar as datas de constituição e encerramento de uma entidade coletiva. Essa informação é crucial para a história administrativa. A orientação da norma, de colocar as datas de início e término das atividades de uma instituição, acarretou a inserção dessa informação nos cabeçalhos de nomes, independentemente se apareciam nas bases autoridades – nomes ou assuntos.

3 – A nota de relação de contexto. A principal função dessa nota para o Serviço de Arquivo é a visualização das possíveis inter-relações entre pontos de acesso de naturezas diferentes, tais como assunto e nome. Tem como principal função explicitar as relações entre os termos ligados pelos contextos histórico, familiar, social, político etc., relevantes para a compreensão do termo. É informado ao usuário que esses descritores identificados como complementares devido ao contexto podem estar relacionados com a pesquisa que está realizando. Também têm como uma de suas funções possibilitar maior autonomia do usuário durante o processo de busca nas bases de autoridades, fornecendo informações que venham a possibilitar uma ampliação em seu campo de pesquisa. São aplicados, por exemplo, quando o nome de pessoa está vinculado a um fato histórico em função do contexto de produção documental.

4 – Anteriormente, a Base Autoridade–Assuntos somente era alimentada com termos que constavam na documentação. A inserção do Serviço de Arquivo no tratamento das informações arquivísticas – mais precisamente ao iniciar o tratamento temático em arquivos pessoais – provoca uma mudança da prática até então adotada. Essa demanda surge com o processo descritivo do arquivo de Ubaldino do Amaral, em 2005, e com a necessidade de contextualizar os documentos para além de seus assuntos. Passou-se, então, a trabalhar com assuntos citados ou implícitos, utilizados como ponto de acesso na fase descritiva. Assim sendo, a base passou a disponibilizar para o usuário termos que não são explicitados nos documentos.

Tanto a terceira quanto a quarta alteração têm como fator motivador explicitar as relações contextuais que aparecem durante o processo descritivo, seja no conjunto documental, seja durante a pesquisa arquivística.

5 – Criação de um elemento que defina o termo específico da classificação arquivística. A sua forma obedece àquela com a qual se apresenta no quadro de classificação de documentos por assunto da Fundação Casa de Rui Barbosa. A criação desse elemento foi necessária visto que o código numérico de classificação, bem como os termos que compõem o plano de classificação, constitui-se em pontos de acesso no caso do acervo institucional, e as linhas da classificação arquivística comportam, muitas vezes, mais de um conceito.

As bases descritivas também sofreram alteração em função das questões anteriormente mencionadas e em busca de uma maior comunicação com o usuário final. A principal delas foi a implementação de um subcampo somente para indicar a categoria de inserção do termo no contexto da descrição. Como disse Oliveira (2007, p. 49), além dos assuntos explícitos, também temos aqueles que são citados e implícitos ao documento.

Os citados são mencionados no próprio documento, e os implícitos são identificados no decorrer do trabalho de pesquisa do contexto arquivístico. Segundo Thomassen, o contexto arquivístico são:

Todos os fatores ambientais que determinam como documentos são gerados, estruturados, administrados e interpretados. Os fatores ambientais que determinam diretamente os conteúdos, formas e estrutura dos registros podem ser diferenciados em contexto de proveniência, contexto administrativo e contexto de uso. Estes fatores são, cada um a seu tempo, determinados pelo contexto sócio-político, cultural e econômico (THOMASSEN, 2006, p. 10).

Segundo Cook (2001), o usuário dos arquivos deve fazer parte do processo de comunicação. Suas necessidades devem ser observadas e os instrumentos de recuperação da informação devem ser elaborados com essa perspectiva: atender ao usuário. A descrição dos acervos e seus pontos de acesso não podem mais ser elaborados no âmbito da visão dos arquivistas, forçando os usuários a se adaptarem.

Atualmente, nas bases descritivas estão disponíveis 19.726 registros sobre fundos e coleções e respectivas séries, subséries, dossiês e itens (quando for o caso), e nas bases autoridades existem 4.517 termos e nomes padronizados<sup>31</sup>. A experiência acumulada indica que o esforço para a ampliação de mecanismos de comunicação com o usuário por intermédio da elaboração de um programa descritivo centrado no usuário não é suficiente.

Apesar da implementação de campos de conteúdo com vistas a minimizar o efeito de fragmentação da representação dos arquivos, de reconstruir o contexto arquivístico, apesar dos esforços na construção de dispositivos de comunicação com o usuário, precisamos que os dispositivos tecnológicos viabilizem esse processo. Os formatos e as normas que buscam uma padronização devem dialogar com os usuários e como os usuários executam suas pesquisas. O reconhecimento das diferenças entre os programas descritivos e práticas arquivísticas devem ser inseridos nos processos. Há mais de dez anos o mercado não oferecia uma solução tecnológica que fosse orientada para a problemática dos acervos arquivísticos e tampouco do contexto arquivístico. Dispomos agora de novos recursos; contudo, o arquivista precisa estranhar suas práticas, centrar-se no usuário e no processo de comunicação em seus distintos aspectos.

## Referência Bibliográficas

BASTOS, Dilza Ramos; CAMPOS, Maria Luiza de Almeida. Análise documentária de crônicas jornalísticas. **Arquivo&Administração**. Rio de Janeiro, v.6,n.1, p.27-42. jan/jun, 2007.

BRASIL, Maria Irene et. al. Vocabulário sistematizado: a experiência da Fundação Casa de Rui Barbosa. Participação de Beatriz Amaral de Salles Coelho, Conchita Sprenger, Maria Luiza de Almeida Campos, Dilza Ramos Bastos. Rio de Janeiro: FCRB, 2002. 12 p. Trabalho apresentado na sessão de comunicações livres e pôsteres, do **1º Congresso Internacional de Arquivos, Bibliotecas, Centros de Documentação e Museus** (São Paulo, 17 a 22/3/2002).

CAMPOS, Maria Luiza de Almeida. **Linguagem documentária**: teorias que fundamentam sua elaboração. Niterói ;RJ: EDUFF, 2001.133p.

CINTRA, Ana Maria et alli. **Para entender as linguagens documentárias**. 2. ed. rev. ampl. São Paulo: Polis, 2002. 130p.

---

<sup>31</sup> Dados de fevereiro de 2009.

- CONSELHO INTERNACIONAL DE ARQUIVOS. **ISAD (G)**: norma geral internacional de descrição arquivística. 2. ed. Rio de Janeiro: Arquivo Nacional, 2003. 110 p.
- CONSELHO INTERNACIONAL DE ARQUIVOS. **ISSAR (CPF)**: Norma internacional de registro de autoridades arquivísticas para entidades coletivas, pessoas e famílias. 2. ed. Rio de Janeiro: Arquivo Nacional, 2004. 99p.
- CONSELHO NACIONAL DE ARQUIVOS. Ata da 39ª Reunião, 05 de dezembro de 2005. Disponível em <<http://www.conarq.arquivonacional.gov.br>>
- \_\_\_\_\_. Ata da 42ª Reunião, 01 de agosto de 2006. Disponível em <<http://www.conarq.arquivonacional.gov.br>>
- \_\_\_\_\_. Ata da 44ª Reunião, 05 de dezembro de 2006. Disponível em <<http://www.conarq.arquivonacional.gov.br>>
- \_\_\_\_\_. NOBRADE: Norma Brasileira de Descrição Arquivística. Rio de Janeiro: Arquivo Nacional, 2006. 124p.
- COOK, Michael. Description standards: the struggle towards the light. **Archivaria**, v. 34, Summer, p. 50-57, 1992.
- COOK, Terry. Fashionable nonsense or professional rebirth: postmodernism and the practice of archives. **Archivaria**, n.51, Spring, 2001
- DAHLBERG, Ingetraut. Teoria do Conceito. **Ciência da Informação**, Rio de Janeiro, 7(2):1001-107, 1978.
- DUCHEIN, Michel. O Respeito aos fundos em arquivística: princípios teóricos e problemas práticos. **Arquivo & Administração**, Rio de Janeiro, 10-14 (1): 14-33, abr. 1982/ago.1986.
- DURANTI, Luciana. **Diplomatica**: usos nuevos para uma antigua ciência. Traducción de Manuel Vázquez. 1ed. Cordoba (Argentina), 1995. 293p.
- Dossiê 5.8.1 Programa de Informatização de Dados 1996-1999 AI 124
- Dossiê 5.8.1 Programa de Informatização de Dados. Sistema de informatização dos acervos da FCRB 1995-2000 AP74
- EASTWOOD, Terry. The theory and practice of description in the digital era. Conferência proferida no **II Encontro de Bases sobre Informações Arquivísticas**. Associação dos Arquivistas Brasileiros. 15 e 16 de março de 2007.
- FCRB. Metadados: base descritivas sobre arquivos pessoais e arquivo institucional (APES e AFCRB).FCRB: Rio de Janeiro, 1998 - .
- FCRB. Metadados base autoridades assunto FCRB: Rio de Janeiro, 1998 - .
- FCRB. Metadados base autoridade nomes FCRB: Rio de Janeiro, 1998 - .
- FERREZ, Helena Dodd; ARAÚJO, J. G.; RONDINELLI, R. C. O desafio da indexação nos arquivos permanentes textuais. **Revista do Patrimônio Histórico e Artístico Nacional**, Rio de Janeiro, v. 22, p. 191-192, 1987.
- FOX, Michael. **Structure and meaning in tools for resource discovery**. 2003. Disponível em: [www.naa.gov.au/recordkeeping/rkpubs/fora/ICA\\_oct03/resource\\_discovery.pdf](http://www.naa.gov.au/recordkeeping/rkpubs/fora/ICA_oct03/resource_discovery.pdf) Acessado em: 25 de mar. de 2005.
- GONZÁLEZ DE GÓMES, Maria Nélide. Novas fronteiras tecnológicas das ações de informação: questões

e abordagens. **Ciência da Informação**, Brasília, v. 33, n. 1, p. 55-67, jan./abr.2004.

JARDIM, José Maria. O acesso à informação arquivística no Brasil: problemas de acessibilidade e disseminação. Caderno de Textos. **Mesa Redonda Nacional de Arquivos**, 1999. Rio de Janeiro: Arquivo Nacional, 1999.

KOBASHI, Nair Yumiko; TÁLAMO, Maria de Fátima Gonçalves Moreira. Informação: fenômeno e objeto de estudo da sociedade contemporânea. **TransInformação**. Pontifícia Universidade Católica de Campinas, Campinas, SP, v.15, Edição Especial, p.7-21, set./dez. 2003.

LANCASTER, F. W. **Indexação e Resumos**: Teoria e Prática. Tradução Antônio Agenaor; 2ª ed. rev. amp. Brasília, DF: Briquet de Lemos/Livros, 2004. 452 p.

MC LAUGHILIN, Derina. **Because that's the sort of thing that users do**: a paper on rules for successful on line access systems.2003. Disponível em: [www.naa.gov.au/recordkeeping/rkpubs/fora/ICA\\_oct03/what\\_users\\_do.pdf](http://www.naa.gov.au/recordkeeping/rkpubs/fora/ICA_oct03/what_users_do.pdf) Acessado em: 25 de março de 2005

MACNEIL, Heather. Picking our text: archival description, authenticity, and the archivist as editor. **The American Archivist**, v. 68, p. 264-278, Fall/Winter, 2005.

OLIVEIRA, Lucia Maria Velloso de. Descrição arquivística: seu papel no processo de comunicação dos conteúdos informacionais dos acervos arquivísticos. In: ENCONTRO DE BASES DE DADOS SOBRE INFORMAÇÕES ARQUIVÍSTICAS, 2., 2007, Rio de Janeiro. **Anais...** Rio de Janeiro: Associação dos Arquivistas Brasileiros, 2007. p.45-50

\_\_\_\_\_. **O usuário como agente no processo de transferência dos conteúdos informacionais arquivísticos**. Orientador: José Maria Jardim. Dissertação (Mestrado em Ciência da Informação)-UFF/IBICT, Rio de Janeiro, 2006.

PÉTILLAT, Christine; PRAX, Hélène. Les Archives Contemporaines ou l'arrivée du flux au quotidien. In: **La pratique archivistique française**, Paris, Archives Nationales, 1993, p. 231-312.

RIBEIRO, Antonia Motta de Castro Memória. **Catálogo de recurso bibliográficos: AACR2r em Marc 21**. 3 ed., rev. e ampli., Brasília: Ed. do Autor, 2006.

Schellenberg, Theodore R. **Arquivos públicos e privados**: arranjo e descrição. Tradução de Manoel A. Wanderley. 2ed., Rio de Janeiro: Editora FGV, 1980. 386 p.

SILVA, Maria dos Remédios da; FUJITA, Mariângela Spotti Lopes. A prática de indexação: análise da evolução de tendências teóricas e metodológicas. **TransInformação**. Pontifícia Universidade Católica de Campinas, Campinas, SP, v.16, n. 2, p.133-162, maio/ago.2004.

SMIT, Johanna Wilhelmina, KOBASHI, Nair Yumiko. **Como Elaborar Vocabulário Controlado para Aplicação em Arquivos**. São Paulo: Arquivo do Estado, Imprensa Oficial do Estado de São Paulo, 2003. 53 p.

SOUSA, Renato Tarciso Barbosa de. A Classificação como função matricial do que-fazer arquivístico. In: **Arquivística temas contemporâneos**, 2007, p.79-162.

THOMASSEN, Theo. **Making archives accessible: increasing pluriformity in pursuing illusions**. Arkistoyndistyksen Julkaisuja, Helsing, 9, p. 31-68, 2004.

\_\_\_\_\_. Uma primeira introdução à Arquivologia. **Revista Arquivo e Administração**. Associação dos Arquivistas Brasileiros, Rio de Janeiro, RJ, v. 5, n. 1, p. 5-16, jan./jun. 2006.

## A Normalização da Descrição Arquivística: a ISAD(G) e os arquivos permanentes

Flávia Carneiro Leão

As primeiras iniciativas de sistematização da descrição remontam ao século XVI. Bonal Zazo comenta que, Felipe II, em sua Instrução de 1588, para o governo do Arquivo de Simancas, ao falar da realização do livro dos inventários, afirmava que “as pessoas que, dali em diante, sucederem o dito cargo de arquivista, deverão ter o particular cuidado de fazer o mesmo, cada um em seu tempo” (Zazo, 2000, p. 56).

Wendy M. Duff e Kent M. Haworth afirmaram em “*The Reclamation of Archival Description: the Canadian Perspective*” que, historicamente os arquivistas sempre usaram algum tipo de sistema para descrever os materiais que chegavam aos seus arquivos, mas até há pouco tempo, não existiam normas para a descrição arquivística. Independentemente do país, ou da instituição em que atuaram, ao invés de usarem normas, eles sucessivamente registraram informações usando manuais escritos por outros arquivistas ou por eles próprios. Estes manuais eram parte integrante no desenvolvimento de abordagens sistemáticas para a descrição e deveriam ser vistos como valiosos precursores para a elaboração de normas e manuais nacionais (Duff e Haworth, 1990, p. 27).

Neste sentido, diversos são os autores<sup>32</sup> a concordarem que o interesse pela normalização descritiva não é algo recente.

Todavia, em 1992, Maria Guercio ainda perguntava aos participantes do XXI Congresso Internacional de Arquivos: “Qual a razão concreta para os arquivistas terem dificuldades em criar e aceitar normas?” (Guercio, 1994, p. 201).

A resposta a esta questão é o que tentaremos abordar, a partir da análise de cinco questões que, na literatura, surgem como justificativas para a resistência à normalização, e que nos parecem derivar, por um lado, da singularidade dos arquivos e, por outro, da teoria arquivística.

A primeira questão a gerar resistência à normalização da descrição talvez possa ser, em parte, atribuída ao caráter único dos arquivos, o que tem levado os arquivistas à criação e à utilização de normas específicas em cada arquivo<sup>33</sup>.

Esta postura, que pode ser confirmada pela diversidade de práticas descritivas verificadas em instituições arquivísticas nacionais e internacionais, evidencia a concepção de que cada arquivo considera seu acervo como sendo único. E, de fato, eles o são. Todavia, é preciso ponderar que, na medida em que o tratamento conferido a um determinado acervo utiliza práticas descritivas específicas e particulares, o desenvolvimento de um trabalho cooperativo, que viabilize o compartilhamento de esforços, de recursos, e o intercâmbio de dados, fica comprometido ou, pelo menos, dificultado.

Parece-nos que tal prática tem sido uma das responsáveis pelo isolamento dos arquivos e pelo tardio desenvolvimento de propostas de normalização, de trabalhos cooperativos etc.

---

<sup>32</sup> Cook, Heredia, Fonseca, Weber, Bellotto, Stibbe, Feliu, Hagen, Cunningham.

<sup>33</sup> González Pedro, La normalización en la descripción de archivos: un movimiento internacional. Madrid, 1992. p. 83-90. Citado por Zazo, 2000, p. 57.

Acreditamos, entretanto, que o caráter único dos acervos não tenha sido o único fator a determinar o alheamento dos arquivistas em relação à normalização, mas tendo influído decisivamente no desenvolvimento de normas, impôs um desafio à padronização.

O segundo ponto a retardar a normalização nos arquivos diz respeito à influência dos usuários no desenvolvimento das práticas normalizadas.

Durante muito tempo, o protótipo do usuário de arquivo foi o do pesquisador erudito e com amplo conhecimento do assunto estudado. Este usuário, “caracterizado pela sua laboriosidade e paciência na pesquisa, em geral não se fiava na informação ou exaustividade dos instrumentos de pesquisa e, geralmente, preferia acessar diretamente a documentação arquivada, sem a intermediação dos inventários, catálogos etc” (Tarraubella i Mirabet, 1998, p. 4).

Contudo, nas últimas décadas, com o aumento do número de pessoas a consultar os arquivos, o perfil dos usuários vem se diversificando. Tal mudança tem demandado não apenas a melhoria dos procedimentos de descrição, que precisam atender às novas demandas informativas, mas, sobretudo a uniformização das práticas.

O terceiro fator a influir sobre a normalização dos arquivos tem sido a diversidade existente no âmbito da Arquivística, “diversidade que também está refletida nas estruturas de organização e funcionamento dos arquivos” (Feliu, 1995, p. 26) e na existência de diferentes realidades históricas, culturais e de formação.

É preciso admitir que a existência de tal diversidade descaracteriza uma das principais funções dos arquivos, que é a sua função social. Se houvesse uma normalização mínima dos procedimentos, prover acesso a um público geral e diversificado, ou seja, ao cidadão, seria um objetivo possível de ser alcançado, ao passo que o atendimento a demandas de acesso particulares inviabiliza, por definição, qualquer perspectiva de normalização.

A quarta causa a refletir negativamente na utilização e na difusão de práticas normativas tem sido a forma como os arquivistas tradicionalmente têm trabalhado que, em geral, é “isolada, independente, e com critérios próprios adaptados às suas necessidades específicas” (Cox, 1994, p. 167).

Não obstante, no que se refere a este aspecto, alguns fatores têm contribuído para o rompimento dessa postura e, conseqüentemente, para a introdução de práticas cooperativas – de práticas que demandam o desenvolvimento de normas. Entre eles está o aumento do número de profissionais na área, a ampliação dos cursos de formação, a projeção que os arquivos vêm obtendo na sociedade, a informática e, sobretudo a influência de disciplinas afins.

Por fim, é oportuno destacar que, a criação de normas técnicas necessita de respaldo econômico, e a Arquivística é uma área que, conhecidamente, dispõe de poucos recursos (Cox, 1994, p. 173).

Em resumo, podemos dizer que nem sempre a normalização arquivística tem integrado as prioridades da Área, surgindo apenas recentemente como uma questão merecedora de atenção, particularmente em função da informática, através da qual a comunicação e a conexão entre arquivos se tornaram possíveis.

## Caracterização das normas

De acordo com o léxico, “norma é aquilo que se estabelece como base ou medida para a realização ou a avaliação de alguma coisa” (Ferreira, 1987, p. 977). Tal definição evidentemente tem um caráter genérico<sup>34</sup>, o que nos impôs a necessidade de buscar critérios que propiciassem uma melhor aproximação do problema.

García Gutiérrez propõe cinco critérios para a distinção de normas: o caráter jurídico (público ou privado), o conteúdo, o âmbito de aplicação, os destinatários e o grau de profundidade na normalização (García Gutiérrez, 1985, p. 59 e 60).

No caso das normas de descrição arquivística, elas também podem ser analisadas segundo a sua origem e, neste sentido, “elas podem ser externas (oriundas de áreas externas à Arquivística) e internas, que são as criadas especificamente para o desenvolvimento de trabalhos arquivísticos” (Zazo, 2000, p. 61). Sendo assim, normas como MARC AMC, MAD, RAD, ISAD(G) ou APPM<sup>35</sup>, são normas de caráter interno, enquanto o AACR2<sup>36</sup>, mesmo não sendo uma norma *stricto sensu*, mas um manual de catalogação que apresenta regras para a descrição de materiais, pode ser considerado como uma “norma” externa, uma vez que é originária da área da Biblioteconomia.

De um outro ponto de vista, podemos dizer que há normas “que permitem normalizar a estrutura dos dados, as que normalizam o conteúdo de cada um dos elementos de descrição, as que padronizam os instrumentos de pesquisa e as que instruem a construção de linguagens de indexação” (Zazo, 2000, p. 59). Estes aspectos são constitutivos de normas de cunho profissional ou técnico, que têm como característica principal sua função normativa, pois regulamentam as práticas e os procedimentos do trabalho arquivístico.

David Bearman<sup>37</sup>, no que diz respeito à profundidade alcançada pela normalização, estabelece quatro tipos de normas:

- a) as que dispõem sobre sistemas de informação, e que definem as funções de cada componente dentro de um sistema de informação global, seja de um arquivo ou de uma rede. No caso de banco de dados referenciais bibliográficos, por exemplo, a norma ISO 2709 é uma referência tida como fundamental;
- b) aquelas que definem a estrutura dos elementos informativos que devem compor um sistema de informação. Este tipo de norma regulamenta o “formato”, ou seja, os dados que devem ser incluídos em um sistema, mas não dispõe sobre o seu conteúdo.
- c) As normas de conteúdo de dados, que são as que estabelecem que informação introduzir em cada elemento da estrutura de um sistema, como o APPM e a ISAD(G), por exemplo; e por fim,
- d) existem as normas sobre o valor dos dados, que são as que estabelecem listas de nomes, códigos alfanuméricos, tabelas de termos ou outros dados utilizados em elementos específicos da estrutura de

---

<sup>34</sup> O objetivo desta discussão é abordar as normas em sentido *latu* e não jurídico.

<sup>35</sup> O MARC AMC (*Machine-Readable Cataloging Record for Archival and Manuscripts Control*) foi adaptado, a partir do formato MARC, pela *Society of American Archivists*, SAA, em conjunto com a Biblioteca do Congresso dos Estados Unidos.

O MAD (*Manual for Archival Description*) foi desenvolvido na Universidade de Liverpool com a participação da *Society of Archivists* e sob a direção de Michael Cook.

A RAD (*Rules for Archival Description*) foi desenvolvida pelo *Bureau of Canadian Archivists* e pelo Conselho Canadense de Arquivos.

A ISAD(G) (*International Standard of Archival Description*) foi desenvolvida pelo Conselho Internacional de Arquivos, CIA.

O APPM (*Archives, Personal Papers, and Manuscripts*) foi desenvolvido por Steven Hensen, com o apoio da *Society of American Archivists* e da Biblioteca do Congresso Americano.

<sup>36</sup> AACR2: *Anglo-American Cataloguing Rules*, 2ª edição.

<sup>37</sup> BEARMAN, David. *Description standards: a framework for action*. *The American Archivist*. v. 52, n. 4, p. 10, 1989. Citado por Zazo, 2000, p. 61.

dados (Zazo, 2000, p. 61). Um exemplo a ser citado é a norma ISO 3166, que dispõe sobre o uso de abreviaturas para os nomes de países.

Estas caracterizações, embora sucintas, indicam que, para normalizarmos dados com o objetivo de criar sistemas de informação arquivística, “devemos ter em conta a dificuldade existente não apenas no nível do conteúdo dos dados, mas também no da estrutura desses dados” (Guercio, 1994, p. 203). Assim, a primeira iniciativa de normalização internacional da descrição arquivística, a ISAD(G), pode ser caracterizada como uma norma voltada para a definição dos elementos constitutivos de um sistema e, também, do conteúdo destes mesmos elementos.

## As primeiras normas de descrição

A iniciativa de normalização descritiva, que teve como resultante a criação da primeira norma de descrição internacional, se iniciou no começo dos anos 80.

Em 1982, Steven Hensen adaptou vários capítulos da segunda edição das *Anglo-American Cataloguing Rules*, AACR2, para a descrição de documentos de arquivo. O resultado deste trabalho foi a obra *Archives, Personal Papers and Manuscripts*, APPM, que, respaldada pela Sociedade dos Arquivistas Americanos, SAA, e pela Biblioteca do Congresso Americano, LC, alcançou natureza normativa nos Estados Unidos.

Poucos anos depois, em 1986, Michael Cook publicou na Grã-Bretanha um manual de descrição com características similares ao de Hensen, embora com um conteúdo mais detalhado, o *Manual of Archival Description*, MAD.

Seguindo um processo semelhante ao do trabalho de Hensen, o MAD surgiu da iniciativa pessoal de seu autor, tendo se convertido, posteriormente, em fonte normativa. Entretanto, a obra de Cook teve uma difusão menor que a de Hensen, mesmo que sua influência teórica tenha sido mais ampla.

No final da década de 1980, os arquivistas canadenses começaram a publicar a terceira grande obra normativa de caráter nacional: as *Règles pour la Description des Documents d'Archives*, RAD, que, assim como o manual de Cook, possuía um corpo extenso e detalhado de instruções normativas e, como o manual de Hensen, sofria uma forte influência biblioteconômica, fruto de sua origem comum: as Regras de Catalogação Anglo-Americanas. Sem dúvida, a singularidade das RAD residia no fato delas terem sido a primeira obra que, embora tivesse nascido com todas as características formais de qualquer “norma”, era o resultado de um consenso profissional, criada por um organismo nacional, com um forte caráter técnico, e de livre aplicação.

O desenvolvimento normativo norte-americano, canadense e britânico foi decisivo para que o Conselho Internacional de Arquivos, CIA, no início da década de 1990, envidasse esforços para a elaboração de uma norma internacional de descrição arquivística e seguramente influenciou a configuração da norma produzida pelo CIA, a ISAD(G).

Entretanto, há que se observar que a principal referência subjacente às obras que fundamentaram a ISAD(G) é originária da Biblioteconomia e não da Arquivística, sendo um manual de catalogação, o AACR2.

As Regras de Catalogação Anglo-Americanas, 2ª edição –AACR2– propõem uma estrutura de dados,

estabelecem os elementos de descrição e como descrevê-los, mas sendo prescritivas e pouco normativas, em diferentes momentos oferecem opções a serem utilizadas para a descrição, como o registro do “nome mais conhecido”, por exemplo, que são bastante discutíveis, além de pensar os documentos isoladamente, ou seja, de acordo com a lógica da Biblioteconomia. Esta característica contraria, a nosso ver, a concepção de qualquer norma, que deve padronizar procedimentos com o objetivo de reduzir variantes e proporcionar meios consistentes e eficazes para o registro de informações, evitando opções e “interpretações” pessoais ou locais.

Neste sentido, normalizar significa “estabelecer, em relação a problemas existentes ou potenciais, prescrições destinadas à utilização comum e repetitiva com vistas à obtenção do grau ótimo de ordem em um dado contexto” (ABNT, s.d.), e o AACR2 não se propõe a isto. Além disto, do ponto de vista arquivístico, não estabelece relações entre os documentos, não trata dos conjuntos documentais, mas aborda os documentos como unidades isoladas e independentes, não contemplando os preceitos fundamentais da Arquivística. Afinal, sua origem e destinação são as bibliotecas.

Por esta razão, causa estranheza que um código de catalogação biblioteconômico tenha sido a principal referência a fundamentar o desenvolvimento de uma norma de descrição arquivística.

## A normalização internacional

A publicação dos manuais de descrição americano, britânico e canadense parecem ter evidenciado a necessidade da elaboração de normas gerais para a descrição de documentos. Todavia, uma questão desafiaria os profissionais incumbidos do desenvolvimento de um padrão descritivo internacional: a dificuldade em unificar as práticas descritivas de diferentes países, as quais apresentavam metodologias de trabalho, instrumentos de pesquisa e terminologias distintas entre si.

Por estas razões, e para que uma norma de descrição pudesse contar com o consenso geral, era apropriado que esta proposta fosse feita por um organismo internacional, como o Conselho Internacional de Arquivos e, sendo assim, o Conselho promoveu o desenvolvimento de um processo institucional, em nível mundial, com a participação de diversos órgãos arquivísticos nacionais.

O início desse processo se deu em 1988, em reunião promovida pelos Arquivos Nacionais do Canadá e pelo CIA, em Ottawa, na qual foi proposta a criação de um grupo de trabalho para a discussão de normas internacionais para a descrição arquivística. Esta sugestão foi acolhida pelo CIA que, em parceria com a UNESCO, passou a organizar uma série de encontros de trabalho para discutir a questão, como o de 1989, em Paris, onde foi realizada uma reunião consultiva para a elaboração de um plano de ação.

Em setembro de 1990, na Alemanha, ocorreu a primeira reunião da Comissão *Ad Hoc* de Normas de Descrição que esboçou uma *Declaração de Princípios*<sup>38</sup>.

“Disseminado internacionalmente para críticas e comentários, esse documento foi revisto em Madri, em janeiro de 1992, passando a se constituir como *Statement of Principles Regarding Archival Description*. Na mesma reunião foi elaborada a primeira versão da ISAD(G). [...] Em 1993, em Estocolmo, a Comissão aprovou o texto da ISAD(G), encaminhando-o para publicação, sendo decidido que num prazo de cinco anos seria novamente revisto [...]” (Fonseca, 2001, p. 4).

---

<sup>38</sup> *Declaration des principes*: documento emitido pelo CIA que define as bases da Norma Internacional de Descrição Arquivística, ISAD(G) (RUFÍ I PAGÈS, 1995, p. 19).

A revisão foi concluída na terceira reunião plenária do Comitê, em Estocolmo, e oferecida para publicação no início de 2000, tendo em vista o XIV Congresso Internacional de Arquivos que ocorreria em Sevilha, em setembro do mesmo ano (CIA, 2000, p. 9).

Embora os primeiros resultados produzidos pelo CIA tenham chegado às mãos dos arquivistas brasileiros em torno de 1995, foi em 2000, com a publicação da tradução brasileira da ISAD(G)<sup>39</sup> pelo Arquivo Nacional, que passamos a dispor de uma versão em português, que permitiu o início das primeiras apreciações da norma.

Como dissemos anteriormente, o processo de normalização descritiva internacional se iniciou em 1988, no encontro de especialistas em Ottawa, e uma das resoluções foi a de que o CIA deveria “avaliar as normas e diretrizes existentes para determinar sua validade para a descrição arquivística” (Rufí i Pagès, 1995, p. 2).

Seguindo este indicativo, o grupo encarregado de elaborar o novo padrão apresentou um plano de trabalho que previa o exame dos padrões existentes em diferentes países. Desta maneira, os membros da Comissão *Ad Hoc* de Normas de Descrição do CIA optaram por analisar “determinadas obras, as quais aparecem no primeiro documento elaborado: o *Statement of principles regarding archival description*” (Rufí i Pagès, 1995, p. 2).

Este documento pretendeu fixar os princípios que serviriam de base ao projeto em curso, elegendo como ponto de partida os já mencionados APPM<sup>40</sup>, RAD<sup>41</sup> e MAD<sup>42</sup>. Três obras surgidas nos anos 80 e que, estimuladas pelo surgimento das novas tecnologias, tiveram o objetivo de padronizar a descrição dos documentos de arquivo em seus países. Mas, vale lembrar, todas originadas de um manual de catalogação biblioteconômico.

## A descrição arquivística na ISAD(G)

A ISAD(G) tem como objetivo o estabelecimento de diretrizes gerais para a preparação de descrições arquivísticas, devendo ser utilizada em conjunto com as normas nacionais (CIA, 2000, p. 1). Tais diretrizes compreendem a identificação do nível de descrição dos documentos, do contexto no qual estão inseridos, das suas características extrínsecas e intrínsecas, e dos seus conteúdos.

Na prática, a Norma estabelece uma estrutura composta por 7 áreas de descrição, subdivididas em 26 elementos. Todos os elementos de descrição compreendem: a) o nome do elemento, b) seu objetivo, c) a(s) regra(s) a ele aplicável(is) e, d) exemplos que ilustram a aplicação da(s) regra(s). Os elementos se encontram distribuídos nas seguintes áreas: Área de identificação, Área de contextualização, Área de conteúdo e estrutura, Área de condições de acesso e uso, Área de notas e Área de controle da descrição.

Diferentemente do que se observa no tocante à descrição arquivística em suas origens, o conceito de descrição adotado pelos responsáveis pela elaboração da ISAD(G) é o da “elaboração de uma acurada representação de uma unidade de descrição e de suas partes componentes, caso existam, por meio da extração, análise, organização e registro de informação que sirva para identificar, gerir, localizar e explicar documentos de arquivo e o contexto e o sistema de arquivo que os produziu. Este termo também se aplica ao produto

---

<sup>39</sup> Acessível em: <http://www.arquivonacional.gov.br/pub/virtual/>.

<sup>40</sup> HENSEN, Steven. **Archives, personal papers and manuscripts**: a cataloguing manual for archival repositories, historical societies and manuscripts libraries. 2 ed. Chicago: Society of American Archivists, 1989.

<sup>41</sup> PLANNING COMMITTEE IN DESCRIPTIVE STANDARDS. **Rules for archival description**. Ottawa: Bureau of Canadian Archivists, 1990.

<sup>42</sup> COOK, Michael e PROCTER, Margaret. **A Manual of Archival Description**. 2. ed. Aldershot: Gower, 1989.

desse processo” (CIA, 2000, p. 4). Tal definição, que rompe com algumas das acepções anteriores, acrescenta aspectos já admitidos pela prática, mas ainda não incorporados pela teoria.

Dentre os aspectos resultantes das concepções adotadas pela ISAD(G) estão: o aumento do número de elementos informativos, a dissociação entre os conceitos de descrição e de instrumentos de pesquisa e o respeito ao princípio de proveniência.

Com respeito ao aumento do número de elementos informativos, é evidente que toda informação que favoreça a identificação da documentação, que explique o seu contexto de produção e que facilite sua compreensão, facilitando o seu uso, deve ser considerada como informação descritiva e, como tal, pode ser incluída na descrição. Assim, depreende-se que, naturalmente, a maioria das normas tende a ampliar o rol de informações a serem descritas.

Quanto à dissociação entre o conceito de descrição e os instrumentos de pesquisa, eles podem ser um dos diferentes resultados decorrentes da atividade descritiva, mas o objetivo da descrição não é o de elaborar instrumentos de pesquisa, mas o de elaborar representações precisas de unidades de descrição; representações que em um contexto automatizado, podem ter diferentes formatos de apresentação e dar origem, portanto, a diferentes instrumentos de pesquisa.

Assim, podemos afirmar que a conceituação da ISAD(G) tem um caráter complementar ou adicional ao da aceção relacionada à elaboração de instrumentos de pesquisa e, neste sentido, “uma descrição poderia ser feita sem se recorrer diretamente ao material descrito, mas a partir dos instrumentos de pesquisa já elaborados” (Rufi i Pagès, 1995, p. 2).

No que se refere à inclusão do princípio de proveniência no conceito de descrição, esta não é uma prerrogativa do conceito adotado pela ISAD(G), pois as normas nacionais que a antecederam, notadamente as RAD e o MAD, também estabeleciam a relação entre a descrição e este princípio arquivístico; do mesmo modo que a inclusão de informação relativa ao contexto de produção documental.

Ao analisar a aplicação do princípio de proveniência pela ISAD(G), Antonia Heredia afirma que “tal princípio pode ser satisfeito simplesmente com o reconhecimento da vinculação de origem, institucional ou pessoal” e, sem se aprofundar na questão, acrescenta que, para tanto “se deve reconhecer dois níveis de aplicação”. Um primeiro, que se refere “à vinculação imediata ou proveniência, identificada com a instituição ou pessoa geradora, pois definitivamente, do fundo em questão”. Um segundo, que tem a ver necessariamente com “uma vinculação imediata, mais próxima, à unidade administrativa ou órgão ou à função ou atividade definida na estrutura da instituição, conhecida como produtor ou sujeito” (Heredia Herrera, 2003 b, p. 7). Neste sentido, é importante distinguir a proveniência do produtor, para uma perfeita identificação na descrição arquivística, pois o “princípio de proveniência configura a metodologia arquivística em suas manifestações essenciais que são a classificação, a ordenação, a descrição” (Heredia Herrera, 2003 b, p. 8).

De acordo com esta perspectiva, é possível reconhecer que o mérito da ISAD(G) reside no fato dela estabelecer a obrigatoriedade desta prática, e de reafirmar a proveniência como um princípio fundamental da Arquivística.

Parece-nos, portanto, que ao definir o acesso ao conteúdo dos documentos de forma contextualizada como o principal objetivo da descrição arquivística, a ISAD(G) promoveu a institucionalização de um conceito.

Em sua introdução, a ISAD(G) determina que, “se o fundo como um todo estiver sendo descrito, ele deverá ser representado numa só descrição. (...) Se for necessária a descrição das suas partes, estas podem ser descritas em separado. (...) A soma total de todas as descrições assim obtidas, ligadas numa hierarquia, representa o fundo e as partes para as quais foram elaboradas as descrições. Tal técnica de descrição é denominada descrição multinível” (CIA, 2000, p. 7).

A descrição multinível não constituiu exatamente uma novidade para os arquivistas, pois a maioria dos manuais de descrição que serviram como base para a elaboração da ISAD(G) de algum modo fazia referência à descrição multinível. “O APPM não mostra uma atenção particular em relação à complexidade da estrutura dos arquivos, ao passo que o MAD e as RAD (...) têm apontado, na teoria dos níveis de descrição, o método mais adequado para dar conta (...) das modalidades de ordenação dos fundos e das relações existentes entre o fundo e as suas partes” (Vitali, 1994, p. 312).

Embora a literatura consultada acuse a referência à técnica de descrição multinível nas obras que fundamentaram a ISAD(G), “na estruturação dos fundos, o caráter hierárquico e físico, rigorosamente unido à organização e à proveniência dos níveis de descrição, se relacionaria mais com as teses do MAD” (Rufí i Pagès, 1995, p. 104). O próprio Michael Cook nos relata, em “*Managing Machine-Readable Archives: Progress with Description and Exchange Standards*” que, “a regra multinível do MAD determina que descrições arquivísticas devem normalmente abarcar mais de um nível de classificação”. E que “isto está completamente de acordo com a regra multinível introduzida na ISAD(G)” (Cook, 1994, p. 9).

De outra parte, o arquivista canadense Hugo Stibbe confirma igualmente que “o princípio fundamental da ISAD(G) de descrição multinível pode ser encontrado, ainda que usando terminologia diferente, no MAD e nas RAD. O APPM dá uma rápida atenção à questão da descrição multinível e não contém muitas regras para aplicá-la” (Stibbe, 1998, p. 3).

Sendo assim, podemos concluir que são, sobretudo o MAD e as RAD, as obras a influírem mais diretamente na definição deste procedimento descritivo chamado multinível.

Além das obras arquivísticas mencionadas, não devemos desconsiderar as duas outras obras que indiretamente influenciaram o conceito de descrição da ISAD(G), o AACR2 e a ISBD<sup>43</sup>, que, de acordo com Mason, se encontra na base das RAD (Mason, s.d.). Neste sentido, observamos que o AACR2 incorporou em sua 2ª edição, de 1978, o disposto pela ISBD (M) em sua edição de 1971, que por sua vez fundamentou as RAD e, conseqüentemente, a ISAD(G) na década de 80.

A ISBD, assim como o AACR2, concebe a descrição bibliográfica a partir de uma perspectiva de segmentação da informação, e estabelece que “a descrição em vários níveis é baseada na divisão da informação descritiva em dois ou mais níveis. O primeiro nível contém as informações comuns ao documento completo ou ao principal. O segundo nível e os níveis seguintes contêm as informações relativas ao volume ou a outra unidade individual. O processo continua para tantos níveis quanto os necessários para descrever a totalidade do documento e as suas partes” (IFLA, 2002, p. 73).

Embora não devamos equiparar a proposta de descrição bibliográfica em vários níveis à concepção arquivística de descrição multinível, por se tratarem de propostas originárias de áreas distintas, com objetivos distintos, e de aplicação idem, é possível traçarmos algumas comparações entre elas.

---

<sup>43</sup> ISBD: International Standard for Bibliographic Description.

Sem pretendermos nos aprofundar em tal análise, mas julgando oportuno tal exercício, notamos que, em primeiro lugar, a descrição bibliográfica tem como resultado uma única descrição, que objetiva dar conta da totalidade de um dado documento, ao passo que no âmbito arquivístico, a descrição multinível objetiva esclarecer as relações existentes entre o fundo e suas partes, sendo elaboradas tantas descrições quantos níveis a serem descritos. Em segundo lugar, há a natureza do material a ser descrito: documentos independentes, no caso das bibliotecas e, no caso dos arquivos, conjuntos documentais.

Além disso, vale ressaltar que as prescrições estabelecidas pela ISBD e pelo AACR2 se referem ao “documento” do ponto de vista biblioteconômico, ou seja, do “documento” como um todo ou das partes que o compõem. No entanto, este todo permanece sendo uma obra individualizada e, portanto, de acordo com a teoria biblioteconômica, deve ser descrita isoladamente. Contudo, do ponto de vista arquivístico, o objeto a ser descrito nunca é o documento individualizado *stricto sensu*, mas sua relação com seu contexto de produção e, por esta razão, podem ocorrer várias descrições, com informações diferentes de acordo com o nível a ser descrito, incluindo o de item documental.

Assim, mesmo que a técnica de descrição multinível possa ser verificada tanto nas concepções arquivísticas quanto nas biblioteconômicas, sua aplicação remete a conceitos diferentes e a aplicações distintas.

Ao que tudo indica, o recurso de segmentação da informação em níveis há algum tempo vem servindo como base para as práticas descritivas, seja na área da Biblioteconomia como na da Arquivística, embora a técnica de segmentação da informação descritiva adquira contornos distintos em cada uma das áreas mencionadas.

Deste modo, parece-nos possível inferir que, assim como a Arquivística incorporou técnicas de descrição oriundas da Biblioteconomia, adaptando-as às suas demandas, a ISAD(G) normalizou uma prática que era já usual, ao menos parcialmente, nos arquivos de todo o mundo. De qualquer maneira, é evidente que para que a norma pudesse ser aceita, ela precisava incorporar os critérios de descrição existentes no contexto internacional e respeitar os princípios básicos das práticas arquivísticas nacionais.

Conseqüentemente, no que diz respeito à origem da técnica de descrição multinível, e de acordo com o que foi exposto até o momento, acreditamos que ela tenha tido uma origem múltipla.

Esta origem é formada, em parte, por aqueles que afirmam que ela provém das técnicas de descrição biblioteconômica, em parte é constituída pelos que acham que a origem se encontra nas obras normativas usadas como fonte para a elaboração de ISAD(G) e, por último, há os que consideram que a descrição multinível sempre foi uma prática comum no âmbito arquivístico.

Se deixarmos a primeira opção de lado, é possível concluirmos que os princípios que inspiraram a proposta de descrição multinível da ISAD(G) já estavam presentes nas práticas descritivas anteriores ao desenvolvimento da Norma, mas nos parece igualmente verdadeiro que os diversos manuais e obras normativas nacionais contribuíram para a definição deste tipo de descrição e para o estabelecimento de sua técnica de aplicação pela ISAD(G).

## Conclusão

Portanto, ao analisarmos alguns dos conceitos subjacentes à proposta de descrição da ISAD(G), suas origens, e as alterações por eles sofridas, concluímos que, o advento da automação nos arquivos provocou uma transformação no modo como os arquivistas vinham cuidando de suas atribuições, sendo que a reflexão

sobre a prática descritiva acabou gerando mudanças expressivas na descrição dos documentos. Longe de anular práticas e princípios que caracterizam a Arquivística, a informática procurou se adaptar às exigências da disciplina, propiciando respostas que pudessem atender a esta nova demanda. Esta interface implicou forçosamente a utilização de normas de trabalho, pois para que fosse possível integrar uma rede ou trabalhar com um aplicativo informático se fez necessário o uso de regras comuns de atuação.

Além disso, se os arquivos usassem normas compatíveis entre si para descrição dos seus documentos, um padrão mínimo, os documentos poderiam ser recuperados com maior facilidade. Porém, não havendo um padrão mínimo de procedimento, as práticas locais parecem ter sido, até então, a regra ao invés da exceção.

Entretanto, mesmo diante da crescente demanda por acesso eletrônico, por rapidez e por facilidade na obtenção de informações arquivísticas, os arquivistas permaneceram por um bom tempo achando que o desenvolvimento e o uso de práticas descritivas unificadas era algo problemático. Tal postura, que retardou consideravelmente o desenvolvimento de uma norma para a descrição dos documentos, encontrou na ISAD(G) uma proposta que se propunha a atender os princípios e práticas arquivísticas, buscando consistência descritiva, fornecendo acesso a informações contextualizadas e permitindo o intercâmbio de dados.

Sendo assim, a análise da primeira proposta internacional de normalização descritiva demonstrou que, a despeito dos diferentes pontos de vista acerca da ISAD(G) e de seus pressupostos, pode ser viável, e diríamos desejável, o estabelecimento e a adoção de um rol de diretrizes que pudesse ser usado por todos. Além disto, a Norma promoveu um deslocamento do conceito de descrição arquivística do arranjo físico, do controle do acervo para o acesso, potencializando, sobretudo o acesso a uma informação contextualizada.

## Bibliografia

ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. [S.d.] **O que é normalização**. Disponível em: <http://www.abnt.org.br/>. Acesso em: set. 2005.

BELLOTTO, H. L. [2002]. **Arquivística: objetos, princípios e rumos**. São Paulo: Associação de Arquivistas de São Paulo. (Scripta, 1).

\_\_\_\_\_. [2004]. **Arquivos permanentes: tratamento documental**. 2 ed. São Paulo: FGV.

BURKE, Peter. [2003]. **Uma história social do conhecimento: de Gutenberg a Diderot**. Rio de Janeiro: Jorge Zahar.

CAMARGO, A. M. A. e BELLOTTO, H. L. (org.). [1996]. **Dicionário de terminologia arquivística**. São Paulo: Associação dos Arquivistas Brasileiros – Núcleo Regional de São Paulo / Secretaria de Estado da Cultura.

CERRI, R. [1992]. Un ulteriore passo avanti verso gli Standard descrittivi per gli archivi: la nuova versione di “Statement of principles” e “ISAD(G)”. **Archivi & Computer**, San Miniato, n. 2, p. 119-127.

\_\_\_\_\_. [2000]. **ISAD(G): Norma geral internacional de descrição arquivística**. 2 ed. Adotada pelo Comitê de Normas de Descrição (Estocolmo, Suécia, 19-22 set. 1999), versão final aprovada pelo CIA. Rio de Janeiro: Arquivo Nacional. Disponível em: <http://www.arquivonacional.gov.br/pub/virtual/mesred.htm>. Acesso em: set. 2002.

- COOK, M. [1992]. Description standards: the struggle towards the right. **Archivaria**: the journal of the Association of Canadian Archivists, Ottawa, n. 34, p. 50-57.
- \_\_\_\_\_. [1994]. Managing Machine-Readable Archives: Progress with Description and Exchange Standards. **IASSIST QUARTERLY**, v. 18, p. 7-10. Disponível em <http://iassistdata.org/publications/iq/iq18/iqvol181-2cook.pdf>. Acesso em: jul. 2005.
- COX, R. [1994] Standardizing archival practices: a tool for the information age. **Archivum**, n. 39, p. 165-180.
- CUNNINGHAM, A. [2000]. Recent development in standards for archival description and metadata. **IASA Journal**, n. 13-16, p. 14-24. \_\_\_\_\_. [2001]. Six degrees of separation: Australian metadata initiatives and their relationships with international standards. **Archival Science**, n. 1, p. 271-283.
- DAVIS, S. E. [2003]. Descriptive standards and the archival profession. **Cataloguing & classification quaterly**, v. 35, n. 3/4, p. 291-308.
- DUFF, W. M. e HAWORTH, K. M. [1990]. The reclamation of archival description: the canadian perspective. **Archivaria**: the journal of the Association of Canadian Archivists, Ottawa, n. 31, p. 26-35.
- DURANTI, L. [1993]. Origin and development of the concept of archival description. **Archivaria**: the journal of the Association of Canadian Archivists, Ottawa, n. 35, p. 47-54.
- FELIU, Miguel C. M. [1995]. ISAD(G): hacia un estándar internacional de descripción archivística. **Métodos de Información**, v. 2, n. 8, p. 27.
- FERREIRA, A. B. H. [1987]. **Novo dicionário da língua portuguesa**. Rio de Janeiro: Nova Fronteira.
- FONSECA, V. M. M. [2001]. **Normalização da descrição arquivística – avanços internacionais e a situação do Brasil**. Rio de Janeiro: Arquivo Nacional. Disponível em: <http://www.arquivonacional.gov.br/pub/virtual/mesred.htm>. Acesso em: mar. 2002.
- GARCÍA GUTIÉRREZ, A. L. [1985]. Normalización general y documental: concepto, historia e instituciones. **Documentación de las Ciencias de la Información**, v. 9, p. 55-96. Disponível em <http://revistas.sim.ucm.es:2004/inf/02104210/articulos/DCIN8585110055.pdf>. Acesso em: jul. 2005.
- GUERCIO, M. [1994]. Models for fostering the development of standard practices within a region or a particular country. **Archivum**, n. 39, p. 200-209.
- GRUPPO di Lavoro sulle ISAAR(CPF). [S.d.]. **Sintesi di un'esperienza**. ANAI – Sezione Lazio. Disponível em: <http://www.anai.org/anai/>. Acesso em: maio 2002.
- HAGEN, A. M. M. [1998]. Algumas considerações a partir do processo de padronização da descrição arquivística. **Ciência da Informação**, Brasília, v. 27, n. 3, p. 293-299, set./dez. Disponível em: <http://www.ibict.br/cionline/270398/27039807.pdf>. Acesso em: jul. 2002.
- HEREDIA HERRERA, A. [1991]. Descripción y normalización. **Boletín de la ANABAD**, Madrid, n. 2, p. 51-58.
- \_\_\_\_\_. [1995 a]. **La norma ISAD(G) y su terminología análisis, estudio y alternativas**. Madrid: ANABAD / Arco.
- \_\_\_\_\_. [1995 b]. Observaciones sobre la norma ISAD(G). **AABADOM**, n. 2, p. 23-26.

- \_\_\_\_\_. [2000]. La nueva versión de la norma ISAD(G). **Archivamos**, n. 36-37, p. 5-11.
- \_\_\_\_\_. [2001 a]. De vuelta a la norma ISAD(G): consenso para um manual. **Boletín de la ANABAD**, Madrid, n. 2, p. 81-87.
- \_\_\_\_\_. [2001 b]. Los niveles de descripción: un debate necesario en la antesala de las normas nacionales. **Boletín de la ANABAD**, Madri, n. 4, p. 41-68.
- \_\_\_\_\_. [2002 a]. **La norma ISAD(G) “Análisis crítico”: norma internacional para la descripción archivística**. Disponível em: <http://infolac.ucol.mx/documentos/isad.pdf>. Acesso em: abr. 2002.
- \_\_\_\_\_. [2002 b]. Las normas de descripción archivística: encuentros y desencuentros. **Boletín del Instituto Andaluz de Patrimônio Histórico**, n. 38, p. 142-151.
- \_\_\_\_\_. [2003 a]. **La norma ISAD(G) - análisis crítico**. São Paulo: Associação dos Arquivistas de São Paulo.
- \_\_\_\_\_. [2003 b]. **El principio de procedencia y los otros principios de la archivística**. São Paulo: Associação dos Arquivistas de São Paulo. (Scripta, 5).
- HILDESHEIMER, F. [1983]. **Directrices metodológicas relativas a la preparación de las guías generales de los archivos nacionales: un estudio RAMP**. Paris: Unesco.
- HOLMES, O. W. [1984]. Archival arrangement. Five different operations at five different levels. In: DANIELS, M. F. e WALCH, T. **A Modern Archives Reader: basic readings on archival theory and practice**. Washington: National Archives Trust Fund Board, p. 162-180.
- INTERNATIONAL FEDERATION OF LIBRARY ASSOCIATIONS AND INSTITUTIONS. [2002]. **ISBD(G): descrição bibliográfica internacional normalizada geral**. IFLA. Disponível em: <http://www.ifla.org/VI/3/nd1/isbdlist-pt.htm>. Acesso em: jun. 2005.
- LOPEZ, A. P. A. [2002 a]. Alcance da descrição arquivística e o processo de automação. **Registro**, Indaiatuba: Fundação Pró-Memória, v. 1, n. 2, p. 27-39.
- \_\_\_\_\_. [2002 b]. International standard of archival description: observações sobre a ISAD(G). **Revista Histórica**, São Paulo: AESP, n. 7, p. 38-46.
- MASON, M. K. [S.d.] The changing tides of the archival descriptive process: where it is taking us on our quest for standardization. S.l.d. Disponível em: <http://www.moyak.com/researcher/resume/papers/archmkm.html>. Acesso em: out. 2005.
- MURTOMAA, E. [2000]. Podría tratarse del inicio de una buena amistad: comparación de la descripción y acceso a objetos de interés entre bibliotecas y archivos Jerusalem: IFLA Council and General Conference, 66th, 13-18 Ago. Disponível em: <http://www.ifla.org/IV/ifla66/papers/>. Acesso em: jun. 2005.
- PEDERSON, A. [1994]. Unlocking hidden treasures through description: comments on archival voyages of discovery. **Archivaria: the journal of the Association of Canadian Archivists**, Ottawa, n. 37, p. 47-63.
- POESNER, E. [1984]. Some aspects of archival development since the French Revolution. In: DANIELS, M. F. e WALCH, T. **A modern archives reader: basic readings on archival theory and practice**. Washington: National Archives Trust Fund Board, p. 3-14.
- ROUSSEAU, J. Y. e COUTURE, C. [1998]. **Os fundamentos da disciplina arquivística**. Lisboa: Publicações Dom Quixote, 1998.

RUFÍ I PAGÈS, J. [1995]. Los modelos y la norma: fuentes de la normalización internacional de la descripción archivística. **Revista Lligall**, Barcelona, 32 p. Disponível em: [http://www.arxivvers.com/html/articulos/pdf/cas/09\\_modelos.pdf](http://www.arxivvers.com/html/articulos/pdf/cas/09_modelos.pdf). Acesso em: jul. 2002.

SCHELLENBERG, T. R. [1980]. **Documentos públicos e privados: arranjo e descrição**. 2 ed. Rio de Janeiro: Arquivo Nacional.

SILVA, J. A. [1998]. Descrição multinível: regras da descrição multinível. In: CONGRESSO BRASILEIRO DE ARQUIVOLOGIA, 10º, 1994, São Paulo: AAB-SP. (cd-rom).

SMIT, J. W. [1998]. **A disponibilização da informação institucionalizada: condicionantes e perspectivas**. São Paulo, 11 p. (mimeo).

STIBBE, H. [1992]. Implementing the concept of fonds: primary access point, multilevel description and authority control. **Archivaria**: the journal of the Association of Canadian Archivists, Ottawa, n. 34, p. 109-137.

\_\_\_\_\_. [1997]. Archival descriptive standards and the archival community: a retrospective, 1996. **Archivaria**: the journal of the Association of Canadian Archivists, Ottawa, n. 41, p. 259-274.

\_\_\_\_\_. [1998]. A experiência canadense e a ISAD(G). Rio de Janeiro. Trab. Apres. no Seminário de Descrição Arquivística, Rio de Janeiro, 1998.

TARRAUBELLA I MIRABET, X. [1998]. Los archivos y sus usuarios. **Revista Lligall**, Barcelona, n. 12, p. 190-204. Disponível em: [http://www.arxivvers.com/idadmin/docs/12\\_15\\_art\\_cast-jNEq27T.pdf](http://www.arxivvers.com/idadmin/docs/12_15_art_cast-jNEq27T.pdf). Acesso em: jul. 2005.

VITALI, S. [1994]. Il dibattito della descrizione: aspetti teorici e prospettive in Italia. **Archivi & Computer**, San Miniato, n. 4, p. 303-323.

WEBER, L. [1989]. Archival description standard: concepts, principles, and methodologies. **American Archivist**, Chicago, n. 52, p. 504-513.

ZAZO, J. L. B. [2000]. La normalización: base del análisis documental en los archivos. **Scire**. n. 6, p. 55-75.

## Interacting with Archival Users: A Status report

Dr. Wendy M. Duff

In 1990, Timothy Ericson called upon archivists to focus more attention on reaching out to the general public and on promoting the use of their archives. He eloquently stated that if “after we brilliantly and meticulously appraise, arrange, describe and conserve our records, nobody comes to use them, .. then we have wasted our time.”<sup>44</sup> Two years later, Richard Cox noted that the dearth of research literature on the use and users of archival material and on archival reference services hindered the discussion of these issues. He reviewed the research literature and identified four broad areas of study: the use of archival records; the accuracy and effectiveness of archival reference; the impact of technology on reference services; and the interaction between the user of archives and the reference archivist. His article discussed the findings of research related to the actual use of records, but he noted that very few studies relating to the other three areas existed. He suggested that “there should be no question that archivists need a significant amount of research ....[and] the nature and breadth of the research leave much to be desired.”<sup>45</sup>

In the last 17 years, researchers have focused much attention on these topics and some have begun to address the gaps identified by Cox. Building on Cox’s study, this paper reviews the relevant research published since the 1992. It discusses the findings of recent doctoral dissertations, as well as completed and ongoing research by academics and practitioners. The paper suggests that researchers have shed new light on some of the research areas highlighted by Cox; however, research in many other areas remains limited. The paper identifies areas needing further research and concludes with suggestions for practitioners who wish to study the use and users of their archives.

Cox’s review of the archival literature notes that the majority of research on reference focused on the use and users of records including how people searched for, and located archival sources. For example, Conway’s research on the use of American presidential libraries indicated that researchers depended on their professional grapevine to help locate material<sup>46</sup> and Beattie’s study on the needs of social historians suggested finding aids did not meet their needs.<sup>47</sup>

Since 1992, many researchers have studied the techniques users employ to locate sources, the types of material they seek, the types of question they ask and the terms they use.

From 2001 to 2002, three studies investigated historians’ use of archival sources. Tibbo<sup>48</sup> and Anderson<sup>49</sup> surveyed and interviewed American and British historians respectively and Duff, Craig and

---

<sup>44</sup> Timothy L. Ericson, “Preoccupied with your own garden”: Outreach and Archivists” *Archivaria* 31 (1990): 114-122.

<sup>45</sup> Richard J. Cox “Researching archival reference as an information function: Observations on needs and opportunities” *RQ* 31, no. 3 (Spring 1992): 387-394?

<sup>46</sup> Paul Conway, “The name assigned to the document by the author. This field may also contain sub-titles, series names, and report numbers. Research in Presidential Libraries: A User Survey” The entity from which ERIC acquires the content, including journal, organization, and conference names, or by means of online submission from the author. *The Midwestern Archivist*, v11, n1 (1986): 35-56

<sup>47</sup> Diane L. Beattie, “An Archival User Study: Researchers in the Field of Women’s History” *Archivaria* 29 (Winter 1989-90): 33-50.

<sup>48</sup> Helen Tibbo, “Primarily History in America: How US Historians Search for Primary Materials at the Dawn of the Digital Age” *American Archivist*: 66, no 1 (2003): 9-50 and “Primarily History: Historians the Search for Primary Source Materials” *Proceedings of the ACM/IEEE Joint Conference on Digital Libraries 2002*. Portland, OR, July 14-18, 2002. New York: ACM, 2002

<sup>49</sup> Ian G. Anderson, “Are You Being Served? Historians and the Search for Primary Sources,” *Archivaria* 58 (Fall 2004) 81-129

Cherry<sup>50</sup> surveyed Canadian historians to gather information on how these users located primary material. American historians used a variety of methods to locate records with ninety-eight percent of respondents following leads in citations (footnote chaining); seventy-nine percent using published bibliographies; fifty-seven percent consulting documentary editions and only seventy-six percent searching published finding aids. On the other hand, ninety percent of respondents indicated they used print finding aids when visiting archives and approximately seventy-five percent visited an archives to get advice from an archivist. Finally, forty-three percent of respondents used the Internet to locate material.<sup>51</sup> Similarly, British historians relied on print based retrieval methods with 97% following references in books and articles, 89% using paper finding aids when visiting the archives, and 58% consulting electronic finding aids. The British historians wanted more folder-level description and more content in the descriptions.<sup>52</sup> Duff, Craig, and Cherry's questionnaire asked respondents to rate the importance of twelve types of sources in becoming aware of, and locating material needed for their research.<sup>53</sup> Sources included both published and unpublished formal sources as well as many informal sources. Table 1 shows the percentage of respondents who rated each source as *Very Important* or *Somewhat Important*.

TABLE 1

Percentage of respondents who consider the following sources as either Very Important or Somewhat Important in becoming aware of and locating needed material

Archival Sources	95%
Archival Finding Aids	94%
Archivists	87%
Footnotes or Other References	87%
Colleagues	78%
Published Bibliographies	67%
Book Reviews	52%
World Wide Web	49%
Indexing Publications	40%
Abstracting Services	29%
Students	27%

Both archival sources and archival finding aids received extremely high ratings: ninety-five percent of the respondents rating archival sources as either *Very Important* or *Somewhat Important* and ninety-four percent of respondents rating archival finding aids this way. Eighty-seven percent of respondents rated footnotes and archivists as either *Very Important* or *Somewhat Important*. In comparison, fewer than thirty percent of respondents rated abstracting services (29%) and students (27%) as *Very* or *Somewhat Important* sources for becoming aware of, or locating material. The

<sup>50</sup> Wendy M. Duff, Barbara Craig and Joan Cherry, "Historians' Use of Archival Sources: Promises and Pitfalls of the Digital Age" *The Public Historian* 26 no.2 (Spring 2004) 7-22

<sup>51</sup> Helen Tibbo, "Primarily History in America: How US Historians Search for Primary Materials at the Dawn of the Digital Age" *American Archivist*: 66, no 1 (2003): 9-50 and "Primarily History: Historians the Search for Primary Source Materials" *Proceedings of the ACM/IEEE Joint Conference on Digital Libraries 2002*. Portland, OR, July 14-18, 2002. New York: ACM, 2002

<sup>52</sup> Ian Anderson, "Are you Being Served?: Historians and the Search for Primary Sources" *Archivaria* 58 (Fall 2004): 80-129

<sup>53</sup> Wendy M. Duff, Barbara Craig and Joan Cherry, "Historians' Use of Archival Sources: Promises and Pitfalls of the Digital Age" *The Public Historian* 26 no.2 (Spring 2004) 7-22

World Wide Web received a rating of either *Very* or *Somewhat Important* from almost half (49%) of the respondents.<sup>54</sup> Though the findings from these three studies vary slightly, the research highlights the important role citations, findings aids and archivists play in helping historians locate archival sources. The studies also indicate that approximately half of the historians consult electronic finding aids and use the Web when seeking information. Historians' use of the Web and electronic finding aids has undoubtedly increased over the last seven year.

Yakel also conducted a study of user behaviour to investigate how undergraduates and scholars locate and use primary material<sup>55</sup>. She reported that the twenty-eight participants in her study relied heavily on finding aids to identify relevant documents; however, many experience difficulty when using these tools. Sweeney's interviewed and observed fourteen archival patrons from three different types of archives for her doctoral research and concluded that finding aids present significant challenges to archival users who have to rely on archival staff for assistance. These participants expected the archives system to work like a library system or an online public access system, and therefore, had problems locating what they needed.<sup>56</sup> Spindler and Pearce-Moses' study of users' understanding of MARC records pointed out that most users cannot understand the information in these descriptions.<sup>57</sup> Duff and Stoyanova's focus groups study investigated the type of information users wanted displayed in online public access catalogues or on the Web. Not only did Duff and Stoyanova agree that users have problems interpreting information regarding physical description and dates of creation, but the participants also indicated confusion when they encountered the word "fonds" in the title of a description.<sup>58</sup> Finally, Christopher Prom conducted an experiment to measure users' interactions with on-line archival finding aids. He studied the interactions of novice users, archivists and computer experts with four collection level systems and four finding aids. He described the results of his analysis.

"Finding aids currently placed on-line are most efficiently used by either archival and computer experts. Nonexperts find them more difficult to work with. They need over ninety seconds on average to find a relevant item when participating in a canned search under observation."<sup>59</sup>

He noted that most of the users in his study reported that they would not spend more than a minute searching for information. He suggested that alphabetical lists, page-top tables of contents, and a single-page search option might improve the searchability of finding aids for both novice and expert users.

Research on the information seeking behavior of genealogists supports Prom's proposal that lists might facilitate retrieval. Duff and Johnson's study involving in-depth interviewed with 10 genealogists found that the interviewees consulted finding aids, but they wanted access to lists of file and item level descriptions and rarely consulted the fonds-level description.<sup>60</sup> One genealogist stated that when consulting a finding aid for the first time she might read the introduction, but subsequently, would consult the file lists

---

<sup>54</sup> Wendy Duff, Barbara Craig and Joan Cherry. "Finding and Using Archival Resources: A Cross-Canada Survey of Historians Studying Canadian History" *Archivaria* 58 (Fall 2004): 51-80

<sup>55</sup> Elizabeth Yakel, "Listening to Users" *Archival Issues* 26, no 2 (2002): 111-27

<sup>56</sup> Shelley Sweeney, "The Source Seeking Cognitive Processes and Behavior of the In-Person Archival Researcher" (Ph. D., University O Texas, 2002),

<sup>57</sup> Robert Spindler and Richard Pearce-Moses, "Does AMC Mean Archives Made Complicated? Patron understanding of USMARC AMC Catalog Records," *American Archivist* 56 (Spring 1993): 330-41

<sup>58</sup> Wendy Duff and Penka Stoyanova, "Archival Displays from a Users' Point of View," *Archivaria* 45 (1998): 44-79.

<sup>59</sup> Christopher J. Prom, "User Interactions with Electronic Finding Aids in a Controlled Setting" *American Archivists* 67, no 2 (2004), p. 263

<sup>60</sup> Wendy M. Duff and Catherine A. Johnson, "Where is the List with All of the Names? Information Seeking Behaviour of Genealogists," *The American Archivist* 66 no.1 (Spring/Summer 2003) 79-95.

or the indices directly. Furthermore, she remembered that as a novice researcher she did not realize the value of a fonds level description. Other professional genealogists in this study explained that they almost never read the introduction preferring instead direct access, as one participant elaborated.

*No I almost never read this preliminary stuff. I want to know, you know, where's the list that has the names of the people or the townships. (G5)*

The genealogists in Duff and Johnson's study wanted information documented in the records; therefore, they wanted direct access to the records or access to file lists which would lead them to records. Fonds level descriptions - with their overview, biographical sketch and scope and content or information about the creation of records did not meet their needs and provided little value in their eyes. As previously noted, historians rated findings aids as Very Important sources for locating information, but British historians wanted more file lists. Lists and pop up table of contents might serve the needs of both genealogists and historians. Duff and Johnson also found that genealogists usually searched by name, place and date when seeking information about their family.

Studies of reference requests asked onsite and remotely, either by mail or email, also identified terms that users employ when describing their reference question. Gagnon-Arguin studied research questions that required the assistance of archivists asked onsite, by phone or by email in Quebec archives<sup>61</sup>. She grouped the questions into subject or theme (41%), Proper name (21%) and type of document (33%) and miscellaneous (4%) and suggested that different types of questions contained different types of terms. In a similar study, Collins examined reference questions asked in two institutions with historical photograph collections and also concluded that these questions contained common elements such as general subject, specific subject, place, date, genre, visual features, physical format, and creator/provenance<sup>62</sup>. Duff and Johnson analyzed 361 email reference requests and categorized the questions as service requests (27%), material findings, or general subject request (17%), fact finding (10%), user education (13%) requests for specific forms of material (8%) and consultation (10%).<sup>63</sup> They identified the terms provided by users as proper name, date, place, subject, form and occasionally event. To improve access to archival material, we need to design systems that support searching with terms people know, i.e. proper name, date, place, subject, visual feature, physical format, documentary form and event.

Based on a study of remote reference questions sent to two archives in 1995 and 1999, Martin reported that novice users tended to ask broader questions and more experienced researchers asked more specific questions<sup>64</sup>. She posited that subject based requests decreased over the years she studied, which she connected to accessible online finding aids. Finally, she observed that many remote users did not provide enough background information for the reference archivist to answer their questions. Duff and Johnson's study of email requests also found that 13% of the request indicated a need for instruction, as previously noted.<sup>65</sup> These studies highlight the importance of providing instruction to users who access virtual archives. Unfortunately, it appears that few remote users get the type of instruction they require.

---

<sup>61</sup> Louise Gagnon-Arguin, "Les questions de recherche comme matériau d'études des usagers en vue du traitement des archives," *Archivaria* 46 (1998): 86-102.

<sup>62</sup> Karen Collins, "Providing Subject Access to Images: A Study of User Queries," *The American Archivist* 61 (Spring 1998): 36-55.

<sup>63</sup> Wendy M. Duff and Catherine A. Johnson, "A Virtual Expression of Need: An Analysis of Archival Reference Questions," *American Archivist* 64 (Spring/Summer 2001): 43-60.

<sup>64</sup> Kristin E. Martin, "Analysis of Remote Reference Correspondence at a Large Academic Manuscripts Collection" *American Archivist* 64 (Spring/Summer 2001): 17-42

<sup>65</sup> Wendy M. Duff and Catherine A. Johnson, "A Virtual Expression of Need: An Analysis of Archival Reference Questions," *American Archivist* 64 (Spring/Summer 2001): 43-60

Duff and Fox interviewed fourteen archivists working in two different archives and concluded that reference archivists spend almost half of their time at the reference desk teaching onsite users how to use archival systems, however, remote reference services rarely involve instruction.<sup>66</sup> Few archives provide chat reference, and yet this service could potentially educate users who have to use archival finding aids and systems remotely. The few archives who offer this service provide it only for a few hours a day<sup>67</sup>. Online tutorials such as the Yale University tutorial can assist some users.<sup>68</sup> The Yale tutorial provide instruction on searching for and using records in the Yale archives, however, because it contains an introduction to archival research, users of other archives might also find it helpful. Notwithstanding, online tutorials do not eliminate the need for personalized one-on-one instruction. Based on a study with 28 participants described earlier, Yakel and Torres delineated the types of knowledge or archival intelligence that users require to use archives: knowledge of archival theory, practices and procedures, strategies for reducing uncertainty and ambiguity, and intellectual skills.<sup>69</sup> If archives provide virtual reference services during nights and weekends users could obtain information about the archives, archival procedures, and the organization and content of finding aids when they visit an archives virtually. This information might help demystify archives to many remote users. Furthermore, answering questions about archival systems and procedures may help many users increase their level of archival intelligence, which in turn, may improve their ability to use archival material.

Archival reference involves both archival instruction and the answering of questions. Although current research has begun to investigate the effectiveness and impact of archival orientation sessions, the effectiveness or accuracy of reference questions, Cox's second area of research remains understudied. However, Anthony's doctoral research which compared the searching behaviour of novice and expert reference archivists sheds light on this topic.<sup>70</sup>

Anthony found that when searching for information novice reference archivists begin by performing a meta-search using the terms taken from a reference question and indicated satisfaction if the retrieved documents contained terms that matched words in the question. They did not review the documents referred to evaluate their relevance and they did not indicate a need to discuss the results or the reference question with the user. On the other hand, when presented with a question expert reference archivists often identified potentially relevant documents, or finding aids that pointed to relevant documents from memory.

After a search, expert reference archivists examined the list of citations retrieved and considered the quality of the sources. They evaluated the list for accuracy and comprehensiveness of the content, determined whether the form or format of the material facilitated skimming, and identified who had processed the collection and created the finding aids. Furthermore, at times, they modified their search strategy and continued to search until they identified accurate, relevant documents.

Anthony reasoned that expert archivists draw on their extensive contextual knowledge; utilizing information about an organization, its functions, how the functions changed over time and the individuals connected with these functions to answer questions. She contends that this contextual or provenance

---

<sup>66</sup> Wendy Duff and Allyson Fox "You're a Guide rather than an Expert": Archival Reference from an Archivists Point of View" *Journal of the Society of Archivists* 27 (Oct 2006): 129-153.

<sup>67</sup> For example users can use the chat reference service at Library and Archives Canada (LAC) "between 2:00 p.m. and 4:00 p.m., Monday to Friday, Eastern Standard Time, except statutory holidays" Available at <http://www.lac-bac.gc.ca/the-public/005-2040-e.html> accessed on October 23, 2008

<sup>68</sup> "Using Manuscripts and Archives: A Tutorial" Available at <http://www.library.yale.edu/mssa/tutorial/tutorial.htm>

<sup>69</sup> Elizabeth Yakel and Deborah A. Torres, "Archival Intelligence and User Experience," *The American Archivist*, 66, n.1 (Spring/Summer 2003) 51-78.

<sup>70</sup> Denise Anthony "Beyond Description: An Exploration of Experienced Archivists' Knowledge and Searching Skills. PhD dissertation, University of Michigan 2006.

information facilitates the retrieval of archival documents. In other words, the expert reference archivists in this study used the provenance method to retrieve relevant documents. Moreover, according to Anthony experts rarely search for subject; unlike novice archivists who often begin their search with a meta-search using topical terms. Archivists need more research that investigates the power of the provenance method of retrieval and they need to identify ways to make searching for archival material more effective. More research on the effectiveness of different methods of searching for information would provide the foundation for the design of better archival retrieval systems. As previously mentioned, research also indicates that novice users and even novice archivists interact with system differently and not as effectively as expert archivists or expert historians such as academic historians. Archives should consider developing systems which supports both novice and expert ways of searching for archival material.

In 1992 when Cox identified the impact of technology on reference as his third area of interest the Internet existed but the Web was still in its infancy. (The Web made its debut in August 1991 with the posting of a short summary about the World Wide Web project on the alt.hypertext newsgroup by Berners-Lee) Some archives had made their finding aids available through gopher, but digital collections, EAD finding aids, chat reference, blogs, wikis, Web 2.0 and a host of other technologies did not yet exist. Few researchers have investigated the changing nature of archives, but Pugh contends that the digital environment has revolutionized reference<sup>71</sup> and Cox et al claim that, “traditional reference activities are in a state of change, greater than many archivists realize.”<sup>72</sup> Though the digital realm has impacted reference services, archives have yet to fully exploit the power of the Web or Web 2.0 technology to serve their users better. In 2008 Samouelian examined 213 archival repositories websites and found that only 38 archives used at least one type of Web 2.0 technology.<sup>73</sup> As previously mentioned, few archives provide chat reference services and “the possibility of using wikis in conjunction with archival reference services has yet to be fully explored by archivists.” Moreover, only one research project has investigated a Web 2.0 enabled access tool from an user’s point of view. Krause and Yakel developed and gathered user feedback on a Web 2.0 finding aid for the Polar Bear Collection.<sup>74</sup> The system functionality enabled users to view paths through the finding aid that other users had taken and to add comments. Over the six months that Kruase and Yakel studied the use of the finding aid, twenty-six users submitted comments, which included 17 corrections. Ten individuals also added information to the finding aids about the people connected with the Polar Bear Expedition. Krause and Yakel concluded that enabling direct and indirect interaction among visitors and archivists, collaborative filtering, and other Web 2.0 features may enrich traditional finding aids and make archival materials more accessible. This study represents a good beginning, but archivists need more research on the impact of these and other Web 2.0 technologies.

## Interaction

Cox’s final area of research was the interaction between archival users and reference archivists. In the last decade three studies have investigated this interaction, one from the user’s point of view and two from the archivists’ point of view.

---

<sup>71</sup> Mary Jo Pugh, *Providing Reference Services for Archives and Manuscripts* Chicago: Society of American Archivists, 2005

<sup>72</sup> Richard J. Cox and the University of Pittsburgh Archives Students, “Machines in the Archives: Technology and the Coming Transformation of Archival Reference,” *First Monday*, 12 (November 2007), available at <http://www.uic.edu/htbin/cgiwrap/bin/ojs/index.php/fm/article/viewArticle/2029/1894>.

<sup>73</sup> Mary E. Samouelian “Embracing Web 2.0: Archives and the Newest Generation of Web Applications A Master’s Paper for the MS in the LS degree, 2008.

<sup>74</sup> Magia Ghetu Krause and Elizabeth Yakel, “Interaction in Virtual Archives: The Polar Bear Expedition Digital Collections Next Generation Finding Aid” *American Archivist* 70 (Fall/Winter 2007): 282-314

Duff and Johnson's study based on interviews with historians posited that historical researchers greatly value close interaction with reference archivists and deliberately establish these relationships to gain access to an archivist's in-depth knowledge about archival resources<sup>75</sup>. Historians make a point of "chatting up" the archivist to access essential archival knowledge. Duff and Johnson claim that the relationship between the archivist and the historian is the social capital of historical researchers which they exploit to gain access to knowledgeable about records.

Trace, and Duff and Fox have also examined the relationship between users and archivists, but from the archivist's point of view. Based on observation and interviews with two archivists, Trace reported that archivists liked to answer 'good' questions, but that they do not agree on what constitutes a 'good' question.<sup>76</sup> According to one archivist 'good' questions are exciting and complex. They cause her think creatively, to conduct in-depth research and to spend time consulting the records. On the other hand, according to the other archivist 'good questions' are "tied to the job that the researcher does in explaining his or her research topic, and the job that the archivist does in drawing out, understanding and expanding that articulation" Reference archivists who answer 'good' questions broaden the user's basic question and assist him/her to carry out complex research.

Duff and Fox posit that reference archivists achieve higher levels of job satisfaction when users express satisfaction with the reference services and when the archivists successfully complete a complicated search.<sup>77</sup> The archivists in Duff and Fox's study also indicated that the best part of reference is serving people, and expressed high level of satisfaction from learning something new, teaching users something new, or doing complex research to answer questions. However, they also reported that providing reference services presented many challenges and the increasing number of email reference questions caused stress. The archivists sometimes felt dissatisfied with their job when they could not find the answer to a question, when the administrative structure separated them from the custodial archivists who arranged and described records, or when they had to provide reference services to difficult users.

## Evaluation

Although the archival field still lacks indepth research in all area of reference, the profession has made great gains in the last decade. However, most of the research discussed in this paper involves academics and doctoral students rather than practitioners. Some practitioners have conducted usability studies of their archival systems. For example, Altman and Nemmers published the findings of the Pepper OnLine Archival Retrieval and Information System (Polaris) project,<sup>78</sup> Proffitt discussed the usability study of RLG's RedLightGreen,<sup>79</sup> Burns described a user study of the University of California's digital image service,<sup>80</sup> and other archivists have undertaken studies on the usability of EAD finding aids.<sup>81</sup> Though these relatively few studies provide insight into the usability of some archival systems in the United States, they remain limited. Archivists need to

---

<sup>75</sup> Catherine A Johnson and Wendy M. Duff, "Chatting up the archivist : Social capital and the archival researcher" *American Archivist* vol. 68, n1 (2005): 113-129

<sup>76</sup> Ciarn B Trace, "for Love of the Game: An Ethnographic Analysis of Archival Reference Work" *Archives and Manuscripts* 32, no 1 124-143.

<sup>77</sup> Wendy Duff and Allyson Fox "You're a Guide rather than an Expert": Archival Reference from an Archivists Point of View" *Journal of the Society of Archivists* 27 (Oct 2006): 129-153.

<sup>78</sup> Burt Altman and John Nemmers, "The Usability of On-line Archival Resources: The Polaris Project Finding Aid," *American Archivist* 64 (Spring-Summer 2001): 121-31.

<sup>79</sup> Merrilee Proffitt, "How and Why of User Studies: RLG's RedLightGreen as a Case Study," *Journal of Archival Organization* 4, no. ½ (2006): 87-110.

<sup>80</sup> Maureen A. Burns, "From Horse-Drawn to Hot Rod: The University of California's Digital Image Experience," *Journal of Archival Organization* 4, no. ½ (2006): 111-139.

<sup>81</sup> Elizabeth Yakel, "Encoded Archival Description: Are Finding Aids Boundary Spanner or Barriers for Users?" *Journal of Archival Organization* 2, no 1/ 2 (2004): 63-77; Christopher J. Prom, "User Interactions with Electronic Finding Aids in a Controlled Setting," *American Archivist* 67 (Fall/Winter 2004): 234-268.

conduct similar studies using different types of systems in different types of archives, if they are to understand better how users interact with their archives.

By contrast, archivists in the United Kingdom (UK) have developed and administered surveys to gain a better understanding of users' opinions of the services they receive. In 1996, British archivists initiated the Public Service Quality Group for Archives and Local Studies (PSQG), a voluntary organization for archivists interested in advancing best practices and improving the quality of service in archives. To help achieve this goal, the Group first carried out the *National Survey of Visitors to UK Archives* in 1998, and has continued to survey users every eighteen months since 2001.<sup>82</sup> This longitudinal study provides a comprehensive and up to date survey of archival users' demographic characteristics and satisfaction within the United Kingdom. For example, the PSQG's comparison of the 2006 and 2004 surveys revealed improvements in archival users' satisfaction with opening hours, IT technologies, and staff assistance. The information provided by the PSQG surveys has helped archives understand better their users and how these users perceive their services and systems.

Evaluating reference services is an essential step in improving services. In a traditional setting, archivists often conduct exit interviews to gather feedback on the user's experience, but in the digital realm, archivists need more formalized methods for gathering this feedback. Tibbo notes that when answering questions remotely archivists lack visual cues; and therefore may not know if users experience difficulty.<sup>83</sup> She suggests that archivists solicit user feedback on their services, and ask users if the reference services met their needs. Archivists should supplement these informal methods of gathering feedback with formal evaluation that request users' opinions on the various aspects of their reference services.

Many North American archivists have indicated interest in gathering user reactions to their services, but they have also expressed concerns about the feasibility of carrying out this research. Duff, et.al carried out focus group session of archivists to gather their thoughts on conducting user-based evaluation research. Duff, et. al. recounted that participants thought evaluation research could result in many benefits including helping them set digitization priorities, understanding remote users and documenting increases in the use and the value of archives. Duff et. Al. report, however, that archives often failed to carry out evaluation studies due to "a lack of money to hire outside experts, lack of time to conduct user-based evaluation research in-house, and a lack of expertise."<sup>84</sup>

To overcome some of these problems, the Archival Metric projects developed and evaluated a set of toolkits<sup>85</sup> to help university archives conduct user-based evaluation. The Archival Metrics Project includes three academic researchers, Yakel, Tibbo and Duff as well as an advisory board of practitioners. The toolkits include questionnaires designed to gather data from teachers, students and general researchers on the various archives services including reference and orientation sessions as well as an archives finding aids and its website. The toolkits also include administration procedures, advice on recruiting participants, and obtaining ethics review, as well as spreadsheets, and sample reports to help with the analysis of the data and distribution of the findings.

A small group of the university archivists who helped test the questionnaire indicated that many planned to make changes to their services based on the results of the survey. One archivist noted

---

<sup>82</sup> See PSQG, "Survey of Visitors to UK Archives 2006," [http://www.ncaonline.org.uk/research\\_and\\_development/survey/](http://www.ncaonline.org.uk/research_and_development/survey/) Accessed April 7, 2007.

<sup>83</sup> Helen R Tibbo "Interview Techniques for Remote Reference: Electronic Versus Traditional Environments" *American Archivist* 58 (Summer 1995): 294-310

<sup>84</sup> Wendy M. Duff, Jean Dryden, Carrie Limkilde, Joan Cherry and Ellie Bogomazonva, "Archivists' Views of User-based Evaluation: Benefits, Barriers and Requirements," *American Archivist* 71 (Spring/Summer 2008): 158.

<sup>85</sup> These are available for free at [www.archivesmetrics.org](http://www.archivesmetrics.org)

that it was wonderful to have hard data and this experience provides a “chance to keep tweaking what I do.” Another indicated that she would add more material to the tutorial on her website to help better educate her users about the archives, and yet another noted that she would modify her orientation sessions and add information about navigating the archives website to overcome some of the problems identified in her survey.

Most of the partners found the information they obtained from the survey very helpful. Finally, many reported that they would probably conduct other surveys in the future. By October 2008 we had 300 downloads from the Archival Metrics Projects website. We are currently working on a survey to evaluate the tools. We have also received funding from the US National Historic Publication Commission (NHPRC) to adapt the tools for government archives. We have a new advisory board of practitioners who work in government archives to help us with this task. I would invite Brazilian archivists to access the site, download the questionnaire, translate it and adapt it to your needs. If you are interested in this tool, we would love to hear from you.

I would also like to extend an invitation to join Ax-snet (Archival eXcellence in Information Seeking Studies Network), an international network of researchers working to improve access to archives by examining how users seek and use archival information. I am confident that if we collaborate we will one day acquire better understanding of how people use archival material; the accuracy and effectiveness of archival reference; the impact of technology on reference services; and the interaction between the user of archives and the reference archivist.

# “Um homem bateu em minha porta e eu abri (?)”: uma avaliação dos serviços de referência e da acessibilidade dos sites de instituições públicas arquivísticas

Leandro Ribeiro Negreiros  
Alan Vasconcelos

## Introdução

A maioria dos órgãos administrativos tem procurado disponibilizar eletronicamente serviços e documentos relativos às suas atividades, como forma de dar maior transparência aos seus trabalhos, aproximar os cidadãos dos órgãos públicos e facilitar o andamento dos trabalhos, garantindo que a agilidade do mundo eletrônico e a facilidade e a disponibilidade de acesso remoto diminuam as exigências da morosidade burocrática.

Com as instituições arquivísticas não seria diferente. Cada vez mais os arquivos públicos têm disponibilizado mecanismos de busca que disponibilizam documentos de valores probatórios e históricos de interesse coletivo. Os *sites* arquivísticos representam o retrato contemporâneo de instituições tradicionais muitas vezes escondidas em construções suntuosas que não despertam a curiosidade dos cidadãos, principalmente devido aos fatores culturais de valorização da memória e do passado.

A simples disponibilização dos documentos na Web, no entanto, não assegura que estes sejam, de fato, acessíveis. Existe uma série de diretrizes voltadas à criação de documentos para web que são regidas pela W3C (World Wide Web Consortium). Esse conjunto de normas foi adotado, também, pelo governo brasileiro, por meio do Ministério do Planejamento com o “Governo Eletrônico”.

Segundo Souza (2005), embora tenha sido projetada para possibilitar o fácil acesso, intercâmbio e a recuperação de informações, a web foi implementada de forma descentralizada e quase anárquica; cresceu de maneira exponencial e caótica; e se apresenta hoje como um imenso repositório de documentos que deixa muito a desejar quando precisamos recuperar a informação de que temos necessidade.

Procurou-se nesse estudo avaliar os *sites* de instituições públicas federais e estaduais como forma de identificar a qualidade dos serviços prestados, sobretudo àqueles relacionados com o atendimento de referência, que colocam os usuários em contato direto com a pesquisa e, conseqüentemente, com os documentos e se os requisitos de acessibilidade para *web* foram empregados, tornando eficaz o acesso a esses documentos e informações pelos cidadãos.

Primeiramente, discorrer-se-á sobre o conceito de arquivos públicos, a importância dos serviços de atendimento público e as facilidades que a disponibilização desses serviços através do ambiente eletrônico da Internet pode trazer. Em seguida, será analisada a eficiência do acesso a esses documentos e o como a legislação brasileira interfere nesse aspecto.

Assim sendo, os procedimentos metodológicos utilizados para a realização das avaliações dos *sites* serão apresentados, bem como os resultados alcançados serão relatados no tópico seguinte. Considerações finais serão apontadas como forma de conclusão do presente trabalho.

## **Sobre a cantiga de roda: contextualizando espaços, sujeitos e ações**

Parafraseou-se a cantiga infantil - que as crianças entoavam para pular corda - para que o verso possa ilustrar os atores, os cenários e as atitudes que se esperam de um serviço de referência e da acessibilidade dos sites de instituições arquivísticas públicas no Brasil

Dessa forma, o “homem” representa o usuário, a comunidade ou público alvo da instituição. Se em bibliotecas e museus a gama de usuários é extremamente variada e de difícil determinação, em arquivos fala-se mais em comunidade ou público-alvo. Ao conjunto desses utilizadores pode-se chamar de comunidade, que poderá ser formada por membros de uma família, quando se tratar de um arquivo pessoal; membros de uma organização, quando se tratar de um arquivo particular ou de uma empresa; ou ainda, membros da sociedade, quando se tratar de um arquivo público.

Em ambiente *web* essa diversidade é ainda mais marcante. Qualquer indivíduo poderá, por motivos de difícil mensuração, em sua “navegação” pela Internet, deparar-se com o *site* de uma instituição arquivística e realizar uma pesquisa através das ferramentas disponíveis.

Os arquivos e seus *sites* estão representados, na cantiga, pela expressão “na minha porta”. A porta remete tanto a imagem daquilo que se abre para seu público quanto para o portal de informações que acessamos através das novas tecnologias. A porta pode ser física, representada pelos prédios que armazenam a documentação arquivística do Estado, quanto pode ser virtual, ou seja, o elo eletrônico entre os cidadãos e o arquivo público. Nesse ponto destacam-se, principalmente, questões de acessibilidade e usabilidade, que definirão se a “porta” é amigável ao usuário ou não.

O “eu” da cantiga infantil representa o profissional que atua em arquivos, seja o arquivista e outras concepções do profissional da informação, ou o especialista da ciência da computação. Nesse ínterim, fala-se de ética, de responsabilidade social e de compromisso com a história e a memória. No ambiente eletrônico, embora o sujeito se “esconda” fisicamente, é seu trabalho que garantirá o sucesso da pesquisa e acesso ao documento e à informação.

O verbo “abri” representa, sobremaneira o atendimento de referência e a disseminação da informação. Em se tratando de sites de instituições arquivísticas, vale ressaltar a importância de definir questões de segurança de acesso, meios de acesso e reprodução de documentos.

Cabe às avaliações dos serviços oferecidos e da acessibilidade definirem se a porta realmente se abre ao homem.

## **Arquivos públicos, o serviço de referência e a acessibilidade**

As definições que circundam o termo “arquivo” caminham por apontar um conjunto de documentos produzidos pelas atividades administrativas de uma pessoa, entidade ou família, possuidores de valores probatórios, legais, administrativos e, conseqüentemente, históricos.

Com os arquivos públicos não é diferente. Os documentos acumulados pelas atividades desenvolvidas pelo Estado constituem arquivos e são de suma importância para que a sociedade civil possa garantir seus direitos, exercer plenamente a cidadania, acompanhar as atividades governamentais, a fim de que haja transparência nas administrações e para que a história do país possa ser preservada. Características como essas fazem com que

os olhares se voltem para as instituições arquivísticas federais, estaduais e municipais, reconhecendo-as como instituições que existem para a referência e pesquisa da população. São, e devem ser sumariamente, instituições respeitadas e que cumpram os seus deveres de formar, informar e garantir direitos a sociedade. A lei nº 8.159, de 8 de janeiro de 1991 – Lei de Arquivos, estabelece, em seu artigo primeiro, que:

É dever do Poder Público a gestão documental e a proteção especial a documentos de arquivo, como instrumento de apoio à administração, à cultura, ao desenvolvimento científico e como elementos de prova e informação (BRASIL, 1991) .

Os arquivos públicos existem primeiramente para aumentar a eficiência governamental:

com o decorrer do tempo, um governo naturalmente acumula tantos documentos que se faz mister tomar providências em relações aos mesmos. Quando tais documentos obliteram as repartições e dificultam a execução dos trabalhos, ocupando espaço valioso, servem como uma advertência diária de que alguma providência precisa ser tomada (SCHELLENBERG, 2002, p. 31).

Em seguida existem por razões de ordem cultural:

Os arquivos públicos constituem um tipo de fonte de cultura entre muitos outros tipos como livros, manuscritos e tesouros de museus. São uma fonte tão importante como os parques, monumentos ou edifícios. Desde que produzidos por um governo são fontes, particularmente, desse governo. (...) O cuidado com os valiosos documentos oficiais é, portanto, uma obrigação pública (SCHELLENBERG, 2002, p. 31).

Aumentar a eficiência governamental significa, antes de tudo, garantir aos cidadãos o cumprimento do direito de acesso a informação:

O grau de democratização do Estado encontra, na sua visibilidade, um elemento balizador: maior o acesso à informação governamental, mais democráticas as relações entre o Estado e sociedade civil. A visibilidade social do Estado constitui um processo de dimensões políticas, técnicas, tecnológicas e culturais, tendo como um dos seus produtos fundamentais a informação “publicizada”. Instrumentalizado para exercer o acesso à informação governamental, o cidadão configura-se como sujeito informativo. Isto é particularmente significativo ao se considerar as diferentes formas de intervenção do Estado na vida social. O acesso à informação governamental, encontra, a partir dos anos 70, em Estados de características democráticas, diversas formas de expressão, traduzidas em duas noções que se referenciam mutuamente: transparência administrativa e direito à informação (JARDIM, 1999<sup>a</sup>, p. 49).

Por outro lado, preocupar-se com as razões de ordem cultural não significa focar os trabalhos arquivísticos em arquivos permanentes, deixando de lado o trabalho com os arquivos correntes e intermediários. Significa realizar um trabalho contínuo, que respeite o ciclo de vida documental, deixando um legado histórico às futuras gerações,

mas não sonhando informações e direitos às gerações atuais. Racionalizar o trabalho arquivístico e otimizar o acesso da população às massas documentais do Estado implica em desenvolver o uso social da informação.

O serviço de referência em arquivos constitui-se basicamente de consultas de documentos em salas especiais, empréstimo e reprodução de documentos, serviços de informações<sup>86</sup> e disponibilização de todos os serviços através da rede, ou seja, *on-line*.

Originados como uma medida de juntar as novas tecnologias à necessidade de, cada vez mais, aumentar o âmbito de ação e atendimento à população, os *sites* de arquivos públicos são grandes exemplos de atendimento de referência:

O *website* de uma instituição arquivística deve ser visto como um instrumento de prestação de serviços – dinâmico e atualizável – e não simplesmente como a reprodução de um folder institucional. Trata-se, na verdade, de um espaço virtual de comunicação com os diferentes tipos de usuários da instituição a ser gerenciado como parte da política de informação da instituição. Dado o potencial e as características da Internet, este espaço, além de redefinir as formas de relacionamento com os usuários tradicionais, poderá atrair outros que, por várias razões, difícil ou raramente procurariam o Arquivo como realidade física (CONARQ, 2000).

Segundo o CONARQ (2000), a concepção do *site* deve prever sua utilidade para usuários de áreas como: atendimento ao cidadão; educação de usuários; pesquisa científica; e atendimento ao governo.

Duas observações de OHIRA; SCHENKEL e SILVEIRA (2003) são pertinentes para configurar esse novo cenário construído pela junção de uma área tradicional – os arquivos públicos – e uma área totalmente inovadora e atual – a utilização das tecnologias de informação:

Com a expansão da Internet, a questão do acesso, e portanto a questão do relacionamento com a clientela da instituição, ganha nova dimensão devido à crescente demanda por serviços automatizados de recuperação de documentos e informações via rede. O acesso remoto nos coloca, num futuro não muito distante, frente a duas situações: por um lado tende a esvaziar as salas de consultas, reduzindo em muito o contato direto com o usuário e por outro lado, abre a possibilidade de ampliação, de modo inimaginável, o leque da clientela [...] A tecnologia de informação surge como um recurso utilizado no processamento eletrônico de dados levando as organizações que têm por finalidade armazenar, tratar e disseminar documentos e informações a rever suas atribuições, seus serviços e seus produtos. A Internet, como canal de comunicação oferece inúmeros recursos que podem ser explorados no sentido de dar acesso à imensa quantidade de informações e documentos armazenados nos Arquivos Públicos Estaduais do Brasil. *Home pages* são criadas pelas instituições arquivísticas que querem divulgar sua imagem, seus serviços e seus produtos e mudar seu padrão de relacionamento com o público.

---

<sup>86</sup> “Este tipo de serviço de referência consiste em fornecer informações extraídas de documentos, ou informações sobre os documentos, distinguindo-se do trabalho de fornecer os próprios documentos, ou cópias, para uso” (SCHELLENBERG, 1999, p. 357)

O Modelo de Acessibilidade do Governo define assim a acessibilidade:

Acessibilidade diz respeito a locais, produtos, serviços ou informações efetivamente disponíveis ao maior número e variedade de pessoas independente de suas capacidades físico-motoras e perceptivas, culturais e sociais. Isso requer a eliminação de barreiras arquitetônicas, a disponibilidade de comunicação, de acesso físico, de equipamentos e programas adequados, de conteúdo e apresentação em formatos alternativos.

A lei decretada pelo presidente da república em dois de dezembro de 2004, um dia antes do dia mundial do deficiente, em seu artigo nº 47, trata exclusivamente da acessibilidade para web. A partir de dois de dezembro de 2005, exatamente doze meses a contar da data de publicação do Decreto, todos os portais e sites da administração pública deveriam ter obrigatoriamente seus conteúdos acessíveis para pessoas portadoras de deficiência visual. Aqueles que não cumprissem as recomendações do governo ficariam sujeitos às penalidades citadas no decreto, como o corte de recursos públicos e empréstimos, sanções administrativas, cíveis e penais cabíveis, previstas em lei, entre outras.

Como se pode perceber, a lei é excludente quanto aos deficientes motores, auditivos e mentais e também não mencionou nada sobre acessibilidade em dispositivos e plataformas.

O Modelo de acessibilidade de governo eletrônico (e-MAG) consiste em um conjunto de recomendações a ser considerado para que o processo de acessibilidade dos sites e portais do governo brasileiro seja conduzido de forma padronizada e de fácil implementação. A primeira versão do e-MAG foi disponibilizada para consulta pública em 18 de janeiro de 2005 e a versão 2.0 já com as alterações propostas, em 14 de dezembro do mesmo ano.

Em 2007, a portaria nº 3, de 7 de maio, institucionalizou o e-MAG no âmbito do sistema de Administração dos recursos de informação e informática (SISP), tornando sua observância obrigatória nos sites e portais do governo brasileiro. Dessa vez, o texto introdutório da portaria foi mais abrangente em relação às pessoas portadoras de necessidades especiais:

Considerando a necessidade de assegurar a todos os interessados, independentemente de suas capacidades físico-motoras, perceptivas, culturais e sociais, o acesso à informação disponível, resguardados os aspectos de sigilo, restrições administrativas e legais, e em respeito a valores republicanos e democráticos de igualdade, respeito e transparência...

Segundo o último censo do IBGE, 14,5% da população brasileira têm algum tipo de deficiência. São maioria os casos de problema de visão: 48,1%. A seguir, vêm os casos de deficiência motora (22,9%), auditiva (16,7%), mental (8,3%) ou física (4,1%). Esses números servem para alertar que seria uma grande irresponsabilidade privar mais de 24 milhões de pessoas do acesso à informação pública e, conseqüentemente, do acesso ao estado.

Já existem no país documentos com recomendações de boas práticas com o objetivo de aprimorar a comunicação e o fornecimento de informações e serviços prestados por meios eletrônicos pelos órgãos do Governo Federal. Essas iniciativas partiram do Comitê Executivo do Governo Eletrônico. São exemplos desses documentos: o e-MAG, o Livro Verde e o e-PING. Além disso, há o ASES, que é um avaliador e simulador de acessibilidade de sites governamentais.

## Procedimentos metodológicos

Neste trabalho, primeiramente, procurou-se avaliar os *sites* de arquivos públicos federal e estaduais em alguns itens que se relacionavam ao atendimento de referência e apresentação de serviços de acesso à informação *on-line*. Para tanto, foi feita uma análise heurística do material disponibilizado na *web* pela instituição.

Para o presente estudo utilizou-se a lista de endereços eletrônicos de arquivos públicos federal e estaduais fornecida pelo Conarq<sup>87</sup>, e as instituições avaliadas foram: Arquivo Nacional, Arquivo Público do Distrito Federal, Arquivo Público do Estado da Bahia, Arquivo Público do Estado do Ceará, Arquivo Público do Estado do Espírito Santo, Arquivo Público Mineiro, Arquivo Público do Pará, Arquivo Público do Estado do Paraná, Arquivo Público do Estado do Rio Grande do Norte, Arquivo do Estado do Rio de Janeiro, Arquivo do Estado de São Paulo e Arquivo Público do Estado de Santa Catarina. Não foram avaliados os Arquivos Municipais listados porque o número de endereços eletrônicos disponíveis era pouco significativo.

Procurou-se quantificar o número de arquivos públicos federais e estaduais que possuísem os seguintes critérios:

- Informações sobre os serviços oferecidos pela instituição;
- Explicação dos fundos, séries, itens documentais etc.;
- Existência de mecanismos de busca;
- Existência de “fale conosco” ou informações para contato com a instituição;
- A apresentação dos registros;
- Disponibilização de documentos digitalizados;
- Fornecimento de *links* relacionados aos acervos que acumulam.

Em seguida, foram aplicados testes automáticos de acessibilidade nos sites pesquisados, utilizando-se uma ferramenta *on-line* chamada DaSilva<sup>88</sup>. Essa ferramenta avalia o nível de acessibilidade das páginas pesquisadas, dentro de um conjunto de regras preconizadas pelo E-gov brasileiro, por meio do e-MAG. Como resultado, os testes produzem automaticamente relatórios detalhados segundo os três níveis de prioridades:

- Prioridade 1: pontos que os criadores de conteúdo web devem satisfazer inteiramente. Se não o fizerem, um ou mais grupos de usuários ficarão impossibilitados de acessar as informações contidas nos documentos;
- Prioridade 2: pontos que os criadores de conteúdo web deveriam satisfazer. Se não o fizerem, um ou mais grupos de usuários terão dificuldades em acessar as informações contidas nos documentos;
- Prioridade 3: pontos que os criadores de conteúdo web podem satisfazer. Se não o fizerem, um ou mais grupos de usuários poderão se deparar com algumas dificuldades em acessar as informações contidas nos documentos.

Os sites que estiverem dentro da padronização, ou seja, que não apresentarem discordâncias com as regras do e-gov nas três prioridades, recebem o selo AAA que poderá ser afixado no próprio site, a fim de identificá-lo com um “site acessível”.

---

<sup>87</sup> Disponível em: <http://www.conarq.arquivonacional.gov.br/cgi/cgilua.exe/sys/start.htm>. Data de acesso: 15/01/2009.

<sup>88</sup> Disponível em: [www.dasilva.org.br](http://www.dasilva.org.br). Acesso em 16 jan. 2009.

## Resultados

No que se refere aos serviços oferecidos, encontrou-se um número maior de opções, que vão de serviços tradicionais de atendimento ao usuário à algumas inovações, originadas a partir da disponibilização dos sites. Nem todos os serviços podem ser utilizados remotamente através do site, mas, na maioria das vezes, o site serve como um primeiro passo para o conhecimento e solicitação dos serviços oferecidos. O quadro a seguir (QUADRO 1) mostra os serviços listados pelos Arquivos Públicos analisados.

QUADRO 1  
Serviços oferecidos pelos Arquivos Públicos

Reprodução (micrográfica, fonográfica etc.);
Atendimento local;
Atendimento à distância (solicitação e emissão de documentos);
Consultas locais ao acervo;
Consultas remotas ao acervo;
Mapotecas;
Bibliotecas de apoio;
Laboratórios fotográficos;
Salas de consultas;
Exposições temáticas;
Notícias;
Bibliotecas digitais;
Expedição de certidões e autenticação de documentos;
Venda de publicações;
Download de publicações e acesso à revistas eletrônicas;
Organização de acervos de outras instituições públicas;
Sala de informática;
Microfilmagem;
Paleografia;
Livraria;
Gestão de documentos.

Fonte: dados da pesquisa.

A tabela abaixo (TAB. 1) apresenta a avaliação dos sites com relação aos serviços oferecidos por eles para auxiliar os cidadãos ao acesso às informações.

**TABELA 1**  
Avaliação dos sites com relação aos serviços de referência

<b>Arquivo Público</b>	Arquivo Nacional	Distrito Federal	Bahia	Ceará	Espírito Santo	Mineiro	Pará	Paraná	Rio Grande do Norte	Rio de Janeiro	São Paulo	Santa Catarina	<b>Total</b>	<b>%</b>
<b>Itens Avaliados</b>														
Explicação de fundos e séries	1	1			1	1		1		1	1	1	8	67
Mecanismo de busca	1					1		1			1	1	5	42
Informações de contato	1	1			1	1	1	1		1	1	1	9	75
Descrições de registro	1	1			1	1		1		1	1	1	8	67
Documentos digitalizados	1	1				1		1		1	1	1	7	58
Links Relacionados	1				1	1	1	1		1	1		7	58

Fonte: Dados da pesquisa.

O Arquivo Nacional, o Arquivo Público do Paraná e o Arquivo Público de São Paulo destacam-se por apresentar todos os itens requeridos na avaliação. Os Arquivos Públicos da Bahia, Ceará e Rio Grande do Norte são instituições vinculadas às secretarias de cultura de seus estados e possuem apenas uma página, cujas informações são insuficientes para a avaliação proposta neste estudo. O Arquivo Público Mineiro possui uma página vinculada à Secretaria de Cultura do Estado de Minas Gerais, entretanto, nesta página há um link para o um site próprio para a pesquisa dos documentos, que foi utilizado para esta avaliação. O Estado do Pará é representado pela Empresa de Processamento de Dados do Estado do Pará (PRODEPA), fundada em 1953. Diferentemente das demais instituições, a PRODEPA está bem voltada para a gestão de documentos.

De maneira geral, as instituições apresentam informações para contato (75%), o que garante o acesso, mesmo que por outros meios de comunicação, aos serviços prestados. A grande maioria delas (67%) descreve em seus sites os acervos que acumulam e preservam e detalha os registros existentes, o que facilita a busca pela informação ao usuário. Infelizmente somente 42% dessas instituições fornecem mecanismos de busca eletrônicos e 58% apresentam os documentos digitalizados (mesmo que não seja a totalidade de documentos existentes).

A tabela seguinte (TAB. 2) apresenta os resultados obtidos com relação a acessibilidade.

**TABELA 2**  
Avaliação dos sites com relação à acessibilidade (Prioridade 1)

<b>Prioridade 1</b>		
<b>Arquivo Público</b>	<b>Erros</b>	<b>Avisos</b>
Arquivo Nacional	1	6
Distrito Federal	14	12
Bahia	1	9
Ceará	13	144
Espírito Santo	2	10
Mineiro	17	93
Pará	1	213
Paraná	14	101
Rio Grande do Norte	27	99
Rio de Janeiro	2	12
São Paulo	58	147
Santa Catarina	16	165

Fonte: Dados da pesquisa.

Apesar de possuírem poucos erros com relação à prioridade 1, os sites dos arquivos públicos Nacional, do Pará e da Bahia trazem problemas graves. O site do arquivo Nacional não possui um endereço eletrônico único para cada página, ou seja, é como se todas as páginas do site tivessem o mesmo endereço. Isso fere um dos princípios básicos da acessibilidade para web descritos na cartilha do e-MAG, que diz que todo documento deve ter um endereço único na Internet, o que força o usuário a ter que, todas as vezes que for recuperar uma informação, começar a pesquisa pela página inicial e navegar até o ponto desejado.

A avaliação dos sites dos Arquivos Públicos da Bahia, do Pará e do Rio Grande do Norte, foi prejudicada, uma vez que essas instituições possuíam uma página apenas a ser analisada pelo instrumento.

As tabelas que se seguem (TAB. 3 e 4) apresentam os resultados obtidos pela avaliação das prioridades 3 e 4:

**TABELA 3**  
Avaliação dos sites com relação à acessibilidade (Prioridade 2)

<b>Prioridade 2</b>		
<b>Arquivo Público</b>	<b>Erros</b>	<b>Avisos</b>
Arquivo Nacional	3	3
Distrito Federal	1	3
Bahia	6	6
Ceará	13	42
Espírito Santo	3	4
Mineiro	3	38
Pará	0	70
Paraná	55	22
Rio Grande do Norte	138	47
Rio de Janeiro	2	4
São Paulo	56	90
Santa Catarina	46	55

Fonte: Dados da pesquisa.

**TABELA 4**  
Avaliação dos sites com relação à acessibilidade (Prioridade 3)

<b>Prioridade 3</b>		
<b>Arquivo Público</b>	<b>Erros</b>	<b>Avisos</b>
Arquivo Nacional	0	9
Distrito Federal	0	10
Bahia	16	9
Ceará	0	100
Espírito Santo	0	10
Mineiro	0	54
Pará	0	110
Paraná	2	60
Rio Grande do Norte	1	67
Rio de Janeiro	1	9
São Paulo	0	49
Santa Catarina	15	58

Fonte: Dados da pesquisa.

Com relação à prioridade 2, a grande maioria dos usuários portadores de deficiências teria dificuldade para acessar os sites das instituições. Mesmo se tratando de pontos que os criadores de conteúdos deveriam atender, o atendimento seria extremamente importante e demonstraria um cuidado especial com esses usuários.

O resultado com a avaliação da prioridade 3 é mais satisfatório, mas, ainda assim, algumas instituições apresentaram problemas com pontos que poderiam ser atendidos. Mesmo as instituições que possuíam somente uma página, não foram capazes de atender os pontos exigidos pelas prioridades.

## **Considerações finais**

De uma forma geral pode-se afirmar que os Arquivos Públicos avaliados nessa pesquisa estão preocupados com a guarda e a organização de documentos permanentes. A necessidade de acesso aos documentos das fases corrente e intermediária garantirá ao cidadão um trabalho maior para recorrer aos diversos órgãos que produziram tais documentos. Se essas organizações realizam o trabalho de lidar com os documentos nas fases iniciais do ciclo documental não o anunciam em seus *sites*.

Um número cada vez maior e diversificado de serviços é oferecido eletronicamente e outros tantos são comunicados virtualmente para que o cidadão tome consciência de seus direitos e das possibilidades de acesso à informação. Os arquivos mostram-se mais preocupados com a guarda e a organização dos documentos e compreendem uma gama interessante de serviços destinados ao atendimento de referência. Disponibilizar documentos digitalizados e atender virtualmente solicitações de usuários são alguns exemplos de serviços oferecidos que facilitariam a pesquisa de um usuário remoto, mas poucas instituições efetivamente os realizam.

Embora exista uma lei que visa garantir o acesso às pessoas portadoras de deficiência aos sites de administração pública desde 2004, o que se percebeu foi uma desobediência tolerada. A grande maioria dos sites dos arquivos públicos não seguiu as normas de acessibilidade constantes no e-MAG. É claro que,

aquisição do selo de acessibilidade, bem como a aprovação nos testes automáticos não assegura que os sites sejam, efetivamente, acessíveis, mas é primordial que os documentos eletrônicos sejam disponibilizados na rede seguindo os padrões estabelecidos de modo a garantir sua interoperabilidade.

Vale ressaltar que os resultados da avaliação com relação à acessibilidade não demonstram que as instituições negligenciaram os requisitos, mas que talvez, os responsáveis pelos sites disponíveis desconhecem a existência dessas normas. Tal constatação exige, o mais rápido possível, que essas normas sejam melhor divulgadas pelo governo federal e, rapidamente, atendidas pelas instituições públicas.

Percebe-se que diminuir a opacidade governamental com relação às informações sobre as atividades governamentais, alertada por JARDIM (1999a), está longe de acontecer, ao menos através dos *sites* avaliados. Tais instrumentos mostram-se como ferramentas para a pesquisa histórica, mais do que como suportes para o exercício dos direitos dos cidadãos. São páginas eletrônicas que confortam a curiosidade e a pesquisa acadêmica e não a necessidade do exercício da cidadania plena.

Caso se atente para o título deste trabalho uma vez mais poderá se perceber que existe um ponto de interrogação ao final. Isso significa que não é ainda certo que as instituições estão prontas para abrirem suas portas virtuais aos cidadãos. Cabem aos profissionais que nelas trabalham reunir esforços para que as normas sejam devidamente atendidas e que os serviços oferecidos sejam eficiente e eficazmente oferecidos.

## Referências

BRASIL. lei nº 8.159, de 8 de janeiro de 1991 – Lei de Arquivos. *Dispõe sobre a política nacional de arquivos públicos e privados e dá outras providências*. Brasília: Casa Civil, 1991. Disponível em: <[https://www.planalto.gov.br/ccivil\\_03/LEIS/L8159.htm](https://www.planalto.gov.br/ccivil_03/LEIS/L8159.htm)>. Data de acesso: 06/01/2006.

BRASIL. *Constituição da República Federativa do Brasil de 1988*. Brasília: Casa Civil, 2006. Disponível em: <[https://www.planalto.gov.br/ccivil\\_03/Constituicao/Constitui%C3%A7ao.htm](https://www.planalto.gov.br/ccivil_03/Constituicao/Constitui%C3%A7ao.htm)>. Data de acesso: 06/01/2006.

CENTRO NACIONAL DE DESENVOLVIMENTO DO GERENCIAMENTO DA INFORMAÇÃO - CENADEM. *Carta para a preservação do patrimônio arquivístico digital: preservar para garantir o acesso*. Disponível em: <[http://www.cenadem.com.br/news\\_carta.php](http://www.cenadem.com.br/news_carta.php)> . Data de acesso: 06/01/2006.

CEPIK, M. Direito à informação: situação legal e desafios. *Revista IP*, v. 2, n. 2, dez. 2000, p. 27-46.

CONSELHO NACIONAL DE ARQUIVOS – CONARQ. *Diretrizes gerais para a construção de websites de instituições arquivísticas*. Rio de Janeiro: CONARQ, 2000. 11 p.

CONSELHO NACIONAL DE ARQUIVOS – CONARQ. *Subsídios para inserção do segmento dos arquivos no programa sociedade da informação no Brasil*. Rio de Janeiro, CONARQ, 2001. Disponível para download em: <[www.conarq.gov.br](http://www.conarq.gov.br)>. Data de acesso: 06/01/2006.

JARDIM, J. M. *Transparência e opacidade no Brasil: usos e desusos da informação governamental*. Niterói: EdUFF, 1999a. 239 p.

JARDIM, José Maria. *O acesso à informação arquivística no Brasil: problemas de acessibilidade e disseminação*.

In: MESA REDONDA SOBRE ARQUIVOS, 1999, Rio de Janeiro. Rio de Janeiro : Arquivo Nacional, 1999b. Disponível em: <<http://www.arquivonacional.gov.br/pub/virtual/>> Data de acesso: 06/01/2006.

MINAS GERAIS. Secretaria de Estado de Planejamento e Gestão – SEPLAG. Resolução nº 72, de 27 de novembro de 2003. *Estabelece diretrizes para estruturação, elaboração e administração de sítios de informação de serviços públicos na Internet dos Órgãos e Entidades da Administração Pública Direta, Autárquica e Funcional do Poder Executivo*. Belo Horizonte: SEPLAG, 2003.

OHIRA, M. L. B.; SCHENKEL, M. B. C.; SILVEIRA, C. *Crerios para avaliaão do conteúdo dos sites dos arquivos públicos estaduais do Brasil*. In: Ciberética - Simpósio Internacional de Propriedade Intelectual, Informação e Ética, II, 2003, Florianópolis. Anais... Florianópolis, 2003. 20 p. Disponível em: <[www.ciberetica.gov.br](http://www.ciberetica.gov.br)> . Data de acesso: 06/01/2006.

SHELLENBERG, T. R. *Arquivos modernos: princípios e técnicas*. 2 ed. Rio de Janeiro: FGV, 2002. 388 p.

## Como integrar sítios de instituições arquivísticas num contexto de baixa padronização de suas descrições?

Galba Di Mambro  
José Guilherme Ribeiro

Nos últimos cinco ou seis anos tem aumentado o número de instituições arquivísticas que mantêm sítio na Web. Esse número tende a crescer ainda mais no futuro. Muitos destes sítios procuram dar acesso à descrição dos acervos das instituições, através de algum sistema de banco de dados, ou outro. Teoricamente, tais descrições deveriam ser elaboradas tendo como referência as normas internacional (ISAD(G)) e nacional (NOBRADE) de descrição arquivística, mas nem sempre esse é o caso. Tais normas nem sempre são adotadas, havendo ainda os casos de descrições legadas, elaboradas em período anterior à sua edição. Delineia-se então o caso de um conjunto disperso de descrições arquivísticas, baseado em sistemas diversificados, com estruturas diversas, difuso na Web.

Nessa situação, em algum momento deverá ser colocada a questão de uma integração desses sítios, e uma possibilidade seria a constituição de um ou mais portais de acesso, oferecendo um ponto de entrada unificado para, pelo menos, uma parte deles. O presente trabalho mostra como esse portal unificado poderia ser constituído, tendo em vista o problema do baixo grau de padronização das descrições.

Um portal como esse poderia ser simples, tendo como modelo os mecanismos de busca da Internet, como Google ou Yahoo!. Podemos imaginar uma página que tivesse uma caixa de texto, onde o usuário digitasse uma expressão de pesquisa. Ao clicar no botão “Pesquisar”, o sistema realizaria uma busca na sua base de dados, e mostraria uma página com resultados. Essa página de resultados ofereceria informações mínimas sobre os elementos encontrados, e possuiria links. Ao clicar num desses links, o usuário seria levado à descrição detalhada, já na página da instituição de origem. Se o leitor observar com atenção, esse roteiro é, essencialmente, uma descrição do comportamento do Google ou do Yahoo!, citados.

Para implementar esse roteiro, precisamos de dois elementos. Em primeiro lugar, precisamos de metadados sobre os registros de origem. Um registro corresponde à descrição de um elemento arquivístico (fundo, série, subsérie, item, etc.), e considera-se que as instituições produziram metadados sobre tais elementos, mesmo que não obedecendo, necessariamente, às normas de descrição. Parte desses metadados seria mostrada ao usuário, na forma de uma página de resultados. Em segundo lugar, precisamos pesquisar a expressão digitada pelo usuário. O procedimento mais simples é usar o texto dos metadados para alimentar um ou mais índices, e fazer uma busca por palavras, contra esses índices, quando solicitado. Existem, portanto, duas funções – visualização e pesquisa – e, para os dois casos, precisamos de metadados.

Um dos modos de obter tais metadados é a coleta automática, ou *harvesting*<sup>89</sup>. Esse modelo supõe que exista uma base de dados centralizada, chamada de *provedor de serviços*, que recolhe, através de um procedimento automatizado, metadados fornecidos pelos *provedores de dados*. O provedor de serviços é responsável por sistematizar os metadados e por oferecer, com base neles, algum serviço de acesso, como a consulta e a pesquisa. Os provedores

---

<sup>89</sup> O modelo de coleta automática de metadados é a base do protocolo OAI-PMH, mencionado mais à frente. Os conceitos de *provedor de dados* e *provedor de serviços* fazem parte do quadro de referência desse protocolo.

de dados fornecem os metadados para coleta. Periodicamente, por meio de um protocolo padronizado, o provedor de serviços consulta os provedores de dados, para saber se houve atualizações ou acréscimos à base de dados. Em caso positivo, uma cópia dos metadados atualizados é importada para o provedor de serviços. Para operar, esse modelo supõe a existência de dois elementos: um protocolo de comunicação e padrões de metadados convencionados.

Um protocolo para a coleta automática de metadados é o protocolo OAI-PMH<sup>90</sup>. Ele surgiu como instrumento de difusão de trabalhos científicos em meio eletrônico (*e-prints*), e tem sido amplamente utilizado por bibliotecas virtuais e por outras iniciativas. Sua adoção ofereceria a vantagem de utilizar um padrão já largamente testado, que possui uma comunidade de usuários ampla e difundida.

Quanto aos metadados, qualquer discussão a esse respeito deve considerar a Norma Brasileira de Descrição – NOBRADE<sup>91</sup>. A NOBRADE é a norma de descrição recomendada para o Brasil pelo Conselho Nacional de Arquivos, e corresponde à adaptação da norma internacional – ISAD(G) – ao contexto nacional. Uma rede que vise o intercâmbio de descrições arquivísticas teria, necessariamente, que considerar a adoção da NOBRADE como padrão.

Um problema vem do fato de que o protocolo OAI exige que os metadados sejam expressos numa linguagem padronizada, a XML<sup>92</sup>. A NOBRADE é um padrão de descrição que não se prende a um formato de dados particular. Uma distinção que pode ajudar a resolver esse problema é a que faz Michael Fox, em certo artigo, entre padrões de estrutura de dados e padrões de comunicação<sup>93</sup>. Os primeiros definem “o que se pode dizer” a respeito de uma coleção ou item – definem um conjunto de dados a serem sistematicamente recolhidos. Os últimos correspondem a formatos através dos quais os metadados podem ser compartilhados entre instituições. Muitos padrões se encaixam em mais de uma categoria, mas, de certo modo, pode-se ver a ISAD(G) como um padrão de estrutura de dados, e a EAD (Encoded Archival Description) como um padrão de comunicação. A rede poderia fixar, então, a NOBRADE como padrão de estrutura de dados e um outro padrão XML, como EAD<sup>94</sup> ou Dublin Core<sup>95</sup>, como padrão de comunicação.<sup>96</sup>

Para todos os efeitos, portanto, os formatos descritivos *ad-hoc*, usados nas descrições institucionais, teriam que se referir à NOBRADE. Aí aparece outra dificuldade, na medida em que o processo de conversão pode ser mais difícil do que se pode esperar.

---

<sup>90</sup> OAI-PMH é a sigla para Open Archives Initiative – Protocol for Metadata Harvesting. Uma visão geral da história, do significado e do funcionamento do protocolo pode ser encontrada em BARRUECO CRUZ, José Manuel, SUBIRATS COLL, Imma. Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH): descripción, funciones y aplicación de un protocolo. **El Profesional de la Información**, 2003, vol. 12, n. 2, p. 99-106. (disponível em <http://eprints.rclis.org/177/> - acesso em 18-01-2009). Mais informações podem ser encontradas no sítio do Open Archives Initiative - <http://www.openarchives.org/> (acesso em 18-01-2009).

<sup>91</sup> CONSELHO NACIONAL DE ARQUIVOS - CONARQ. Câmara Técnica de Descrição Arquivística. **NOBRADE**. Norma Brasileira de Descrição Arquivística. Rio de Janeiro: CONARQ, 2006. 123 p.

<sup>92</sup> A XML é uma linguagem, proposta em 1998 pelo consórcio internacional W3C, que tem servido, entre outras funções, ao desenvolvimento de padrões para a troca de informações entre sistemas de computadores. Um sítio que contém uma grande relação de recursos sobre a XML (livros, bibliografias, material on-line, etc.) é a página do W3C (<http://www.w3.org/XML/> - acesso em 10-07-2007). O leitor pode, ainda, recorrer ao livro de Harold & Means sobre o tema - HAROLD, Elliott Rusty, MEANS, W. Scott. **XML in a nutshell: a desktop quick reference**. 2.ed. Sebastopol (USA): O'Reilly, 2002.

<sup>93</sup> Na classificação de M. Fox, os padrões internacionais podem ser divididos em 4 categorias: padrões de estrutura de dados, de conteúdo, de valores de dados e de comunicação. A função de armazenamento (“data storage”) também é mencionada. O leitor interessado pode consultar FOX, Michael J. Stargazing: locating EAD in the descriptive firmament. In: PITTI, Daniel V., DUFF, Wendy M.(Ed.) **Encoded archival description on the Internet**. New York: The Haworth Press, 2001. p. 61-74. (Volume publicado simultaneamente no **Journal of Internet Cataloging**, v. 4, n. 3-4, 2001. Cópias do artigo podem ser compradas em <http://www.HaworthPress.com> - acesso em 15-09-2007).

<sup>94</sup> A EAD – Encoded Archival Description – é um padrão mantido pela Biblioteca do Congresso dos EUA e pela Associação do Arquivistas Americanos, e que serve para a codificação de descrições arquivísticas. No site oficial da EAD podem-se encontrar textos de referência sobre a mesma, bem como *links* para outros sites que usam a EAD e/ou que contém mais material sobre ela. (Cf. <http://www.loc.gov/ead/> - acesso em 23-04-2007). Muito material de referência pode ser encontrado no site Cover Pages sobre a EAD (<http://xml.coverpages.org/ead.html> - acesso em 12-07-2007).

<sup>95</sup> O Dublin Core Element Set é um conjunto de metadados para a descrição de objetos digitais. É mantido pela Dublin Core Metadata Initiative, uma iniciativa internacional que visa o desenvolvimento de mecanismos para a pesquisa de recursos em ambiente eletrônico. Mais informações sobre Dublin Core podem ser obtidas na página do Cover Pages - <http://xml.coverpages.org/dcmi.html> (acesso em 18-01-2009). O site oficial da iniciativa é <http://dublincore.org/> (acesso em 18-01-2009). O formato Dublin Core é independente da linguagem XML, ou de qualquer outro formato específico. No entanto, existem esquemas XML recomendados para representá-lo naquela linguagem. Para mais detalhes, consulte <http://dublincore.org/schemas/xmls/> (acesso em 18-01-2009).

<sup>96</sup> O papel do padrão de comunicação não deve ser subestimado. O modelo de dados da rede deverá brotar de um exame cuidadoso de todos os padrões envolvidos, de modo a levar em conta as especificidades de cada um. O que estamos afirmando aqui é que, nesse contexto, deve caber à NOBRADE um papel primordial.

Analisando o problema em termos abstratos, temos que, dadas duas estruturas de descrição – a NOBRADE, de um lado, e uma descrição *ad-hoc*, de outro – surgem três possibilidades. Na primeira, um ou mais dos campos da descrição não possuem correspondentes na NOBRADE. Nesse caso, ou os elementos excedentes são incorporados a algum outro campo definido pela Norma, ou ficam de fora do escopo da rede. Num outro caso, um ou mais campos da descrição teriam que ser desmembrados entre os campos da Norma. O desmembramento de um campo entre dois outros (ou mais) é um procedimento delicado e dificilmente automatizável. Finalmente, tem-se a possibilidade inversa, em que alguns campos da descrição *ad-hoc* devam ser fundidos num certo campo da NOBRADE.

O surgimento desses problemas – não-correspondência, desdobramento ou fusão de campos – implica numa dificuldade, num ambiente de rede, na medida em que o seu tratamento exige a consideração de cada caso. Essa questão, em geral, exige o estudo e o questionamento sobre o campo semântico dos elementos da Norma e da descrição – algo que demanda tempo e pessoal qualificado. Em todos os casos, surge um obstáculo à expansão da rede, que se torna mais acentuado quanto menor for a padronização do metadados. Este é justamente o caso que estamos analisando.

Para solucionar esse problema, é possível imaginar alguns procedimentos. O mais imediato seria a discussão aprofundada sobre a norma de descrição, de modo a reduzir o hiato entre as descrições efetivas e a norma, e aprofundar o consenso sobre o significado de cada um de seus elementos. Trata-se de solução que só renderia resultados no médio e longo prazos. Uma alternativa seria a adoção de uma estrutura de transição, que pudesse ser utilizada no curto prazo, para facilitar a operação da rede em condições de baixa padronização, e que pudesse ser removida posteriormente, sem grandes efeitos colaterais, quando a aderência às normas de descrição fosse mais estrita.

Essa abordagem pode partir da retomada das duas funções dos metadados na rede: visualização e pesquisa. A visualização corresponde à criação de uma página com o sumário dos resultados de uma consulta. Nem todos os metadados precisam constar no sumário: apenas aqueles que dêem uma visão preliminar ao internauta, que poderá seguir links para ver, no site original, as informações detalhadas. Esses elementos podem ser reduzidos ao mínimo. O mais importante é o título da unidade de descrição (correspondente ao elemento NOBRADE 1.2). Esse seria o único elemento realmente obrigatório. Ao seu lado, seriam desejáveis outros elementos, como datas (elemento 1.3), nível de descrição (1.4) e código de referência (1.1). Todos os outros elementos da NOBRADE poderiam ser considerados, do ponto de vista do sumário dos resultados, opcionais. Essa disposição – apenas um elemento obrigatório – reduz as necessidades de metadados ao mínimo imprescindível para o funcionamento da rede.

A função de pesquisa seria executada por meio de um ou mais índices.<sup>97</sup> Um índice pode ser visto como uma estrutura de dados que relaciona uma palavra aos documentos em que ela ocorre. Mecanismos de busca como o Google ou Yahoo! não distinguem o local de ocorrência de um termo no interior da página (documento) que estão indexando. Se o usuário entra com uma palavra como critério de pesquisa, o Google retorna as páginas em que essa palavra ocorre, indiferente se essa ocorrência se dá numa tabela, formulário ou divisão específica. Por outro lado, não é possível formular uma busca com o critério “retorne todas as páginas em que a palavra tal apareça na linha

---

<sup>97</sup> Essa passagem ocorreu-nos, inicialmente, do estudo do protocolo SRU e de sua linguagem de consulta, a CQL (sobre esses elementos, v. <http://www.loc.gov/standards/sru/> - acesso em 18-01-2009). Ela é compatível com o funcionamento de mecanismos de busca como o Lucene (sobre o Lucene, v. <http://lucene.apache.org/> - acesso em 18-01-2009, v. tb. GOSPODNETIC, Otis, HATCHER, Erik. **Lucene in action**: a guide to the Java search engine. Greenwich (USA): Manning, 2005.). Para uma descrição dos mecanismos de busca em geral, v. BAEZA-YATES, Ricardo, RIBEIRO-NETO, Berthier. **Modern information retrieval**. New York: ACM Press, 1999.

de título”. O índice trata o documento todo – a página HTML – como um texto contínuo, sem diferenciar entre suas seções ou campos.

Esse princípio poderia ser aplicado à indexação de uma ficha descritiva, estruturada em campos. Para um índice no estilo Google, que trata uma descrição como um todo, interessa que uma determinada palavra (por exemplo, o termo “fotografia”) foi usada em algum lugar da ficha descritiva, não sendo importante determinar se ela ocorreu em “título” ou em “dimensões e suporte”. Para se constituir um índice como esse, pode-se simplesmente abstrair a estrutura dos campos da descrição, adicionando ao índice apenas o conteúdo de cada campo, numa certa ordem. O texto que alimenta o índice seria a concatenação de todos os campos da ficha descritiva, numa ordem pré-determinada, qualquer que fosse a estrutura da ficha.<sup>98</sup>

A conclusão dessa análise é que, numa arquitetura como a que definimos, os índices podem ser alimentados de duas formas alternativas. A primeira, importando do provedor de dados para o provedor de serviços uma ficha estruturada em campos, e adicionando o conteúdo desses campos ao índice. A segunda, extraindo, ainda no provedor de dados, o conteúdo dos campos, adicionando-os numa cadeia de indexação e repassando essa cadeia – que é independente da estrutura de dados utilizada – para o provedor de serviços. Essa forma é independente da estrutura da ficha descritiva, e é a mais adequada à condição de baixa padronização das descrições.

A solução que propomos repousa, então, no uso de dois formatos de dados. Cada registro em um provedor de dados seria exportado duas vezes. A primeira, no formato XML escolhido para representar a NOBRADE. Nesse caso, apenas o campo “título” seria obrigatório, sendo desejável constarem também os campos “datas”, “código de referência” e “nível”. O segundo formato seria o formato de indexação. Nesse caso, seria constituída uma cadeia de indexação para cada índice da base de dados, correspondente à concatenação dos campos em que se define aquele índice. Sugerimos que haja, inicialmente, um único índice “geral”, correspondente à concatenação de todos os campos da ficha descritiva, semelhante ao dos mecanismos de busca da Web.

Esse modelo, com dois formatos de metadados, poderia ser mantido enquanto o uso da NOBRADE não fosse muito difundido, ou enquanto o hiato entre os esquemas de representação *ad-hoc* e a Norma fosse muito acentuado. Na medida em que essa distância se reduzisse, seria possível voltar a trabalhar com apenas um formato de metadados, para as duas funções. O formato de indexação deixaria de ser utilizado, e se ampliaria o número de campos obrigatórios da ficha NOBRADE – de modo a se fazer a indexação com esse mesmo conjunto de metadados. A rede passaria a funcionar, então, com o seu modelo de arquitetura definitivo.

A solução que propomos não resolve, do ponto de vista teórico, o problema de integrar sites institucionais em condições de incipiente padronização. Ela contorna o problema, criando uma estrutura de transição que funcionaria enquanto não fossem encontradas condições ideais. Sua vantagem está em repousar no uso de um modelo aberto, o OAI-PMH, amplamente testado e utilizado. Ela pode ser vista como uma solução provisória, a qual se pode reverter, sem grandes atropelos, quando necessário. Ela parece ser uma solução relativamente simples e barata, para um problema complexo. Isso nos motivou apresentá-la nesse Encontro.

---

<sup>98</sup> Esse mesmo procedimento poderia dar origem a mais de um índice. Cada índice corresponde ao conteúdo de um ou mais campos. Para criar um novo índice, basta concatenar o conteúdo dos campos referentes àquele índice. Isso permitiria criar qualquer tipo de índice, desde que baseado num conteúdo textual. Não obstante, os autores crêem que, numa situação de baixa padronização, talvez seja mais interessante limitar-se a um único índice, que trate o conteúdo das fichas descritivas como um todo.

## Referências Bibliográficas

BAEZA-YATES, Ricardo, RIBEIRO-NETO, Berthier. **Modern information retrieval**. New York: ACM Press, 1999.

BARRUECO CRUZ, José Manuel, SUBIRATS COLL, Imma. Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH): descripción, funciones y aplicación de un protocolo. **El Profesional de la Información**, 2003, vol. 12, n. 2, pp. 99-106. (disponível em <http://eprints.rclis.org/177/> - acesso em 18-01-2009).

CONSELHO NACIONAL DE ARQUIVOS - CONARQ. Câmara Técnica de Descrição Arquivística. **NOBRADE**. Norma Brasileira de Descrição Arquivística. Rio de Janeiro: CONARQ, 2006.

**Cover Pages – Dublin Core Metadata Initiative (DCMI)**. Disponível em: <http://xml.coverpages.org/dcmi.html> (acesso em 18-01-2009).

**Cover Pages - Encoded Archival Description (EAD)**. Disponível em: <http://xml.coverpages.org/ead.html>. (acesso em 12-07-2007).

**Dublin Core Metadata Initiative**. Disponível em: <http://dublincore.org/> (acesso em 18-01-2009).

**Encoded Archival Description (EAD)**. Disponível em: <http://www.loc.gov/ead/> (acesso em 23-04-2007).

**Extensible Markup Language (XML)**. Disponível em: <http://www.w3.org/XML/> (acesso em 10-07-2007).

FOX, Michael J. Stargazing: locating EAD in the descriptive firmament. In: PITTI, Daniel

V., DUFF, Wendy M.(Ed.) **Encoded archival description on the Internet**. New York: The Haworth Press, 2001. p. 61-74. (Volume publicado simultaneamente no **Journal of Internet Cataloging**, v. 4, n. 3-4, 2001).

GOSPODNETIC, Otis, HATCHER, Erik. **Lucene in action**: a guide to the Java search engine. Greenwich (USA): Manning, 2005.

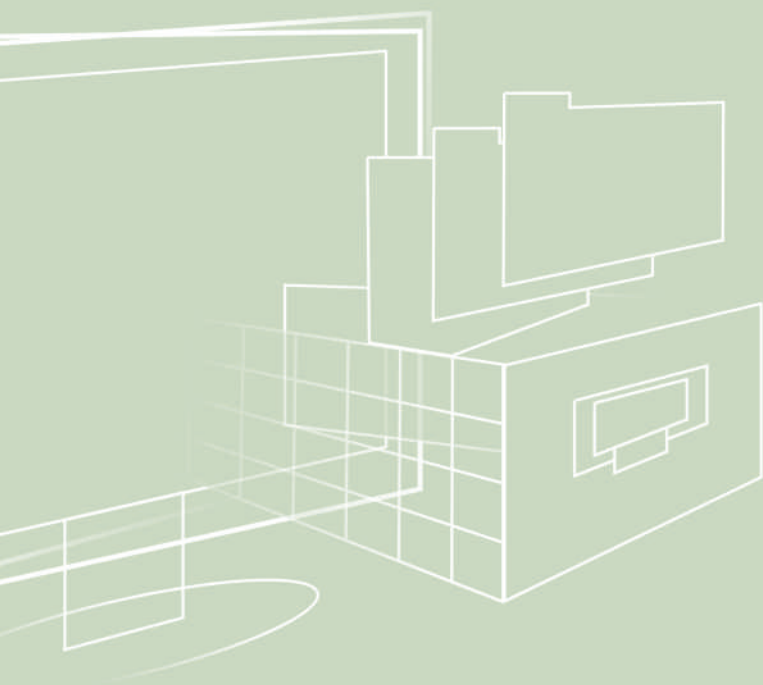
HAROLD, Elliotte Rusty, MEANS, W. Scott. **XML in a nutshell**: a desktop quick reference. 2.ed. Sebastopol (USA): O'Reilly, 2002.


**Lucene**. Disponível em: <http://lucene.apache.org/> (acesso em 18-01-2009).

**Open Archives Initiative**. Disponível em: <http://www.openarchives.org/> (acesso em 18-01-2009).

**SRU: Search/Retrieval via URL**. Disponível em: <http://www.loc.gov/standards/sru/> (acesso em 18-01-2009).





 **Associação dos  
Arquivistas  
Brasileiros**

apoio:

