

CONSTRUÇÃO DE SISTEMAS WEB PARA ACESSO A REPRESENTAÇÕES DE INFORMAÇÃO ARQUIVÍSTICA PERMANENTE: ALGUMAS INDICAÇÕES DE CRITÉRIOS E COMPONENTES

Ricardo Sodré Andrade¹

Resumo

A representação digital da informação arquivística de acervos permanentes demanda a utilização de sistemas apropriados, que possam promover o acesso por meio da Internet. Este trabalho apresenta algumas considerações sobre o assunto, indicando componentes para o desenvolvimento de um sistema *web* que atenda aos seguintes critérios: acesso remoto, interoperabilidade e preservação digital. Os componentes indicados são tecnologias específicas ou apropriadas para a representação e acesso de informação arquivística em meio digital. Eles foram divididos em três grupos: normalização, codificação e software livre. Os componentes citados são: Norma Brasileira de Descrição Arquivística (Nobrade), Descrição Arquivística Codificada (EAD-DTD), Pleade, Cheshire 3 Information Framework e Navimages. O objetivo do trabalho foi selecionar os componentes, colocando-os em evidência, de forma a fomentar a crítica destes ou de outros semelhantes com vistas ao desenvolvimento dos referidos sistemas de representação e acesso.

Palavras-chave: Representação de informação arquivística; Metadados; Interoperabilidade; Preservação digital; Descrição arquivística.

¹ Graduando em Arquivologia pelo ICI/UFBA; Assessor de Tecnologia do Arquivo Público da Bahia (Fundação Pedro Calmon), Ladeira de Quintas, nº 50, Salvador, Bahia, Brasil;
E-mail: ricardo@feudo.org; Website: www.feudo.org

Introdução

A informação tornou-se um recurso valorizado na sociedade atual, seu valor estratégico é indiscutível para as interações culturais, econômicas e sociais. As tecnologias da informação e comunicação proporcionam os meios pelos quais ocorrem as transmissões entre pessoas e recursos informacionais. Nossa sociedade se reorganiza a cada novo avanço, descoberta ou criação.

A informação arquivística é entendida como aquela que existe nos documentos produzidos e recebidos por organismos, pessoas ou instituições durante o exercício das suas atividades. Essa informação pode ser encontrada nos ofícios, portarias, correspondências, atas e demais documentos que atendem ao requisito apresentado.

Ainda sobre as características da informação arquivística, essas possuem um determinado valor jurídico ou administrativo que, com o passar dos anos, expira. Após isso, existem duas possibilidades para essa informação: ela adquire valor histórico-cultural, que significa que deve ser mantida permanentemente, ou é descartada. Na gestão da informação em suportes tradicionais, essa afirmação se traduz numa prática relativamente simples, bastando apenas observar a legislação em vigor, as orientações da teoria existente e as necessidades do organismo em questão. Todavia, em um contexto tecnológico, os procedimentos de avaliação e descarte ou custódia tornam-se problemáticas, visto que as variáveis encontradas em meio digital podem impedir a simples transposição da prática anterior à nova perspectiva tecnológica.

As tecnologias atuais permitem a criação de sistemas de recuperação de informação baseadas em formatos, padrões e/ou normas específicos e utilizando conceitos como *acesso remoto*, facilitando à sociedade o acesso às informações relativas à história e cultura, *interoperabilidade*, possibilitando a interligação virtual e cooperativa entre instituições arquivísticas geograficamente distantes e, pela natureza da tecnologia em questão, demandando ações de *preservação digital*, de modo a permitir que os arquivistas possam “transferir informações para os computadores do futuro”

(WALCH, 1994). Se tais conceitos não são novidades na literatura, métodos práticos e viáveis de aplicação destes, na realidade brasileira, não são facilmente localizados.

O objetivo desse trabalho é apresentar alguns componentes compatíveis entre si, que podem ser utilizados no desenvolvimento de ambientes de acesso a representações de informação arquivística, tanto do contexto quanto do conteúdo, atendendo a critérios como interoperabilidade, acesso remoto e preservação digital.

Conceitos

Justificar a criação de ambientes utilizando as ferramentas e normas que serão apresentadas depende da compreensão de alguns conceitos. A observação e aplicação desses conceitos no desenvolvimento dos ambientes de acesso farão a diferença entre uma simples disponibilização de representação de informação arquivística na *web* de uma metodologia alinhada com fundamentos arquivísticos tradicionais e sob a crítica dos impactos e possibilidades que o “paradigma” digital impôs ao fazer arquivístico.

Os conceitos a seguir não foram exaustivamente trabalhados, porém, espera-se que sejam suficientes para o entendimento: representação da informação, preservação digital, interoperabilidade e acesso remoto.

Representação da Informação

O termo representação da informação pode ser entendido como um processo e também como o(s) produto(s) derivado(s) deste processo. O processo de representação busca estabelecer uma correspondência entre o objeto a ser representado e o padrão de representação utilizado (YAKEL, 2003, p. 1). Essa correspondência é um produto informacional que representa aspectos relevantes do objeto representado.

Segundo Hagen (1998, p. 2), em citação a Cook, a representação da informação arquivística deve ser realizada em relação à substância e estrutura da documentação. Consideramos que a substância deve abranger o conteúdo

e o contexto da informação. A estrutura se relaciona com o suporte e os meios de acesso a este.

Utilizaremos o entendimento de que representação da informação abrange desde o processo de arranjo documental até a elaboração de instrumentos de acesso destinados ao uso por pesquisadores, tais como guias, inventários e repertórios.

Em meio digital, representar a informação é uma atividade que não pode prescindir do uso de conjunto de metadados.

Metadado é um termo que remete às tradicionais atividades de representação da informação (catalogação, descrição arquivística e outros), aquelas com as quais os profissionais da informação já estavam às voltas antes da chegada das atuais tecnologias. Usar conjuntos de metadados para descrever um objeto informacional nada mais é que afirmar acerca das características do objeto informacional, daí vem a definição de “dados sobre dados” ou “dados acerca de dados”, que, por não ajudarem no entendimento do conceito, devem ser evitadas, sendo devidamente substituídas por definições mais esclarecedoras.

Preservação Digital

A preservação digital consiste em estabelecer meios para a manutenção do acesso a objetos informacionais em longo prazo. Esses meios precisam ser definidos de forma que obstáculos conhecidos sejam superados ou, pelo menos, amenizados.

Como comentado em trabalho anterior (ANDRADE, BORGES e JAMBEIRO, 2006), diversos estudos e iniciativas surgiram na busca de respostas para os questionamentos que a preservação digital suscitou. Pesquisadores se debruçaram sobre questões como a obsolescência do hardware, a dependência do software e a deterioração da mídia, em busca de soluções para esses aspectos negativos do uso das novas tecnologias. Esses três aspectos são, resumidamente, os obstáculos enfrentados pelos gestores

da informação digitalizada que pretendem manter a integridade dessas a longo prazo.

Em sua pesquisa bibliográfica, Arellano considera que a preservação de documentos digitais dependerá “principalmente da solução tecnológica adotada e dos custos que ela envolve” (2004, p. 15). Essa afirmação é uma das justificativas para o uso preferencial de software livre nas iniciativas de preservação digital, lembrando que software livre não é sinônimo de software gratuito (*freeware*), conforme Stallman (2000) salienta, pois ser *livre* implica em outras *liberdades* além da gratuidade do software.

Interoperabilidade

Sistemas interoperáveis são conceituados por Marcondes e Sayão (2002, p. 46) como aqueles “passíveis de serem consultados simultaneamente”, ainda, segundo esses autores, redes que congreguem sistemas com essa característica são alcançadas “mediante adoção de um conjunto de especificações técnicas e princípios organizacionais bastante simples, porém potencialmente poderosos e de grande alcance, no objetivo de integração”.

O conjunto de especificações citado anteriormente são protocolos ou padrões que permitiriam criar aplicativos como meta-buscas, capazes de recuperar documentos com descritores semelhantes em acervos de instituições diferentes, representados digitalmente em sistemas compatíveis com o referido conjunto de especificações. Dessa forma, seria possível, por exemplo, pesquisar documentos acerca de uma personalidade em arquivos municipais de todo o país, ou apenas naqueles escolhidos dentre as opções existentes, facilitando o levantamento de fontes para a pesquisa.

Para ilustrar, podemos observar o Holmes (<http://www.holmes.feudo.org/>), um sistema de meta-busca que possui, em um único mecanismo, a possibilidade de consultar trabalhos acadêmicos em Ciência da Informação depositados em 28 repositórios e periódicos mantidos por diversas instituições ao redor do mundo. O Holmes é possível graças a um

conjunto de especificações, o protocolo OAI-PMH, suportado nos diferentes softwares adotados por estas instituições (ODDONE & ANDRADE, 2006).

Existem algumas iniciativas para a criação de redes de arquivos cujos sistemas informatizados estejam baseados na utilização de padrões. Podemos citar o projeto da Rede de Arquivos de Minas Gerais, que pretende utilizar a ISAD(g) e Descrição Arquivística Codificada (EAD-DTD) para codificar os instrumentos de acesso de diversas instituições arquivísticas daquele estado. Os instrumentos codificados seriam depositados em um único sistema de recuperação de informação (RIBEIRO, DI MAMBRO, PARRELA e ANDRADE, 2005).

Outra iniciativa semelhante, ainda em fase inicial, é a do Arquivo Público da Bahia (APB), que pretende fazer uso de uma metodologia alinhada com os conceitos apresentados neste trabalho. Um dos objetivos é evitar o manuseio excessivo da documentação, além de possibilitar o fortalecimento de uma rede de arquivos municipais na Bahia, por meio de um sistema colaborativo de busca baseado no uso da Nobrade e da Descrição Arquivística Codificada (EAD-DTD).

Acesso Remoto

A criação da *web* permitiu que conteúdos multimídia armazenados em um local pudessem ser acessados por interessados em qualquer parte do planeta, contando com um dispositivo de acesso à teia.

Diversos softwares foram desenvolvidos e estão sendo utilizados por instituições arquivísticas, porém, grande parte deles não permite que usuários possam acessar as bases de dados remotamente.

O acesso remoto é o critério que inclui a telecomunicação como elemento essencial para a implantação de serviços arquivísticos. As possibilidades são muitas, baseadas nas várias vantagens que as novas tecnologias oferecem. É importante lembrar que a Sociedade da Informação pressupõe o acesso a conteúdos por meio da Internet e as instituições

arquivísticas precisam considerar isso na promoção do acesso às informações contidas em seus acervos.

Componentes para o desenvolvimento do ambiente

Respeitando o caráter introdutório que esse texto possui e considerando os conceitos apresentados anteriormente, foram selecionados alguns componentes que poderiam ser utilizados para a construção de um ambiente de acesso a representações de informação arquivística que fosse acessível pela *web*. Uma única instância desse ambiente poderia ser utilizada por uma ou mais instituições no objetivo de disponibilizar instrumentos de acesso e objetos digitais multimídia produzidos a partir do acervo.

Os componentes foram divididos conforme o que seriam as etapas principais durante a criação do ambiente: normalização, codificação e software livre.

Normalização

Normalizar significa padronizar, aplicar critérios acordados a um conjunto de elementos que possuem a mesma natureza. As normas a que nos referimos aqui são as de descrição arquivística.

Uma série de iniciativas pelo mundo contribuiu para a normalização das práticas arquivísticas, principalmente as relacionadas ao tratamento documental. Como exemplo, podemos citar a tradicional função de descrição de documentos, que até então foi praticada pelos profissionais de arquivo utilizando critérios dos mais diversos, principalmente pela falta de uma norma própria. Essa lacuna só foi preenchida de forma notável pelo esforço que resultou na Norma Geral Internacional de Descrição Arquivística – ISAD(g).

No Brasil, o Arquivo Nacional (www.arquivonacional.gov.br) promoveu o desenvolvimento da Norma Brasileira de Descrição Arquivística (Nobrade), a partir da ISAD(g). Considerando essa o padrão proposto pelo Arquivo Nacional para os arquivos brasileiros, indicamos essa norma como ideal para a construção do ambiente.

Então, o primeiro passo em direção a um ambiente web de acesso a representação de informações é a criação de instrumentos de acesso utilizando a Nobrade, ou a adaptação dos instrumentos de acesso disponíveis de acordo com a referida normativa.

Codificação

Uma vez criados, os instrumentos de acesso deverão ser codificados em meio digital. Para isso, consideramos que o padrão chamado Descrição Arquivística Codificada (EAD-DTD)² deve ser utilizado.

Da necessidade de um padrão ou conjunto de metadados próprios para fins arquivísticos, surge a *Encoded Archival Description* (EAD), que teve sua primeira versão publicada pela *Network Development and MARC Standards Office* da *Library of Congress* (EUA), em parceria com a *Society of American Archivists*, em 1998, sendo atualizada em 2002. A EAD é uma definição de tipo de documento³ que utiliza a *Extensible Markup Language* (XML), ou linguagem de marcação extensível. Então, a EAD é uma DTD específica para a codificação de instrumentos de acesso para documentação arquivística. Por proporcionar o desenvolvimento de instrumentos legíveis por máquina e humanos, permite que seja possível disponibilizar tais instrumentos na *World Wide Web* utilizando softwares compatíveis com ela.

Por ser tecnologicamente neutra, isto é, não necessita de nenhum software ou hardware específico para seu funcionamento, a XML permite que instrumentos codificados com EAD-DTD possam acompanhar as migrações e atualizações de suas plataformas tecnológicas ao longo do tempo, sem prejuízo ao esforço empreendido inicialmente quando da criação das codificações.

² <http://www.loc.gov/ead>

³ Document Type Definition - DTD

A EAD-DTD é compatível com várias normas de descrição arquivística, incluindo a Norma Brasileira de Descrição Arquivística, Nobrade e a Norma Geral Internacional de Descrição Arquivística, ISAD(g).

Neste momento, os instrumentos de acessos serão codificados com a EAD-DTD e estarão prontos para serem disponibilizados por meio de um sistema *web* compatível.

Software Livre

Após uma busca por softwares livres que pudessem suportar a disponibilização das codificações criadas na Internet, foram encontradas algumas possibilidades interessantes, dentre as quais foram selecionadas duas. As buscas foram facilitadas pela existência do repositório de projetos *open source* denominado SourceForge (www.sf.net). Esses softwares não podem ser considerados ideais para toda e qualquer instituição, antes, devem ser entendidos como exemplos de aplicação dos critérios anteriormente apresentados e de integração dos componentes que estão sendo comentados neste trabalho.

O primeiro software é o *Cheshire 3 Information Framework* (www.cheshire3.org), um mecanismo de busca para bases em XML, construído com a linguagem de programação Python e bibliotecas C e possuindo, entre suas funcionalidades, compatibilidade com os protocolos Z39.50 e OAI. Esse software é utilizado pelo projeto britânico *Archives Hub* (www.archiveshub.ac.uk), que disponibiliza o acesso a cerca de 20.000 codificações EAD-DTD de acervos custodiados por universidades do Reino Unido.

O segundo é o *Pleade* (www.pleade.org), um projeto *open source* que desenvolve uma plataforma para publicação de codificações EAD-DTD e EAC-DTD⁴, na *web*. O Pleade é construído com o uso da linguagem de programação Java e de padrões como XML e XSL.

⁴ Outra definição XML, desta vez relacionada com a ISAAR(CPF)

O *Pleade* possui ainda uma interessante compatibilidade com outro projeto, *Navimages* (<http://sdx.archivesdefrance.culture.gouv.fr/gpl/navimages/>), voltado para o tratamento e disponibilização de imagens de documentos na *web*. Há formas de criar ligações (hiperlinks) em trechos da codificação a uma imagem ou série de imagens contidas na base do *Navimages*.

A integração desses componentes, resultando em um sistema, permitiria que usuários interessados nos acervos pudessem consultar as codificações dos instrumentos de acesso e, possivelmente, os objetos multimídia criados a partir da digitalização dos documentos, *hiperlinkados* nas codificações. Essa última possibilidade depende de outros conhecimentos e ferramentas específicas, como o *Navimages*.

Considerações finais

O objetivo desse trabalho foi indicar componentes específicos para a criação de um sistema *web* interessante para as instituições arquivísticas. Apesar de não detalhar os componentes, a intenção foi colocar os citados componentes em evidência, sem detalhá-los, na tentativa de provocar a análise e discussão crítica desses e de outros componentes.

As relações existentes entre os conceitos apresentados e os componentes indicados não foram estabelecidas nesse trabalho, mas é importante apontar que as tais relações são cruzadas e múltiplas. Como exemplo, podemos apontar o papel na interoperabilidade e na preservação digital que a EAD-DTD pode desempenhar ou da importância para o acesso remoto e interoperabilidade que o *Cheshire 3 Information Framework* representa, perceptível na observação direta de sua aplicação real, o *Archives Hub*.

A respeito das questões técnicas que envolvem o uso de alguns dos componentes citados, reconhecemos que os profissionais de informação envolvidos com acervos arquivísticos precisam se tornar competentes em diversos temas relacionados com a tecnologia da informação, então, apenas

dessa forma, poderiam superar o problema de recursos humanos oriundos da área de Informática, considerando que este é um dos problemas que muitas instituições arquivísticas enfrentam (MÁRDERO ARELLANO & ANDRADE, 2005; ANDRADE, 2006).

Ainda sobre dificuldades, considerando a constante falta de recursos financeiros encontrados nas instituições arquivísticas, percebe-se a necessidade de uma metodologia que utilize os componentes apresentados, ou outros de natureza semelhante, sobre uma plataforma operacional livre, como o GNU/Linux. Esse detalhe pode ajudar na eliminação de custos de aquisição e de atualização de software.

Possivelmente há demanda por demonstrações de como esses componentes devem ser combinados na prática, então, é provável que próximos trabalhos contemplem o detalhamento da implantação dessas ou de outras tecnologias de igual utilidade em cenários reais. Salientamos que há uma vasta documentação sobre cada componente nos seus respectivos *websites*.

Um sistema que atenda aos critérios apresentados permite que as instituições arquivísticas melhorem a qualidade de seu atendimento. Os usuários poderão ter acesso a informações descritivas do acervo sem necessidade de uma visita presencial, isso representa economia em várias ordens, tanto para o usuário quanto para a instituição de custódia.

Referências

- ANDRADE, Ricardo Sodré. Tecnologia, memória e a formação do profissional arquivista. **Arquivística.net**, vol. 2, n. 1, 2006.
- ANDRADE, Ricardo Sodré; BORGES, Jussara; JAMBEIRO, Othon. Digitalizando a memória de Salvador: Nossos presente e passado têm futuro?. **Perspect. ciênc. inf.**, Belo Horizonte, v.11 n.2, p. 243-254, mai./ago. 2006.
- HAGEN, Acácia Maria Maduro. Algumas considerações a partir do processo de padronização da descrição arquivística. **Ci. Inf.**, V. 27, n. 3, 1998. Disponível em: <<http://www.ibict.br/cienciadainformacao/viewarticle.php?id=350>>. Acesso em: 24 out. 2006.
- MARCONDES, Carlos Henrique; SAYÃO, Luis Fernando. Documentos digitais e novas formas de cooperação entre sistemas de informação em C&T. **Ci. Inf.**, Brasília, v. 31, n. 3, p. 42-54, set./dez. 2002.
- MÁRDERO ARELLANO, Miguel Ángel; ANDRADE, Ricardo Sodré. Preservação digital e os profissionais da informação. **DatagramaZero**, v. 7, n. 5, out, 2005.
- MÁRDERO ARELLANO, Miguel Angel. Preservação de documentos digitais. **Ci. Inf.**, Brasília, v. 33, n. 2, p. 15-27, maio/ago. 2004.
- ODDONE, Nanci; ANDRADE, Ricardo Sodré. Acesso à informação baseado em Open Archives: a experiência do Holmes. In: SEMINÁRIO NACIONAL DE BIBLIOTECAS UNIVERSITÁRIAS, 14, 2006, Salvador. **Anais...** Salvador, 2006.
- RIBEIRO, José Guilherme; DI MAMBRO, Galba; PARRELA, Ivana; ANDRADE, Ricardo Sodré. A criação da Rede de Arquivos de Minas Gerais - RAMG. In: CONGRESSO DE ARQUIVOLOGIA DO MERCOSUL, 6, 2005, Campos do Jordão. **Anais...** São Paulo: Associação dos Arquivistas de São Paulo, 2005.
- STALLMAN, Richard. O Projeto GNU. **DatagramaZero**, n. 1, fev, 2000.
- WALCH, Victoria Irons. **Standards for Archival Description: A Handbook**.1994. Disponível em: <<http://www.archivists.org/catalog/stds99/index.html>>. Acesso em: 17 set. 2006.
- YAKEL, Elizabeth. Archival representation. **Archival Science**, n. 3, p. 1-25, 2003.