

**Formatos de Arquivo e a Preservação de Documentos Digitais**  
**Bodê, Ernesto**  
**Arquivista – Superior Tribunal de Justiça**

### **1 - Objetivos e relevância científica**

O cenário moderno é rico em assuntos relacionados ao grande tema documentos eletrônicos. Entre eles está a função arquivística preservação, sobre a qual, Conway observa:

“Há muito tempo, as bibliotecas e os arquivos têm a responsabilidade de reunir, organizar e proteger a documentação da atividade humana. A ética da preservação, como gerenciamento consciente e coordenado, é um fenômeno mais recente. Bibliotecários e arquivistas – assim como os clérigos e escribas que vieram antes deles – aumentaram as chances de preservação das evidências, provas de nossas vidas, pensamentos e conquistas. A preservação tradicional, como forma responsável de resguardar essas informações, funciona somente quando a prova tem uma forma física, quando o seu valor é superior aos custos de sua manutenção, e quando os papéis desempenhados pelos seus criadores, responsáveis por sua guarda e usuários são mutuamente reforçados.” (CONWAY, 2001, p. 11)

De um paradigma que prevaleceu até o final do séc. XX, relacionado a problemas de conservação e restauração dos documentos em suportes tradicionais como o papel, papiro e pergaminho com foco na preservação do suporte físico, o que, em grande parte, ainda predomina. Passamos a um contexto onde a preservação do suporte material é apenas parte de uma problemática maior. Hoje, a questão da preservação dos documentos eletrônicos oferece muitos outros desafios. Desta forma e, em função da necessidade de obter respostas, como arquivista, aos acervos formados por documentos eletrônicos e digitais, nossa pesquisa, após uma sondagem inicial, direcionou nosso foco de atenção para o tópico Formatos de Arquivo, ou, a estrutura da codificação digital gravada nas mídias.

“Para que um documento possa ser legível no futuro, duas condições devem ser respeitadas. Primeiro, os bits que constituem o documento precisam ser legíveis da mídia para a memória do computador. Então, é preciso software disponível para interpretar os dados” (LORIE, 2001, tradução nossa)

Identificamos, portanto, dentro do **assunto** Preservação, no **tema** Documentos Eletrônicos, nosso problema. Qual seja, responder *Qual o papel dos Formatos de Arquivo para a preservação dos documentos eletrônicos e digitais ?*

Nosso objetivo neste trabalho é caracterizar e apresentar os objetos Formatos de Arquivo, a partir da perspectiva de preservação arquivística e Gestão Documental e relacioná-los com a problemática da preservação de documentos digitais.

Os profissionais envolvidos com Acervos Documentais, entre eles, e talvez principalmente os Arquivistas, estão sofrendo uma enorme pressão em função dos Documentos Eletrônicos. Os avanços tecnológicos das últimas décadas, têm colocado novos desafios relacionados à **Gestão de Documentos** e todo o conhecimento inserido nos mesmos. Surgiram novos desafios relacionados à indexação e classificação, descrição, controle de sua

produção, avaliação e preservação, ou seja, as funções arquivísticas tradicionais. A função Preservação, neste ambiente, depara-se com desafios que vão além dos cuidados com o suporte físico. Para ilustrar esta afirmação, basta nos lembrarmos da experiência de utilização de disquetes de 8 polegadas com arquivos do formato produzido pelo aplicativo WordStar, em uso há apenas duas décadas. São enormes hoje as dificuldades para obtenção de hardware e software adequados para sua utilização. Estamos falando da questão da *obsolescência tecnológica* de *hardware* e *software*.

Além da dimensão acima exposta, relacionada ao avanço científico teórico na área da preservação de documentos, outra dimensão que justifica este trabalho se refere ao quanto o problema da preservação de documentos eletrônicos está cada vez mais em evidência, desde o final do século XX até este início de século XXI, o que pode ser comprovado pela existência de várias iniciativas de pesquisa que estão sendo levadas a cabo por instituições com o *National Archives* no Reino Unido ou a *Library of Congress* nos EE.UU, entre outros, inclusive no Brasil com a *Câmara Técnica de Documentos Eletrônicos* de nosso Arquivo Nacional.

Parece haver uma evidente falta de estudos e trabalhos de pesquisa sobre **Formatos Eletrônicos de Arquivo** no que cabe a sua preservação por longos períodos. Com países como EE.UU, continente Europeu e Austrália, dominando o cenário de pesquisa. Nossa base teórica de consulta utilizou basicamente os relatórios e artigos produzidos nos centros de pesquisa destes países. No Brasil, já encontramos vários trabalhos relacionados à preservação de documentos eletrônicos, de forma geral, mas não especificamente e com foco na questão dos Formatos de Arquivo. Em 1999, publicado pelo *Council on Library and Information Resources* estadunidense, surge um relatório (BECK, 1999) de Ingrid Beck que “conta uma história” (o projeto CPBA) sobre um projeto de tradução de textos para nosso vernáculo, a publicação, no que diz respeito à preservação no universo digital menciona a tradução de um excelente texto de Paul Conway de 1996, *Preservation in a Digital World*. Em 2001, Marcelo Leone Sant’Anna publica “Os Desafios da Preservação de Documentos Públicos Digitais” (SANT’ANNA, 2001). No início de 2004 temos o trabalho “A preservação digital e o modelo de referência Open Archival Information System (OAIS)” de Kátia P. Thomaz e Antônio J. Soares 2004 (THOMAZ; SOARES, 2004). Em meados de 2004 Miguel A. Arellano, publica uma excelente “pesquisa bibliográfica”, “Preservação de Documentos Digitais” (ARELLANO, 2004).

A última dimensão que levantamos como justificativa para esta pesquisa e, talvez, a mais importante, refere-se aos aspectos de preservação da memória institucional ou nacional. Com relação ao aspecto cultural da memória, lembra Bellotto:

“Sendo a função primordial dos arquivos permanentes ou históricos recolher e tratar documentos públicos, após o cumprimento das razões pelas quais foram gerados, são os referidos arquivos os responsáveis pela passagem desses documentos da condição de “arsenal da administração” para a de “celeiro da história”, na conhecida acepção do consagrado arquivista francês Charles Braibant” (BELLOTTO, 2004, p. 23)

Nesse sentido, todo o imenso conteúdo informacional que encontramos hoje em documentos eletrônicos. E os avanços tecnológicos têm aumentado o volume deste conteúdo. Precisam de cuidados especiais, pois uma parcela deste conteúdo certamente tem grande importância cultural para a humanidade.

## **2 – Metodologia**

Nossa metodologia de pesquisa que objetivou trazer respostas para o problema da preservação de documentos digitais, com foco nos Formatos de Arquivo, consistiu na prospecção da literatura produzida sobre o tema, no Brasil e internacionalmente, objetivando desenhar o Estado da Arte atual sobre o tema. As principais fontes de pesquisa foram os grandes centros mundiais, que atualmente desenvolvem projetos relacionados à preservação de documentos eletrônicos e que também incluem pesquisas e dados sobre Formatos de Arquivo.

Inicialmente desenvolvemos uma pesquisa sobre o histórico da evolução dos documentos na humanidade, objetivando caracterizar melhor o que é um documento eletrônico. O resultado deste levantamento foi muito profícuo e nos possibilitou caracterizar adequadamente o que é um Documento Eletrônico e Digital, o que é particularmente importante em função de alguns problemas terminológicos existentes, como o termo documento eletrônico ou documento digital, nem sempre equivalentes na prática.

O resultado destas revisões bibliográficas foi confrontado então com os desafios que encontramos como arquivista responsável por Acervos Documentais, tanto em suportes tradicionais como aqueles em ambiente tecnológico, como bases de dados e documentos em hipertexto, fotografias digitais e etc.

## **3 - Resultados alcançados**

### **3.1 - Um breve histórico do documento**

Parece óbvia a existência de uma forte ligação entre a evolução do documento e a própria evolução dos sistemas de escrita. Desta forma, analisaremos este aspecto, tentando dar

foco também nos suportes utilizados nestes documentos, o que será importante mais adiante como comparação entre épocas.

Segundo Steven R. Fischer (2003), até que a humanidade obtivesse sistemas de escrita completos como os atuais, ela fez uso de símbolos gráficos e mnemônicos de vários tipos para armazenar informações, sobre um dos mais antigos artefatos encontrados, escreveu:

“Artefatos desenterrados em Bilzingsleben, Alemanha, datadas de, pelo menos, 412.000 anos atrás [...] foram interpretados por seus descobridores como entalhes intencionais (algum tipo de símbolos gráficos). É evidente que os entalhes são marcas; o que significam e se significam algo, não está claro.” (FISCHER, 2003, p. 16, tradução nossa)

Ainda segundo Fischer (2003), a humanidade utilizou então sistemas pictográficos (como as representações em cavernas). Em um segundo momento, passou a utilizar símbolos gráficos para representar objetos reais como vasos, animais e etc., até o grande salto da fonetização, quando um símbolo gráfico representa um som da linguagem local. Tal invenção surgiu na Mesopotâmia entre 6.000 e 5.700 anos atrás.

Como suportes para registros de informações, a humanidade utilizou os mais diversos materiais, segundo Dard Hunter (1978), foram utilizados a madeira, metais, pedras, troncos, tecidos, o papiro (*Cyperus papyrus*), pergaminho e finalmente o papel.

O uso de argila em tábuas é particularmente importante, pois ao que parece o primeiro sistema completo de escrita (por volta de 2.500 AC) utilizou este material como suporte. Estas tábuas estiveram presentes desde antes do desenvolvimento da escrita cuneiforme (que foi utilizada por vários povos) e cujo princípio, símbolos representando sons, foi a base dos hieróglifos egípcios, que deu origem ao primeiro alfabeto da humanidade.

No Egito, o uso do papiro rivalizou com o uso de tábuas de argila em outros povos. Na verdade os egípcios desenvolveram diferentes sistemas de escrita para diferentes aplicações, desde rituais a contabilidade. Para cada aplicação havia um sistema de escrita e suportes específicos como paredes, ouro e etc.

O uso de pergaminho também foi um fato importante para o registro de informações, “*O rei de Pérgamo (197-159 AC) normalmente recebe os créditos pela invenção e acredita-se que esteja relacionada com o desejo de produzir um material de escrita que rivalizasse com o papiro egípcio*”. (HUNTER, 1978, p. 12, tradução nossa)

Finalmente, a invenção do papel possibilitou um grande salto na produção de documentos, pois tratava-se de um material de fácil fabricação e menor custo, além da qualidade em relação a outros suportes. A data normalmente atribuída à invenção do papel é a de 105 DC, na China (HUNTER, 1978).

Vários outros suportes foram utilizados para o registro de documentos e em determinados períodos históricos alguns competiram entre si, como o papel e o pergaminho. O tipo de papel próximo do que é utilizado hoje só existiu a partir do século XIX (DOCTORS, 1999).

Temos então, até o século XIX, uma produção documental, registrada basicamente em papiro, pergaminho e papel (principalmente). Documentos do gênero textual, de diferentes naturezas, de inventários de bens a literatura e filosofia. Em meados do século XIX surge uma invenção que aumenta a diversidade dos acervos documentais, trata-se da fotografia, “*A invenção da fotografia foi anunciada oficialmente em 19 de agosto de 1839, pelo francês Louis Jacques M. Daguerre (1787-1851), sob a forma do daguerreótipo*” (SMIT; GONÇALVES, 2005, p. 9). Esta invenção passaria por um processo de evolução tecnológica que culminaria, no final do século XX com o advento da fotografia digital, a qual, por si só, tornou-se uma nova revolução. Também com tecnologia bastante próxima dos registros fotográficos, apesar da aplicação diferente, encontramos também o microfilme como meio para registro documental, ainda hoje bastante utilizado. E não poderíamos deixar de mencionar o cinema, no final do séc. XIX, invenção que é uma aplicação direta da fotografia. Na verdade uma filmagem e exibição cinematográfica em película é o resultado de 24 fotografias por segundo.

No final do século XIX, vários inventos para registro do som culminaram no início do século XX com os discos de áudio, e logo depois, o uso também de fitas magnéticas. Estas últimas, após um período de evolução, também passaram a ser utilizadas para gravação de vídeo (os primeiros programas televisivos gravados). No final do século XX surgiram os discos do tipo CD (Compact Discs), inicialmente para gravações de áudio (CD-ROM) e depois surgindo os modelos específicos para vídeo (DVD's). Hoje, há uma miríade de opções de CD e DVD - além do tipo ROM, os outros formatos mais conhecidos são os do tipo R (*Recordable*) e RW (*Rewritable*). Estes seriam logo aproveitados juntamente com a tecnologia de computadores.

O próximo grande passo seria dado pelo uso dos computadores,

“Os primeiros computadores **modernos** apareceram na década de 40; embora tenha havido muitas contribuições individuais para o avanço da tecnologia, esta cresceu e se desenvolveu, na América, especialmente graças à associação entre militares, universidade e firmas.” (KIDDER, 1982, p. 13, grifo nosso)

O uso, cada vez maior de computadores, inicialmente pelas grandes corporações. E, a partir da década de 80 do século XX, também pelo cidadão comum, representou um grande salto para o registro, armazenamento e recuperação de documentos. Estas máquinas,

em função da exigência de cada vez mais espaço para registro de seus *bits* (codificação digital), passaram a aproveitar diferentes tecnologias, desde as fitas magnéticas, passando por discos magnéticos, ópticos e diversos outros. Hoje, no início do século XXI, as novidades incorporadas ao conjunto de mídias são os *tocadores de áudio* e *iPods*.

O que se observa também é que, a partir dos documentos com conteúdos basicamente textuais, basicamente em suportes tradicionais, encontrados até meados do século XX. Passamos, neste início do século XXI, com o uso dos documentos eletrônicos e digitais, a encontrar uma grande diversidade de conteúdos nos mesmos, desde imagens fixas, imagens em movimento, registros sonoros, até a combinação de todos estes elementos. Além dos documentos, cada vez mais importantes, que surgiram especificamente com o uso da informática, como bancos de dados, planilhas eletrônicas e etc.

### **3.2 – A caracterização do Documento Eletrônico**

Uma comparação entre os documentos disponíveis e suas características até o final do século XIX com os documentos eletrônicos e digitais atuais, revela as características contrastantes e peculiares dos últimos, elencamos aqui, pelo menos quatro delas:

**Legibilidade por Máquina.** A necessidade de se utilizar máquinas para que seja possível o acesso ao conteúdo destes documentos nos parece ser a característica mais marcante. Apenas no final do século XIX e durante o século XX surgem documentos que não podem ser acessados diretamente pelo homem, sem a ajuda de máquinas. E hoje, também sem a presença de software.

**Independência entre suporte e conteúdo.** Tradicionalmente, os suportes físicos da informação registrada em um documento, não podem ser removidos, sem danos ao documento em si. O documento tradicional, constituído por suporte com informações registradas, forma um conjunto indissociável. O documento eletrônico digital, por sua vez, ainda será formado pela dupla suporte e conteúdo e sempre necessitará de um suporte físico para completar esta dupla. No entanto, percebe-se que o suporte físico pode ser facilmente substituído, sem danos no que cabe ao conteúdo das informações registradas.

**Codificação digital.** A informação registrada em um documento tradicional, utilizará diferentes linguagens, desde a composição visual e artística até o vernáculo particular de cada grupo social. Nos documentos eletrônicos e digitais, apesar da visualização de diferentes linguagens quando do acesso ao conteúdo destes documentos, em última análise, a linguagem utilizada na gravação destes documentos será sempre digital, independentemente do vernáculo utilizado, tipo de imagem, cor ou características sonoras do documento.

**Diversidade de conteúdos.** O tipo de conteúdo, até o advento dos documentos eletrônicos e digitais, é diferente para cada tipo de documento. Assim, temos os documentos com conteúdos textuais, que são diferentes daqueles com conteúdos de imagens fixas (fotografias, por exemplo), documentos fotográficos, ou os documentos sonoros (com conteúdo de áudio) e etc. Por outro lado, os documentos eletrônicos e digitais, formam um grupo único, e são capazes de decodificar, a partir da linguagem digital utilizada para sua gravação, diferentes tipos de conteúdo, como o som, o texto, imagens e diversos outros.

### 3.3 - Os Formatos de Arquivo

Neste trabalho, nosso foco não está nos problemas relacionados aos suportes físicos de documentos eletrônicos e digitais. Mas sim, na forma e estrutura como as informações estão gravadas nestes documentos. Diversas propostas de atividades relacionadas à preservação destes documentos têm como base esta forma e estrutura, a qual, no jargão técnico constitui os Formatos de Arquivo.

A partir de um projeto implementado por pesquisadores da Universidade de Leeds no Reino Unido e subsidiado pelo *Joint Information Systems Committee* (JISC), foi produzido um relatório a partir de um *survey* sobre Formato de Arquivo e problemas relacionados, que assim define um Formato de Arquivo:

“No nível mais básico, objetos digitais são seqüências de zeros e uns que representam dados codificados. Diferentes **Formatos de Arquivo** especificam como estes códigos representam o conteúdo intelectual criado por um autor do objeto digital. Um exemplo disto é o formato Microsoft Word. Este formato é uma especificação para armazenamento de dados textuais, bem como informações de formatação. Muitos Formatos de Arquivo são incrivelmente complexos, de maneira que os códigos podem ficar ininteligíveis para um observador humano. Para que este objeto digital tenha sentido, um software será necessário para interpretar e exibir [ou renderizar] os dados para o usuário.” (UNIVERSITY OF LEEDS, 2003, tradução e grifo nossos)

Podemos agrupar os Formatos de Arquivo em algumas categorias mais comuns, para efeitos didáticos, são elas os formatos de (com algumas extensões reais do nome do formato):

- Texto (.txt, .rtf, .doc)
- Imagens fixas (.tif, .jpg, .gif, .png, .bmp)
- Imagens em movimento (.mpeg, .avi)
- Som (.mp3, .wav)
- Bancos de Dados (.db, .sql)
- Planilhas eletrônicas (.wr1, .xls)

Para cada Formato de Arquivo produzido por determinado software, existirá uma especificação técnica, na verdade, haverá também uma especificação para cada **versão** de um

determinado Formato, por exemplo a especificação TIFF 5.0 e a TIFF 6.0, cada uma, com seu detalhamento técnico. Dependendo do Formato de Arquivo, tal especificação técnica pode ser extremamente diferente para cada versão de um mesmo formato.

As especificações de cada Formato de Arquivo são de caráter bastante técnico e estão no escopo de desenvolvedores de software em geral. Tais documentos, explicam, detalhadamente, como as seqüências de bits no arquivo devem ser estruturadas e onde cada tipo de dado deve ser gravado. Para cada formato de arquivo haverá diferenças marcantes entre as especificações.

A figura 1 mostra um trecho da especificação JPEG versão 1.02, extraída do manual disponibilizado pelas instituições responsáveis pela especificação. Assim como esta, diversas outras especificações estão disponíveis para consulta.

Length	(2 bytes)	Total APP0 field byte count, including the byte Count value (2 bytes), but excluding the APP0 Marker itself
Identifier	(5 bytes)	= X'4A', X'46', X'58', X'58', X'00' This zero terminated string ("JFXX") uniquely Identifies this APP0 marker. This string shall have zero parity (bit 7=0)
Extension_code	(1 byte)	= Code which identifies the extension. In this version, the following extensions are defined: = X'10' Thumbnail coded using JPEG = X'11' Thumbnail stored using 1 byte/pixel = X'13' Thumbanil stored using 3 bytes/pixel
Extension_data	(variable)	= The specification of the remainder of the JFIF extension APP0 marker segment varies with the extension. See below for a specification of extension_data for each extension.

Figura 1 – Parte da especificação JPEG, fonte: manual de 1992

Um ponto crucial sobre Formatos de Arquivo, e que está diretamente ligado a problemas com sua preservação, refere-se ao fato de se tratar de um formato proprietário ou não. Em outras palavras, por diversos motivos, principalmente interesses comerciais, os detalhes técnicos de um formato de arquivo podem não estar disponíveis em momento algum ao público em geral. Sobre este problema:

“Um formato é freqüentemente controlado como propriedade intelectual de uma entidade comercial, a qual, tipicamente tem grande interesse em esconder o código base. A competição direciona freqüentes mudanças no formato individual, tanto quanto nas empresas que os controlam; as tecnologias da informação também impõem contínuas transformações. Esta combinação de opacidade e mudança significa que não há segurança de que a tecnologia futura irá suportar os formatos de hoje. De fato, o cenário digital de amanhã será repleto de objetos grandemente difíceis de preservar, acessar e interpretar.” (LeFURGY, 2003, tradução nossa)

O que LeFurgy sintetizou tão bem, é o principal problema que relaciona as atividades de preservação de documentos eletrônicos e digitais à problemática Formato de Arquivo. O termo obsolescência de software é freqüentemente utilizado para se referir a este problema. Como os dados gravados neste tipo de documento, não podem ser lidos diretamente



pelo usuário, mas sim através de equipamentos e software, ao longo do tempo, necessitaremos manter os documentos e todo o aparato necessário para sua leitura, como periféricos de leitura, computador, sistema operacional e programas aplicativos originais, no mínimo. Ou, caso se conheça como foi feita a codificação (como são os detalhes técnicos do Formato de Arquivo), no futuro, será possível ler e renderizar novamente o conteúdo originalmente gravado. Note-se que para isto, é fundamental que se conheçam os detalhes do Formato de Arquivo. Sobre isto:

“A ameaça à era da informação digital ultrapassou o perigo das mídias instáveis e obsolescência de hardware. Os problemas mais prementes confrontando os gestores de coleções digitais são o formato de arquivo e a obsolescência de software” (LAWRENCE et al, 2000, p. 1, tradução nossa)

Existem várias propostas para tentar manter, no futuro, o acesso às informações de um documento gravado através de um determinado Formato de Arquivo, como a emulação e a migração. De qualquer, forma, todas as propostas dependem do conhecimento sobre Formatos de Arquivo para que possam ser executadas com sucesso, em maior ou menor grau.

Diante de todo este quadro, todos os profissionais que se confrontam com acervos de documentos eletrônicos e digitais, têm diante de si o desafio de utilizar ou não formatos de arquivo conhecidos (padrão aberto) ou proprietários. De qualquer forma, um segundo desafio que ocorrerá no futuro, mas que já deve ser planejado, diz respeito a como identificar qual o Formato de Arquivo (e sua versão) de um determinado documento. Apesar de hoje, utilizando-se os programas aplicativos que geram tais documentos, isto parecer óbvio, certamente não será óbvio num futuro com outro sistema operacional e outros aplicativos (que não necessariamente e muito provavelmente não serão compatíveis com a tecnologia atual). Este problema fica maior ainda quanto mais longe olhamos para o futuro. Como identificar então um Formato de Arquivo ? E qual a versão deste Formato de Arquivo ?

### **3.4 - Identificação de Formatos de Arquivo**

Primeiro, vamos enfatizar melhor o problema da **versão** do Formato de Arquivo, “[...] arquivos com dados do Word 6.0 não são legíveis pelo Word 5.0, apesar dos arquivos do Word 6.0 e Word 5.0, ambos representarem documentos de processamento de palavras produzidos por duas versões diferentes do mesmo produto” (OCKERBLOOM, 1998, p. 1, tradução nossa). Considerando este exemplo, possui a mesma importância a identificação do Formato de Arquivo e a **versão** utilizada, para

efeitos de preservação futura do acesso aos documentos que utilizaram determinado formato.

Brown (2005) observa que, além de ser desejável que o processo de identificação de um Formato de Arquivo seja automatizado, propõe como meio para isto a identificação de uma *assinatura*, ou seja, uma seqüência de bits que representam códigos e que estão presentes em cada tipo de Formato de Arquivo. Estas assinaturas, podem ser externas ou internas.

As assinaturas externas “*abrangem todos os indicadores que estão externos à seqüência de bits do objeto digital, como os data forks do Macintosh e as extensões de arquivo do windows.*” (BROWN, 2005, p. 7, tradução nossa)

No entanto, ainda o mesmo autor neste mesmo trabalho, enumera algumas desvantagens na identificação de um Formato de Arquivo pelas assinaturas externas, como extensões do nome do arquivo. Como o fato de que uma extensão *não é necessariamente única* para um determinado formato, elas não permitem, a *identificação da versão do formato* [apesar de existirem exceções no mercado]. Ou o fato de que extensões *podem ser definidas ou alteradas* pelos usuários. Eu acrescentaria também o fato de que a extensão do arquivo, pelo menos em ambiente de Sistema Operacional Windows (o mais comum), só se torna relevante para a identificação quando um aplicativo correspondente está disponível e instalado no Sistema Operacional, caso contrário o usuário será solicitado a indicar um aplicativo correspondente para utilizar aquele formato de Arquivo.

Por outro lado, as assinaturas internas ao arquivo parecem ser um método bem mais promissor, “*Por definição, a especificação de um Formato de Arquivo impõe uma estrutura específica para o conteúdo da seqüência de bits, que é consistente entre todos os objetos daquele formato*” (BROWN, 2005, p. 7, tradução nossa)

A partir, basicamente, do uso de assinaturas digitais e utilizando-se de diversas técnicas diferentes para extrair dados sobre um determinado arquivo e comparar com uma assinatura previamente armazenada, várias instituições ligadas à questão da preservação de objetos digitais, tem disponibilizado software para identificação de Formatos de Arquivo. Ressalte-se que além do próprio software para identificação de um formato específico, é necessário um banco de dados, previamente armazenado, com a relação das assinaturas internas de Formatos de Arquivo, para cada versão específica.

Duas iniciativas neste sentido são:

O pacote de aplicativos **JHOVE** (<http://hul.harvard.edu/jhove>), (JSTOR/Harvard Object Validation Environment). O qual é capaz de, além de executar a identificação de formatos, também sua validação (“o processo de determinar o nível de conformidade de um objeto com a especificação original”) e caracterização (“quais são as propriedades deste arquivo e formato, como tamanho em bytes, data de atualização e outras”);

E o projeto **PRONOM** (<http://www.nationalarchives.gov.uk/pronom>) que disponibiliza o aplicativo *DROID*. Este aplicativo bastante simples de se utilizar, permite após sua instalação, atualizações na base de dados de assinaturas de formatos, de maneira a incorporar futuros formatos de arquivo e/ou novas versões.

Existem outras iniciativas com propostas semelhantes, como alimentar um banco de dados com informações sobre formatos de arquivo, possibilitando consultas sobre informações técnicas, as quais, futuramente podem subsidiar ações de preservação.

Os cuidados com o uso de determinados Formatos de Arquivo, não podem se restringir somente à disponibilização de informações técnicas sobre este ou aquele formato específico, na verdade, tais cuidados devem ser tomados ainda na criação dos documentos eletrônicos e digitais em questão. A partir de uma publicação original de Neil Beagrie e Maggie Jones o DPC – *Digital Preservation Coalition*, mantém um *handbook* que funciona como um guia para a gestão de recursos digitais objetivando manter o acesso aos mesmos ao longo do tempo. Com relação a Formatos de Arquivo, este trabalho enumera algumas recomendações para sua escolha e utilização (DPC, 2006, tradução nossa), na relação abaixo fica patente a importância do objeto Formato de Arquivo:

- Utilizar **formatos de arquivo** não proprietários, com código aberto e bem documentados, sempre que possível;
- Utilizar também **formatos de arquivo** que foram amadurecidos, tenham sido largamente adotados e se tornaram padrão *de facto* no mercado;
- Identificar **formatos** que possam ser aceitáveis para os propósitos de transferência, armazenamento e distribuição aos usuários (podem ser distintos);
- Minimizar o número de **formatos de arquivo** a serem gerenciados, dentro dos possível ou desejável;
- Não utilizar criptografia ou compressão para arquivamento de arquivos, se possível.

#### **4 – Conclusão**

O problema da preservação de documentos eletrônicos e digitais tem recebido, cada vez mais atenção da comunidade internacional, no Brasil as iniciativas ainda são tímidas comparando-se com o quadro lá fora. As principais instituições à frente de projetos ligados a

este tema são Arquivos Nacionais e Bibliotecas ou rede de Bibliotecas. Vários consórcios e associações foram criados com este fim.

Parece um fato consumado que este problema não pode mais passar sem a devida atenção. As organizações, tanto públicas quanto privadas, estão se deparando com objetos digitais não mais legíveis, em função da obsolescência de software ou hardware, por exemplo. Ou estão se deparando com a eminência deste problema.

Por se tratar de um tema relativamente novo, internacionalmente as primeiras iniciativas surgem em meados da década de 90, existem muitas propostas de possíveis soluções ainda em debate. Apenas o processo normal de discussão e pesquisas, inclusive práticas, proporcionará um quadro mais estável e seguro que sirva de guia para os profissionais envolvidos com tais problemas.

Neste trabalho, procuramos apresentar uma visão geral sobre um tópico específico dentro da miríade de temas que têm sido discutidos nacional e internacionalmente, os **Formatos de Arquivo**. Acreditamos que conhecer melhor este aspecto do grande tema Preservação de Documentos no Universo Digital será fundamental para as políticas de Gestão Documental. Em contraste com as atividades de preservação de documentos em suportes tradicionais, baseados em papel, principalmente, os documentos digitais necessitam de medidas preventivas antes mesmo de sua criação, como a escolha adequada de um dos inúmeros formatos de arquivo disponíveis no mercado. Além disto, conhecer quais são os formatos de arquivo em uso institucionalmente, também é fundamental para definir como, e os custos envolvidos (inclusive financeiros) para sua eventual manutenção a longo prazo e até mesmo a curto prazo, já que em algumas décadas é comum que formatos e softwares associados já apresentem problemas de manutenção no acesso.

Um dos resultados apresentados nesta pesquisa, foi uma melhor caracterização sobre o que é um Formato de Arquivo e os tipos mais comuns de conteúdos registrados nestes. Estes conteúdos (som, imagem e etc.) podem ser associados aos diversos documentos em suportes tradicionais utilizados ao longo da história, que inicialmente abordamos em um breve histórico.

Um outro resultado, refere-se à necessidade e possíveis entraves sobre como pode ser feita a identificação de um Formato de Arquivo e a versão utilizada neste formato. Em se tratando da preservação do acesso ao longo do tempo, identificar para determinar o quanto se conhece (se é que se conhece) um formato de arquivo, é uma questão fundamental.

Deste conhecimento decorrem várias recomendações para escolha de um determinado Formato de Arquivo, expusemos algumas, com base em trabalhos sobre o assunto.

Procuramos, nesta linha, fundamentar o que consideramos um elemento essencial e básico para várias abordagens sobre preservação de documentos eletrônicos e digitais.

Em complementação a este assunto, há outros aspectos que não pudemos desenvolver nesta oportunidade e que parecem ser de fundamental importância para a preservação de objetos digitais. Entre eles, estão o uso e definição de Metadados, estes descritores de um documento eletrônico, além de serem utilizados para a descrição do conteúdo do documento, também podem descrever aspectos técnicos que auxiliam na continuidade do acesso ao longo do tempo, como informações sobre os aplicativos utilizados para renderizar um arquivo ou o percurso institucional e histórico do documento. Muito próximo da questão do uso de Metadados está o uso de Repositórios para armazenar e, através de diversas técnicas, manter o acesso a objetos digitais inseridos em seu conteúdo ao longo do tempo. Ambos são assuntos que merecem aprofundamento.

Há também questões sobre direitos autorais, que poderão ser relevantes no futuro, caso se tente desenvolver maneiras de acessar um documento gravado em um formato de arquivo que não seja de domínio público.

## **Bibliografia**

ARELLANO, Miguel A. Preservação de documentos digitais. **Ciência da Informação**, Brasília, v.33, n.2, p. 15-27, mai/ago 2004.

BECK, Ingrid. **Building preservation knowledge in Brazil**. Washington: CLIR, 1999. Disponível em <<http://www.clir.org/pubs/reports/>>. Acesso em 10 de abril de 2006.

BELLOTTO, Heloísa Liberalli. **Arquivos permanentes: tratamento documental**. 2 ed. Rev. e ampl. Rio de Janeiro: Editora FGV, 2004.

BROWN, Adrian. **Automatic format identification using pronom and droid**. National Archives: UK, 2005. Disponível <<http://www.pro.gov.uk/about/preservation/>>. Acesso em 10 de abril de 2006.

CONWAY, Paul. **Preservação no universo digital**. 2 ed. Rio de Janeiro: Projeto Conservação Preventiva em Bibliotecas e Arquivos : Arquivo Nacional, 2001.

Digital Preservation Coalition – DPC. The handbook. Disponível em <http://www.dpconline.org/graphics/handbook/>. Acesso em 05 de abril de 2006.

DOCTORS, Márcio. **A cultura do papel**. Rio de Janeiro: Casa da Palavra, 1999.

HUNTER, Dard. **Papermaking: the history and technique of na ancient craft**. NY: Dover, 1978.

KIDDER, Tracy. **A alma da nova máquina**. São Paulo: Melhoramentos, 1981.

LAWRENCE, Gregory W. et al. **Risk management of digital information: a file format investigation**. Washington: Council on Library and Information Resources, 1999. Disponível em <<http://www.clir.org/pubs/reports>>. Acesso em 10 de abril de 2006.

LeFURGY, William G. PDF/A: Developing a file format for long-term preservation. **RLG News**, NY, v. 7, n. 6, 2003. Disponível em: <http://www.rlg.org>. Acesso em: 10 novembro 2005.

LORIE, Raymond A. A project on preservation of digital data. **RLG News**. Nova York, v. 5, n. 3, 2001. Disponível em: <http://www.rlg.org>. Acesso em 10 de novembro de 2005.

OCKERBLOOM, John. **Mediating among diverse data formats**. Carnegie Mellon University: Pittsburg, 1998. Tese de doutorado.

SANT'ANNA, Marcelo L. Os desafios da preservação de documentos públicos digitais. **Revista IP**, ano 3, n. 2, dez. 2001. Disponível em <[http://www.ip.pbh.gov.br/revista0302/res\\_ip0302santanna.html](http://www.ip.pbh.gov.br/revista0302/res_ip0302santanna.html)>. Acesso em 13 de abril de 2006.

SMIT, Johanna; GONÇALVES, Cássia Camargo. Como organizar arquivos fotográficos: **projeto como fazer**. São Paulo: AASP, 2005. Apostila do curso

FISCHER, Steven Roger. **A history of writing**. London: Reaktion Books, 2003.

THOMAZ, Kátia; SOARES, A. José. A preservação digital e o modelo de referência open archival information system (OAIS). **Datagramazero**, Rio de Janeiro, v. 5, n. 1 fev. 2004.

UNIVERSITY OF LEEDS. Survey and assesment of sources of information on file formats and software documentation. Final Report.