



UNIVERSIDADE FEDERAL DE SANTA CATARINA
Centro de Ciências da Educação
CURSO DE GRADUAÇÃO EM BIBLIOTECONOMIA



IVAN PAES JOSÉ

A interoperabilidade dos documentos editáveis disponíveis nos sítios das universidades públicas de Santa Catarina.

Florianópolis, 2012

IVAN PAES JOSÉ

A interoperabilidade dos documentos editáveis disponíveis nos sítios das universidades públicas de Santa Catarina.

Trabalho de Conclusão do Curso de Graduação em Biblioteconomia, do Centro de Ciências da Educação da Universidade Federal de Santa Catarina, requisito parcial à obtenção do título de Bacharel em Biblioteconomia. Orientação de: Prof. Dr. Moisés Lima Dutra.

Florianópolis, 2012

Ficha catalográfica elaborada pelo graduando em
Biblioteconomia/UFSC Ivan Paes José

J83p JOSÉ, Ivan Paes. 1982 –

A interoperabilidade dos documentos editáveis disponíveis nos
sítios das universidades públicas de Santa Catarina./ Ivan Paes
José. Florianópolis, 2012.

67 f.

Orientador: Moisés Lima Dutra.

Trabalho de Conclusão de Curso (Graduação em
Biblioteconomia) - Universidade Federal de Santa Catarina,
Centro de Ciências da Educação, Florianópolis, 2012.

1. Preservação de Documentos Digitais. 2. Interoperabilidade.
3. ODF. I. Título

CDU – 02:004.93

Esta obra é licenciada por uma licença *Creative Commons* de atribuição, de uso não comercial e de compartilhamento pela mesma licença 3.0



Você pode:

- copiar, distribuir, exibir e executar a obra;
- criar obras derivadas.

Sob as seguintes condições:

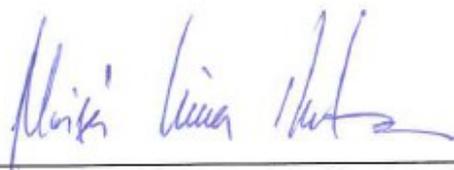
- Atribuição. Você deve dar crédito ao autor original.
- Uso não-comercial. Você não pode utilizar esta obra com finalidades comerciais.
- Compartilhamento pela mesma licença. Se você alterar, transformar ou criar outra obra com base nesta, somente poderá distribuir a obra resultante com uma licença idêntica a esta.

Acadêmico: Ivan Paes José

Título: A interoperabilidade dos documentos editáveis disponíveis nos sítios das universidades públicas de Santa Catarina.

Trabalho de Conclusão de Curso apresentado ao Curso de Graduação em Biblioteconomia, do Centro de Ciências da Educação da Universidade Federal de Santa Catarina, como requisito parcial à obtenção do título de Bacharel em Biblioteconomia, aprovado com nota 8,5.

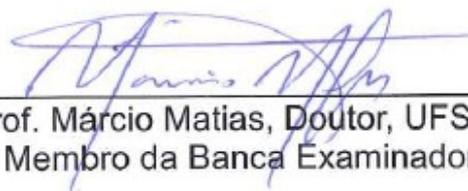
Florianópolis, 18 de Dezembro de 2012.



Prof. Moisés Lima Dutra, Doutor, UFSC
Professor Orientador



Prof. Angel Freddy Godoy Viera, Doutor, UFSC
Membro da Banca Examinadora



Prof. Márcio Matias, Doutor, UFSC
Membro da Banca Examinadora

Dedico este trabalho aos meus pais Ana Lúcia e Eraldo (in memoriam), com todo o amor do mundo.

À minha amada esposa Jaqueline, por seu amor, carinho e respeito.

Em especial, ao meu sogro, Nicolau Ricardo (in memoriam), pela sua dedicação à leitura e desejo de conhecimento.

AGRADECIMENTOS

A Deus, por tudo que tem me proporcionado: saúde, família, amigos e pela força nesta caminhada.

Ao meu pai Eraldo (in memoriam), pela vida.

À minha mãe Ana Lúcia; exemplo de luta e por sempre incentivar os meus estudos.

À minha esposa Jaqueline por todos os momentos compartilhados, pelo amor, pela confiança depositada e apoio incondicional. Te amo!

Ao meu irmão Vânio pelo seu companheirismo.

Às minhas tias Nilda (in memoriam) e Anísia (in memoriam) por alimentar essa vontade de estudar e ressaltar a importância dos estudos na vida.

À minha tia Rosária que foi responsável por terminar meus estudos e de propiciar a mim realização de um sonho, ter cursado uma universidade. Obrigado, sem seu apoio nada disso seria possível.

Aos meus sogros Clementina e Nicolau (in memoriam) por me acolherem em sua casa e de darem a esperança de terminar essa caminhada.

Ao Professor Moisés Lima Dutra, pela paciência e contribuição na realização deste trabalho como meu orientador e a todos os professores do curso por compartilharem seus conhecimentos e experiências.

À Universidade Federal de Santa Catarina pela oportunidade de ensino concedida e ao Curso de Graduação em Biblioteconomia pelo rico aprendizado nesses 4 anos.

Aos amigos conquistados nesta jornada: Cintia Darzicia Wílbirt e Evanir Badziak Raduvanski.

A todos os colegas de classe.

Ao meu amigo Cleidson Neves, pois em um momento do curso que estava para desistir veio até mim e compartilhou comigo sua experiência de vida. Fazendo com que eu levantasse a cabeça e prosseguisse em frente. Muito obrigado!

E a todas as pessoas que de alguma forma contribuíram para a concretização de mais essa etapa em minha vida.

JOSÉ, Ivan Paes. **A interoperabilidade dos documentos editáveis disponíveis nos sítios das Universidades públicas em Santa Catarina**. 2012. 67 f. Trabalho de Conclusão de Curso (Graduação em Biblioteconomia) - Universidade Federal de Santa Catarina, Florianópolis, 2012.

RESUMO

Este trabalho faz uma análise dos documentos digitais editáveis disponíveis nos sítios da Universidade Federal de Santa Catarina (UFSC) e Universidade do Estado de Santa Catarina (UDESC). Nesta análise, objetiva-se verificar se os documentos digitais editáveis estão em conformidade com as orientações do Governo Brasileiro, que sugere/prefere o uso do formato aberto *Open Documento Format* (ODF) para a criação de documentos digitais editáveis em toda a esfera pública nacional. Para tanto, utilizou-se o método exploratório-descritivo e quali-quantitativo no que tange ao enfoque do problema. O corpus da pesquisa foi obtido por meio dos operadores *ext* e *site*, aplicados no buscador Google (www.google.com.br), tendo sido recuperados 14.695 documentos no sítio da UDESC e 32.677 documentos no sítio da UFSC. Após isso, repetiu-se a mesma busca dentro dos sítios que hospedam os cursos de Biblioteconomia das duas instituições, tendo sido recuperados 524 documentos na UDESC e 57 documentos na UFSC. Além disso, realizou-se uma avaliação da interoperabilidade, que consistiu em analisar se as informações contidas nos documentos coletados fora recuperada sem nenhum tipo de perda, independentemente de extensão e *softwares* de escritório utilizados. Para esta pesquisa, foram selecionados 50 documentos, que foram abertos em dois *softwares* de escritório: o *Microsoft Office 2010* e o *LibreOffice 3.6*. O sistema operacional utilizado como plataforma nessa pesquisa foi o *Windows 7 Ultimate 64 bits Service Pack 1*. Os resultados foram apresentados por meio de gráficos e ilustrações que mostram a situação atual implementada nas duas instituições. Finalmente, constatou-se na pesquisa, que tanto a UFSC quanto a UDESC não estão adotando as orientações propostas pelo Governo Brasileiro. Quanto à interoperabilidade, constatou-se que ocorreram perdas significativas nas informações dos documentos, com alguns documentos, inclusive, tendo a sua leitura completamente impossibilitada.

Palavras-Chave: Documentos digitais editáveis. Interoperabilidade. Padrão ODF. Extensões Proprietárias. Extensões não-proprietárias.

JOSÉ, Ivan Paes. **Interoperability of editable documents available on the websites of the public universities in Santa Catarina**. 2012. 67 pages. Completion of Course Work (Undergraduation in Library Science) - Federal University of Santa Catarina, Florianópolis, 2012.

ABSTRACT

This work undertakes a survey on the editable digital documents available on the websites hosted by the Federal University of Santa Catarina (UFSC) and by the State University of Santa Catarina (UDESC). This work aims to check whether the editable digital documents are in accordance or not with the guidelines of the Brazilian Federal Government, which suggests/prefers the use of Open Document Format (ODF) for creating editable digital documents throughout the national public sector. To accomplish such a task, we have used the exploratory descriptive method and the qualitative-quantitative approach to deal with the research problem. The *corpus* of this research was obtained through the use of operators *ext* and *site* on the Google search engine (www.google.com.br), through which 14,695 documents have been recovered on the UDESC website and 32,677 documents on the UFSC website. After that, the same analysis was repeated for the Library Science websites, also hosted in both institutions. This time we have then recovered 524 documents on the UDESC website and 57 documents on the UFSC website. Furthermore, a survey on interoperability was carried out, which consisted of analyzing the consistence of the documents that have been retrieved in both websites, concerning any loss of information, regardless of the files' extensions and the used softwares. For this purpose, 50 documents were selected out of the total retrieved, which were opened by two office softwares: Microsoft Office 2010 and LibreOffice 3.6. The operating system used as a platform was Windows 7 Ultimate 64-bit Service Pack 1. The obtained results have been presented in charts and figures that show us the current status of interoperability politics in the two institutions. Finally, we can consider that both UFSC and UDESC are not adopting the guidelines proposed by the Brazilian Federal Government. Regarding the interoperability, we have found out that there were significant losses of information in the documents, and in some cases they were even completely unable to be opened and read.

Keywords: Editable digital documents. Interoperability. ODF. Proprietary extensions. Non-proprietary extensions.

LISTA DE GRÁFICOS

Gráfico 1 – Extensões editáveis no domínio udesc.br.....	50
Gráfico 2 – Extensões editáveis no domínio ufsc.br.....	51
Gráfico 3 – Extensões editáveis no sitio faed.udesc.br	52
Gráfico 4 – Extensões editáveis no sítio dptcin.paginas.ufsc.br.....	52
Gráfico 5 – Quantidade de documentos selecionados por extensões.....	53
Gráfico 6 – Quantidade de extensões Proprietárias e Não-Proprietárias.....	54
Gráfico 7 – Documentos com problemas X Documentos sem problemas	56
Gráfico 8 – Quantidade de problemas utilizando o <i>MS Office 2010</i>	56
Gráfico 9 – Porcentagem de documentos abertos utilizando o MS Office 2010.....	57
Gráfico 10 - Quantidade de problemas utilizando o LibreOffice	58

LISTA DE TABELAS

Tabela 1 – Proporção de domicílios com computador.....	23
Tabela 2 – Proporção de domicílio com acesso à Internet (excluindo o acesso por telefone celular).....	23
Tabela 3 – Extensões proprietárias e não-proprietárias	26
Tabela 4 – Tabela com as informação de compatibilidade do sistemas da Microsoft Corporation.....	28
Tabela 5 – Relação de arquivos utilizados na avaliação da interoperabilidade	54

LISTA DE FIGURAS

Figura 1 – Dependências do software proprietário	29
Figura 2 – Ilustração de Interoperabilidade	33
Figura 3 – Documento com extensão .docx aberto no LibreOffice	34
Figura 4 – Documento com extensão .docx aberto no MS Office 2010	34
Figura 5 – Documento com extensão .odp aberto no LibreOffice	35
Figura 6 – Documento com extensão .odp aberto no Microsoft Office 2010	35
Figura 7 – Tentativa de abertura de documento com extensão não-proprietária num software proprietário.....	36
Figura 8 – Dependências do software não-proprietário.....	37
Figura 9 – Ilustração das técnicas de Migração, Encapsulamento e Emulação.....	42
Figura 10 – Tela da pesquisa com os operadores utilizados	48

LISTA DE SIGLAS

CCJC	Comissão de Constituição e Justiça e de Cidadania
CCP	Coordenação de Comissões Permanentes
CETIC	Centro de Estudos sobre as Tecnologias da Informação e da Comunicação
CIN	Ciência da Informação
CITRA	Conférence Internationale de la Table Ronde des Archives
CONARQ	Conselho Nacional de Arquivos
e-PING	Padrões de Interoperabilidade de Governo Eletrônico
EPUB	Eletronic Publication
FAED	Faculdade de Educação
GB	Gigabyte
HTML	HyperText Markup Language
IBGE	Instituto Brasileiro de Geografia e Estatísticas
ICA	International Council on Archives
IEC	International Electrotechnical Commission
IP	Internet Protocol
ISO	International Organization for Standardization
MB	Megabyte
MHz	Megahertz
MS	Microsoft
NASA	National Aeronautics and Space Administration
OAIS	Open Archival Information System
ODF	Open Document Format
OOXML	Office Open Extensible Markup Language
PC	Personal Computer
PL	Projeto de Lei
PNAD	Pesquisa Nacional por Amostra de Domicílios
SMIL	Synchronized Multimedia Integration Language
SO	Sistema Operacional
SVG	Scalable Vector Graphics
RAM	Random Access Memory
TCP	Transmission Control Protocol
TIC	Tecnologia de Informação e Comunicação

UDESC	Universidade do Estado de Santa Catarina
UFSC	Universidade Federal de Santa Catarina
XML	Extensible Markup Language
XSL	Extensible Style Language

SUMÁRIO

1	INTRODUÇÃO	14
1.1	Motivação e Justificativa	15
1.2	Objetivo Geral	19
1.3	Objetivos específicos	20
1.4	Estrutura do Texto	20
2	FUNDAMENTAÇÃO TEÓRICA	22
2.1	Documentos Digitais	25
2.1.1	Extensões proprietárias	27
2.1.2	Extensões não-proprietárias	30
2.2	Interoperabilidade de Dados	32
2.3	Preservação Tradicional	38
2.4	Preservação Digital	39
2.5	Legislação Brasileira sobre documentos digitais	43
3	PROCEDIMENTOS METODOLÓGICOS	46
3.1	Etapas da Pesquisa	47
4	DISCUSSÃO DOS RESULTADOS	50
5	CONCLUSÕES	60
5.1	Trabalhos futuros	61
	REFERÊNCIAS	63
	APÊNDICE	67

1 INTRODUÇÃO

Desde a segunda guerra mundial o mundo vem apresentando mudanças significativas no modo de vida da população. Uma dessas mudanças foi a invenção do computador que, num primeiro momento, é utilizado apenas em projetos militares para, logo depois, passar a ser usado também por instituições acadêmicas.

Com o passar dos anos e da facilitação do acesso aos computadores, empresas e instituições passaram a investir em máquinas e interligá-las, criando sua própria rede e facilitando assim suas atividades e estudos. Devido à falta de protocolos padrões de comunicação na época, todas as redes de computadores que existiam eram isoladas, ou seja, só se comunicavam com redes que possuíam a mesma tecnologia de comunicação. Pensando nessa dificuldade de comunicação, foi desenvolvido o protocolo TCP/IP¹, que possibilitou a interconexão das redes no mundo todo, possibilitando posteriormente, a criação da Internet como conhecemos hoje.

Assim como a imprensa foi responsável pelo aumento e disseminação do registro em papel no século XVIII, a Internet foi a responsável pelo aumento em escala exponencial na produção e disseminação do documento digital no fim do século XX.

Devido a esse crescimento na produção e uso de documentos digitais, os profissionais da área de Ciências da Informação, já em meados dos anos 1960, alertavam que a produção científica, apenas no meio digital, se perderia, pois não havia a devida preocupação com a criação desses documentos e nem estratégias de preservação. Segundo Thomaz (2005, p. 11),

Apesar de o tema já ter sido abordado pela primeira vez em 1964, considera-se o marco inicial das discussões sobre os desafios do documento eletrônico para os arquivos, o relatório preparado pelo professor Robert Henri Bautier para o *International Council on Archives - ICA*, apresentado na *Conférence Internationale de la Table Ronde des Archives - CITRA*, realizada em Bonn, Alemanha, em 1971.

¹ Transmission Control Protocol - Protocolo de Controle de Transmissão e o Internet Protocol - Protocolo de Interconexão.

O relatório do professor Bautier apresentava as seguintes conclusões;

1. Crescimento no uso das técnicas de computação pelos arquivos no final dos anos 1960;
2. Crescimento da quantidade de pesquisadores, utilizando os computadores para processar seus dados no período de 1965-1971, levando ao aumento da demanda por documentos eletrônicos na administração pública; e
3. Eliminação de documentos em suporte digital por parte de algumas instituições no mesmo período sem a intervenção de arquivistas em função de três fatores básicos: 1) falta de conhecimento sobre computação; 2) escassez de recursos para preservação do acervo e 3) obstáculos legais para o recolhimento desse tipo de documento (BAUTIER, 1971 *apud* THOMAZ, 2005, p. 12).

Após mais de 40 anos da publicação do relatório do professor Bautier, ainda é enfrentado o mesmo problema exposto na terceira conclusão. Conforme afirma Innarelli (2011, p. 83),

Atualmente, por conta da ignorância, de interesses ou negligência com a documentação, os profissionais da Tecnologia da Informação e Comunicação atuaram como responsáveis pela gestão e preservação da documentação digital, pois os profissionais da Ciência da Informação e suas áreas afins desconheciam, não só a importância das novas tecnologias para a gestão documental, mas a existência da documentação digital, neste meio tempo, os documentos digitais foram geridos por profissionais que não têm competência para entender a teoria em relação aos documentos gerados e gerenciados pelos sistemas de informação que desenvolveram e desenvolvem.

Esse problema no Brasil, segundo Innarelli (2011), ainda ocorre, pois os grupos que vêm discutindo acerca da preservação digital não têm a visibilidade, o investimento e a importância que deveriam ter.

1.1 Motivação e Justificativa

O interesse em abordar esse tema surgiu quando o autor realizava um estudo sobre o que escrever para o trabalho de conclusão de curso, sobre o tema preservação de documentos digitais, e se deparou com uma reportagem de Ricardo Cesar (2007), na revista Exame, intitulada “O esperanto dos arquivos”, que

apresentava as dificuldades existentes no acesso dos documentos digitais que eram encaminhados para as equipes de resgate que auxiliaram as vítimas do tsunami de 2004.

Além disso, o autor estuda e é adepto do *software* livre há vários anos e, de certa, forma acredita assim como outros, que a alternativa para essa dificuldade tecnológica seja a adoção de um padrão aberto e público, onde todos possam ter acesso, estudar e usufruir das informações e orientações sobre *software* livre. O movimento do *software* livre prega quatro preceitos, que ficaram conhecidos como “as quatro liberdades²”:

- A liberdade de executar o programa, para qualquer propósito;
- A liberdade de estudar como o programa funciona e adaptá-lo para as suas necessidades;
- A liberdade de redistribuir cópias de modo que você possa ajudar o seu próximo;
- A liberdade de aperfeiçoar o programa, e liberar os seus aperfeiçoamentos, de modo que toda a comunidade se beneficie deles.

Outra motivação deste trabalho advém da experiência pessoal do autor durante as aulas do curso de Biblioteconomia na UFSC, quando muitos colegas de turma reportavam dificuldades em abrir determinados documentos. Geralmente o que ocorria era que se constatava que a versão do *software* utilizada por eles para abrir tais documentos, apesar de ser da mesma empresa, era mais antiga e já não conseguia mais trabalhar com um documento criado num formato mais recente do mesmo *software*.

Para a sociedade a motivação de utilizar um padrão aberto é de garantir a interoperabilidade dos documentos em diferentes *softwares* de escritório, possibilitando com isso a longevidade do documento para gerações futuras. Para a área de biblioteconomia a motivação é de despertar o interesse dos profissionais da informação no suporte eletrônico em particular os documentos editáveis, pois se observa uma migração para o digital.

Vamos utilizar uma analogia para explicar a razão da existência de um

² <http://www.gnu.org/philosophy/free-sw.pt-br.html> - Acesso em: 17 nov 2012.

padrão aplicado a documentos digitais. Suponhamos um avião, que para ser desenvolvido custou em torno de R\$ 35 bilhões, tem seu valor de compra no mercado avaliado em R\$ 750 milhões, e pode chegar tranquilamente a 20 anos de voo.

Imaginemos que o manual desse avião já foi escrito todo em meio digital e que, durante esse tempo, a arquitetura computacional e o *software* utilizados na edição desse manual ficaram defasados e não foi pensada nenhuma estratégia para a sua preservação digital.

Suponhamos ainda que é detectado no avião, depois de 10 anos de operação e durante uma manutenção de rotina, um problema com um de seus componentes. Num primeiro momento não se sabe ao certo qual é a origem deste problema. Os técnicos e engenheiros responsáveis analisam as informações existentes e não conseguem chegar a uma solução sem consultar o manual da aeronave.

No momento em que se tenta consultar o manual ocorre uma desagradável surpresa, pois o mesmo não está acessível com a estrutura tecnológica disponível, ou seja, além do problema no avião surgiu outro problema, que é o do acesso para leitura do documento que contém o manual. Uma possível solução seria entrar em contato com a empresa que desenvolveu o *software* que gerou esse documento, com o intuito de solicitar uma solução por parte da mesma. Sem dúvida a empresa irá conseguir abrir esse documento, afinal ela é a proprietária da extensão do documento. Porém, é de se esperar que um preço bastante elevado seja cobrado por este serviço. Outra solução seria simplesmente descartar o avião e vendê-lo, com apenas metade da sua vida útil percorrida, e comprar um novo.

Utilizando essa mesma analogia, se fosse utilizada uma extensão de *software* não-proprietária a solução seria simples, bastaria “baixar” na Internet um *software* que tenha suporte a essa extensão, acessar o documento e as informações necessárias para solução do problema do avião, e deixá-lo em condições de voo novamente.

Pode-se concluir, portanto, que um padrão aberto, público e interoperável seria a melhor opção para a utilização em sistemas ou ambientes que necessitem de segurança, durabilidade e longevidade.

Innarelli também afirma que:

Nossa sociedade preserva sua cultura através da transmissão do conhecimento e de sua conseqüente apropriação, sendo que essa apropriação, quando registrada e transmitida, gera um novo estado de conhecimento, ciclo que garante nossa evolução sócio-cultural. (INNARELLI, 2011, p. 73)

Arellano complementa, dizendo que:

A natureza dos documentos digitais está permitindo ampla produção e disseminação de informação no mundo atual. É fato que na era da informação digital se está dando muita ênfase à geração e/ou aquisição de material digital, em vez de manter a preservação e o acesso a longo prazo aos acervos eletrônicos existentes (MÁRDERO ARELLANO, 2008, p. 44).

Podemos perceber a evidente preocupação dos autores Innarelli e Arellano quanto à importância da preservação, transmissão, durabilidade e longevidade da informação para os usuários, garantindo desta forma, a preservação do conhecimento da humanidade.

Indo ao encontro desta questão, a recuperação da informação é um ponto central na preservação digital de documentos. Pois, apesar de todos os avanços das técnicas de recuperação da informação de documentos digitais, este é ainda é um problema com o qual ainda nos deparamos consecutivas vezes.

[...] as fotos de satélites da NASA dos anos 70, que documentavam as condições da floresta amazônica brasileira estão irremediavelmente perdidas, assim como dados do censo americano de 1960 (BORGHOFF et al, 2003 *apud* ALMEIDA, CENDÓN, SOUZA, 2012, p. 128).

Outro exemplo foram os problemas enfrentados pelas equipes de resgate no Tsunami de 2004, no Sudeste Asiático:

A operação de auxílio às vítimas do tsunami no final de 2004 foi marcada por todo tipo de dificuldade. Percalços seriam inevitáveis em uma tragédia dessa proporção, que deixou um rastro de 200 000 mortes no Sudeste Asiático. Mas as equipes enviadas por mais de 60 países depararam com um problema inesperado: os relatórios e as planilhas trocados entre elas muitas vezes não abriam ou sofriam perdas de configuração que comprometiam seu entendimento. Como se não bastasse a dimensão da tragédia, os trabalhos de resgate e apoio às vítimas tiveram de enfrentar as complicações da tecnologia (CESAR, 2007).

Na contramão da tendência mundial, a biblioteca da Virgínia, nos Estados Unidos da América, tomou uma decisão surpreendente para arquivar seus documentos públicos.

Como não há normas nacionais para a permanência do meio em que os registros eletrônicos são mantidos (o meio é não considerado permanente), registros em formatos eletrônicos não são aceitáveis para armazenamento de arquivos. Se ou quando os padrões de permanência de mídia eletrônica se tornarem uma realidade, esta política será revista em conformidade. Até esse momento, registros eletrônicos avaliados como de arquivo permanente devem ser mantidos na agência ou localidade convertidos em microfilme, papel, ou outro meio aceitável que atendam às normas nacionais para a permanente retenção de arquivo antes de serem transferidos para o Arquivo (BIBLIOTECA DA VIRGÍNIA, 1999, tradução nossa).

Com uma atitude como a da biblioteca da Virgínia, a manutenção do seu acervo dependerá de investimento muito alto, ou seja, não se tem mais o problema de recuperar as informações nos documentos digitais, porém existem vários outros problemas, como afirma Arellano:

O suporte físico da informação, o papel e a superfície metálica magnetizada se desintegram ou podem se tornar irrecuperáveis. Existem, ademais, os efeitos da temperatura, umidade, nível de poluição do ar e das ameaças biológicas; os danos provocados pelo uso indevido e o uso regular, as catástrofes naturais e a obsolescência tecnológica (ARELLANO, 2004, p.15).

Diante desse cenário, nosso olhar se volta para o Brasil, onde procuramos verificar se estamos tendo a preocupação de preservar nossos documentos de forma que no futuro estes possam ser utilizados sem nenhum problema.

1.2 Objetivo Geral

Este trabalho analisa os documentos digitais editáveis disponíveis nos sítios da Universidade Federal de Santa Catarina e Universidade do Estado de Santa Catarina para verificar se os mesmos estão em conformidade com as orientações do Governo Brasileiro, que sugere o uso de formato aberto.

1.3 Objetivos específicos

- Quantificar os documentos editáveis que estão disponíveis para o usuário nos domínios ufsc.br e udesc.br;
- Efetuar um levantamento sobre as extensões de documentos existentes nos referidos sítios, apontando igualmente quais extensões são proprietárias e quais são não-proprietárias, categorizando os documentos editáveis por *softwares* de escritório;
- Testar a interoperabilidade dos documentos com *softwares* de escritório proprietários e não-proprietários;
- Verificar o grau de adaptação dos referidos domínios às orientações do governo federal para a criação de documentos.

1.4 Estrutura do Texto

A presente pesquisa está estruturada em cinco capítulos. No Capítulo 1 apresenta-se a introdução, a importância do tema, as justificativas, o problema da pesquisa e os objetivos propostos.

O Capítulo 2 mostra a fundamentação na literatura para o desenvolvimento da pesquisa, abordando os temas: Documentos Digitais, Extensões proprietárias, Extensões não-proprietárias, Interoperabilidade de Dados, Preservação Tradicional, Preservação Digital e Legislação Brasileira sobre Documentos Digitais.

O Capítulo 3 aborda os procedimentos metodológicos que foram utilizados para o desenvolvimento da pesquisa, contendo a descrição de como a mesma foi realizada, incluindo-se aí os operadores e sítios utilizados para obtenção dos dados coletados. No Capítulo 4 é apresentada a análise dos resultados. A partir da quantidade de documentos coletados no domínio das duas instituições e nos sítios dos cursos de Biblioteconomia, são apresentadas e discutidas as extensões que são utilizadas, o tipo das extensões (proprietárias ou não-proprietárias), quais destas extensões apresentam algum tipo de problema de interoperabilidade, e que problemas são esses.

O Capítulo 5 traz as conclusões do estudo. Ao final, são apresentadas as

referências bibliográficas utilizadas no desenvolvimento da pesquisa.

2 FUNDAMENTAÇÃO TEÓRICA

Nos últimos anos tem sido observado no Brasil um aumento nos número de computadores conectados à Internet. Entre outras coisas, isso ocorre devido aos recentes investimentos na área de telecomunicações e a um maior acesso das pessoas a produtos e serviços de tecnologia. Segundo o sitio³ do Ministério das Comunicações, os investimentos do Governo Brasileiro passaram de uma média de R\$ 17 bilhões/ano nos últimos dez anos para R\$ 21 bilhões em 2011, tendo uma previsão de chegar a R\$ 25 bilhões por ano em 2012. O público-alvo desses investimentos é o usuário doméstico. Portanto, quanto maior a participação deste tipo de usuário na Internet, maior será a quantidade de busca por informações on-line.

No Brasil, em 2008, segundo Pesquisa Nacional por Amostra de Domicílios (Pnad), realizado pelo Instituto Brasileiro de Geografia e Estatísticas – IBGE, 31% das residências possuíam computadores e, destas, 24% possuíam acesso à Internet.

Em 2008, 17,95 milhões de domicílios brasileiros (31,2%) possuíam microcomputador, contra 26,5% em 2007. Desses, 13,7 milhões (23,8%) têm acesso à internet, contra 20% em 2007. Mais da metade dos domicílios com computador (10,2 milhões) estavam no Sudeste em 2008, dos quais 7,98 milhões tinham com acesso à internet. (IBGE, 2008)

Em 2011, um estudo realizado pelo Centro de Estudos sobre as Tecnologias da Informação e da Comunicação (CETIC) atualizou os dados de domicílios com computador e acesso à Internet.

³ <http://www.mc.gov.br/telecomunicacoes/noticias-telecomunicacoes/25355-portaria-do-minicom-vai-incentivar-investimentos-no-setor-de-comunicacoes>

Tabela 1 – Proporção de domicílios com computador

Percentual (%)		Sim	Não	Não sabe/ Não respondeu
TOTAL BRASIL		45	55	0
ÁREA	URBANA	51	49	0
	RURAL	16	84	0
REGIÃO	SUDESTE	57	43	0
	NORDESTE	25	75	0
	SUL	56	44	0
	NORTE	32	68	0
	CENTRO-OESTE	47	53	0
RENDA FAMILIAR	Até 1 SM	10	90	0
	Mais de 1 SM até 2 SM	28	72	0
	Mais de 1 SM até 2 SM	48	52	0
	Mais de 3 SM até 5 SM	67	33	0
	Mais de 5 SM até 10 SM	84	16	0
	Mais de 10 SM	93	7	0
CLASSE SOCIAL	A	98	2	0
	B	84	16	0
	C	45	55	0
	DE	8	92	0

Percentual sobre o total de domicílios. (base: 25.000 domicílios)

Fonte – CETIC - <http://cetic.br/usuarios/tic/2011-total-brasil/index.htm>

Tabela 2 – Proporção de domicílio com acesso à Internet (excluindo o acesso por telefone celular).

Percentual (%)		Sim	Não	Não sabe / Não respondeu
TOTAL BRASIL		38	62	0
ÁREA	URBANA	43	57	0
	RURAL	10	90	0
REGIÃO	SUDESTE	49	51	0
	NORDESTE	21	79	0
	SUL	45	55	0
	NORTE	22	78	0
	CENTRO-OESTE	39	61	0
RENDA FAMILIAR	Até 1 SM	6	94	0
	Mais de 1 SM até 2 SM	21	79	0
	Mais de 1 SM até 2 SM	38	62	0
	Mais de 3 SM até 5 SM	58	42	0
	Mais de 5 SM até 10 SM	76	23	0
	Mais de 10 SM	91	9	0
CLASSE SOCIAL	A	96	3	0
	B	76	24	0
	C	35	65	0
	DE	5	95	0

Percentual sobre o total de domicílios. (base: 25.000 domicílios)

Fonte – CETIC - <http://cetic.br/usuarios/tic/2011-total-brasil/index.htm>

Analisando as tabelas 1 e 2 é possível perceber um aumento de 14% no número de domicílios que possuem computador fazendo uma relação entre os anos de 2008 e 2011. Este mesmo crescimento também é identificado nos domicílios que têm acesso à Internet, em relação aos mesmos anos. Na área urbana, mais da metade dos domicílios possuem computador e, destes, 43% estão conectados à Internet.

O estudo do IBGE se concentra nas pessoas de 10 anos ou mais de idade que utilizaram a Internet no período de referência dos últimos três meses, classificando-as por finalidade de acesso à Internet e segundo as Unidades da Federação e as Regiões Metropolitanas. Os objetivos de acesso que emergiram nos resultados desta pesquisa foram:

- O acesso por Educação e aprendizado (de 71,7% em 2005 para 65,9% em 2008) uma queda de 5,8%, falta de investimento em divulgação dos serviços;
- Comunicação com outras pessoas (de 68,6% em 2005 para 83,2% em 2008) alta de 14,6%, devido à explosão das redes sociais e dos comunicadores como Twitter, Orkut, Facebook, MSN e Skype;
- Atividade de lazer (de 54,3% em 2005 para 68,6% em 2008) alta de 14,3% com investimentos em telecomunicações e melhoramento da rede (Internet) foi possível disponibilizar jogos online e disponibilizar eventos ao vivo para os usuários desses serviços;
- Leitura de jornais e revistas (de 46,9% em 2005 para 48,6% em 2008) alta de 1,7% devido à disponibilização em meio digital das principais revista e jornais do país;
- Comprar ou encomendar bens ou serviços, transações bancárias e financeiras (de 25,2% em 2005 para 20,9% em 2008) queda de 4,3% devido à insegurança, notícias de invasão a sites de banco e desconhecimento de como utilizar esse tipo de serviço;
- Interação com autoridades públicas ou órgãos do governo, buscar informações e outros serviços (de 42,9% em 2005 para 34,8% em 2008) queda de 8,1% devido a falta de divulgação dos seus serviços.

A partir de uma análise por estados (ver Apêndice), pode-se constatar que

o estado de Santa Catarina teve um aumento de 4,3% no que diz respeito à interação com autoridades públicas ou órgãos do governo e à busca de informações e outros serviços (subindo de 35,1% em 2005 para 39,4% em 2008). Na direção oposta, o estado do Maranhão teve uma perda de 22,1% com relação à pesquisa de 2005 (indo de 41,2% em 2005 para 19,0% em 2008).

Santa Catarina se destaca em relação à tendência nacional de queda (8,1%) por busca de informações relacionadas ao governo no meio eletrônico. Portanto, é evidente a necessidade de se investigar se as instituições governamentais e públicas do estado de Santa Catarina estão se preparando para esta crescente demanda eletrônica. Em especial, esse trabalho de conclusão de curso irá focar nas instituições de educação superior UFSC e UDESC.

2.1 Documentos Digitais

Documentos digitais são documentos que têm sua origem no meio digital ou eletrônico. Todo documento digital gera um arquivo de computador, que por sua vez tem uma extensão e esta extensão serve de indicação ao sistema operacional do programa que deverá abrir o arquivo. Conforme afirmam Curty e Gama (2007, p. 115):

Como é sabido, os documentos digitais podem ser tomados como aqueles originalmente criados em computadores ou aqueles advindos de processos de conversão e digitalização, representados por estruturas binárias legíveis por computador.

Nos dias atuais temos uma variedade de extensões que estão associadas a arquivos de documentos digitais. Como esse trabalho tem por objetivo estudar os documentos editáveis por *softwares* de escritório, são apresentadas abaixo as principais extensões utilizadas, por esses *softwares*, para documentos editáveis.

Tabela 3 – Extensões proprietárias e não-proprietárias

Tipo	Extensão	Proprietária	Não-Proprietária
Documento de Texto	.odt		X
Modelo de documento de texto	.ott		X
Documento de Texto do OpenOffice.org 1.0	.sxw		X
Modelo de documento de Texto do OpenOffice.org 1.0	.stw		X
OpenDocument Text (Flat XML)	.fodt		X
Unified Office Format text	.uot		X
Rich Text Format	.rtf	X	
Microsoft Word 2007/2010	.docx	X	
Microsoft Word 6.0/95/97/2000/XP/2003	.doc	X	
Documento habilitado para macro	.docm	X	
Modelo	.dotx	X	
Modelo habilitado para macro	.dotm	X	
Planilha de cálculo	.ots/.ods		X
Apresentação de slides	.otp/.odp		X
Pasta de trabalho	.xlsx	X	
Pasta de trabalho habilitada para macro	.xlsm	X	
Modelo	.xltx	X	
Modelo habilitado para macro	.xltn	X	
Pasta de trabalho binária não-XML	.xlsb	X	
Suplemento habilitado para macro	.xlam	X	
Apresentação	.pptx / ppt	X	
Apresentação habilitada para macro	.pptm	X	
Modelo	.potx	X	
Modelo habilitado para macro	.potm	X	
Suplemento habilitado para macro	.ppam	X	
Apresentação de slides	.ppsx	X	
Apresentação de slides habilitada para macro	.ppsm	X	
Slide	.sldx	X	
Slide Habilitado para macro	.sldm	X	
Tema do Office	.thmx	X	

Fonte: Microsoft Corporation - <http://office.microsoft.com/pt-br/word-help/introducao-a-novas-extensoes-de-nome-de-ficheiro-e-formatos-office-xml-HA010006935.aspx> – Com adaptação do autor.

Essas extensões em sua grande maioria são **proprietárias**, ou seja, para uma aplicação ter o suporte a essas extensões, será necessário o pagamento de *royalties*. No entanto, em muitas situações, quem tem a patente da extensão não a disponibiliza, obrigando desta forma, a compra de seus aplicativos. Além disso, existem as extensões que são chamadas de **não-proprietárias**, ou seja, para uma aplicação/*software* suportar essas extensões, é necessário seguir as normas que as

regulam.

2.1.1 Extensões proprietárias

Conforme Wensing (2010), extensões proprietárias são as extensões que foram desenvolvidas por uma empresa, a qual detém os direitos de utilizar, comercializar e modificar os tipos dos documentos que utilizam essas extensões.

Segundo Bodê (2007) e Wensing (2010) (*apud* MELLO, 2011, p. 29),

[...]possuem restrições para acesso às especificações técnicas proprietárias, por motivos legais, não podem ser utilizadas por outra empresa a fim de desenvolver *software* que leia documentos digitais gerados a partir dessas mesmas especificações[...]

Quando algum desenvolvedor se interessa em utilizar essa extensão em seu *software*, o proprietário da extensão pode ceder o direito de usá-la, desde que o interessado pague por isso. No entanto, quem pagou para uso dessa extensão não tem o direito de modificá-la para ajustar ao seu *software*, pelo contrário, o *software* é que terá que se adaptar à extensão.

Existem várias extensões proprietárias disponíveis no mercado. Com relação às utilizadas em *e-books*, por exemplo, o cenário torna-se ainda mais complexo, pois além de o usuário ter que utilizar a aplicação proprietária para abrir o documento, o *hardware* também tem que ser da mesma empresa. Um exemplo de extensões para *e-books* são as desenvolvidas pela *Amazon*⁴: *.azw* e *.azw1*. As extensões *.azw* e *.azw1* são utilizadas apenas pelo *Kindle*, que é o leitor de *e-books* da *Amazon*. Porém, a *Amazon* disponibilizou uma *app*, ou seja, uma aplicação de *software* para ser instalada em *tablets* e *smartphones*, que torna possível a leitura dos arquivos da *Amazon* nesses dispositivos. Estas extensões possuem o dispositivo DRM (*Digital Rights Management*) que impossibilita que o arquivo seja copiado para outro dispositivo que não seja o *Kindle* da *Amazon*.

Neste trabalho trataremos das extensões proprietárias através da suíte de escritório *Microsoft*⁵ *Office*, da empresa *Microsoft Corporation* na sua versão 2010. A escolha dessa suíte se deve ao fato de ser a mais utilizada no mundo.

⁴ <http://www.amazon.com/Kindle-Store-Brasil/lm/R1OJ9TYOBQHEZ1>

⁵ <http://www.microsoft.com/pt-br/default.aspx>

O *Microsoft Office 2010* oferece suporte a inúmeras extensões, porém as extensões-padrão deste *software* são:

- ✓ **.docx** – extensões são utilizadas para documentos de texto (*Microsoft Word*);
- ✓ **.xlsx** – extensões utilizadas para documentos de planilha de cálculo (*Microsoft Excel*);
- ✓ **.pptx** – extensões utilizadas para documentos de apresentações (*Microsoft PowerPoint*).

Quando escolhemos tais extensões para salvar documentos digitais, geramos uma cadeia de dependências, conforme indicadas na Tabela 4. Analisando a mesma, observamos que, na utilização das extensões proprietárias, existe a necessidade de se ter um requisito mínimo de *hardware* e *software* para que se possa utilizar a aplicação desejada, no caso o *Microsoft Office 2010*.

Tabela 4 – Tabela com as informações de compatibilidade do sistemas da *Microsoft Corporation*.

Processador	Sistema operacional	Memória RAM	Espaço em Disco
Processador de 500 MHz ou mais rápido	<i>Microsoft Windows XP</i> (é necessário ter SP3) (32 bits), <i>Microsoft Windows 7</i> , <i>Microsoft Windows Vista</i> com <i>Service Pack</i> ⁶ (SP) 1.	256 MB de RAM; 512 MB	3 GB de espaço disponível em disco

Fonte: *Microsoft Corporation* – <http://technet.microsoft.com/pt-br/library/ee624351.aspx>

A *Microsoft Corporation* liberou uma documentação⁷ sobre o seu formato de extensão proprietária, porém, segundo a empresa, os dados só serão interpretados com segurança por meio da utilização de suas aplicações. Conforme Silva (2008, p. 24),

A documentação sobre o formato proprietário do *Microsoft Office 2003* foi recentemente publicada pela *Microsoft Corporation* (15 de Fevereiro de 2008). De acordo com esta documentação, os dados armazenados no formato proprietário são armazenados em formato binário (não inteligível a humanos) e segundo a empresa, a documentação descreve todas as estruturas necessárias à manipulação destes documentos, ou seja, a organização dos bytes de dados dentro dos arquivos.

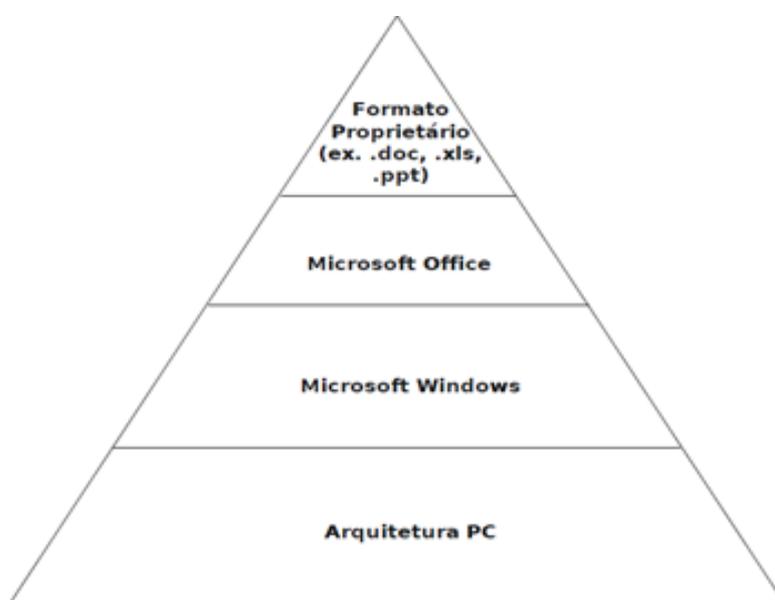
⁶ Um Service Pack (SP) é uma atualização do Windows, em geral combinando atualizações anteriores, que ajuda a deixar o Windows mais seguro. <http://windows.microsoft.com/pt-BR/windows/service-packs-download#sptabs=win7>. Acessado em 15 de nov. 2012.

⁷ <http://msdn.microsoft.com/en-us/library/cc313118.aspx> acesso: 12 de novembro de 2012.

Apesar da recente publicação da documentação do formato pela *Microsoft Corporation*, o acesso às informações contidas em documentos armazenados através do formato proprietário só podem ser seguramente acessadas através das ferramentas da *Microsoft Corporation*.

Abaixo, a Figura 1 ilustra, de forma sintetizada, a cadeia de dependência que o padrão proprietário traz para poder abrir um documento utilizando o seu padrão.

Figura 1 – Dependências do *software* proprietário



Fonte: Silva (2008)

No topo da pirâmide temos os dados propriamente ditos, o conteúdo dos documentos digitais, como esse trabalho aqui antes da sua impressão, por exemplo. Para salvá-los, escolhemos uma extensão **.doc**. Na parte de baixo, temos a aplicação utilizada para gerar esses dados, ou seja, o *Microsoft Office* com o seu editor de texto *Microsoft Word*. Na camada imediatamente inferior temos o sistema operacional a ser utilizado com o editor de texto, no caso o *Microsoft Windows*, que, por sua vez, para funcionar necessita de uma arquitetura PC.

Conclui-se então que, para abrir as extensões proprietárias dependemos de uma determinada suíte de escritório, de um determinado sistema operacional e de uma determinada arquitetura de computador. Vale lembrar que para se utilizar a suíte de escritório e o sistema operacional mostrados na figura anterior é

necessária ainda a aquisição de licenças. Ou seja, a utilização de uma extensão proprietária está configurada numa verdadeira cadeia de interdependências.

2.1.2 Extensões não-proprietárias

Diferentemente das extensões proprietárias, as extensões não-proprietárias não possuem um “dono”, ou seja, quem as desenvolve disponibiliza a documentação com suas diretrizes e especificações para quem tiver interesse em utilizá-las em seus *softwares*, sem a necessidade de ter que pagar por licenças ou direito de uso. Um exemplo disso é o formato ODF (*Open Document Format*). Nas palavras de Wensing (2010, p. 51),

O formato ODF foi desenvolvido por uma grande variedade de organizações, sendo possível usar livremente as respectivas especificações. Isto significa que o ODF pode ser implementado em qualquer sistema seja ele de código aberto ou não, sem ser necessário efetuar qualquer tipo de pagamento ou estar sujeito a uma licença de uso restrito. O ODF constitui-se numa alternativa aos formatos de documentação que são propriedade de empresas privadas, sujeitos às licenças de uso restrito ou onerosas, permitindo a organizações e indivíduos escolherem o *software* que mais lhes convém para lidar com os arquivos guardados nesse formato.

Além disso, elas quebram o paradigma de que o *software* é que precisa se ajustar às extensões, pois os desenvolvedores de posse da documentação podem modificar livremente o funcionamento das extensões, desde que também disponibilizem seus códigos modificados.

Um exemplo de extensão não-proprietária é o HTML (*HyperText Markup Language*), que é utilizado na criação de sites e que se tornou um padrão para a Internet. Há também o EPUB (*Electronic Publication*), que é um formato padrão de arquivo específico para *e-books*. O EPUB é uma extensão livre e aberta desenvolvida pelo *International Digital Publishing Forum* e que veio para substituir o *open e-book*, pois este novo formato fornece um método simplificado de preparar uma publicação digital em um único arquivo, tornando assim mais fácil para os editores, revendedores e outros oferecerem e utilizarem materiais de leitura digital. Diferentemente do *open e-book*, que é um pacote de arquivos XHTML (*eXtensible Hyper Text Language Markup*) que dependem de diferentes ferramentas para serem interpretados.

Outro exemplo é o XML (*eXtensible Markup Language*), que pode ser considerado um divisor de águas, pois foi criado para suportar a interoperabilidade entre aplicativos, e acabou por ser utilizado em várias aplicações, não apenas em documentos editáveis. O formato ODF foi pensado para ser utilizado em documentos editáveis tendo como base o XML.

Segundo Wensing (2010, p. 51),

O desenvolvimento do ODF contou e conta com a participação de diversas empresas e especialistas do mundo todo, garantindo assim a sua neutralidade tecnológica [...].

Outra extensão que utiliza o XML como base é o OOXML (*Office Open eXtensible Markup Language*) desenvolvido pela *Microsoft Corporation*. Segundo Suca e Silva (2011, p. 119),

O OOXML foi desenvolvido pela Microsoft em 2008 e tornou-se um padrão aberto ISO (ISO/IEC 29500:2008). OOXML foi projetado para representar o corpus preexistente de documentos de processamento de texto, apresentações e planilhas que são codificados pela *Microsoft*. A especificação OOXML contém material normativo e informativo estruturado em aproximadamente 6546 páginas.

Assim como foi feito com relação às extensões proprietárias, utilizaremos neste trabalho uma suíte de escritório, o *LibreOffice*, para explicar as extensões não-proprietárias.

As extensões utilizadas, por padrão, no *LibreOffice* são as não-proprietárias, tomando-se o formato ODF como base. Entre estas, as extensões não-proprietárias utilizadas nesse trabalho são:

- ✓ **.odt** – extensão utilizada para documentos de texto (*LibreOffice Writer*);
- ✓ **.ods** – extensão utilizada para documentos de planilha de cálculos (*LibreOffice Calc*);
- ✓ **.odp** – extensão utilizada para documentos de apresentação (*LibreOffice Impress*).

2.2 Interoperabilidade de Dados

Interoperabilidade é a capacidade dos sistemas digitais trocar informações sem que exista perda de dados, possibilitando a compreensão da informação e, com isso, facilitando a geração e a disseminação do conhecimento.

Na literatura existem várias definições de interoperabilidade,

Interoperabilidade é uma propriedade que se refere à habilidade de diversos sistemas e organizações em trabalhar de forma conjunta (ou interoperar). De acordo com o Dicionário Padrão de Computação do IEEE2, interoperabilidade é a habilidade de dois ou mais sistemas ou componentes, de trocar informação e utilizar a informação que foi trocada (SILVA, 2008, p. 6).

Segundo Ukoln (2005) e Miller (2000) (*apud* SAYÃO e MARCONDES, 2008, p. 136),

a interoperabilidade pode ser considerada como o processo contínuo de assegurar que sistemas, procedimentos e cultura de uma organização sejam gerenciados de tal forma que possibilitem a maximização das oportunidades para intercâmbio e reuso de informação.

Para Moura (2009, p. 61),

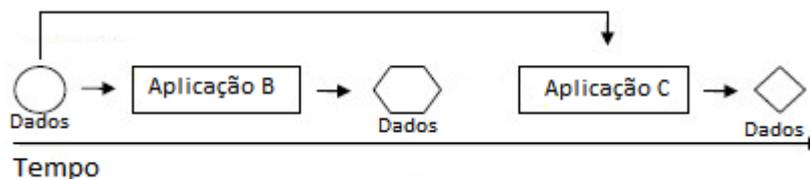
[..] a interoperabilidade é considerada como a capacidade de sistemas autônomos (informatizados ou não) comunicarem de modo transparente entre si, devido à adoção de padrões comuns e protocolos que permitem o uso compartilhado de informações.

Já Shah e Kesan (2009 *apud* SUCA e SILVA, 2011, p. 119) afirmam “que não existem implementações que ofereçam 100% de compatibilidade (avaliação da leitura dos documentos) dentro das implementações de OOXML e ODF”. Porém, Shah e Kesan (2009, p. 59, tradução nossa) dizem que “a única maneira de assegurar a interoperabilidade é usar sempre às versões mais estáveis e homologadas”.

Um exemplo para explicar o que é interoperabilidade seria podermos criar um documento, com qualquer extensão, e conseguirmos abri-lo em qualquer ambiente computacional. As suas características não deveriam ser perdidas, ou seja, a informação contida no texto teria que poder ser utilizada para a geração de

novos textos/documentos.

Figura 2 – Ilustração de Interoperabilidade



Fonte: Keep, [2011?]. (apud DIAS, 2011) – adaptado pelo autor

A Figura 2 nos mostra como é o processo de interoperabilidade. No início temos um arquivo criado em uma aplicação A, que possui dados característicos, como macros, animações, entre outras coisas. Quando este arquivo é aberto em uma aplicação de terceiros, por exemplo, ao se utilizar a aplicação B, que pertence a uma determinada empresa, algumas características se perdem. Na figura acima os dados originais são representados por um círculo, que, depois de abertos com a aplicação B se transformam em um hexágono, perdendo algumas das características originais. Isso não significa, no entanto, que este arquivo não poderá ser lido, pois ainda será possível interpretar os dados existentes no arquivo após a abertura do mesmo pela aplicação B. Por fim, quando a aplicação C (de outra empresa) abre esse mesmo arquivo, as características se perdem ainda mais. Ou seja, imaginemos que a aplicação C não pagou os *royalties* necessários, a exemplo da aplicação B, para abrir o documento com o mínimo de formatação, e acabou por transformar o círculo original em losango, apesar deste não bloquear completamente a interpretação dos dados originais, igualmente da mesma forma que a aplicação B.

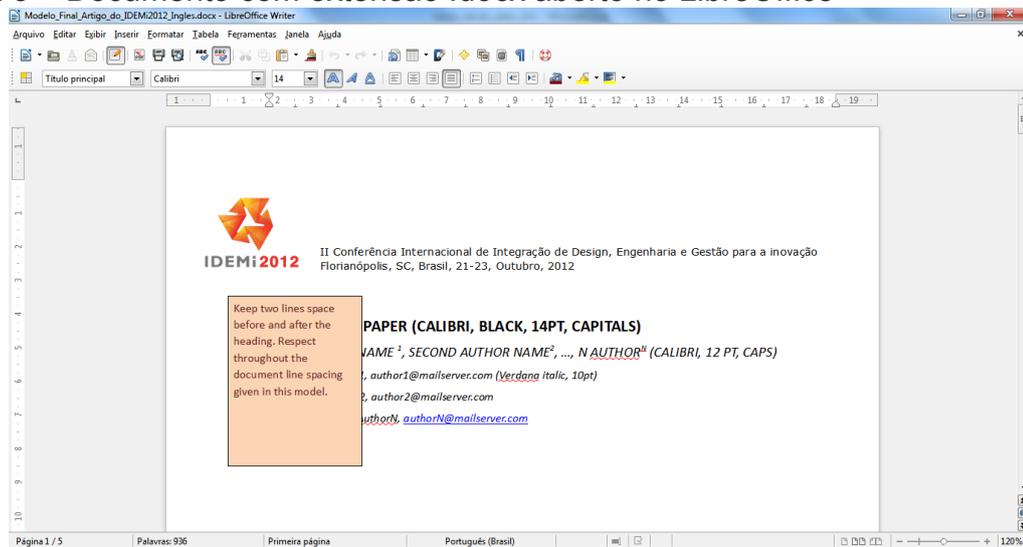
Portanto, na contemporaneidade, onde o poder de processamento dos computadores, celulares e *tablets* cresce continuamente, a melhor forma de salvaguardar um documento sem que existam perdas ou dificuldades de interpretá-los é a utilização de um padrão para armazenamento de documentos digitais editáveis. Desta forma será possível recuperar o documento em qualquer período histórico, independentemente do *software* ou *hardware* que estivermos utilizando.

O *LibreOffice*⁸, que atualmente está na sua versão 3.6, está disponível Internet e consegue abrir praticamente todas as extensões existentes não-proprietárias e proprietárias. Porém, algumas dessas extensões proprietárias não liberam informações sobre sua estrutura. Para contornar este problema, o

⁸ <http://pt-br.libreoffice.org/baixe-ja/>

LibreOffice utiliza um artifício técnico conhecido como engenharia reversa, ou seja, a partir do documento final possuidor de uma extensão proprietária tenta-se chegar a uma estrutura que melhor o represente. Muitas vezes o resultado não sai como o esperado e as informações ficam truncadas e desformatadas (Figura 3).

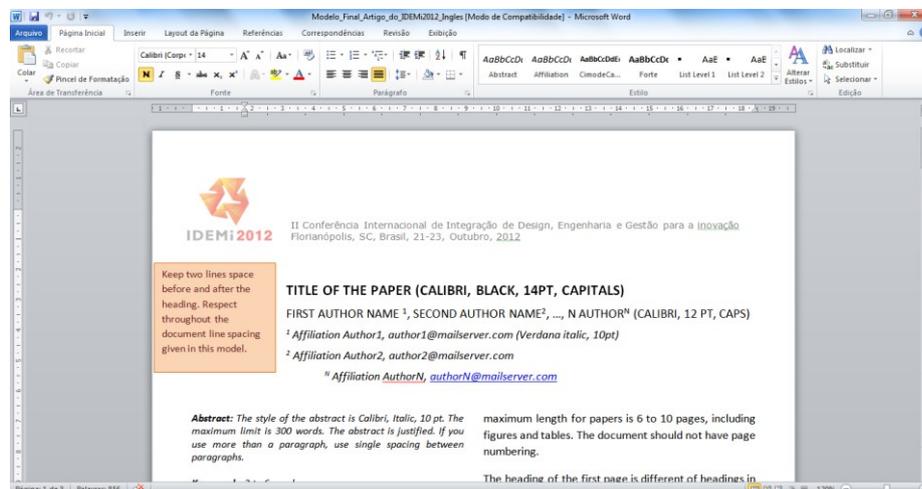
Figura 3 – Documento com extensão .docx aberto no LibreOffice



Fonte – Elaborado pelo autor

A Figura 3 apresenta uma tentativa de abertura de um documento com extensão proprietária no *LibreOffice*. Apesar de estar desformatado, é possível compreender os dados contidos no documento. Na Figura 4 é apresentado o mesmo documento, porém agora sem o problema de formatação, já que desta vez o *software* utilizado é o mesmo que o criou.

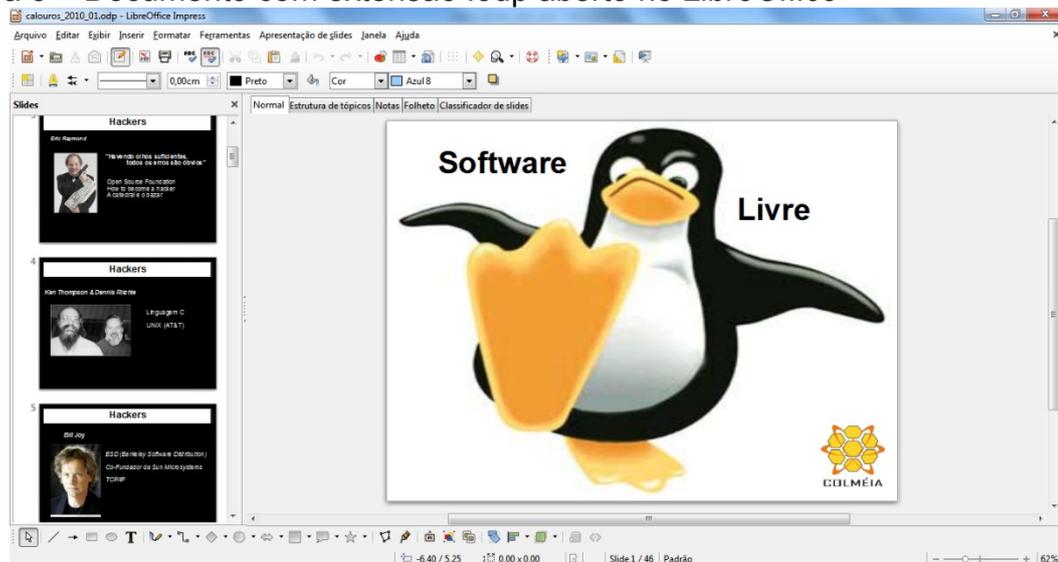
Figura 4 – Documento com extensão .docx aberto no MS Office 2010



Fonte – Elaborado pelo autor

A Figura 5 apresenta um exemplo de arquivo criado com extensão não-proprietária e aberto no *software* de escritório *LibreOffice*. Observa-se que nenhum problema de interoperabilidade é apresentado.

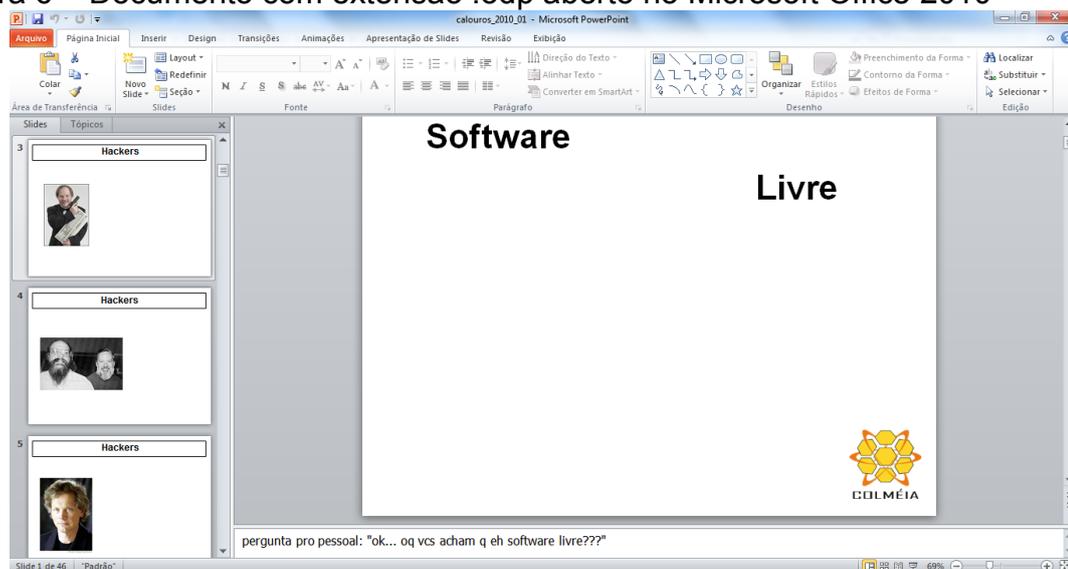
Figura 5 – Documento com extensão .odp aberto no *LibreOffice*



Fonte – Elaborado pelo autor

Já na Figura 6, quando o mesmo arquivo é aberto no *Microsoft Office 2010*, o problema de interoperabilidade fica evidente.

Figura 6 – Documento com extensão .odp aberto no Microsoft Office 2010

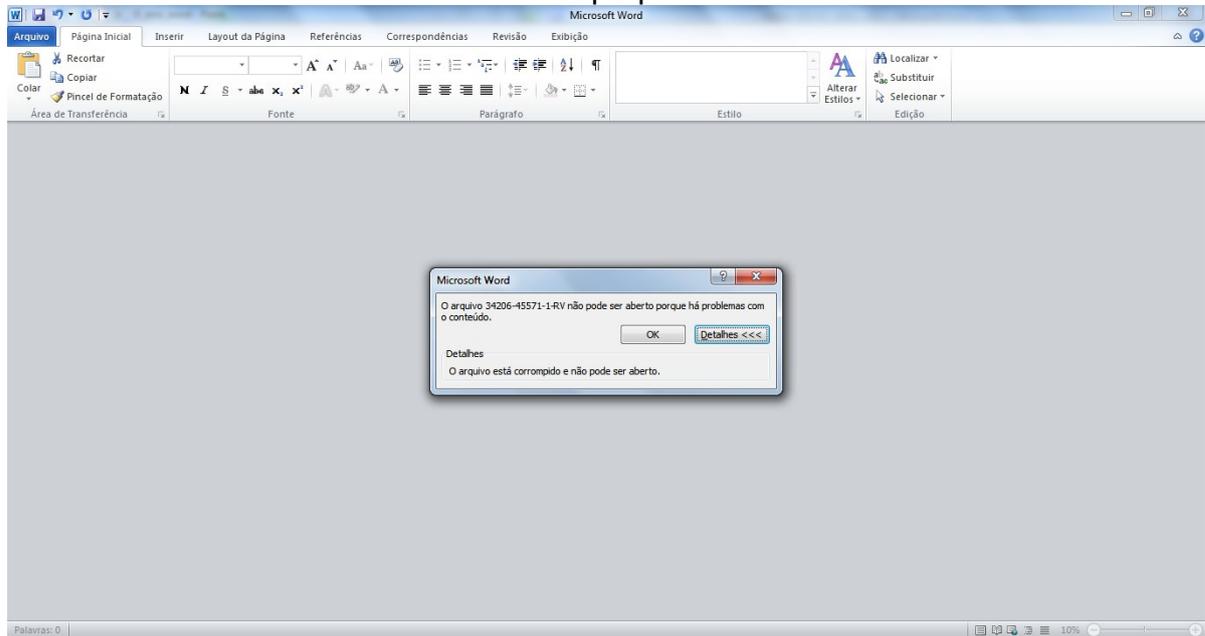


Fonte – Elaborado pelo autor

Na Figura 7 é apresentada a tentativa de se abrir um documento de

extensão .odt no *Microsoft Office 2010*.

Figura 7 – Tentativa de abertura de documento com extensão não-proprietária num *software* proprietário



Fonte – Elaborado pelo autor

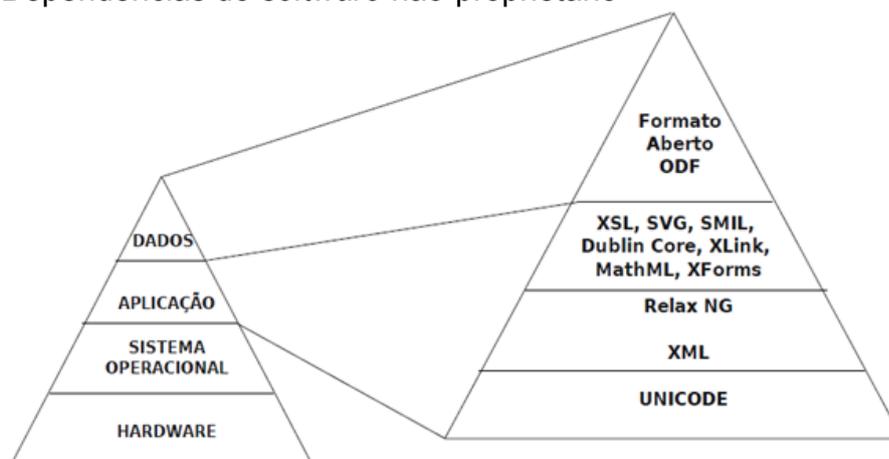
A engenharia reversa de *software* consiste em analisar um determinado sistema para criar representações do próprio em um nível mais alto de abstração. Também pode ser encarada como “Voltar atrás no ciclo de desenvolvimento do *software*”. Na prática, existem dois tipos de engenharia reversa de *software*:

- No primeiro caso, o código-fonte já está disponível, mas os aspectos mais globais, talvez documentação escassa ou não válida, têm que ser descobertos.
- No segundo caso o código-fonte do *software* não está disponível, e todos os esforços para descobrir uma possível fonte do código para o *software* são considerados como engenharia reversa (CANHOTA JUNIOR et al., 2005).

Diferentemente das extensões proprietárias, as dependências encontradas nas extensões não-proprietárias se restringem apenas à categoria da aplicação. Por exemplo, se a aplicação seguir as normas que estão descritas na NBR 26300/2008, que é a especificação do formato ODF para aplicação de escritório, conseguirá abrir documentos que possuem as extensões ODF.

A Figura 8 ilustra, de forma sintetizada, a cadeia de dependência do padrão aberto ODF.

Figura 8 – Dependências do *software* não-proprietário



Fonte: Silva (2008)

Conclui-se que para o dado ser interpretado por uma aplicação, é necessário que a mesma possua suporte a padrões dos quais o ODF depende, como, por exemplo,

- *XLS (eXtensible Style Language)* esta especificação define as características e a sintaxe para o Extensible Stylesheet Language (XSL), uma linguagem para expressar folhas de estilo;
- *SVG (Scalable Vector Graphics)* esta especificação define as características e a sintaxe para Scalable Vector Graphics (SVG) Versão 1.1, uma linguagem modularizada para descrever gráficos vetoriais bidimensionais e gráficos mistos vetor/raster em XML;
- *Extensible Markup Language*, abreviadamente XML, descreve uma classe de objetos de dados chamada de documentos XML e parcialmente descreve o comportamento dos programas de computador que irão processá-los.
- Dublin Core é um vocabulário de quinze propriedades para a utilização em descrição de recursos. O nome "Dublin" é devido a sua origem em 1995 um workshop em Dublin, Ohio; "core" porque os seus elementos são amplos e genéricos, utilizáveis para a descrição de uma ampla gama de recursos.

Com essas dependências satisfeitas, é possível abrir o arquivo independentemente do sistema operacional e da arquitetura computacionais utilizados, o que se configura, portanto, como independência de fornecedor de *hardware* e de *software* (sistema operacional).

2.3 Preservação Tradicional

Para entendermos a importância da preservação temos primeiramente que saber o que significa “preservar”. Segundo o dicionário on-line Michaelis (2009), preservar é: “pôr(-se) ao abrigo de algum mal, dano ou perigo futuro; defender(-se), resguardar(-se)”.

O dicionário Aurélio (1999) diz,

preservar

[Do lat. praeservare, ‘observar previamente’.]

Verbo transitivo direto.

1. Livrar de algum mal; manter livre de corrupção, perigo ou dano; conservar: A Monarquia brasileira preservou a unidade nacional. Verbo transitivo direto e indireto.

2. Livrar, defender, resguardar: “Preservai, juízes de amanhã, preservai vossas almas juvenis desses baixos e abomináveis sofismas.” (Rui Barbosa, Oração aos Moços, p. 68.). Verbo pronominal.

3. Defender-se, proteger-se, resguardar-se. (Aurélio)

Segundo Ferreira (2006, p.17),

Desde a invenção da escrita que existe uma manifesta preocupação pela preservação dos artefactos que resultam de processos intelectuais e criativos do ser humano. A preservação desses artefactos permite às gerações futuras compreender e contextualizar a história e a cultura dos seus povos.

Desta maneira, os profissionais da informação têm a missão de escolher a melhor estratégia de preservação para esses artefatos ficarem disponíveis aos usuários.

Conforme explicam Moreno, Lopes e Chiara (2011, p. 5), preservar é:

[...] disponibilizar um espaço adequado que propicie as condições necessárias para a conservação dos diferentes suportes dos documentos, bem como, utilizar mobiliários e invólucros apropriados para que os documentos se mantenham conservados e tenham uma vida útil prolongada.

Com o processo de evolução das tecnologias da informação e comunicação, especialmente a Internet, cada vez mais as informações estão sendo produzidas e disponibilizadas no meio digital. Ainda segundo Ferreira (2006, p. 17),

Nos dias de hoje, uma parte significativa da produção intelectual é realizada com o auxílio de ferramentas digitais. A simplicidade com que o material digital pode ser criado e disseminado através das modernas redes de comunicação e a qualidade dos resultados obtidos são factores determinantes na adopção deste tipo de ferramentas.

No entanto, é observado que as antigas técnicas de preservação para documentos tradicionais não se aplicam a esta nova realidade que estamos vivendo, na qual os documentos digitais são cada vez mais utilizados. Na próxima seção serão apresentadas as técnicas de preservação utilizadas para os documentos digitais.

2.4 Preservação Digital

A preservação digital não se diferencia muito da preservação tradicional no que tange à conservação, pois também aqui é necessário se traçar a melhor estratégia.

Hedstrom (1996, *apud* ARELLANO, 2004, p. 17), afirma que a preservação digital se refere ao “[...] planejamento, alocação de recursos e aplicação de métodos e tecnologias para assegurar que a informação digital de valor contínuo permaneça acessível e utilizável [...]”.

Para assegurar essa interoperabilidade do documento digital, autores como Arellano, Ferreira, Lee e Sayão (*apud* DIAS, 2011) afirmam que as estratégias são divididas em três classes fundamentais:

- a) Emulação;
- b) Migração;
- c) Encapsulamento.

a) Emulação

Emulação é basicamente a técnica de virtualizar um sistema operacional ou *software* em um *software/hardware* mais atual. Segundo Ferreira, emulação é:

As estratégias de emulação baseiam-se essencialmente na utilização de um *software*, designado emulador, capaz de reproduzir o comportamento de uma plataforma de *hardware* e/ou *software*, numa outra que à partida seria incompatível. A grande vantagem desta abordagem está na capacidade de preservar, com um elevado grau

de fidelidade, as características e as funcionalidades do objecto digital original (FERREIRA, p. 33, 2006).

Outros autores sugerem que não se emule apenas o *software* que será necessário para leitura do arquivo digital, mas que emule também o seu *hardware* como sugere Rothenberg,

um modelo teórico capaz de emular plataformas actuais em computadores futuros. O modelo consiste na conservação do objecto digital juntamente com todo o *software* necessário à sua execução/apresentação (incluindo o sistema operativo) e na criação de uma especificação abstracta da plataforma de *hardware* que suporta a execução desse *software*. Essa especificação deverá ser escrita numa linguagem independente da plataforma e ser suficientemente rica para que um emulador possa ser construído automaticamente num qualquer computador do futuro (ROTHENBERG, 1999 *apud* FERREIRA, 2006, p. 34).

Já Handley tem uma posição mais conservadora,

[...]a emulação apenas deveria ser utilizada em contextos em que a comunidade de interesse valoriza a preservação do ambiente tecnológico original ou ainda em situações em que os objectos digitais não são passíveis de ser convertidos para formatos contemporâneos. Outros autores consideram potencialmente arriscado confiar no *software* original como forma de preservar objectos digitais, uma vez que este pode ser portador de vírus ou bugs que poderão, no futuro, resultar em perdas substanciais de informação (HANDLEY, 1998 *apud* FERREIRA, 2006, p. 34).

A estratégia de emulação enfrentará, em alguns anos, vários problemas, além dos que afirma Handley, pois as gerações futuras não entenderiam a interface dos programas antigos. Em *softwares* antigos, por exemplo, a imagem do ícone que representa o salvamento de um documento é a de um disquete. Novas gerações, que não conheceram esse periférico, terão dificuldade em compreender que clicando ali o documento será salvo. Será difícil manipular esses documentos, a não ser que se tenha investimentos em cursos e treinamentos específicos. Além do mais, pode-se considerar um retrocesso ensinar algo obsoleto para as gerações futuras.

b) Encapsulamento

Quanto ao encapsulamento, esta é uma estratégia de armazenamento

para períodos curtos, pois consiste em “congelar” os aparatos tecnológicos usados na criação do documento, utilizando apenas este recurso de *software* e *hardware* utilizado na criação do documento original. “Na prática, os dados podem ser encapsulados junto com a aplicação de *software* utilizado na sua criação, assim como uma descrição do ambiente de *software* e *hardware* requerido para seu funcionamento.” (MÁRDERO ARELLANO, 2008, p.68).

Essa estratégia pode sofrer os mesmos problemas da emulação com um agravante, pois, como comenta Castells (1999), baseado na “Lei de Moore” de 1965, o periférico de um computador se torna obsoleto em 18 meses. Portanto, essa estratégia se aplica somente a casos bem específicos.

Além do mais, se ocorrer uma falha no processo de encapsulamento, ou seja, se o *hardware*, já obsoleto, apresentar falhas, não existirá no mercado componente para substituição, e, então todo documento que foi encapsulado para ser usado nessa arquitetura de *hardware* se torna “lixo eletrônico” e passa a não ter mais serventia. Portanto, essa é técnica bastante arriscada de ser posta em prática.

c) Migração

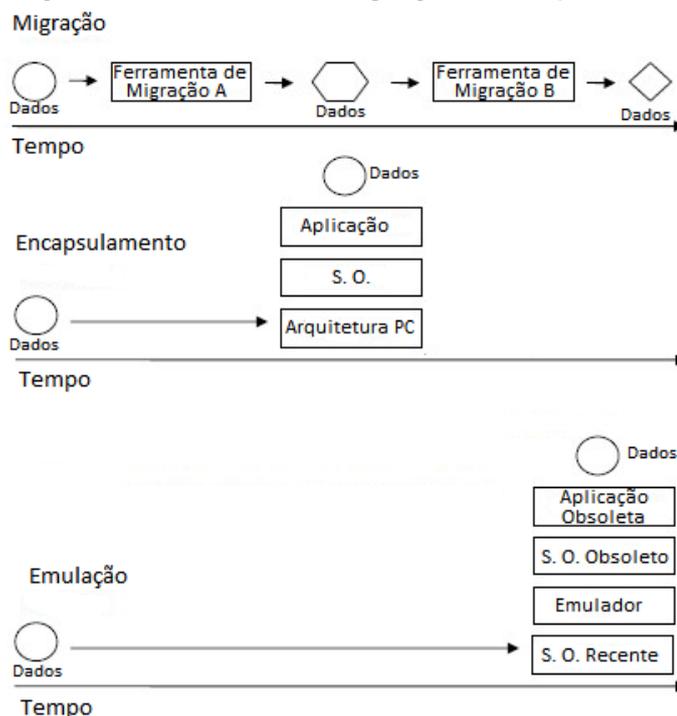
A última das três classes fundamentais, a migração, é a menos suscetível a falhas, pois tem como objetivo preservar a obra intelectual a longo prazo. Conforme afirmam Ferreira, Batista e Ramalho (2005), “Na migração os objetos digitais não são conservados nos seus formatos originais. Esta estratégia tem como objetivo fundamental preservar o conteúdo intelectual do objeto e não a sua estrutura”.

Ainda segundo Ferreira,

Podemos dizer que a migração de um modo geral não é totalmente segura, não há garantia que os objetos digitais sejam transferidos de forma que o conteúdo mantenha-se intacto, isto acontece quando os documentos são convertidos em novos formatos, já que muitas vezes o conversor utilizado não consegue converter o objeto digital de forma satisfatória (FERREIRA, 2006 p.36 *apud* DIAS, 2011).

A Figura 9 nos mostra a diferença entre as três técnicas apresentadas nesse trabalho.

Figura 9 – Ilustração das técnicas de Migração, Encapsulamento e Emulação



Legenda: S. O. - Sistema Operacional

Fonte: Keep, [2011?]. (apud DIAS, 2011) – adaptado pelo autor

É possível observar que a migração possui total independência de *hardware* e parcial independência de *software*, pois durante migração para outra aplicação há a necessidade de se padronizar os *softwares* utilizados para esta atividade, já que isso se faz necessário para garantir a integridade dos dados.

Vale salientar que nas etapas de migração, passando pela ferramenta de migração A, o arquivo passa para o formato hexágono, porém há não perda de informações importantes no arquivo. A situação se repete quando o arquivo passa por outra ferramenta de migração, no caso a B, ficando no formato de um losango.

Já o encapsulamento, conforme mencionado anteriormente, é o menos indicado para períodos longos de armazenamento dos dados. Isso pode ser observado na Figura 9, que mostra que para abrir um documento, depois de passado determinado período de tempo, é necessário montar toda a arquitetura computacional da época. Devido ao desgaste natural e à falta de suporte de *software* e principalmente de *hardware*, chegará um momento que não será mais possível montar esta mesma arquitetura computacional.

A emulação, diferentemente do encapsulamento, é mais independente na questão do *hardware*, que sempre será o mais recente, porém ficará dependente

de uma solução de *software* que seja capaz de dar as condições ideais para a recuperação do documento.

2.5 Legislação Brasileira sobre documentos digitais

A legislação brasileira sobre documentos digitais possui estreita relação com o direito de acesso à informação, o que nos remete à Declaração Universal dos Direitos Humanos, proclamada em 10 de dezembro de 1948 na Assembleia Geral da Organização das Nações Unidas – ONU, que diz:

Toda pessoa tem direito à liberdade de opinião e expressão; este direito inclui a liberdade de, sem interferência, ter opiniões e de procurar, receber e transmitir informações e idéias por quaisquer meios e independentemente de fronteiras. (Artigo 19)

Toda pessoa tem igual direito de acesso ao serviço público do seu país. (Artigo 21.2) (DECLARAÇÃO UNIVERSAL DOS DIREITOS HUMANOS - 1948)

Apesar de esta declaração ter sido redigida num período em que a computação, os documentos digitais e a comunicação não tinham a expressão e importância que possuem na atualidade, suas diretrizes se encaixam perfeitamente na questão do acesso à informação.

Nos tempos atuais, para acessar a informação que desejamos, temos diversos dispositivos de *hardware* e *software* para essa tarefa. Por causa dessa diversidade de dispositivos, as empresas privadas e públicas, órgãos públicos, instituições de ensino, entre outros, adotam alguns desses dispositivos como padrão, que nem sempre são os mesmos utilizados por seus usuários, o que acaba por causar assim um desgaste desnecessário na recuperação da informação.

Levando-se em consideração este cenário, com o intuito de mitigar este problema, e tendo como base a declaração universal dos direitos humanos, os congressistas brasileiros apresentam, já há mais de 10 anos, projetos de leis tais como: PL 2269/1999, PL-3051/2000, PL-4275/2001, PL-7120/2002, PL-2152/2003, PL-3280/2004, PL-3070/2008, que englobam a utilização de sistemas operacionais, suítes de escritório e padrão de arquivos editáveis em código aberto.

Esse trabalho tem foco nos padrões de documentos editáveis, portanto, está intimamente relacionado com o projeto de lei 3070/2008:

PROJETO DE LEI Nº 3070 DE 2008

Dispõe que os órgãos e entidades da Administração Pública Direta, Indireta, Autárquica, bem como os órgãos autônomos e empresas sob o controle estatal adotarão, preferencialmente, formatos abertos de arquivos para criação, armazenamento e disponibilização digital de documentos (TEIXEIRA, 2008).

O PL 3070/2008, assim como a maioria dos outros, está na Comissão de Constituição e Justiça e de Cidadania (CCJC), na Câmara dos Deputados, e o projeto de lei PL 2269/1999 encontra-se na Coordenação de Comissões Permanentes (CCP), também da Câmara.

Ainda em 2008, o Governo Brasileiro desenvolveu o e-PING – Padrões de Interoperabilidade de Governo Eletrônico – que

define um conjunto mínimo de premissas, políticas e especificações técnicas que regulamentam a utilização da Tecnologia de Informação e Comunicação (TIC) na interoperabilidade de Serviços de Governo Eletrônico, estabelecendo as condições de interação com os demais Poderes e esferas de governo e com a sociedade em geral (BRASIL, 2008).

No âmbito privado, foi desenvolvido pelo Governo Brasileiro em parceria com empresas privadas o Protocolo de Brasília, que é um documento com diretrizes para a adoção do padrão aberto para documentos digitais editáveis por empresas que demonstrem interesse.

Art. 1 – Em consonância com as recomendações propostas pelo Governo Federal através do documento de referência para padrões de interoperabilidade, e-PING, e visando atuar de forma cooperativa na aplicação de soluções que gerem maior independência tecnológica e economia para o erário, os órgãos e entidades da Administração Pública Direta, Indireta, Autárquica e Fundacional, bem como as empresas privadas, abaixo relacionados, propõem-se a adotar formatos abertos de arquivos para criação, armazenamento e disponibilização digital de documentos dos tipos texto, planilha e apresentação (PROTOCOLO DE BRASÍLIA, 2008).

O texto do PL 3070-2008 cita o estado do Paraná como o pioneiro na adoção de um padrão aberto para seus documentos digitais editáveis, de acordo com a Lei Nº 15742, de 18 de dezembro de 2007. O último a adotar foi o estado do Rio de Janeiro, com a Lei Nº 5978, de 24 de maio de 2011.

Além disso, o Conselho Nacional de Arquivos (CONARQ), com base na Carta de Preservação do Patrimônio Digital da UNESCO, desenvolveu um documento visando atender às especificações do documento digital de caráter arquivístico que se chama Carta de Preservação do Patrimônio Arquivístico Digital⁹.

Guedes (2006, p. 39) diz,

O CONARQ tem aberto várias frentes: a elaboração de glossário de termos relacionados a documentos eletrônicos, a especificação de requisitos funcionais e metadados para sistemas eletrônicos de gestão arquivística de documentos que atendam às especificações da legislação e práticas arquivísticas brasileiras, o desenvolvimento de normas nacionais de modelo de repositório de objetos eletrônicos, a definição de diretrizes para os sistemas legados e para recolhimento às instituições arquivísticas, a definição de infraestrutura tecnológica e de recursos humanos; editou a Resolução nº 20, de 16 de julho de 2004, que “Dispõe sobre a inserção dos documentos eletrônicos em programas de gestão arquivística de documentos dos órgãos e entidades integrantes do Sistema Nacional de Arquivos”.

As definições técnicas do CONARQ estão sendo pautadas em padrões abertos, como é o caso do modelo de referência OAIS – Open Archival Information System.

Neste capítulo foram apresentados os conceitos que norteiam este trabalho, buscando entender a importância de se existir um padrão para documentos digitais editáveis, os formatos proprietários e não-proprietários, a interoperabilidade entre diferentes formatos e as diretrizes do Governo Brasileiro no que tange à utilizados de documentos digitais. No próximo capítulo serão apresentados os procedimentos metodológicos utilizados neste trabalho, que teve por intuito evidenciar a necessidade de se pensar em um padrão livre, aberto e interoperável para salvaguardar documentos digitais editáveis.

⁹ <http://www.conarq.arquivonacional.gov.br/Media/publicacoes/cartapreservpatrimarqdigitalconarq2004.pdf> - Acesso em 15 de nov. de 2012.

3 PROCEDIMENTOS METODOLÓGICOS

Para Richardson (1999, p. 16), “a pesquisa, como ferramenta para adquirir conhecimento, pode ter os seguintes objetivos: resolver problemas específicos, gerar teorias ou avaliar teorias existentes”. Gil (2010, p. 1) afirma “Pode-se definir pesquisa como o procedimento racional e sistemático que tem como objetivo proporcionar respostas aos problemas que são propostos”.

Barros e Lehfeld (1986 *apud* PEREIRA, 2011, p. 20) “indicam que o termo metodologia designa a área de estudo que descreve e analisa os métodos, que são os processos ordenados empregados ao longo de um caminho para se chegar a um fim”.

Esta é uma pesquisa exploratória-descritiva, no aspecto dos objetivos e quali-quantitativa em relação ao enfoque do problema. Exploratória porque segundo, Cervo e Bervian (1996, p. 49), “tem por objetivo familiarizar-se com o fenômeno ou obter nova percepção do mesmo e descobrir novas ideias”. Descritiva na medida em que procurou descrever as características (GIL, 2002, p. 42) dos documentos editáveis disponíveis nos domínios da UFSC e UDESC, tendo como suporte as técnicas estatísticas, para a organização dos resultados em tabelas e gráficos com o intuito de resumir, descrever e apresentar os dados.

É construída de forma qualitativa e quantitativa, sendo que a qualitativa tem o intuito de privilegiar as particularidades e descrever os significados e descobertas, e a quantitativa o de fornecer dados na forma numérica para a análise proposta.

Segundo os ensinamentos de Goode e Hart (1973):

A pesquisa moderna deve rejeitar como uma falsa dicotomia a separação entre estudos “qualitativos e “quantitativos”, ou entre pontos de vista “estatístico ou “não estatístico”. [...] Além disso, não importa quão precisas sejam as medidas, o que é medido continua a ser uma qualidade (GOODE e HART, 1973, p. 398).

Embora a pesquisa quantitativa seja de certo modo qualitativa, de acordo com Richardson (1999), a forma como se pretende analisar o problema, ou, por assim dizer, o enfoque adotado é que, de fato, exige uma metodologia qualitativa ou quantitativa.

O método qualitativo, que para Richardson (1999), justifica-se por ser

uma forma adequada para entender a natureza de um fenômeno social, será utilizado na análise das informações contidas em nosso estudo quantitativo: interpretamos os dados colhidos e analisamos a sua representação no contexto os quais estão inseridos. Assim, podemos dizer que a coleta de dados é a etapa da pesquisa que exige mais tempo e trabalho para reunir informações, comportando algumas normas que dependem e se ajustam ao tipo de estudo que se pretende.

3.1 Etapas da Pesquisa

Com base nos objetivos propostos, a pesquisa foi realizada por meio do acesso aos sítios das Instituições de Ensino Superior Universidade Federal de Santa Catarina (UFSC) e Universidade do Estado de Santa Catarina (UDESC), com o intuito de averiguar de que forma essas instituições estão liberando seus documentos editáveis para os usuários que irão à busca desses documentos.

Para a análise do conteúdo foram utilizadas as técnicas de Bardin (1977):

1. A pré-análise: fase de organização, no plano da análise. Os documentos que farão parte do corpus da pesquisa são escolhidos, bem como são elaboradas, se necessário, hipóteses, além dos objetivos e indicadores que fundamentarão a interpretação final. Tais procedimentos se concretizarão ao longo da leitura flutuante e contato exaustivo com os documentos.
2. A exploração do material: trabalhosa fase de codificação, enumeração, entre outros procedimentos previamente definidos, do material que compõe o corpus.
3. O tratamento dos resultados obtidos e interpretação: fase de transformar os dados brutos em informações significativas e válidas. Usam-se operações estatísticas e esquemas visuais para condensar o conteúdo trabalhado. A partir daí realizam-se interpretações e inferências, além de comparar os resultados com os objetivos previstos, sendo que estes poderão coincidir ou apresentar descobertas inesperadas.

Na pré-análise foi utilizado o buscador Google (www.google.com.br) como ferramenta para coleta de dados nos domínios udesc.br e ufsc.br, por meio da utilização dos operadores **ext** (busca arquivos específicos) e **site** (busca apenas em um *website* ou *domain* pré-determinado).

A Figura 10 apresenta os resultados da coleta de dados utilizando-se a expressão `ext:odt site:ufsc.br`. Neste procedimento foram recuperados 133 documentos com a extensão `.odt` no domínio `ufsc.br`.

Figura 10 – Tela da pesquisa com os operadores utilizados

The screenshot shows a Google search interface with the following elements:

- Navigation bar: +Você, Pesquisar, Imagens, Mapas, Play, YouTube, Notícias, Gmail, Disco, Agenda, Mais -
- Search bar: ext:odt site:ufsc.br
- Search filters: Web (selected), Imagens, Mapas, Shopping, Vídeos, Mais ▾, Ferramentas de pesquisa
- Results summary: Aproximadamente 133 resultados (0,40 segundos)
- Search results:
 - [ODF] Nosso acervo no formato texto .odt (248KB) - LDA - UFSC**
www.lda.ufsc.br/acervo.odt
 Formato do arquivo: OpenDocument
 Registro. Título. Autor. Palavra Chave. 0. APPX516. appx524. SILVA, Ana Claudia Lorenzi da. appx526. appx537. BITTENCOURT, Júlia Lins. appx544 ...
 - [ODF] RELATÓRIO MENSAL - MAIO 2009 O Serviço de Controle de ...**
www.hu.ufsc.br/ccih/relatorio05_09.odt
 Formato do arquivo: OpenDocument - Ver em HTML
 SERVIÇO PÚBLICO FEDERAL. UNIVERSIDADE FEDERAL DE SANTA CATARINA – UFSC. HOSPITAL UNIVERSITÁRIO PROF. POLYDORO ERNANI DE SÃO ...
 - [ODF] Requerimento 2a chamada.odt - Departamento de Matemática ...**
www.mtm.ufsc.br/Graduandos/Requerimento_2a_chamada.odt
 Formato do arquivo: OpenDocument - Ver em HTML
 DEPARTAMENTO DE MATEMÁTICA. REQUERIMENTO PARA REALIZAÇÃO DE AVALIAÇÃO. Orientações: De acordo com o Art. 74, da Resolução 17/CUn/97, ...
 - [ODF] odt - Departamento de Matemática da UFSC**
www.mtm.ufsc.br/ensino/segunda_chamada.odt
 Formato do arquivo: OpenDocument - Ver em HTML
 DEPARTAMENTO DE MATEMÁTICA. REQUERIMENTO PARA REALIZAÇÃO DE AVALIAÇÃO. Orientações: De acordo com o Art. 74, da Resolução 17/CUn/97, ...

Fonte: google.com.br

A partir dos resultados da pré-análise, foi realizada a separação dos documentos digitais pela sua extensão proprietária ou não-proprietária. Nos mesmos moldes foi realizada este procedimento nos sítios dos cursos de Biblioteconomia da UFSC e UDESC para identificar quais extensões estão sendo utilizadas.

Foi avaliada também a interoperabilidade dos documentos digitais, por meio da abertura dos mesmos pelas suítes de escritórios *Microsoft Office 2010* e *LibreOffice 3.6*, com o intuito de averiguar o comportamento dos documentos. Para escolha da amostra foi utilizada a amostragem aleatória simples (GIL, 2002), a onde foi colocando todos os documentos coletados dos quatro sítios em uma pasta, no computador, e de forma aleatória foram selecionados 50 documentos entres proprietários e não-proprietários, e o sistema operacional utilizado foi o *Windows 7 Ultimate 64 bits Service Pack 1*.

Após a abertura dos documentos digitais pelos *softwares* de escritório, analisou-se a disponibilização da informação, ou seja, se existiu perda de

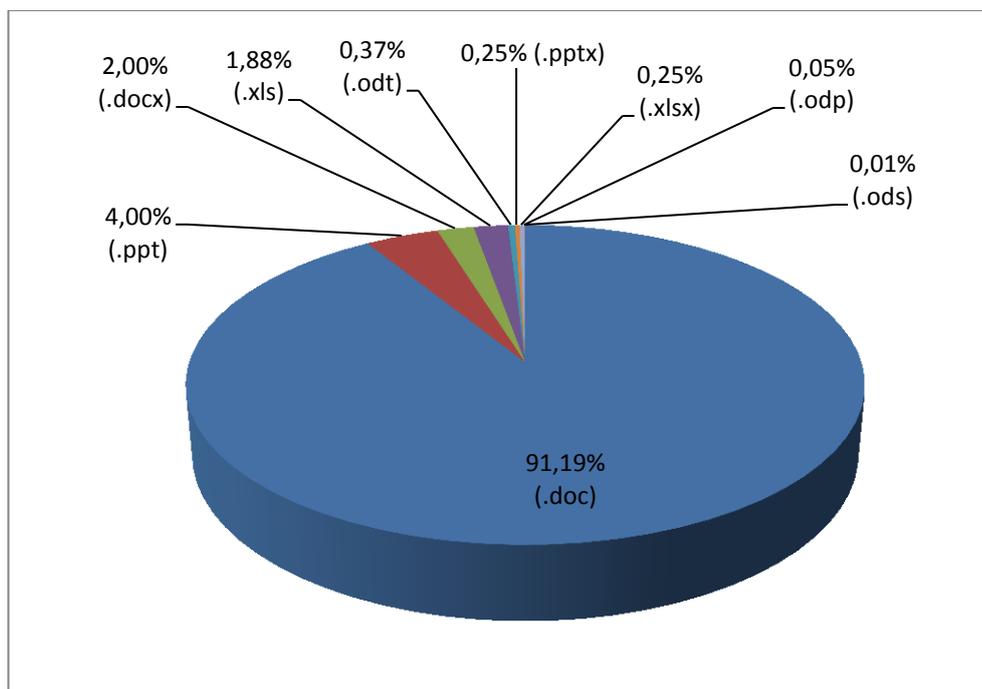
informação do documento, se ocorreu algum erro no momento da abertura e se algum *software* de escritório não conseguiu abrir algumas das extensões encontradas.

Os resultados obtidos na pesquisa são apresentados e discutidos no próximo capítulo.

4 DISCUSSÃO DOS RESULTADOS

Apresentamos inicialmente o *corpus* da pesquisa, realizada nos sítios das instituições UFSC e UDESC. A partir da utilização dos operadores de busca **ext** e **site** do buscador Google (www.google.com.br) nos domínios udesc.br e ufsc.br obteve-se os seguintes resultados.

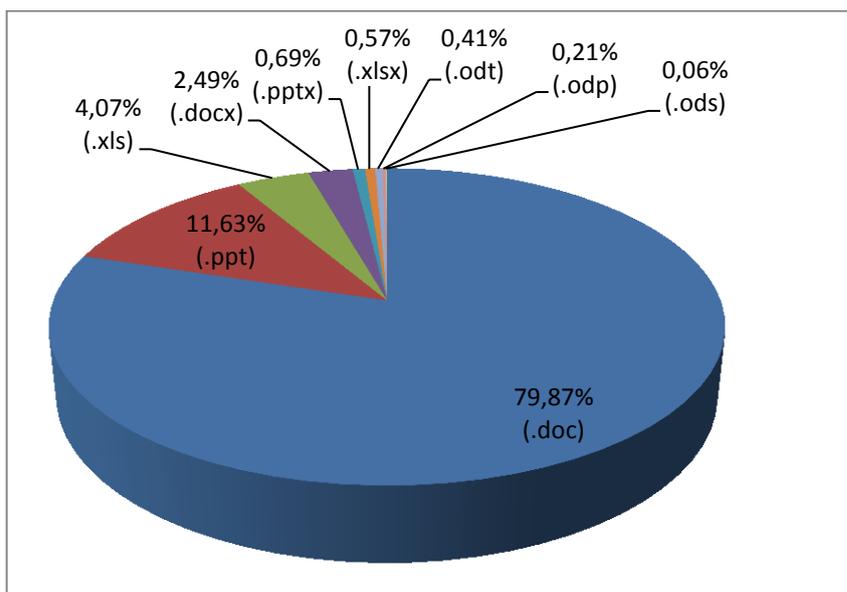
Gráfico 1 – Extensões editáveis no domínio udesc.br



Fonte – Elaborado pelo autor.

O Gráfico 1 mostra a distribuição das extensões encontradas no domínio da UDESC, onde dos 14.695 documentos encontrados estão assim distribuídos: 91,19% *.doc* (13.400), 4% *.ppt* (588), 2% *.docx* (294), 1,88% *.xls* (276), 0,37% *.odt* (54), 0,25% *.pptx* (37), 0,25% *.xlsx* (37), 0,05% *.odp* (8) e 0,01% *.ods* (1). Observamos que 99,57% (14.632) das extensões são proprietárias e 0,43% (63) são extensões não-proprietárias.

Gráfico 2 – Extensões editáveis no domínio ufsc.br

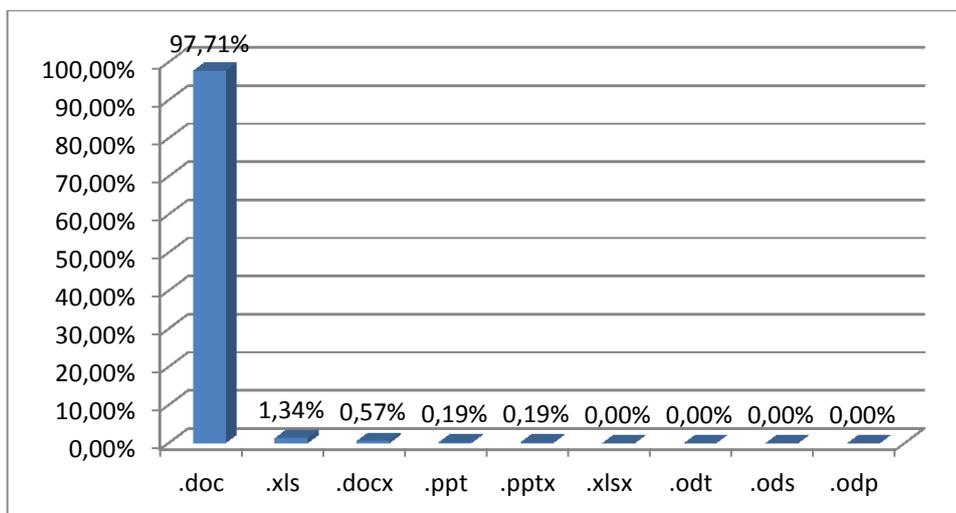


Fonte – Elaborado pelo autor.

Os resultados da UFSC não são muitos diferentes dos obtidos na UDESC, embora exista uma quantidade bem superior de documentos disponível aos usuários da UFSC. Pois, a quantidade de documentos com extensões proprietária e não-proprietária se assemelham. Existem 32.677 documentos editáveis disponíveis no domínio ufsc.br: 79,87% .doc, 11,63% .ppt, 4,07% .xls, 2,49% .docx, 0,69% .pptx, 0,57% .xlsx, 0,41% .odt, 0,21% .odp e 0,06% .ods. Da mesma forma que no site da UDESC, aqui também é observada uma enorme diferença entre a quantidade de extensões proprietárias (99,32%) e não-proprietárias (0,68%).

Na UDESC, o sítio do curso de Biblioteconomia está situado dentro da Faculdade de Educação, Centro de Ciências Humanas e da Educação, cujo endereço eletrônico é faed.udesc.br. Na UFSC, o sítio equivalente está dentro do departamento de Ciência da Informação (CIN), cujo endereço é dptcin.paginas.ufsc.br. Após a coleta de dados em ambos os sítios, obteve-se os seguintes resultados.

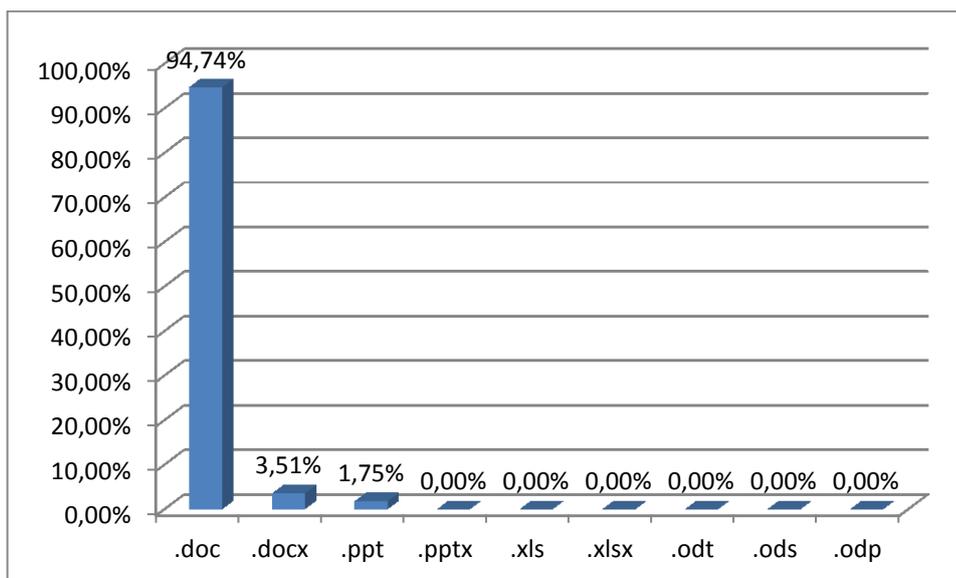
Gráfico 3 – Extensões editáveis no sitio faed.udesc.br



Fonte – Elaborado pelo autor

Observa-se que o sítio da FAED/UEDESC não possui nenhuma extensão não-proprietária, sendo as extensões proprietárias 100% dos documentos. Os 524 documentos encontrados no sítio da UEDESC estão distribuídos da seguinte forma: 512 são .doc (97.71%), 7 .xls (1,34%), 3 .docx (0,57%), 1 ppt (0,19%) e 1 pptx (0,19%).

Gráfico 4 – Extensões editáveis no sitio dptcin.paginas.ufsc.br



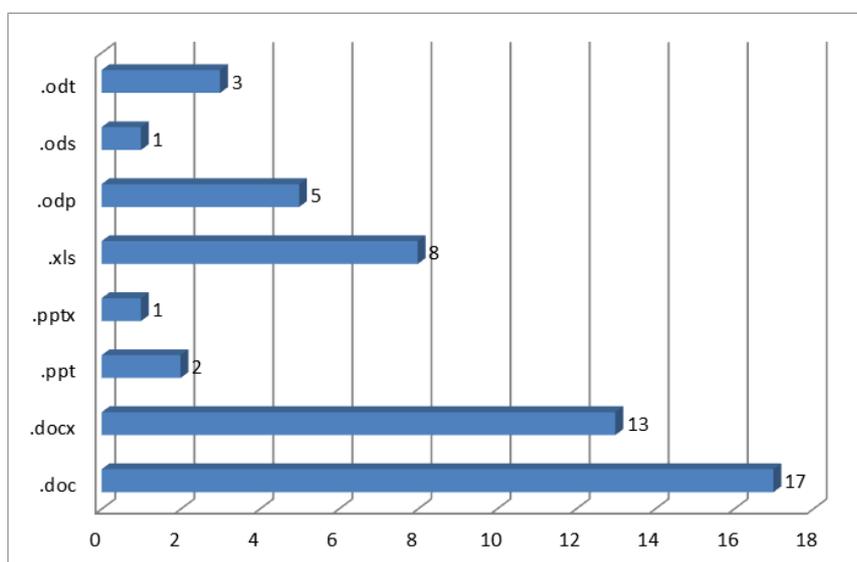
Fonte – Elaborado pelo autor

O departamento de Ciência da Informação da UFSC, a exemplo do site

da FAED/UEDESC, não possui nenhuma extensão não-proprietária. Os 57 documentos encontrados neste sítio estão assim distribuídos: 54 .doc (94.74%), 2 .docx (3,51%), 1 ppt (1,75%).

Foram separados 50 documentos aleatoriamente dos sítios: ufsc.br, udesc.br, faed.udesc.br e dptcin.paginas.ufsc.br para serem utilizados na avaliação da interoperabilidade, na qual todos os documentos foram abertos pelos *softwares* de escritório da *Microsoft Office 2010* e *Libreoffice*, executados no sistema operacional *Microsoft Windows 7 Ultimate SP 1 64 bits*.

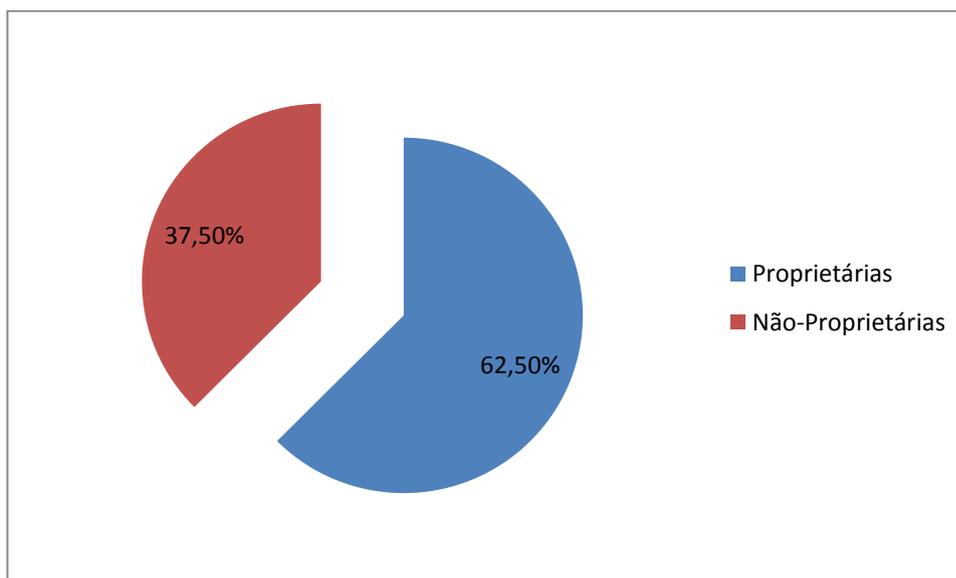
Gráfico 5 – Quantidade de documentos selecionados por extensões



Fonte – Elaborado pelo autor

O Gráfico 5 mostra a quantidade de documentos selecionados por extensão para realização desta pesquisa. O conjunto selecionado possui a seguinte distribuição por extensão: 17 (34%) .doc, 13 (26%) .docx, 8 (16%) .xls, 5 (10%) .odp, 3 (6%) .odt, 2 (4%) ppt, 1 (2%) pptx e 1 (2%) .ods fazendo um total de 50 (100%) documentos.

Gráfico 6 – Quantidade de extensões Proprietárias e Não-Proprietárias



Fonte – Elaborado pelo autor

O Gráfico 6 mostra a relação de extensões obtida a partir da avaliação da interoperabilidade. Foi identificado 37,50% de extensões não-proprietárias e 62,50% de proprietárias.

A Tabela 5 mostra os nomes dos documentos e suas extensões, utilizados na avaliação da interoperabilidade.

Tabela 5 – Relação de arquivos utilizados na avaliação da interoperabilidade

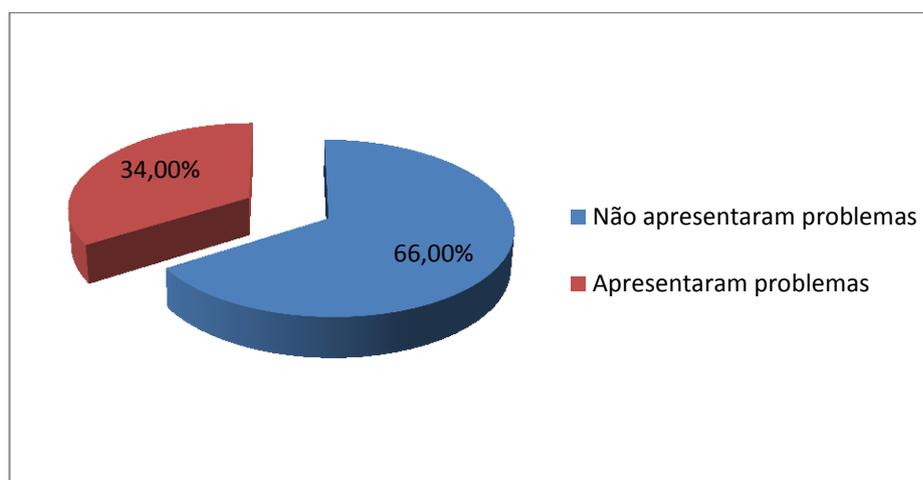
Nome do Arquivo	Sítio
Edital__02-2011_PPGEEL.doc	udesc.br
edital_001_2009_ppg_edi__remanescentes.doc	udesc.br
CRONOGRAMA_DE_ATIVIDADE.docx	udesc.br
Modelo_Final_Artigo_do_IDEMi2012_Ingles.docx	udesc.br
NovidadesESRI-Parte2.ppt	faed.udesc.br
Apresentacao_Intro.pptx	faed.udesc.br
acao.xls	faed.udesc.br
despesas_2010.xls	faed.udesc.br
despesas_2011.xls	faed.udesc.br
divu.xls	faed.udesc.br
formcadastro_capes_ds.xls	faed.udesc.br
cronograma_gestao_arquivos.doc	dptcin.paginas.ufsc.br
folder_arquivos.doc	dptcin.paginas.ufsc.br
FORMULARIO DE PAGAMENTO DE DIARIAS E PASSAGENS CED.doc	dptcin.paginas.ufsc.br
Lidia_Universidade_Aveiro_Portugal_maio_2009.doc	dptcin.paginas.ufsc.br

relatorio de viagem FORMULARIO CED.doc	dptcin.paginas.ufsc.br
seminario_arquivologia.doc	dptcin.paginas.ufsc.br
fotografias_10_10_2008.ppt	dptcin.paginas.ufsc.br
34206-45571-1-RV.odt	ufsc.br
ata-30-11-11.odt	ufsc.br
ESQUEMA DE PLANO DE AULA-openvas.odt	ufsc.br
747_tsesaotenti.xls	ufsc.br
qualis_periodicos_AUD1-sites.xls	ufsc.br
REUNI_2010_02_Tabela-de-Relatório-por-Modalidade.xls	ufsc.br
7413_sss.docx	ufsc.br
8348_DOCUMENTO- Taxonomia microbiana.docx	ufsc.br
o_uso_da_cadeirinha_para_criancas_em_veiculos_e_a_resolucao_277.docx	ufsc.br
calc.odp	udesc.br
calouros_2010_01.odp	udesc.br
CursoAPIxml.odp	udesc.br
IntroXML.odp	udesc.br
solicitar_ajuste.ods	udesc.br
UDESC_Curso_SL_Motivacional_v3.odp	udesc.br
governo_eletronico.docx	ufsc.br
3Modelo_Convenio_UFSC-AGI-Concedente3.doc	ufsc.br
7Modelo_TermoRescisaoConvenio.doc	ufsc.br
Extensão-Cronograma-e-Edital-2012.1.doc	ufsc.br
Extensão-Cronograma-e-Edital-2012.-2.doc	ufsc.br
ENR5105-Hidroponia.doc	ufsc.br
estagio_no_exterior-procedimentos.doc	ufsc.br
editaldoutoradacienciasmedicas-2011.doc	ufsc.br
Calendário-2012.2.doc	ufsc.br
edital_completo.doc	ufsc.br
ArtigocientificoAtualizado2012.docx	ufsc.br
CriandoummosaiconoGE.docx	ufsc.br
LivrosEletronicos.docx	ufsc.br
MODELOabstract.docx	ufsc.br
MODELOcapa.docx	ufsc.br
MODELOreferencias.docx	ufsc.br
2012-2-violeta.docx	ufsc.br

Fonte – Elaborado pelo autor

Os arquivos apresentados na Tabela 5 foram todos abertos pelos softwares de escritório *Microsoft Office 2010* e *LibreOffice 3.6*. A seguir são apresentados detalhes sobre estas operações.

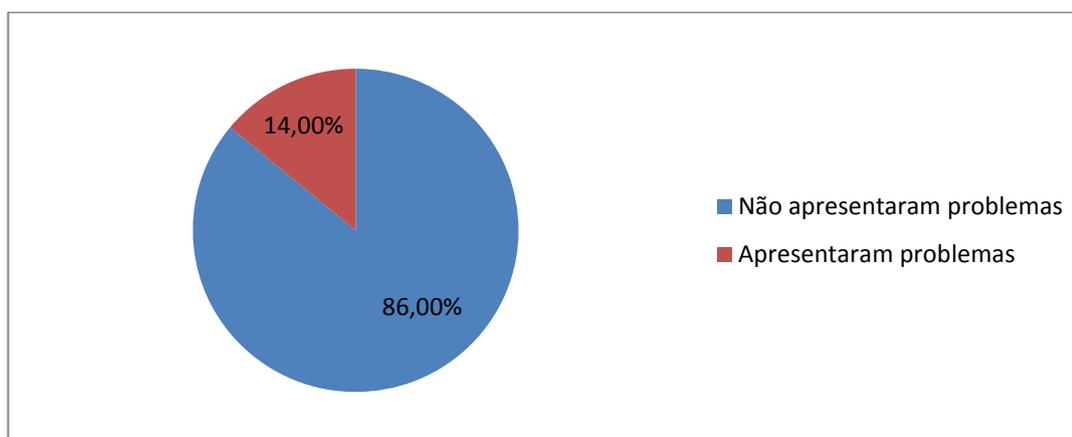
Gráfico 7 – Documentos com problemas X Documentos sem problemas



Fonte – Elaborado pelo autor

O Gráfico 7 traz a relação de documentos que apresentaram algum tipo de problema, não apenas de abertura/acesso, mas todos os problemas de formatação, animações que não funcionaram da forma esperada, imagens que não foram carregadas, fundo sem a cor desejada e textos apagados.

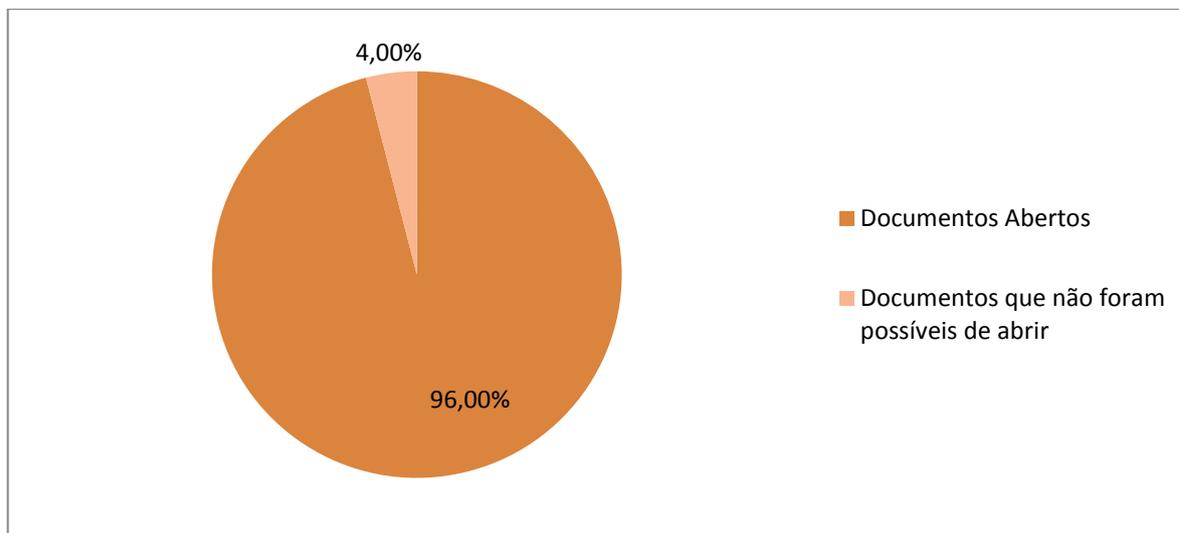
No total, 34% desses documentos apresentaram os problemas acima relatados. Esses dados se referem aos documentos abertos nos dois *softwares* de escritório utilizados na pesquisa. Isso não significa que os documentos apresentaram problemas nos dois programas ao mesmo tempo. Por exemplo, em alguns casos ocorreram problemas com um arquivo no *Microsoft Office 2010* e nenhum no *LibreOffice* e vice-versa.

Gráfico 8 – Quantidade de problemas utilizando o *MS Office 2010*

Fonte – Elaborado pelo autor

O Gráfico 8 mostra a quantidade de problemas encontrados durante a utilização do *software* de escritório *Microsoft Office 2010*. Os dados considerados englobam todos os tipos de problemas apresentados.

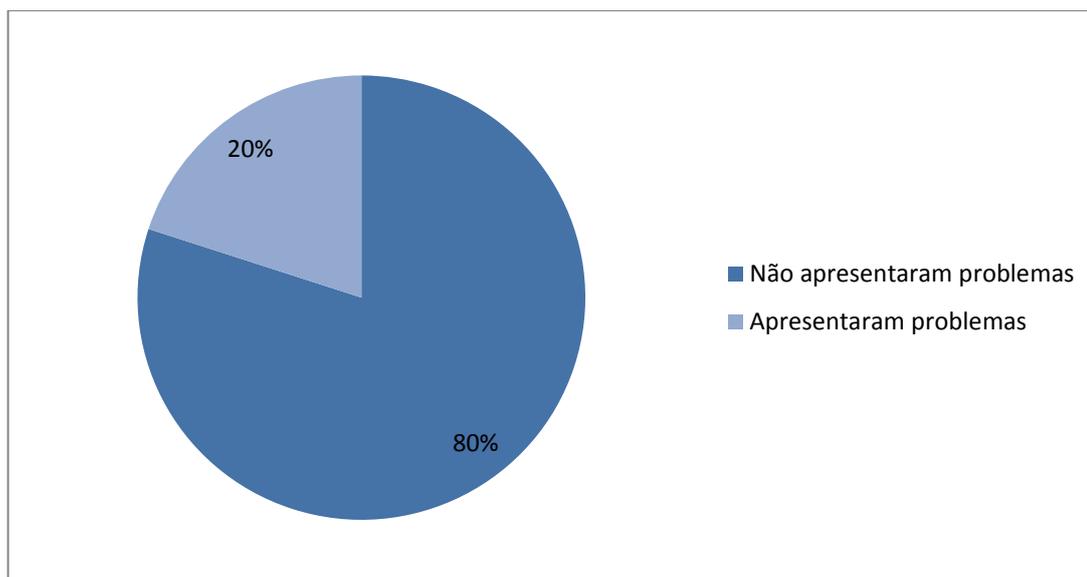
Gráfico 9 – Porcentagem de documentos abertos utilizando o *MS Office 2010*



Fonte – Elaborado pelo autor

O Gráfico 9 apresenta a relação dos documentos abertos pelo *Microsoft Office 2010* e a indicação de que 4% desse total não foi possível de ser aberto.

Pode-se dizer que 4% é um número relativamente baixo, porém, se utilizarmos a analogia do avião descrita no início do trabalho, e supondo que todos os 50 documentos digitais abertos por esse *software* de escritório fossem manuais de 50 aviões distintos, neste contexto, 4% seria uma perda bem elevada.

Gráfico 10 - Quantidade de problemas utilizando o *LibreOffice*

Fonte – Elaborado pelo autor

O Gráfico 10 apresenta a quantidade de problemas encontrados utilizando-se o *software* de escritório *LibreOffice*. Observa-se que este programa apresenta uma quantidade maior de problemas. Compreende-se este fato por meio da informação de que a quantidade de documentos com extensões proprietárias é bem superior a de documentos com extensões não-proprietárias, que são o foco do *LibreOffice*.

Apesar disso, todos os documentos foram abertos por essa suíte de escritório, ou seja, todos os dados ficaram legíveis e foram possíveis de serem interpretados. Diferentemente do *Microsoft Office 2010*, que apresentou problemas na abertura de 4% dos documentos, e indisponibilizou boa parte dos dados originais.

Fazendo uma relação com as recomendações do governo brasileiro foi possível observar que as instituições pesquisadas possuem documentos em formato aberto, mais propriamente dito o ODF, mas que a quantidade destes ainda é muito pequena nos domínio das duas instituições, não chegando a 1% do total. A situação é mais crítica quando voltamos nossos olhares para os sites que hospedam os cursos de Biblioteconomia, pois nestes não foi recuperado nenhum documento em formato aberto. Diante disso, detectou-se uma inobservância tanto pela UFSC quanto pela UDESC das diretrizes proposta pelo governo para a utilização de documentos digitais em instituições públicas.

Neste capítulo foram apresentados os resultados da pesquisa. De

maneira geral, os resultados estão dentro daquilo que era esperado pelo autor, devido à sua experiência pessoal e profissional. Observa-se que é nítida a necessidade de pensarmos na utilização de extensões cada vez mais interoperáveis, independentemente de serem ou não proprietárias, pois o que está em jogo é a evolução da capacidade da humanidade de armazenar seus conhecimentos para as gerações futuras.

5 CONCLUSÕES

Esta pesquisa teve como objetivo analisar os documentos digitais editáveis disponíveis nos domínios Internet das instituições de ensino superior públicas de Santa Catarina UFSC e UDESC e verificar se os mesmos estão em conformidade com as orientações do Governo Brasileiro sobre a utilização de documentos digitais editáveis.

Este trabalho teve como alicerce a literatura de diversas áreas do conhecimento: biblioteconomia, tecnologia, direito (legislação municipal e estadual, projetos de leis) e diretrizes do governo federal sobre a questão da preservação dos documentos digitais. Utilizando o método de pesquisa exploratória-descritiva foi possível realizar a coleta dos dados e entender o cenário analisado.

Percebe-se, a partir da análise dos resultados, que a preservação digital torna-se essencial para a disseminação e geração do conhecimento necessários para garantir a evolução da sociedade em que vivemos, a partir da maneira pela qual esta repassa seus conhecimentos produzidos às gerações futuras.

A pesquisa mostrou que as instituições UFSC e UDESC estão disponibilizando diversos documentos editáveis em seus sítios. Porém, pesquisando-se diretamente nos sítios que hospedam os cursos de Biblioteconomia das referidas instituições, não foi observado o mesmo cenário que se apresentou nas pesquisas dos domínios. Pois, dos 14.695 documentos editáveis recuperados no domínio da udesc.br, 3,57% pertencem ao sítio FAED/UDESC, enquanto que dos 32.677 documentos editáveis recuperados no domínio ufsc.br menos de 1% (0,17%) pertencem ao sítio CIN/UFSC.

Este trabalho mostrou ainda que, na sua maioria, as extensões proprietárias estão presentes em mais de 98% dos documentos das duas instituições, sendo as extensões não-proprietárias muito pouco utilizadas. Esse dado é preocupante, pois significa que a maioria das informações armazenadas dependem de uma empresa para serem interpretadas de forma que se tenha 100% de garantia de que o documento que as contém será recuperado sem perda de dados.

Pôde-se avaliar também que a interoperabilidade dos documentos dos

referidos sítios é possível e está sendo feita, apesar de que alguns documentos que utilizam extensões não-proprietárias não foram abertos no *software* de escritório da *Microsoft Corporation*. Porém, todos os documentos foram abertos pelo *software* de escritório *LibreOffice*. No entanto, alguns documentos com extensões proprietárias perderam algumas de suas características, como por exemplo, num documento de extensão .ppt, que é de apresentação de lâminas, algumas animações das lâminas não funcionaram. Mas, apesar de tudo, características como essas não impossibilitam completamente a recuperação das informações dos documentos.

Outra conclusão a que se chega é a de que as instituições pesquisadas não estão em conformidade com as diretrizes do Governo Brasileiro para criação de documentos editáveis. O Governo Brasileiro tem como orientação – e preferência – a utilização das extensões ODF. Conforme verificou-se, muito pouco ainda é utilizado deste tipo de extensão. Diante disso, se faz necessário um alerta a essa questão nas universidades analisadas e nas instituições de ensino públicas em geral, pois deixar quase todo o seu conhecimento e trabalho nas “mãos” de uma empresa não é a melhor estratégia de preservação a ser adotada.

Também é sabido que não podemos ficar reféns apenas de um único padrão, e faz-se necessário estarmos sempre atentos a novas tecnologias e estratégias de preservação. Já tivemos períodos na história não muito distante nos quais a língua “universal” era o Latim. Na contemporaneidade muito pouco ou quase nada se ouve falar deste idioma e a língua “universal” passou a ser o Inglês.

Portanto, espera-se que a presente pesquisa seja útil para gestores de unidades de informação públicas, e que ofereça parâmetros para os profissionais orientarem seus usuários para a melhor forma, e conforme as orientações do Governo Brasileiro, de salvaguardar seus documentos digitais.

5.1 Trabalhos futuros

Ao término desse trabalho constatou-se que o que foi realizado é apenas um pouco daquilo que ainda pode ser explorado sobre o tema. Tema este que para a contemporaneidade é muito importante, haja vista que os documentos em sua grande maioria estão sendo criados em meio digital, e, saber como deixar

essas informações acessíveis ao usuário requer mais atenção e cuidado do que apenas disponibilizá-las on-line.

Portanto, sugerem-se novos trabalhos onde se realizem estudos em unidades de informações que adotaram uma determinada extensão para salvar seus documentos e depois de algum tempo tiveram que migrar para outra extensão. Estudar o porquê dessa necessidade, o que ocorreu para que tal atitude fosse tomada, por que se escolheu a nova extensão, quais os critérios de escolha, além de se verificar se ocorreu perda de documentos com a migração realizada.

REFERÊNCIAS

- ADVANCING OPEN STANDARDS FOR THE INFORMATION SOCIETY. **OASIS OpenDocument Format (ODF) Adoption TC**. Disponível em: <http://www.oasis-open.org/committees/tc_home.php?wg_abbrev=odf-adoption >. Acesso em: 03 dez. 2011.
- ALMEIDA, Maurício Barcellos; CENDÓN, Beatriz Valadares; SOUZA, Renato Rocha. Metodologia para implantação de programas de preservação de documentos digitais a longo prazo. **Encontros Bibli: Revista Eletrônica de Biblioteconomia e Ciência da Informação**, Florianópolis, v. 17, n. 34, p. 103-130, 2012. Disponível em: <<http://www.brapci.ufpr.br/download.php?dd0=20369>>. Acesso em: 15 de set. 2012
- ARELLANO, Miguel Angel. **Preservação de documentos digitais**. Ci. Inf. [online]. 2004, vol.33, n.2, pp. 15-27. Disponível em: <<http://www.scielo.br/pdf/ci/v33n2/a02v33n2.pdf> >. Acesso em: 15 de set. 2012
- BARDIN, Laurence. **Análise de conteúdo**. Lisboa: Edições 70, 1977.
- BRASIL. Câmara dos Deputados. **Projeto de Lei PL 3070/2008**. Dispõe que os órgãos e entidades da Administração Pública Direta, Indireta, Autárquica, bem como os órgãos autônomos e empresas sob o controle estatal adotarão, preferencialmente, formatos abertos de arquivos para criação, armazenamento e disponibilização digital de documentos. Disponível em: <http://www.camara.gov.br/proposicoesWeb/prop_mostrarintegra?codteor=547009&filename=PL+3070/2008>. Acesso em: 14 de mar. 2012. Texto Original.
- CASTELLS, Manuel. **A sociedade em rede**. 9. ed. São Paulo: Paz e Terra, 2006. 698p.
- CENTRO DE ESTUDOS SOBRE AS TECNOLOGIAS DA INFORMAÇÃO E DA COMUNICAÇÃO – CETIC. **Tecnologias da Informação e da Comunicação Domicílios e Usuários 2011**. Disponível em: <<http://cetic.br/usuarios/tic/2011-total-brasil/index.htm> >. Acesso em: 20 out 2012.
- CERVO, Amado Luiz; BERVIAN, Pedro Alcino. **Metodologia Científica**. 4. ed. São Paulo: Makron Books, 1996. 209 p.
- CESAR, Ricardo. O esperanto dos arquivos. In: **Grande Decisões**. Revista Exame. Edição 889. 2007. Disponível em: <<http://exame.abril.com.br/revista-exame/edicoes/0889/noticias/o-esperanto-dos-arquivos-m0125023>>. Acesso em: 03 dez 2011.
- CURTY, Renata Gonçalves; GAMA, Fernando Alves da. Conjugando Diplomática e XML: aproximação possível no contexto da proveniência de documentos jurídicos digitais. **Arquivística.net**, Rio de Janeiro, v. 3, n. 2, p. 104-120, jul./dez. 2007.

DIAS, Rafael Cobbe, **PRESERVAÇÃO DIGITAL**: análise dos periódicos da base BRAPCI. Florianópolis, 2011. Disponível em: <http://repositorio.ufsc.br/xmlui/bitstream/handle/123456789/203/Rafael%20Cobbe%20Dias_2011_Preserva%C3%A7%C3%A3o%20Digital.pdf?sequence=1>. Acesso em: 20 maio 2012.

DICIONÁRIO ONLINE MICHAELIS. 2009. Disponível em: <<http://michaelis.uol.com.br/>>. Acessado em: 20 ago 2012.

FERREIRA, Aurélio Buarque de Holanda. **Dicionário Aurélio básico da língua portuguesa**. Rio de Janeiro, RJ: Nova Fronteira, 1999. xii, 687 p.

FERREIRA, Miguel. **Introdução à preservação digital**: conceitos, estratégias e actuais consensos. Guimarães, Portugal: Escola de Engenharia da Universidade do Minho, 2006. Disponível em: <<http://repositorium.sdum.uminho.pt/bitstream/1822/5820/1/livro.pdf>>. Acesso em: 20 maio. 2012.

FERREIRA, Miguel; BAPTISTA, Ana Alice; RAMALHO, José Carlos. Avaliação Automática de Migração em Redes Distribuídas de Conversores. In: CONFERÊNCIA DA ASSOCIAÇÃO PORTUGUESA DE SISTEMAS DE INFORMAÇÃO, 6., 2005, Bragança. **Anais...** Bragança: APSI, 2005. Disponível em: <<http://hdl.handle.net/10229/100645>>. Acesso em: 6 maio 2012.

GIL, Antonio Carlos. **Como elaborar projetos de pesquisa**. 4. ed. São Paulo: Atlas, 2002. 176 p.

GIL, Antonio Carlos. **Como elaborar projetos de pesquisa**. 5. ed. São Paulo: Atlas, 2010. 184 p.

GOODE, William J.; HATT Paul K. **Métodos em Pesquisa Social**. São Paulo: Companhia Editora Nacional, 4ª Edição, 1973.

GUEDES, Mário Augusto Muniz. **FATORES DE RISCO DE PERDA DE DOCUMENTOS ELETRÔNICOS DE CARÁTER ARQUIVÍSTICO EM UMA INSTITUIÇÃO PÚBLICA**: um estudo de caso na câmara dos deputados. Brasília, 2006. Disponível em: <http://bd.camara.gov.br/bd/bitstream/handle/bdcamara/1522/fatores_risco_guedes.pdf?sequence=1>. Acesso em: 01 nov 2012.

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICAS – IBGE. **Pesquisa Nacional por Amostra de Domicílios (Pnad)**. 2008. . Acesso: 01 maio 2012.

INNARELLI, Humberto Celeste. PRESERVAÇÃO DIGITAL: A INFLUÊNCIA DA GESTÃO DOS DOCUMENTOS DIGITAIS NA PRESERVAÇÃO DA INFORMAÇÃO E DA CULTURA. **Revista Digital de Biblioteconomia e Ciência da Informação**, Campinas, v. 8, n. 2, p.72-87, Não é um mês valido! 2011. Semestral. Disponível em: <http://www.sbu.unicamp.br/seer/ojs/index.php/sbu_rci/article/viewFile/487/330>. Acesso em: 20 maio 2012.

MÁRDERO ARELLANO, Miguel Ángel. **Critérios para a preservação digital da informação científica**. 2008. 354 f. Tese (Doutorado em Ciência da Informação)– Universidade de Brasília, Brasília, 2008. Disponível em: <http://eprints.rclis.org/bitstream/10760/12649/1/Tese_Miguel_%C3%81ngel_M%C3%A1rdero_Arellano.pdf>. Acesso em: 15 maio 2012.

MELLO, Tania Ramos de, **PERIÓDICOS CIENTÍFICOS**: processo de migração para a plataforma OJS. Florianópolis, 2011. Disponível em: <http://eprints.rclis.org/retrieve/28587/Tania_Tcc_2011.pdf>. Acesso em: 20 maio 2012.

MORENO, Nádina Aparecida; LOPES, Maria Aparecida; CHIARA, Ivone Guerreiro Di. A CONTRIBUIÇÃO DA PRESERVAÇÃO DE DOCUMENTOS E A (RE) CONSTRUÇÃO DA MEMÓRIA. **Biblionline**, João Pessoa, v. 7, n. 2, p. 03-11, 2011. Disponível em: <<http://www.brapci.ufpr.br/download.php?dd0=19727>>. Acesso em 01 nov 2012.

MOURA, Maria Aparecida. INFORMAÇÃO, FERRAMENTAS ONTOLÓGICAS E REDES SOCIAIS AD HOC: a interoperabilidade na construção de tesouros e ontologias. **Informação & Sociedade: Estudos**, João Pessoa, v. 19, n. 1, p. 59-73, jan./abr. 2009. Disponível em: <<http://www.brapci.ufpr.br/download.php?dd0=11393>>. Acesso em: 01 nov 2012.

ORGANIZAÇÃO DAS NAÇÕES UNIDAS – ONU. **Declaração Universal dos Direitos Humanos**, 1948. Acesso: 01 de maio de 2012. Disponível em: <http://portal.mj.gov.br/sedh/ct/legis_intern/ddh_bib_inter_universal.htm>

PARANÁ. Assembleia Legislativa. **Lei 15742/2007**. Dispõe que os órgãos e entidades da Administração Pública Direta, Indireta, Autárquica e Fundacional do Estado do Paraná, bem como os órgãos autônomos e empresas sob o controle estatal adotarão, preferencialmente, formatos abertos de arquivos para criação, armazenamento e disponibilização digital de documentos. Disponível em: <<http://www.legislacao.pr.gov.br/legislacao/pesquisarAto.do?action=exibir&codAto=30&indice=1&totalRegistros=1>>. Acesso em: 14 de mar. 2012. Texto Original.

RICHARDSON, Roberto Jarry; PERES, Jose Augusto de Souza. **Pesquisa social: métodos e técnicas**. 3. Ed. rev. Ampliada. São Paulo: Atlas, 1999.

RIO DE JANEIRO. Assembleia Legislativa. **Lei 5978/2011**. Dispõe que os órgãos e entidades da Administração Pública Direta, Indireta, Autárquica e Fundacional do Estado do Rio de Janeiro, bem como os órgãos autônomos e empresas sob o controle estatal adotarão, preferencialmente, formatos abertos de arquivos para criação, armazenamento e disponibilização digital de documentos. Disponível em: <alerjln1.alerj.rj.gov.br/contlei.nsf/f25edae7e64db53b032564fe005262ef/1b49ea86f53a8056832578a0005bba40?OpenDocument>. Acesso em: 14 de mar. 2012. Texto Original.

SAYÃO, Luís Fernando; MARCONDES, Carlos Henrique. O desafio da interoperabilidade e as novas perspectivas para as bibliotecas digitais. **Transinformação**, Campinas, v. 20, n. 2, p. 133-148, maio/ago. 2008. Disponível

em: < <http://www.brapci.ufpr.br/download.php?dd0=5629>>. Acesso em: 11 ago 2012.

SILVA, Jomar. **A independência de fornecedor de tecnologia propiciada pelo padrão ODF – OpenDocument Format (NBR ISO/IEC 26300)**. São Paulo: [s.n.], 2008. 38 p.

SUCA, Erika Guetti, SILVA; Flávio Soares Corrêa da. **Interoperabilidade e portabilidade de documentos digitais usando ontologias**. São Paulo, 2011. Disponível em: <http://ceur-ws.org/Vol-776/ontobras-most2011_paper11.pdf>. Acesso em: 01 nov 2012.

STILLE, Alexander. **A Destruição do passado**: como o desenvolvimento pode ameaçar a história da humanidade. São Paulo (SP): Arx, 2005. 423p.

THOMAZ, Katia P.. Gestão e preservação de documentos eletrônicos de arquivo: Revisão de literatura - Parte 1 / Electronic record: problem evolution – Part 1. **Arquivística.net**, Rio de Janeiro, v. 1, n. 2, p. 8-30, jul./dez. 2005.

VIRGINIA STATE LAWS ON OPTICAL IMAGES. **The Librery of Virginia** [Internet]. Virginia - EUA: The Librery of Virginia; 1999 [Acesso em: 20 de out. 2012]. Disponível em: <<http://www.archiveindex.com/laws/law-va.htm>>.

WENSING, Jairo. **Preservação e recuperação de informação em fontes de informações digitais**: estudo de caso do Greenstone. 2010. 219 p. Dissertação (Mestrado em Ciência da Informação) – Programa de Pós Graduação em Ciência da Informação, Universidade Federal de Santa Catarina, Florianópolis, 2010.

APÊNDICE

Unidades da Federação e Unidades da Federação e Regiões Metropolitanas	2005 (%)	2008 (%)	Diferença em (%) 2008 e 2005	Conclusão
Santa Catarina	35,08	39,38	4,30	Alta
Rondônia	21,36	22,69	1,34	Alta
Tocantins	34,16	33,85	-0,30	Baixa
Roraima	28,06	26,39	-1,68	Baixa
Piauí	31,02	27,05	-3,98	Baixa
Paraíba	36,58	31,78	-4,80	Baixa
Amapá	36,43	30,85	-5,58	Baixa
Paraná	42,64	36,98	-5,66	Baixa
São Paulo	44,04	38,15	-5,89	Baixa
Rio Grande do Sul	41,75	35,58	-6,17	Baixa
Minas Gerais	40,58	33,95	-6,63	Baixa
Região Metropolitana de Curitiba	49,35	42,32	-7,03	Baixa
Amazonas	34,04	26,81	-7,23	Baixa
Região Metropolitana do Rio de Janeiro	51,00	42,45	-8,55	Baixa
Sergipe	41,91	33,09	-8,82	Baixa
Rio de Janeiro	48,59	39,75	-8,84	Baixa
Região Metropolitana de São Paulo	50,64	41,35	-9,29	Baixa
Mato Grosso do Sul	35,95	26,57	-9,38	Baixa
Distrito Federal	57,24	47,68	-9,56	Baixa
Região Metropolitana de Belo Horizonte	51,22	41,53	-9,69	Baixa
Região Metropolitana de Recife	50,46	39,85	-10,61	Baixa
Região Metropolitana de Salvador	53,05	42,07	-10,98	Baixa
Região Metropolitana de Porto Alegre	48,58	37,36	-11,22	Baixa
Bahia	44,81	33,58	-11,22	Baixa
Pernambuco	45,97	34,27	-11,71	Baixa
Goiás	40,39	28,16	-12,23	Baixa
Espírito Santo	37,31	24,95	-12,35	Baixa
Alagoas	36,01	23,47	-12,54	Baixa
Região Metropolitana de Belém	49,43	36,71	-12,72	Baixa
Mato Grosso	41,57	28,24	-13,34	Baixa
Rio Grande do Norte	45,92	31,46	-14,46	Baixa
Acre	46,14	31,42	-14,72	Baixa
Ceará	39,57	24,35	-15,22	Baixa
Região Metropolitana de Fortaleza	43,68	27,95	-15,73	Baixa
Pará	44,86	28,31	-16,56	Baixa
Maranhão	41,18	19,04	-22,14	Baixa