



**UNIVERSIDADE ESTADUAL DA PARAÍBA – UEPB  
CENTRO DE CIÊNCIAS BIOLÓGICAS E SOCIAIS APLICADAS – CCBSA  
COORDENAÇÃO DE ARQUIVOLOGIA**

**GLAUCCO RANNIERE DE SOUZA PONTES**

**ARQUIVANDO NAS NUVENS:**  
Um recurso estratégico para preservação de documentos  
arquivísticos digitais

**João Pessoa – PB  
2014**

**GLAUCCO RANNIERE DE SOUZA PONTES**

**ARQUIVANDO NAS NUUVENS:**

Um recurso estratégico para preservação de documentos  
arquivísticos digitais

Trabalho de Conclusão de Curso  
apresentado ao Curso de Bacharel em  
Arquivologia, da Universidade Estadual da  
Paraíba, como requisito parcial para a  
obtenção do grau de **Bacharel** em  
Arquivologia, no Centro de Ciências  
Biológicas e Sociais Aplicadas.

**Orientadora:** Prof<sup>a</sup> Dr<sup>a</sup> Francinete Fernandes de Sousa

**João Pessoa – PB  
2014**

É expressamente proibida a comercialização deste documento, tanto na forma impressa como eletrônica. Sua reprodução total ou parcial é permitida exclusivamente para fins acadêmicos e científicos, desde que na reprodução figure a identificação do autor, título, instituição e ano da dissertação.

P813a Pontes, Glaucio Ranniere de Souza

Arquivando nas nuvens [manuscrito] : um recurso estratégico para a preservação de documentos arquivísticos digitais / Glaucio Ranniere de Souza Pontes. - 2014.

126 p. : il.

Digitado.

Trabalho de Conclusão de Curso (Graduação em Arquivologia) - Universidade Estadual da Paraíba, Centro de Ciências Biológicas e Sociais Aplicadas, 2014.

"Orientação: Profa. Dr. Francinete Fernandes de Souza, Departamento de Arquivologia".

1. Preservação digital 2. Documentos em nuvem. 3. DSpace.  
I. Título.

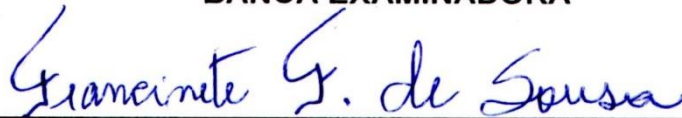
21. ed. CDD 025.174

**GLAUCCO RANNIERE DE SOUZA PONTES**

**ARQUIVANDO NAS NUVENS: um recurso estratégico para a preservação de documentos arquivísticos digitais.**

Trabalho de Conclusão de Curso apresentado ao Curso de Bacharel em Arquivologia, da Universidade Estadual da Paraíba, como requisito parcial para a obtenção do grau de **Bacharel** em Arquivologia, no Centro de Ciências Biológicas e Sociais Aplicadas.

**BANCA EXAMINADORA**



---

Profª Drª Francinete Fernandes de Sousa/ UEPB  
Orientadora



---

Profa. Drª Eliete Santos UEPB



---

Profa. Msª Maria Meliane Vieira Rocha UFPB

**À minha amada mãe, por ser a minha primeira  
docente a transmitir o conhecimento e a me  
educar com amor, disciplina e carinho,**

**DEDICO.**

## **AGRADECIMENTOS**

Ao Grande Arquiteto Do Universo, por permitir a expansão do meu microcosmo ávido por cognição que ajudará na ampliação do macrocosmo em evolução.

A professora Dr<sup>a</sup> Francinete Fernandes de Sousa, pela dedicação, amor à profissão, carinho pelos seus alunos e pela maestria nas suas orientações, a qual nos ajudou a compor a estrutura deste trabalho.

Ao professor Dr. Josemar Henrique de Melo, por explanar, brilhantemente, as suas aulas sobre Documento Digitais e por nos ter orientado no início deste trabalho.

Aos professores Ms.<sup>a</sup> Maria José Cordeiro de Lima e Dr. José Washington Medeiros, Coordenadora e Vice-coordenador do curso de Arquivologia, respectivamente, pela capacidade, empenho e competência.

Aos professores do Curso de Arquivologia da UEPB, Dr<sup>a</sup>. Francinete Fernandes de Sousa, Dr<sup>a</sup>. Eliete Correia, Ms<sup>a</sup>. Esmeralda Porfírio, Dr. Jimmy Lellis, Ms<sup>a</sup> Maria José Cordeiro de Lima (Mara), Dr. José Washington Medeiros, Dr. Vancarder Brito, Esp. Eutrópio Bezerra, Dr<sup>a</sup>. Acácia Maria, Ms<sup>a</sup>. Manuela Maia, Dr. Roberto Jorge, Dr<sup>a</sup>. Suerde Brito, Dr<sup>a</sup>. Jacqueline Barrancos, Dr. Josemar Henrique de Melo, Ms. Henrique França que poliram esta pedra bruta para reluzir o conhecimento.

Aos colaboradores da UEPB que desempenham um exímio trabalho, em especial, Daniela Duarte, secretária do Curso de Arquivologia, sempre prestativa para com todos.

Ao Professor Dr. Alberto Banal, por ceder gentilmente a entrevista em seu escritório, pelo seu carisma e pelas belas colocações durante a entrevista.

A minha amada mãe Albina de Souza Pontes, por ser minha grande mentora e por me amparar nos momentos críticos. Ao meu pai Dr. Leonardo Batista de Pontes.

A minha namorada Taís Grebos, pelo intenso e constante apoio, principalmente, nos momentos em que o cansaço parecia me dominar.

Aos muitos colegas do Curso de Arquivologia que se transformaram em pouquíssimos amigos.

*“Deixe o futuro dizer a verdade, e avaliar cada um de  
acordo com seus trabalhos e suas conquistas”.*

*(Nikola Tesla)*

## RESUMO

Este trabalho fez uma reflexão sobre os processos, métodos, técnicas e estratégias utilizadas pelos profissionais da informação e, em suas considerações, foram discutidas os conhecimentos arquivísticos adquiridos, de um passado não tão distante, pelo profissional arquivista que fará a preservação dos documentos digitais. O referido estudo pretendeu, além de estreitar os laços entre as novas tecnologias e os profissionais da informação, em especial o arquivista, apresentar um novo recurso estratégico de preservação digital a ser aplicado em documentos arquivísticos e nos mais variados objetos digitais. As questões abordadas por autores como: Hedstron (2003), Katia Thomaz (2004), Miguel Ferreira (2006), Miguel Arellano (2008), José Grácio (2012), e a forma desafiadora com a qual eles e outros profissionais da informação encaram a preservação digital, permitiu a descrição dos processos mais adotados para preservar a documentação digital arquivística e a explicação das potências virtuais e a estrutura funcional de uma nova tecnologia denominada de computação em nuvem. Vislumbrando que a preservação digital atinge todo ou qualquer objeto digital, o trabalho foi direcionado para um acervo fotográfico digital do projeto social: Quilombos da Paraíba, que detêm mais de 25.000 (vinte e cinco mil) fotografias digitais, o que proporcionou a materialização do objeto de estudo e facilitou a demonstração de um novo recurso estratégico de preservação digital. O trabalho, também, fez uso de varias ferramentas as quais, quando integradas, acomodam um significativo leque de opções para o desenvolvimento de um sistema de gerenciamento e preservação de documentos arquivísticos digitais, ou seja, repositórios digitais. O DSpace, repositório referenciado na pesquisa, foi descrito como peça fundamental para a aplicação da preservação digital em nuvem, pois a sua atuação mantém os objetos digitais seguros, disponíveis, atualizados e acessíveis por múltiplas plataformas, mas isto se dará se, e somente se, for implementado em uma estrutura de computação em nuvem, provendo desta forma a longevidade, não apenas dos documentos do acervo Quilombos da Paraíba, mas de todo documento arquivístico digitalizado ou produzido digitalmente.

**Palavras-chave:** Preservação digital; Documentos em nuvem; *DSpace*; Computação em nuvem



## ABSTRACT

This paper made a reflection on the processes, methods, techniques and strategies used by information professionals and their considerations, the archival knowledge acquired in the not so distant past, the professional archivist who will make the preservation of digital documents were discussed. This study sought, in addition to strengthening ties between new technologies and information professionals, especially the archivist, presenting a new strategic resource for digital preservation to be applied to records and in various digital objects. The issues addressed by authors such as: Hedstron (2003), Katia Thomaz (2004), Miguel Ferreira (2006), Miguel Arellano (2008), José Grácio (2012), and challenging way in which they and other information professionals face the digital preservation, allowed the description of the processes adopted to preserve more digital archival documentation and explanation of virtual powers and functional structure of a new technology called cloud computing. Envisaging that the digital preservation affects all or any digital object, the work was directed to a digital photographic archive of the social project: Quilombo of Paraiba, which hold more than 25,000 (twenty five thousand) digital photographs, which provided the materialization of the object study and facilitated the demonstration of a new strategic resource of digital preservation. The work also made use of several tools which, when integrated, accommodating a significant range of options for the development of a system for managing and preserving digital records, or digital repositories. DSpace, referenced in the research repository, has been described as essential to the implementation of digital preservation cloud piece because its actions keep them safe, available, updated and accessible by multiple platforms digital objects, but this will occur if and only if, is implemented in a framework of cloud computing, thereby providing longevity, not just the collection of documents Quilombos of Paraiba, but all scanned or digitally produced archival document.

**Palavras-chave:** Digital preservation; Documents cloud; DSpace; Cloud Computing.

## LISTA DE FIGURAS

Figura 1 – Estrutura por camadas para criação do documento digital.....	31
Figura 2 – Componentes gerando o modelo funcional do OAIS.....	41
Figura 3 – Modelo de Refrescamento.....	46
Figura 4 - Recuperação de um objeto digital por meio de emulação.....	48
Figura 5 - Preservação baseada na emulação do conceito Rothenberg.....	49
Figura 6 - Degradação do objeto digital ao longo de sucessivas migrações.....	52
Figura 7 – Migração a-pedido.....	53
Figura 8 – Esquema de passagens entre Real e Virtual.....	58
Figura 9 – Representação da criação de Servidores ou Máquinas Virtuais.....	62
Figura 10 – Virtualização de Hardwares.....	65
Figura 11 – Esquema de Paravirtualização.....	66
Figura 12 – Esquema da Virtualização Completa.....	68
Figura 13 – Esquema de funcionamento de virtualização de aplicativo (Java Virtual Machine).....	69
Figura 14 – Esquema básico da infraestrutura da computação em nuvem.....	74
Figura 15 – Representação dos serviços sem e com aplicação de virtualização.	79
Figura 16 – Processo sistemático funcional do Dspace.....	96
Figura 17 – Infraestrutura do recurso estratégico conceitual de preservação digital em nuvem.....	105

## LISTA DE IMAGENS

Imagem 1 – Visão aérea da Data Center Facebook.....	76
Imagem 2 – Servidores da Data Center do Facebook.....	76
Imagem 3 – Dspace Institucional da Universidade Estadual da Paraíba.....	95
Imagem 4 – Dspace modificado para atender as necessidades de preservação da instituição.....	95

## LISTA DE ABREVIATURAS E SIGLAS

API	<i>Application Programming Interface</i>
AIP	<i>Archival Information Package</i>
AIT	<i>Advanced Intelligent Tape</i>
BLU-RAY	Disco óptico de nova geração
CCSDS	<i>Consultive Committee for Space Data</i>
CD	<i>Compact Disk</i>
CD-ROM	<i>Compact Disk Read-Only Memory</i>
CD-RW	<i>Compact Disk – ReWritable</i>
CONARQ	Conselho Nacional de Arquivo
DAT	<i>Digital Audio Tape</i>
DDS	<i>Digital Data Storage</i>
DIP	<i>Dissemination Information Package</i>
DNS	<i>Domain Name System</i>
doc	Abreviação de <i>document</i> – Extensão do <i>Microsoft Office Word</i>
DVD	<i>Digital Versatile Disc</i>
DVD-R	<i>Digital Versatile Disc – Read only Memory</i>
DVD-RW	<i>Digital Versatile Disc – ReWritable</i>
EBC	Empresa brasileira de Comunicação
ERD	<i>Emergency Repair Disk</i>
GB	Gigabyte
HD	<i>Hard Disk</i>
HP	<i>Hewlett-Packard</i>

HTTP	<i>Hypertext Transfer Protocol</i>
IaaS	Infraestructure as a Service
IBICT	Instituto Brasileiro de Informações em Ciência e Tecnologia
IBM	<i>International Business Machines</i>
InterPARES	<i>International Research on Permanent Authentic Records in Electronic Systems</i>
IP	<i>Internet Protocol</i>
ISO	<i>International Organization Standardization</i>
jpeg	<i>Joint Photographics Experts Group</i>
JVM	<i>Java Virtual Machine</i>
LAN	<i>Local Area Network</i>
MAN	<i>Metropolitan Area Network</i>
MB	Megabyte
MIT	<i>Massachussets Institute of Tecnology</i>
NAS	<i>Network Attached Storage</i>
NIST	<i>National Institute of Standard and Tecnology</i>
OAI	<i>Open Archive Initiative</i>
OAI-PMH	<i>Open Archive Initiative Protocol for Metadata Harvesting</i>
OAIS	<i>Open Archival Information System</i>
OCLC	<i>Online Computer Library Center</i>
ODF	<i>Open Document Format</i>
PaaS	<i>Plataform as a Service</i>
PDF	<i>Portable Document Format</i>
POM	<i>Project Object Model</i>
RAID	<i>Redundant Array of Independent Drives</i>

RODA	Repository Object Digital Authentic
SaaS	<i>Software as a Service</i>
SD	<i>Secure Digital</i>
SERPRO	Serviço Federal de Processamento de Dados
SGBD	Sistema de Gerenciamento de Bando de Dados
SIP	<i>Submission Information Package</i>
SOA	Service-Oriented Architecture
SSD	<i>Solid State Drive</i>
TI	Tecnologia da Informação
TIC	Tecnologia da Informação e Comunicação
UEPB	Universidade Estadual da Paraíba
UNESCO	Organização das Nações Unidas para a Educação, a Ciência e a Cultura
VoIP	<i>Voice over Internet Protocol</i>
WAN	<i>Wide Area Network</i>
Web	Sistema que opera sobre a Internet
XML	eXtensible Markup Language

## SUMÁRIO

<b>1. INTRODUÇÃO</b> .....	17
<b>2 ASPECTOS METODOLOGICOS</b> .....	21
2.1 CARACTERIZAÇÃO DA PESQUISA.....	21
2.2 PROBLEMATIZAÇÃO.....	23
2.3 OBJETIVOS.....	25
2.3.1 Geral.....	25
2.3.2 Específicos.....	25
2.4 UNIVERSO E AMOSTRAGEM.....	26
2.5 INSTRUMENTO DE COLETA DE DADOS.....	27
<b>3 PRESERVAÇÃO DIGITAL</b> .....	28
3.1 DOCUMENTO DIGITAL.....	30
3.2 SUPORTES E DISPOSITIVOS DE ARMAZENAMENTO.....	34
3.2.1 Tipos de suporte, dispositivos de armazenamento e mídias.....	34
3.2.1.1 Magnéticos.....	34
3.2.1.2 Ópticos.....	36
3.2.1.3 Eletrônicos.....	37
3.3 SISTEMA E ESTRATÉGIAS DE PRESERVAÇÃO.....	38
3.3.1 Sistema de informação de arquivos abertos – OAIS.....	39
3.3.2 Estratégias de preservação.....	44
3.3.2.1 Refrescamento.....	46
3.3.2.2 Emulação.....	47
3.3.2.3 Migração de dados.....	50
<b>4 DO VIRTUAL A VIRTUALIZAÇÃO</b> .....	55
4.1 DESCREVENDO O VIRTUAL.....	55
4.2 O QUE É VIRTUALIZAÇÃO.....	60
4.3 O FUTURO DA VIRTUALIZAÇÃO.....	70
<b>5 COMPUTAÇÃO EM NUVEM</b> .....	71
5.1 O CONCEITO DE COMPUTAÇÃO EM NUVEM.....	72
5.2 INFRAESTRUTURA.....	75
5.2.1 Modelos de serviços.....	76
5.2.2 Tipos de nuvem.....	79
5.3 UTILIZANDO A NUVEM.....	82
<b>6 PRESERVAÇÃO EM NUVEM</b> .....	84
6.1 MECANISMOS DE REFERÊNCIA PARA PRESERVAÇÃO EM NUVEM.....	85
6.2 ANALISANDO O ACERVO FOTOGRÁFICO DIGITAL.....	90

<b>6.3 REPOSITÓRIO DIGITAIS: O caminho para nuvem.....</b>	<b>92</b>
<b>6.4 O RECURSO ESTRATÉGICO PARA PRESERVAÇÃO DE DOCUMENTOS ARQUIVÍSTICOS DIGITAIS EM NUVEM.....</b>	<b>100</b>
<b>7 CONSIDERAÇÕES FINAIS.....</b>	<b>109</b>
<b>REFERÊNCIAS.....</b>	<b>112</b>
<b>APÊNDICES.....</b>	<b>118</b>
<b>ANEXOS.....</b>	<b>125</b>



## 1 INTRODUÇÃO

Encarado como um grande desafio, este trabalho foi fomentado a partir da disciplina “documentos digitais”, ofertada pela Universidade Estadual da Paraíba e ministrada pelo professor Dr. Josemar Henrique de Melo, no sexto período do curso de graduação em Arquivologia. Os fatores que corroboraram para a elaboração desta produção foram as questões levantadas sobre os procedimentos adequados para preservar o documento digital e como o profissional da informação, em especial o arquivista, utiliza seu conhecimento para elaborar métodos, técnicas ou estratégias que irão garantir a preservação do documento digital por um longo prazo.

A Arquivística, Ciência e disciplina que objetiva gerir, organizar e delimitar políticas de acesso a todos os documentos expedidos e recebidos por órgãos de caráter físico e jurídico, independente do seu suporte, não obstante, as informações que possam ser tramitadas e arquivadas em sistemas de informação e/ou repositórios digitais, emprega princípios, normas, técnicas e procedimentos diversos que são aplicados nos processos de composição, coleta, análise, identificação, organização, processamento, desenvolvimento, utilização, disseminação, fornecimento, circulação, armazenamento, preservação e recuperação da informação.

Ao abordarmos fenômenos da tecnologia da informação, temos de permanecer em constante atualização, pois a esta é dinâmica e a sua obsolescência torna-se volátil. Além de a tecnologia trazer consigo o dinamismo, ela nos oferece a comodidade para soluções de problemas encontrados, cotidianamente, nos mais variados campos sociais sejam em residências com a criação de documentos simples em um editor de texto, ou mesmo em grandes instituições governamentais que circulam uma grande quantidade de processos documentais por dia.

Cada documento requer cuidados, principalmente o digital, pois este é a atualização da obsolescência do papel. Sendo assim, a preservação digital é um tema vigente que atinge os campos das tecnologias e da Ciência da Informação, além de sempre ser repercutido com grande entusiasmo e debatido calorosamente em inúmeros eventos profissionais e acadêmicos. Isto tem ocorrido pelo forte

impacto causado pelas demandas tecnológicas que aumentam a cada dia em todo o mundo. Este impacto faz a sociedade aderir, de forma natural, aos avanços tecnológicos, provocando de uma forma consequencial, o surgimento do “dilúvio de informação” e acúmulo dos documentos orgânicos e digitais. (LÉVY, 1996)

O processo de preservação está atrelado com o do acesso, uma vez que este fator ocorre como uma condição “*sine qua non*”, onde a preservação é a ação para um acesso futuro. Para Arellano (2007), “a preservação digital deve ser uma preocupação dos editores científicos e dos centros de informação”, pois a preservação traz consigo fatores consideráveis para a sociedade, tais como a memória, valor probatório, valor histórico, valor cultural, entre outros. As tecnologias desenvolvidas para o ambiente de preservação procuram facilitar, não apenas o acesso, mas encurtar a distância entre o usuário e o objeto, isto é, dinamiza a disseminação e ainda preserva o objeto digital em seu meio e possibilita a implementação dos recursos de segurança a estes objetos preservados.

Como produto das novas tecnologias, os documentos digitais trouxeram uma problemática para os vanguardistas da Ciência da Informação, com o mesmo peso em que a sociedade necessita das novas tecnologias para executar suas atividades cotidianas. Entretanto, algumas possíveis soluções para o devido problema da preservação foram desenvolvidos naturalmente, como são os casos dos disquetes, fitas magnéticas, CD-R, DVD-R, BLU-RAY e dentre outras mídias que são criadas todos os dias, mas que se tornam obsoletas em um período de tempo muito curto, sendo consideradas frágeis e passíveis de extravio para o armazenamento, (INNARELLI, 2007). Por fim, ainda temos as questões que abrangem os softwares, suas licenças de uso e incompatibilidade de *softwares* que trazem altos custos anuais para as instituições ou para usuários comuns. (MARCONDES; SAYÃO, 2002)

No percurso destes pensamentos delimitamos esta pesquisa que veio com o propósito de expor as potencialidades das novas tecnologias e dentre variadas inovações tecnológicas, destacar a Computação em Nuvem, sendo esta, a infraestrutura promissora de armazenamento, compartilhamento, acessibilidade, maior interoperabilidade entre *softwares* e outros benefícios impostos por essa tecnologia que adquire, a cada dia, mais destaque no cenário nacional e internacional.

Outro fator para o fomento desta pesquisa foi a perspectiva de reestruturação dos processos de preservação de documentos e informação produzidos no ambiente digital, pois harmonizando as novas tecnologias, em especial a infraestrutura virtualizada de computação em nuvem, atrelada aos princípios arquivísticos, ocasionará resultados consideráveis para o campo da Ciência da Informação e da tecnologia, trazendo assim soluções não apenas para as questões da esfera acadêmica e corporativa, mas para toda sociedade. Então, uma das questões norteadora da pesquisa foi: Qual a melhor estratégia para preservar um objeto digital e como preservar? Muitas questões vão além dos mundos citados anteriormente, pois envolve todo universo social.

Além de a pesquisa ter indicado soluções práticas inéditas, como preservar os objetos digitais em repositórios embarcados em uma infraestrutura de computação em nuvem, pode a nosso ver acarretar inúmeras provocações ao campo de estudo teórico arquivístico, ocasionando em uma aliança ou um estreitamento com o campo tecnológico, trilhando, desta forma, um caminho para a Arquivologia e a Ciência da Informação, o qual vem buscando alguns estudiosos mais progressistas, na área em questão.

Sendo assim, esta pesquisa reflete sobre a construção das novas bases para a preservação digital e para a arquivística. Os seus princípios são embarcados em um ambiente tecnológico, renovando todo o sentido de armazenamento, gestão, acesso, controle, fluxo, segurança, disseminação, análise e preservação das informações e dos documentos, originando, dessa forma, a vanguarda arquivística tecnológica. Para apresentar uma experiência concreta do nosso modelo, trouxemos um modelo para a organização de um acervo de 25.000 (vinte e cinco) mil fotos.

No tocante a estrutura deste trabalho, para dar maior clareza à investigação, ele foi dividido da seguinte forma. No primeiro capítulo, temos a introdução que explica a escolha do tema. No segundo capítulo, abordamos os procedimentos metodológicos que nos direcionaram na pesquisa. No terceiro capítulo, versamos sobre os princípios, conceitos e técnicas de virtualização. No quarto capítulo, descrevemos a preservação digital, os mecanismos e seus procedimentos técnicos já utilizados. No quinto capítulo, explicamos o funcionamento da infraestrutura de computação em nuvem. No sexto, apresentamos o repositório digital, o recurso

estratégico de preservação em nuvem e a sua análise após a aplicação no acervo fotográfico digital. Por último, o sétimo capítulo que trata das considerações finais do trabalho.

## 2 ASPECTOS METODOLÓGICOS

Nesta etapa faremos uma descrição de quais passos e como os mesmos foram dados para a consecução da pesquisa.

### 2.1 CARACTERIZAÇÃO DA PESQUISA

Preservação em Nuvens é um recurso estratégico de armazenamento de objetos digitais que utiliza os recursos da computação em nuvem e os mecanismos de preservação digital para prolongar o acesso e uso de documentos arquivísticos digitais e outros objetos digitais. Este recurso estratégico é constituído por pessoas e camadas de sistemas, que aprofundam a responsabilidade de preservar por um longo prazo a informação ou documentos digitais e torná-los disponíveis para o usuário.

Leva-se em consideração sobre a computação em nuvem, a evolução da virtualização dos recursos informáticos e computacionais, possibilitando que o usuário obtenha o acesso, produção, gerenciamento, armazenamento dos objetos ou documentos digitais, independentemente da plataforma que o usuário utilize, ou da necessidade da instalação de aplicativos em seu dispositivo de acesso, salvo os *softwares* navegadores (Web browser).

Os mecanismos de preservação digital, o modelo referencial OAIS (Sistema de Informações de Arquivos Abertos), o repositório digital e a computação em nuvem, são os quatro elementos que podem ser transformados em uma ferramenta que trará, de fato, os resultados que indicarão a funcionalidade deste recurso estratégico. No decorrer da pesquisa, analisamos individualmente estes elementos que irão compor o recurso estratégico de preservação. Tal recurso demonstrou que os mecanismos de preservação digital podem deixar de possuir o caráter mecanicista para adquirir aspectos holísticos, ou seja, passarem a ser integrados e interoperáveis, sendo ferramentas sistêmicas para a gestão e preservação da documentação digital.

Considerando o tema abordado, optamos pela pesquisa teórica. Este tipo de pesquisa é indicado a um sentido de renovar teorias, remodelar quadros de referência, rever e alterar se necessário, as condições explicativas da realidade, fomentar polêmicas e gerar discussões pertinentes.

A pesquisa teórica não provoca de imediato à intervenção na realidade, porém isto não deixa de ser importante, pois sua função é decisiva na criação de categorias para a intervenção da teoria. A pesquisa teórica, segundo Demo (2000, p. 20), “é a pesquisa dedicada a reconstruir teoria, conceitos, ideias, ideologias, polêmicas, tendo em vista, em termos imediatos, aprimorar fundamentos teóricos”.

Na relevância do tema e da caracterização da pesquisa, Demo (1995, p. 36), ressalta a seguinte visão para o aspecto teórico, “o conhecimento teórico adequado acarreta rigor conceitual, análise acurada, desempenho lógico, argumentação diversificada, capacidade explicativa”.

A opção pela utilização da pesquisa teórica partiu da necessidade de pensar desde a graduação em renovar os métodos de preservação digital de documentos e estreitar os laços da arquivística com as novas tecnologias, pois a união destas duas subáreas visa uma ampliação dos estudos sobre a documentação e informação no espaço digital, no entanto utilizamos um exemplo concreto para deixar mais claras as postulações feitas ao longo do trabalho.

Ressaltamos que não tratou-se de um experimento. Isto pode ser feito em uma pós- graduação. É apenas a explicação concreta de nossas convicções teóricas sobre a possibilidade de fazermos um armazenamento em nuvem, cumprindo itens de organização e preservação, técnicas já muito discutidas em suportes físicos como o papel, mas agora possível nas nuvens.

Considerando a abordagem, a pesquisa fundamentou-se na esfera qualitativa, uma vez que, mais se adaptou e que nos fez ter uma melhor observação do objeto estudado, nos ajudando a descrever sua funcionalidade, como é afirmado por Neves (1996, p. 36), a pesquisa qualitativa é: “[...] um conjunto de diferentes técnicas interpretativas que visam a descrever e a decodificar os componentes de um sistema complexo de significados. Tendo por objetivo traduzir e expressar o sentido dos fenômenos do mundo social [...]”.

A pesquisa qualitativa procura tratar as atividades da ciência, visando construir a realidade, porém preocupando-se com as ciências sócias em determinado ponto da realidade que não pode ser quantificado. A pesquisa qualitativa é explicada por Gil (1996, p. 159), da seguinte forma:

Considera que há uma relação dinâmica entre o mundo real e o sujeito, isto é, um vínculo indissociável entre o mundo objetivo e a subjetividade do sujeito que não pode ser traduzido em números. A interpretação dos fenômenos e a atribuição de significados são básicos no processo de pesquisa qualitativa. Não requer o uso de métodos e técnicas estatísticas. O ambiente natural é a fonte direta para coleta de dados e o pesquisador é o instrumento chave. É descritiva. Os pesquisadores tendem a analisar seus dados indutivamente. O processo e seu significado são os focos principais de abordagem.

Desta forma, a pesquisa se fortalece e respalda cientificamente nas técnicas e teorias para a coleta dos dados, para que estes sejam compilados e Interpretados, garantindo assim, o sucesso e conclusão do tema.

## 2.2 PROBLEMATIZAÇÃO

Com o forte uso das Tecnologias da Informação e Comunicação (TIC) que vem revolucionando o cenário mundial da informação, no que toca o âmbito da produção, reprodução e uso dos documentos nos ambientes digitais, vem atrelando consigo algumas dúvidas sobre estas novas tecnologias, tais como: segurança, disponibilidade, custo, fácil utilização, entre outros. Frente a essas demandas tecnológicas, o Brasil se insere obtendo destaque no mundo cibernético, uma vez que os estudos relacionados às questões das tecnologias da informação surgiram no país de forma relevante e produtiva, fomentando as pesquisas no campo da Ciência da Informação, desde a consolidação da Web, por volta de 1995.

Em meados dos anos 90, no século passado, alguns autores como Pierre Lévy, Luciano Floridi, dentre outros, discutiam sobre esse “dilúvio” informacional digital, sendo uns otimistas e outros pessimistas na aceitação das novas

tecnologias, mas todos contribuindo com estudos que ajudaram a desenvolver o campo da Ciência da Informação.

Atualmente, os profissionais que lidam com documentação e informação digital se encontram com problemas, que se iniciam desde a utilização dos computadores e com o advento da web, tais como: organização, disseminação, compartilhamento, produção e qualidade da informação, segurança da informação, Infodesign, Internet e Preservação Digital. Este último, objeto do nosso estudo, tem fomentado calorosas discussões e, conseqüentemente, aumentado a produção científica e bibliográfica.

A preservação de documentos digitais causa desentendimento até os dias atuais, dividindo-se em temas e fomentando questões como: hardware, software, políticas de preservação, sistemas de gerenciamento, migração, meta-dados, replicação, lixo digital, durabilidade, confiabilidade, suporte e outros. A inserção das instituições no ambiente web veio embarcar toda sua produção e tramite documental para este ambiente, trazendo mais dinamicidade e conforto para os usuários. Por outro lado, trouxe problemas como: obsolescência, licença de software, acondicionamento, gestão dos dados, segurança, qualificação de pessoas, preservação, dentre outros.

Levou-se em consideração alguns modelos definidos ao longo de anos de discussão sobre a fidedignidade do suporte eletrônico, segurança e a preservação digital. Desta forma, após a compreensão dos mecanismos de preservação, do modelo OAIS e do repositório digital, analisou-se a possibilidade da aplicação do modelo referencial às nuvens e observamos as vantagens e desvantagens que foram convertidas em valores qualitativos.

A conexão entre os elementos utilizados para o novo recurso estratégico trará um alicerce teórico complexo, mas que provocará mudanças significativas na área tecnológica e científica e, sendo mais específico, no campo da preservação e processos de gerenciamento digitais.

Faz-se necessário refletir sobre os antigos métodos políticos e estratégicos de preservação aplicados aos objetos digitais, relacionando-os com a pós-



implementação do modelo OAIS e da computação em nuvem, isto é, o caso específico deste trabalho sobre preservação digital.

O problema de investigação da pesquisa foi: Qual a estratégia mais eficaz, nos dias atuais, para se preservar as informações, documentos e/ou objetos digitais por um longo prazo?

## **2.3 OBJETIVOS**

### **2.3.1 Objetivo Geral**

Apresentar um recurso estratégico arquivístico de preservação de documentos digitais, utilizando um repositório digital integrado à infraestrutura de computação em nuvens.

### **2.3.2 Objetivos Específicos**

- i. Descrever os processos de preservação digital;
- ii. Explicar o funcionamento da infraestrutura de computação em nuvens;
- iii. Apresentar um recurso estratégico de preservação de objetos digitais, utilizando um modelo de infraestrutura de computação em nuvens aplicado a um acervo de 25 mil fotos digitais.

## 2.4 UNIVERSO E AMOSTRAGEM

Para tornar concreta a realização da referente pesquisa, fez-se necessário para o desenvolvimento do trabalho delimitar o universo e a amostragem. Segundo (LAKATOS; MARCONI, 2008, p. 225)

[...] universo ou população é o conjunto de seres animados ou inanimados que apresentem pelo menos uma característica em comum. Sendo N o número total de elementos do universo ou população, o mesmo pode ser representado pela letra latina maiúscula X, tal que  $XN = X1; X2 X3; [...]; XN$ . A delimitação do universo consiste em explicitar que pessoas ou coisa, fenômenos etc. serão pesquisados, enumerando suas características comuns, como, por exemplo, sexo, faixa etária, organização a que pertencem, comunidade onde vivem etc.

Entendendo o universo como sendo o conjunto de fenômenos que obtêm fatos apresentados com uma característica em comum, e população como um conjunto de números obtidos pela realização da mensuração dos atributos dos fenômenos que compõem um universo, incidimos que a composição do universo da nossa pesquisa abrange 150.000 (cento e cinquenta mil) fotos do acervo Quilombos da Paraíba.

O elemento de amostragem foi um acervo fotográfico digital com mais de 25.000 (vinte e cinco mil) fotos. Para Lakatos e Marconi (2007 p. 225) amostragem é “uma parcela convenientemente selecionada do universo (população); é um subconjunto do universo”. Inicialmente, a pesquisa não foi aplicada diretamente no acervo, uma vez que esta pesquisa teórica trata apenas de um recurso estratégico de preservação. Portanto, o produto deste trabalho pode ser adotado pelo gestor do acervo, que se demonstrou interessado, após a entrevista.

## 2.5 INSTRUMENTOS DE COLETA DE DADOS

Para Marconi e Lakatos (2003, p.165), a coleta de dados “é a etapa que se inicia a aplicação dos instrumentos elaborados e das técnicas selecionadas, a fim de efetuar a coleta dos dados previstos”.

A técnica escolhida para complementar com a pesquisa foi a entrevista. Esta mostrou-se mais ajustável para esclarecer como uma organização ou instituição realiza a preservação dos objetos digitais. Segundo Marconi e Lakatos (2003, p. 165) “são vários os procedimentos para realização de coleta de dados, que variam de acordo as circunstâncias ou com a investigação”.

A entrevista utilizada para estruturar os dados foi a “focalizada”, pois como afirma Marconi e Lakatos (2003, p.197), é “um roteiro de tópicos relativos ao problema que se vai estudar e o entrevistador tem liberdade para fazer as perguntas que quiser: sondar razões, motivos e dá esclarecimentos, não obedecendo, a rigor, a uma estrutura formal”.

Portanto, a entrevista se ateve em questões que faziam alusão aos recursos utilizados para armazenamento digital, procedimentos adotados para o armazenamento e recuperação, experiência tecnológica, perda de dados, gestão arquivística, dificuldades e facilidades em gerir o acervo e recursos financeiros para preservar os objetos digitais.

### 3. PRESERVAÇÃO DIGITAL

Desde a sua origem, a humanidade desenvolveu técnicas que preservam as atividades de hoje para o uso no amanhã, assegurando a continuação das atividades, progressão social, registro cultural e, evidentemente, para o resgate da memória factual em um futuro próximo ou remoto. Le Goff (1991, p. 404), diz que “a história viveu uma revolução documental e que a memória arquivista foi revolucionada pelo aparecimento de uma nova memória: o banco de dados.”.

Haja vista que o desenvolvimento da tecnologia da informação foi a mola propulsora para a referente revolução, trazendo incontáveis soluções para o dia a dia, mas também nos enveredando em uma problemática que se perpetua ano após ano, isto é, a preservação digital.

Para Ferreira (2006, p. 20), a preservação é definida como “a capacidade de garantir que a informação digital permaneça acessível e com qualidade de autenticidade para que possa, no futuro, ser interpretada numa plataforma tecnológica diferente daquela utilizada em sua criação.”.

Com o desenvolvimento da Tecnologia da Informação, os métodos e técnicas de execução das atividades se tornaram mais flexíveis, pois o homem se adapta, facilmente, àquilo que lhe proporciona maior conforto. Por este fator, a humanidade se torna cada vez mais dependente das novas tendências das Tecnologias da Informação e Comunicação geradas pela indústria tecnológica (TURBAN; LEIDNER, et al, 2008)

Ao mesmo tempo em que a indústria tecnológica nos fornece inúmeras soluções geniais para executarmos dinamicamente as atividades diárias, surge o grande desafio de preservar o produto final dessas atividades, ou seja, o documento digital. Essas atividades são, redundante e conseqüentemente o número de documentos digitais produzidos são maiores que as atividades designadas, pois de uma atividade pode-se gerar mais de um documento e em formatos distintos. (THOMAZ, 2004)

Com isso, há grandes preocupações dos profissionais da informação, no que toca o documento digital e sua preservação, pois muito se fala da insegurança, obsolescência de *software*, *hardwares*, volatilidade, suporte de mídias, custos com dispositivos de *back-up*, cópia desnecessária, entre outros. (GRÁCIO, 2012)

Essas e outras preocupações não estão ligadas somente aos profissionais da informação, mas sim a todos que fazem uso das tecnologias para a solução das suas atividades, pois a necessidade de preservar é instintiva do ser humano. Com a preservação o homem projeta o presente sempre para o futuro e o futuro sempre é esclarecido pelo passado.

A preservação digital designa ser o conjunto de atividades ou processos responsáveis por garantir o acesso ininterrupto às informações e ao patrimônio cultural em formatos digitais a longo-prazo. A preservação digital garante que a informação digital continue a ser acessível, interpretável, autêntica e ainda que seja executada em uma plataforma tecnológica distinta da que a gerou. (ARELLANO, 2008).

A preservação digital se torna um desafio para o profissional da informação, mas esse desafio é superado quando o profissional detém conhecimento dos recursos tecnológicos necessários para a preservação digital, conforme a carta de preservação do patrimônio arquivístico digital. (CONARQ, 2004, p. 2)

A preservação digital de documentos arquivísticos tem por objetivo garantir a autenticidade e integridade da informação, enquanto o acesso depende dos documentos estarem em condições de serem utilizados e compreendidos. O desafio da preservação de documentos arquivísticos digitais está em garantir o acesso contínuo a seus conteúdos e funcionalidades, por meio de recursos tecnológicos disponíveis a época em que ocorre a sua utilização.

A resolução nº 20 de 16 de Julho (CONARQ, 2004), também diz que, para preservar os documentos digitais são necessários requisitos que auxiliem a gestão e o prolongamento dos objetos digitais, vejamos a seguir.

A gestão arquivística de documentos digitais deverá prever a implantação de um sistema eletrônico de gestão arquivística de documentos, que adotará requisitos funcionais, requisitos não funcionais e metadados estabelecidos pelo Conselho Nacional de Arquivo, que visam garantir a integridade e a acessibilidade de longo prazo dos documentos arquivísticos.

Com estas considerações sobre a preservação digital, iniciamos a próxima seção na qual iremos abordar a preservação digital partindo da sua gênese, o documento digital. Contextualizaremos as principais atividades mais eficazes para a preservação digital. Entre os mais variados assuntos sobre preservação digital, destacamos os suportes midiáticos, o sistema de informação e o Sistema de Informação para Arquivos Abertos – OAIS e veremos as estratégias tecnológicas que podem ser embarcadas e utilizadas para preservar o documento neste sistema.

### 3.1 DOCUMENTO DIGITAL

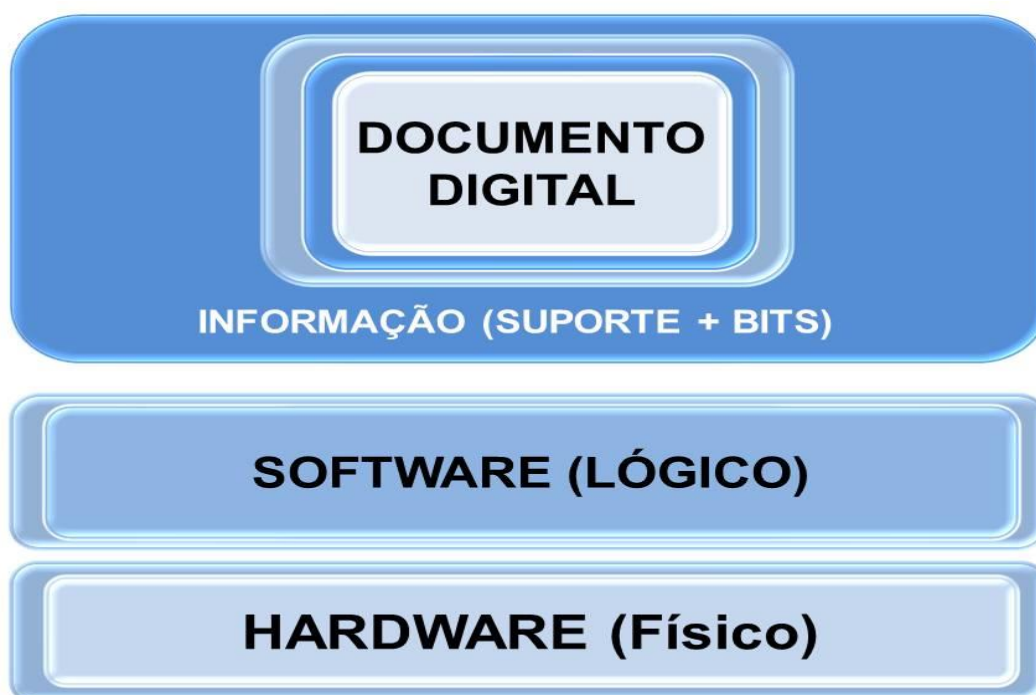
Antes de iniciarmos a discussão, vale salientar que quando nos referirmos à palavra “objeto digital” estamos fazendo alusão a qualquer Informação que possa ser gerada ou convertida para uma sequência de dígitos binários, armazenada e possível de ser recuperada sobre controle de um computador ou dispositivo eletrônico e que é tratada como uma unidade, do ponto de vista da informação, contendo os elementos essenciais à sua preservação e acesso continuado. (Le Coadic, 1996).

A primeira reflexão que vem em nossas mentes a respeito dos documentos digitais, resume-se em algo digitado em editor de texto como o Office Word da Microsoft ou o Write do Libre Office. Evidente que essa é a forma atual de produzir os documentos formais e informais por qualquer indivíduo que detenha os conhecimentos da escrita e da técnica de manipulação do *software*.

Entretanto, o que muitos não conseguem enxergar é que antes de um documento ser analógico, ele é produzido em um ambiente totalmente digital, tendo a sua realização a partir do momento da sua impressão. Uma vez impresso, a própria holística tecnológica se encarrega de possibilitar a guarda desse documento em um espaço não visualizado pelo seu criador, mas é compreensível que a sua existência esteja no referido espaço para ser utilizado posteriormente. (INNARELLI, 2007)

Essas questões paradigmáticas entre o universo real e virtual, serão esclarecidas nos capítulos posteriores deste trabalho. O que queremos realçar é a compreensão do documento digital que vem se desenvolvendo e evoluindo desde os meados dos anos 80, data que marcou o início da era dos “Escritórios sem papeis” (*Paperless Office*). Para iniciarmos o estudo, segue abaixo a figura criada por Innarelli (2007), mas modificadas por nós para compreendermos melhor a composição do documento digital.

**Figura 1 – Estrutura por camadas para criação do documento digital.**



Fonte: Innarelli, adaptado pelos dados da pesquisa, 2013.

Segundo Innarelli (2007, p. 26-27), a construção de um documento digital dar-se por essas três camadas acima e sua produção pode ser por três formas: “por meio de sistemas informatizados através de dados contidos em sistemas gerenciadores de bancos de dados (SGBD), por processos de digitalização e/ou diretamente com uso de um *software* ou sistema específico”.

Com essas múltiplas formas de produção de documentos, podemos dizer que cada *software* gera seu documento específico, tratando de deixar os resquícios da

fonte geradora. Esses resquícios são metadados<sup>1</sup> de reconhecimento para a plataforma que produziu o documento identificá-lo como código compatível e também permitir o usuário encontrar o documento e ter acesso à informação contida no objeto digital. Podemos até pensar que os problemas de compatibilidade já estejam sanados, pelo fato das atualizações das plataformas, mas o mau uso ou má compreensão das ferramentas tecnológicas pelo homem ainda é o maior problema. (YATES; NETO, 2013)

Para melhor compreensão, descrevemos as três formas citadas por Innarelli (2007) para ampliarmos o nosso conhecimento sobre os documentos digitais.

Nos *softwares* ou sistemas específicos, cada documento digital recebe sua extensão de acordo com o *software* e a plataforma que disponibiliza o formato desejado, pois um *software* ou sistema específico pode gerar mais de uma extensão como é o caso do *Word* que pode gerar documentos com extensão: .doc, que é a sua extensão oficial, ou gerar .jpeg, .pdf, http e outras que são compatíveis com o *Word* e outros *softwares*. Isto é, podemos dizer que os documentos digitais podem ser potencializados para outros formatos, sem que percam seu caráter intrínseco do documento, a informação.

Na digitalização, o documento analógico é reproduzido para o ambiente digital por meio de um *software* conectado a um *hardware*, especificamente, a um Scanner que fará a captura do documento que receberá uma extensão, ou seja, o seu formato para ser reconhecido na plataforma de um software compatível. Esse processo, poderíamos chamá-lo de captura de imagem, pois é, relativamente, o mesmo realizado pelas câmeras fotográficas e filmadoras digitais. Esta é uma simples analogia sobre o que é dito nas recomendações para digitalização de documentos dos arquivos permanente. (CONARQ, 2010, p. 6)

De acordo com a natureza do documento arquivístico original, diversos dispositivos tecnológicos (*hardware*) e programas de computadores (*software*) serão utilizados para converter em dados binários o documento original para diferentes formatos digitais. No entanto, o produto dessa conversão não será igual ao original e não substitui o original que deve ser preservado.

---

<sup>1</sup> São Dados estruturados e codificados, que descrevem e permitem acessar, gerenciar, compreender, e/ou preservar outros dados ao longo do tempo. Dicionário Brasileiro de Terminologia Arquivística. Rio de Janeiro: Arquivo Nacional, 2005, 232 p.



Nos sistemas de gerenciamentos de banco de dados, a construção do documento ocorre pela quantidade de dados fornecidos pelo usuário ao sistema de informação conectado com o SGBD. Para que uma informação seja gerenciada por um SGBD, o usuário terá que conectá-lo a um *software* ou sistema de informação que irá fazer a leitura dos dados armazenados no SGBD e apresentá-lo de acordo com a configuração que o administrador predefiniu no sistema de informação. Este sistema de informação ou *software* irá fazer todo alinhamento e padronização dos dados fornecidos, formando o corpo documental. (INNARELLI, 2007)

Estes são os três meios de produção de documentos digitais. Isto nos faz perceber que todos os *softwares* geram e/ou reproduzem arquivos ou documentos direta e/ou indiretamente. Embora cada documento receba sua extensão, ou seja, seu formato no meio digital, sua essência continua sendo a mesma codificação.

Entretanto, diante de uma complexa explicação, como poderíamos resumir o que viria a ser um documento digital? Um documento digital é uma sequência de dígitos binários<sup>2</sup> transcritos em uma tecnologia com memória que interpreta os números em dados, constituindo um corpo de informação a partir dos dados reunidos. Este corpo é gerado através de um *software* que se encarrega de disponibilizar o seu formato ou extensão, obtendo a sua mensuração quantificada por bytes.<sup>3</sup> (CONARQ, 2010)

Essa definição esclarece o núcleo gerador do documento digital, mas nos faz questionar sobre o seu armazenamento ou suporte físico, pois uma vez gerados, qual seria o melhor suporte apropriado para armazenar os documentos digitais? Essa e outras questões latentes serão demonstradas a seguir.

---

<sup>2</sup> É um sistema de numeração posicional em que todas as quantidades se representam com base em 0 ou 1.

<sup>3</sup> É unidade de medida para especificar o tamanho do objeto digital ou capacidade de armazenamento de um determinado dispositivo ou objeto digital.

## 3.2 SUPORTES E DISPOSITIVOS DE ARMAZENAMENTO

Os dispositivos de armazenamento ou suporte de dados sempre foram uma das preocupações, tanto dos especialistas, quanto dos usuários em geral. Esta preocupação é ocasionada pela busca da mídia perfeita, pois todos os dispositivos de armazenamento são frágeis demais, passíveis a extravio e falhas sinistras ou entram rapidamente na obsolescência. (FERREIRA, 2006)

A obsolescência é ocasionada pelo desenvolvimento tecnológico e ao mesmo tempo em que os *softwares* e *hardwares* foram avançando na tecnologia, os dispositivos também foram evoluindo na sua capacidade de armazenamento, durabilidade, transporte e usabilidade.

A necessidade de manter os dados mais próximos do seu produtor e permitir que esses dados sejam moveis fez que emergisse, do mundo tecnológico, os dispositivos e mídias de armazenamento. Um dispositivo de armazenamento pode guardar e processar os dados. Já a mídia, apenas armazena os dados, necessitando de um dispositivo para realizar a recuperação dos dados armazenados.

Enveredamos, nos capítulos e itens que se seguem por esclarecer alguns dos tipos de suportes e mídias de armazenamento que foram e são mais utilizados para preservar a informação e documentos digitais.

### 3.2.1 Tipos de suporte, dispositivos de armazenamento e mídias

Segundo Innarelli (2007, p. 30), os dispositivos de armazenamento se dividem em três tipos: “Por meio Magnéticos, Ópticos e Eletrônicos”. Os referentes suportes vêm evoluindo rapidamente, isto é, adquirem mais espaço para armazenamento de objetos digitais e esse é o fator que atrai o consumidor para obter este suporte e abarrotá-lo com dados, dando-lhe uma falsa sensação de preservação. Para

compreender mais sobre os suportes de armazenamento, veremos abaixo cada um deles.

### 3.2.1.1 Magnéticos

Os dispositivos magnéticos foram os primeiros a serem utilizados e explorados para o armazenamento de dados. As fitas magnéticas foram as primeiras mídias de armazenamento a registrar dados. Logo depois vieram os discos rígidos, mais conhecidos como *HD's*, estes se consolidaram como um dispositivo muito eficiente, tendo a sua capacidade de armazenamento extremamente alta. Esses dispositivos continuam renovando-se e perdurando até hoje. (INNARELLI, 2007)

Para a locomoção dos dados de um ponto a outro ou armazenar uma quantidade limitada de dados, foi criado o disco flexível que permitia armazenar até 1,44 MB de dados para o compartilhamento e/ou cópia de segurança. Os primeiros disquetes obtinham uma capacidade muito pequena de armazenamento, isto é, em relação com a sua época, no que toca a diferenciação do poder computacional de hoje, este dispositivo armazenava vários documentos. (PAIVA, 2008)

Outros suportes pouco difundidos, mas muito utilizados por instituições de grande porte são as fitas: "DAT, AIT e DDS". Estas fitas são utilizadas apenas para *Backup* e sua capacidade de armazenamento ultrapassa os terabytes<sup>4</sup>. Uma desvantagem no uso deste suporte é o alto custo financeiro das fitas e da manutenção do equipamento. (INNARELLI, 2007).

A fragilidade dos suportes magnéticos sempre foi ocasionada pela aproximação a um campo eletromagnético ou a umidade. Mas isto só é aplicado para os discos flexíveis e para alguns discos rígidos mais antigos, pois o contato desses dispositivos com um campo magnético provocava a perda dos dados e impossibilitava a sua recuperação, restando apenas formatar o dispositivo para utilizá-lo novamente.

---

<sup>4</sup> Unidade de informação ou memória que corresponde a 1<sup>12</sup>.

Quanto à umidade, aplica-se aos discos flexíveis, pois a resistência da película magnética era baixa e com isso a umidade facilitava o surgimento de mofo na película, danificando-a. Hoje os discos rígidos não são vulneráveis a campos magnéticos de baixa intensidade, porém uma grande carga de pulso eletromagnético não faria apenas a perda dos dados, mas também danificaria os circuitos dos discos rígidos. (INNARELLI, 2007)

### 3.2.1.2 Ópticos

Nos anos 90, o CD chegou ao mercado com o intuito de ser utilizado apenas para o armazenamento de áudio. Entretanto, esta nova tecnologia foi considerada, relativamente à produção de outros dispositivos, de custo baixo. Com isso, os CDs passaram a ser utilizados para o armazenamento de dados de multimídias, passando então a surgir o CD-ROM. Sua superioridade ao disquete é percebida pela capacidade de armazenamento que atinge os 650 MB a 700 MB e a velocidade de acesso aos dados. Portanto, o CD-ROM foi o sucessor imediato do disco flexível. (INNARELLI, 2007)

Há dois tipos de CD's-ROM: O CD-R (*read*), que permite a leitura e apenas uma gravação e o CD-RW (*read/write*), que permite a leitura e regravações. (RONDINELLI, 2002)

Outro dispositivo óptico foi o DVD-ROM, que “trata de um dispositivo similar ao CD-ROM”. Obtendo, obviamente, a diferença na sua capacidade de armazenamento que pode atingir 15.9 GB, isto é, utilizando a mídia de dupla camada. “A mídia de camada única é a mais utilizada, tendo sua capacidade de 4.7 GB”. (MORLEY; PARKER, 2010, p.107).

Assim como o CD-ROM, o DVD-ROM possui dois tipos: “O DVD-R, que possui as mesmas características do CD-R”, leitura e única gravação. Quanto ao DVD-RW, também utiliza o mesmo princípio do CD-RW, leitura e permite regravações. (MORLEY; PARKER, 2010, p. 107).

Embora essas mídias sejam de custo benefício para qualquer empresa ou instituição, sua vulnerabilidade para extravio e, conseqüentemente, para perda de dados é grande. Qualquer ranhura na sua película de registro danificará a mídia e, possivelmente, o dado que se localiza na área atingida não será encontrado ou será identificado como um dado corrompido.

A umidade é outro fator vulnerável, pois há películas que não oferecem resistência a umidades. Isto pode ocasionar mofo, que ao longo do tempo desgasta a película da mídia danificando-a. (INNARELLI, 2007).

### 3.2.1.3 Eletrônicos

Dentre outros nomes, os dispositivos eletrônicos são mais conhecidos como: Pen-drive, Cartão SD, Cartão de Memória, Mini ou micro SD, entre outros. Embora esses dispositivos tenham o seu tamanho reduzido, sua capacidade de armazenamento é superior à capacidade de armazenamento das mídias ópticas. No entanto, estes dispositivos não foram criados para preservar os dados por uma longa data, mas sim para transportar os dados de um determinado ponto a outro. (VERMAAT, 2013)

Os dispositivos eletrônicos de armazenamento apresentam-se, até o momento, como os mais seguros, dinâmicos e versáteis. Eles ajudaram a impulsionar o desenvolvimento de dispositivos cada vez menores e isto tudo se deve a Nanotecnologia. (VERMAAT, 2013)

Conhecidos também como SSDs (Dispositivo de estado sólido), esses dispositivos não possuem partes rotatórias como os dispositivos magnéticos e ópticos, apenas circuitos eletrônicos que não necessitam se movimentar para que ocorra a sua leitura ou gravação de dados, isto é, tudo funciona a partir da eletricidade. (O'BRIAN; MARACAS, 2010)

Quanto à sua fragilidade, isto nos força dizer que depende, totalmente, do manuseio do usuário. Se tratando de dispositivos eletrônicos, o usuário deve manuseá-los de modo coerente para que não ocorra a perda dos dados. Outro ponto

relevante é o uso desses dispositivos em computadores de terceiros, dos quais não se conhecem as procedências do uso seguro. A utilização desses dispositivos sem as medidas de segurança para proteção dos dados poderá incidir em falhas, perdas ou na formatação dos dados do dispositivo.

### 3.3 SISTEMAS E ESTRATÉGIAS DE PRESERVAÇÃO DIGITAL

No momento em que decidimos preservar um documento ou dados informacionais, devemos ter o discernimento para a escolha da melhor mídia ou dispositivo de armazenamento, pois a escolha de um dispositivo frágil ou que se torne obsoleto rapidamente pode ocasionar em mais gastos futuros na atualização ou migração das informações contidas nesses dispositivos ultrapassados para os novos dispositivos.

A Arquivologia, junto com outras Ciências que tornam a grande área da Ciência da Informação mais eficiente, trabalha para o desenvolvimento de métodos e sistemas que dinamizam a gestão e a preservação dos documentos digitais e dados informatizados. O impulso para o desenvolvimento desses sistemas se deve a publicação da carta de preservação digital da UNESCO, em 2003. Vejamos um breve comentário de Pinto (2005, p. 54), sobre a carta da UNESCO em artigo sobre a preservação digital.

A Carta da UNESCO é muito clara no que concerne à caracterização de uma Sociedade da Informação na qual é inquestionável que cidadãos e organizações, sejam elas públicas ou privadas, produzem e transformam, em ritmo crescente, informação exclusivamente em formato digital (textos, bases de dados, folhas de cálculo, imagens fixas ou em movimento, gravações sonoras, material gráfico, páginas web, programas informáticos, realidade virtual, entre muitas outras possibilidades), sendo diversos os meios e os formatos usados para o seu registo e complexas as ligações entre «objectos» num mesmo documento.

No mesmo ano de 2003, a Organização Internacional para Padronização (ISO) em conjunto com o Comitê Consultivo de Sistemas de Dados Espaciais (CCSDS), expôs a ISO 14721:2003<sup>5</sup>, que normaliza os dados espaciais e sistemas

---

<sup>5</sup> Norma específica de padronização um Sistema de Informação de Arquivo Aberto.

de transferência de informação - Sistema de Informação Aberto de Arquivo. Este modelo de referência é mais conhecido como OAIS (Open Archival Information System). Outro estudo aprofundado sobre a preservação digital é a InterPARES - Pesquisa Internacional sobre Registros Permanentes Autênticos em Sistemas Eletrônicos, com sede no Canadá e coordenado pela professora Dr<sup>a</sup>. Luciana Duranti.

Tendo como norma e referência para o desenvolvimento de sistemas que visam à preservação digital, ressaltaremos as principais características do OAIS para preservação do objeto digital.

### **3.3.1 Sistemas de Informação de Arquivos Abertos - OAIS**

O OAIS, segundo Ferreira (2006, p. 27), “é um modelo conceitual que visa identificar os componentes funcionais que deverão fazer parte de um sistema de informação dedicado à preservação digital”. Esse modelo foi desenvolvido pelo CCSDS em conjunto com a ISO, devido a crescente produção de documentos digitais e pela preocupação de como preservar essa exorbitante quantidade de documentação gerada todos os dias.

Innarelli (2007, p. 37), destaca que “a aplicação do OAIS em arquivos consiste na organização de pessoas e sistemas”, tendo como responsabilidade a preservação e o acesso da informação à comunidade interessada e, como foco principal a informação digital. Além de preservar a informação digital e torná-la acessível a uma comunidade, o OAIS também realça duas funções principais de um Arquivo: A primeira é preservar a informação, assegurando a sua persistência em longo prazo. A Segunda é tornar a informação arquivada acessível de forma inteligível e de acordo com as necessidades dos consumidores ou usuários futuros.

O objetivo do OAIS é aumentar o grau de consciência e compreensão dos conceitos relevantes para o arquivamento de documentos digitais, especialmente entre instituições não arquivísticas. (ARELLANO, 2008). A implantação dos conceitos de arquivamento no mercado provocaria a participação das instituições

que não estão habituadas em gerir os documentos, a preservar sua documentação. Para tanto, essa instituição teria que contratar o profissional que conheça os princípios de arquivamento e preservação de documentos para realizar a gestão documental de forma adequada. (SCHELLENBERG, 2004)

A expansão deste modelo referencial e a contratação do profissional para manipulá-lo, ampliariam a consciência e a compreensão sobre os conceitos relevantes da Arquivística que são pertinentes para a preservação e o acesso a longo-prazo. Isto aplica-se não apenas aos documentos digitais, mas também para os documentos analógicos.

Portanto, o OAIIS pode ser aplicado em qualquer arquivo público ou privado, porém a sua especificidade essencial é voltada para organizações que contêm a responsabilidade de tornar a informação disponível por um longo prazo.

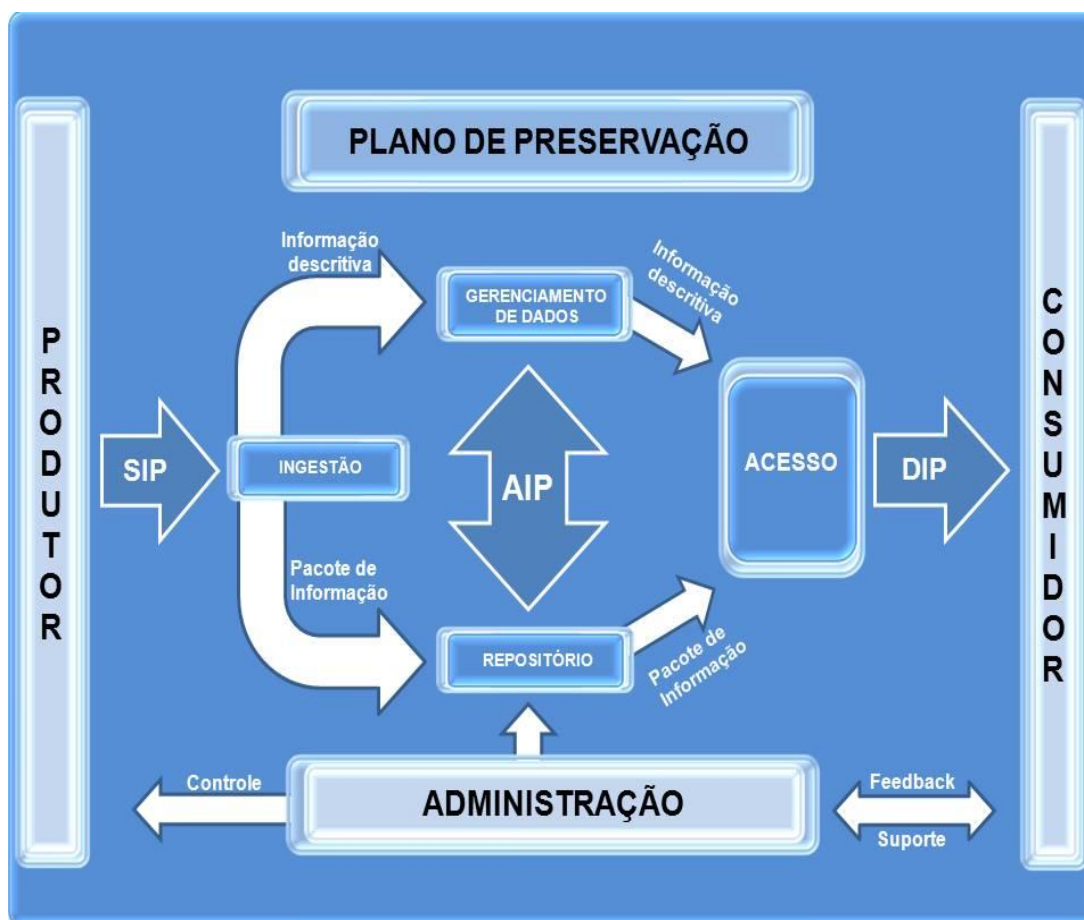
O inerente sistema pode ser facilmente implementado nas instituições, pois o OAIIS é um modelo referencial básico para a construção de repositórios e/ou outros sistemas de armazenamentos que possam ser administrados para controlar o fluxo, não apenas da documentação produzida internamente, mas também recebida de instituições externas ou de entidades que sejam vinculadas. Vejamos agora como Arellano (2008, p. 92) descreve a utilidade do OAIIS para as instituições.

Com a norma OAIIS, as instituições podem configurar sistemas responsáveis pela preservação da informação de longo prazo e torná-la disponível para uma comunidade de usuários. A definição dos componentes necessários para a criação desses sistemas que o modelo proporciona serve para suportar uma gama de serviços de preservação. Essa descrição está relacionada com as atividades de submissão, organização e gerenciamento e acesso contínuo de um repositório em funcionamento.

Como afirma Arellano (2008), o modelo é configurável e dirigido para instituições que detenham a responsabilidade de preservar a informação por um longo prazo e que os riscos da perda dos dados sejam mínimos. Abaixo, segue uma figura que exemplifica, claramente, o alusivo modelo OAIIS.



Figura 2 – Componentes gerando o modelo funcional do OAIS.



Fonte: Ferreira (2006, p. 28)

Como ilustra a figura acima, o referido modelo envereda-se em uma teoria sistêmica, onde podemos enxergar uma entrada, processo e saída. Arquivisticamente, poderíamos interpretá-la como: Produção, Arquivamento, gerenciamento e Consulta. É importante ressaltar que, apenas, o Produtor, Administração e Consumidor são entidades externas do sistema, ou seja, são os gerenciadores do sistema. Entretanto, iremos explicar o modelo de forma que venhamos a compreender todo o processo sistêmico do modelo OAIS.

O **produtor**, como já citado, deve ser visto como entidade externa de todo o sistema. Sua responsabilidade é pela submissão de documetação ou informação digital para a alimentação do sistema, formando assim um **Pacote de Informação de Submissão** (SIP), que vem do inglês: *Submission Information Package*.

Na passagem do processo de submissão, titulado neste exemplo como **Ingestão**, é realizado o processo de verificação da integridade da informação. Após a verificação, o documento e/ou informação é validado e distribuído em duas vias: a primeira como Informação Descritiva que é enviada para o Gerenciamento de Dados, e a segunda como **Pacote de Informação de Arquivo - *Archival Information Package*** (AIP), que é encaminhado para o Armazenamento, ou seja, todo o pacote é armazenado e preservado pelo repositório.

A **Informação Descritiva**, conhecida mais como “meta-informação”, segundo Ferreira (2009), pode ser descrita pelo produtor ou criada durante o processo de recolhimento. Essa meta-informação é enviada para o gerenciamento de dados onde será armazenada. Depois de armazenada, essa meta-informação, organizada pelo **Gerenciamento de Dados**, fará a conexão com o repositório, onde se encontra todo AIP. Quando o sistema for requisitado pelo consumidor, o índice de informação a ser apresentado será mais eficiente pelo fato das metas-informações definidas na indexação. Portanto, a informação descritiva estabelece a relação maior com o pacote de informação de arquivo.

O repositório é o protagonista do alusivo sistema, pois além de preservar a informação digital, ele é responsável por gerenciar a camada de armazenamento, satisfazer os pedidos de acesso por parte do consumidor e garantir que as informações não sejam alteradas, tanto pelos suportes onde estão sendo armazenadas as informações, quanto pelas partes de nível lógico do sistema. O repositório é o responsável por oferecer a função de preservação e recuperação de dados nos casos de desastre. O repositório também pode ser aplicado em um sistema de RAID, que realiza cópias de segurança de todos seus dados. (EMC, 2009). O OAIS é de grande importância para a preservação, pois este modelo conceitual é responsável por fomentar muitos sistemas e estratégias, entre tantos destaca-se o repositório digital.

[...] Os repositórios digitais funcionará não apenas como garantias da viabilidade em longo prazo de materiais sob sua custódia, mas também como portais de acesso. Cumprir essa dupla missão exige que os processos de preservação operem de forma transparente ao lado de serviços de acesso. [...] a preservação não deve impedir o acesso ou reduzir a possibilidade de compartilhamento de informações. [...] Sistemas de gestão de conteúdo digital devem encontrar maneiras de integrar as ferramentas de preservação e serviços em seus ambientes. (LAVOIE; DEMPSEY, 2004, tradução nossa)

O acesso é responsável pela conexão entre consumidor, que é uma entidade externa, é o sistema. É de competência do acesso: Processar os pedidos realizados ao OAIIS e retorná-los como **Pacote de Informação de Disseminação – Dissemination Information Package (DIP)**.

No **Plano de Preservação** se estabelece as políticas de preservação, determinando a classificação das informações para o Acesso. Além de estabelecer as políticas de preservação, o Plano de preservação tem como responsabilidade desenvolver e aplicar estratégias de preservação da informação, permanecer atualizado ao desenvolvimento de novas tecnologias e elaborar relatório periódicos sobre a análise de riscos.

Sua função inclui avaliar os conteúdos do documento e recomendar, periodicamente, migrações da informação arquivada, desenvolver recomendações para padrões políticos do arquivo e monitorar as mudanças no ambiente tecnológico, nas demandas de serviço e na base de conhecimento da comunidade alvo.

A entidade Planejamento de Preservação esboça, ainda, modelos de pacotes de informação e realiza adaptações do SIP e AIP para submissões específicas. É responsável, também, pelo desenvolvimento de planos detalhados de migração, protótipos de software, e planos de teste para permitir a implementação das metas de migração da entidade Administração do Sistema.

Por último, a **Administração** é uma entidade híbrida, pois trabalha interna e externamente. Responsável por regular as tarefas do OAIIS e por coordenar as atividades das outras seis entidades internas: Ingestão, Gerenciamento de Dados, Repositório, Acesso, Plano de Preservação e Administração. É da competência da administração interagir com as três entidades externas, isto é, com os produtores, para formatação das submissões, com os consumidores, oferecendo suporte e mantendo um *feedback* para saber o grau de satisfação sobre o uso do sistema, e também com o administrador que além de fomentar as políticas de preservação, fica encarregado de supervisionar as operações dos sistemas de arquivo e acesso e conferir o desempenho do sistema.

Em suma, o modelo OAIIS abrange três entidades externas e seis entidades funcionais internas que em conjunto compõem um sistema de arquivamento e de

preservação da informação em longo prazo. Tornando a informação acessível aos consumidores, disseminando-a para a comunidade alvo ou toda sociedade. Todo sistema que realize a preservação digital terá que seguir este modelo conceitual de preservação definido por padrões internacionais.

### **3.3.2 Estratégias de preservação**

Para um excelente funcionamento do OAIS ou de qualquer sistema de preservação de informação digital é imprescindível à formulação de estratégias. As estratégias são fomentadas, geralmente, pelas comissões técnicas ou pelos administradores e aplicadas neste ou em outros sistemas para que as informações sejam sustentadas, oferecendo o acesso para os futuros consumidores ou usuários.

Para tanto, a escolha da estratégia de preservação deve ser baseada em características como: custo, qualidade, formato da informação, ambiente computacional, disponibilidade, riscos e segurança. Estes requisitos são, relativamente, obrigatórios para os sistemas que dedicam a esta prestação de serviço. Essa relatividade é ajustada de acordo com o grau de relevância da informação a ser preservada. (FERREIRA, 2006)

Ferreira (2009, p. 26), propõe uma definição para estratégia de preservação digital como sendo “toda abordagem técnica que garante o acesso continuado à informação existente em formatos digitais”. Todo plano estratégico de preservação digital, se tratando de um modo técnico, requer um conhecimento especializado para realizar um diagnóstico daquilo que será preservado. Isto é, todo acervo documental e fluxo informacional, antes que seja preservado, terá que estar de acordo com as políticas estabelecidas pelas entidades administrativas e depois avaliadas pela comissão técnica de preservação.

As técnicas de preservação foram evoluindo de acordo com a tecnologia e com o seu formato de produção. Embora existam vários formatos e, conseqüentemente, varias técnicas de preservação, a informação digital traz consigo a sua essência de sistema binário de codificação que é transformada em informação

digital e para preservar este tipo de informação destacamos três formas que suprem a necessidade de preservá-la: **Refrescamento, Migração de Dados e Emulação**. Essas três técnicas foram destacadas pelo fato da sua facilidade para compreensão e aplicação na produção de documentos que sejam reservados para a preservação. (FERREIRA, 2009). Esses três métodos acima

A eficácia de um documento arquivístico depende da qualidade e do rigor dos procedimentos de produção e manutenção realizados pelas organizações produtoras de documentos. Entretanto, como a informação em formato digital é extremamente suscetível à degradação física e à obsolescência tecnológica – de hardware, software e formatos –, essas novas facilidades trazem consequências e desafios importantes para assegurar sua integridade e acessibilidade. (CONARQ, 2004, p. 02)

Assim como a carta de preservação, a declaração da UNESCO em Vancouver (2012, p. 2-3) trouxe mais relevância para a memória do mundo na era digital. Este evento reuniu mais de 500 participantes de 110 países que concordaram em várias recomendações das quais destaca-se:

- a. Tomar um papel ativo em influenciar publicamente para tornar os cenários e as práticas de preservação digital uma realidade, através da promoção da administração de objetos digitais e de sua preservação em todas as formas apropriadas, e inclusive trabalhar com outras agências das Nações Unidas, fundos e programas; b. apoiar o trabalho da comunidade internacional de arquivo, biblioteca e museu para assegurar um cenário legal internacional para os direitos autorais, assim como para exceções e limitações, de forma a assegurar a preservação e o acesso de patrimônio cultural em formato digital, e aquisição e acesso a este patrimônio de uma maneira culturalmente apropriada; c. colaborar com associações profissionais internacionais e com outros organismos internacionais para desenvolver um currículo acadêmico voltado para a digitalização e para a preservação digital, assim como implementar programas de treinamento e abordagens educacionais globais que promovam as capacidades do pessoal dos arquivos, bibliotecas e museus para administrar e preservar a informação digital.

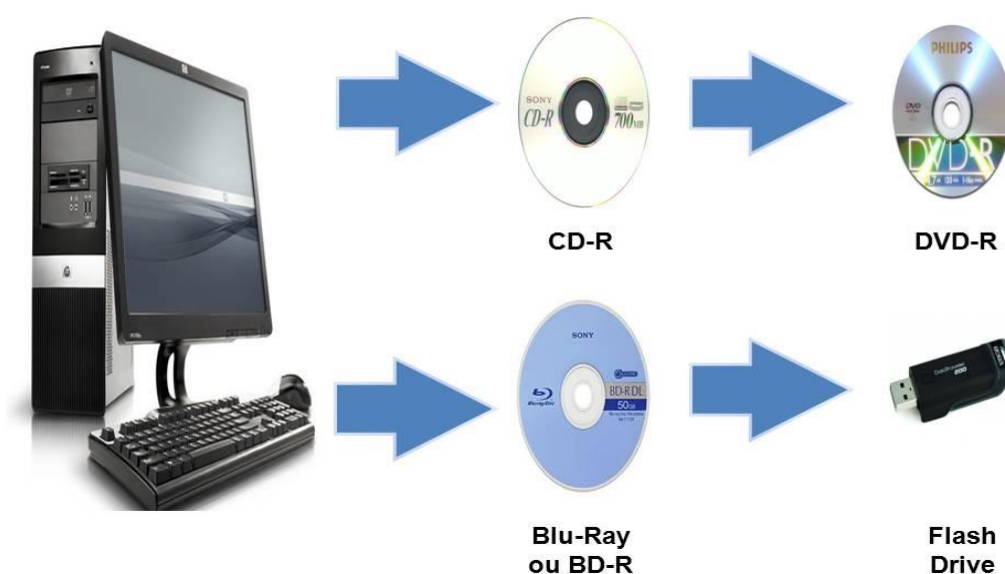
Sabemos que o e-ARQ Brasil (2006, p. 102), realça a “preservação da tecnologia e conservação dos dados”, mas entendemos que para preservar a informação ou documento digital necessitará apenas das três alusivas técnicas que serão descritas com mais detalhes a seguir.

### 3.3.2.1 Refrescamento

O refrescamento incide na transferência cíclica da informação digital registrada em um suporte ou dispositivo de armazenamento digital (CD-ROM, DVD, HD) para outro suporte mais atual, de forma a preservar a informação contra a obsolescência e/ou a deterioração física dos suportes. A integridade do dispositivo de armazenamento é o que garantirá o acesso à informação, pois qualquer desgaste no suporte poderá afetar as informações nele contida.

Outro ponto para a utilização do refrescamento é a não fabricação dos periféricos para leitura dos dispositivos, pois não havendo periféricos para a leitura, logo a informação contida não poderá ser acessada. Isto é também visto como: obsolescência de *hardware*. (FERREIRA, 2006)

**Figura 3 – Modelo de Refrescamento.**



**Fonte: Dados da pesquisa, 2013.**

Esta técnica de transferência possibilita que a informação seja preservada em um período de curto para médio prazo, pois a volatilidade dos suportes faz com que a transferência seja periódica para que o suporte não entre na zona de

obsolescência. Esta cíclica transferência acarreta em um gasto demasiado em dispositivos de armazenamento, pois a capacidade desses dispositivos é limitada. O administrador que gerir o sistema de preservação deve estar em alerta no que diz respeito à periodicidade de transferência, as novas tecnologias, e novos suportes para que garanta o sucesso da técnica.

Percebemos que o refrescamento é de total competência humana. Por isso o investimento em recursos humanos é mais vantajoso, pois a técnica de refrescamento não se enquadra em um modelo informatizado de preservação, mas sim como um suporte para o sucesso de qualquer estratégia de preservação. (FERREIRA, 2009)

### 3.3.2.2 Emulação

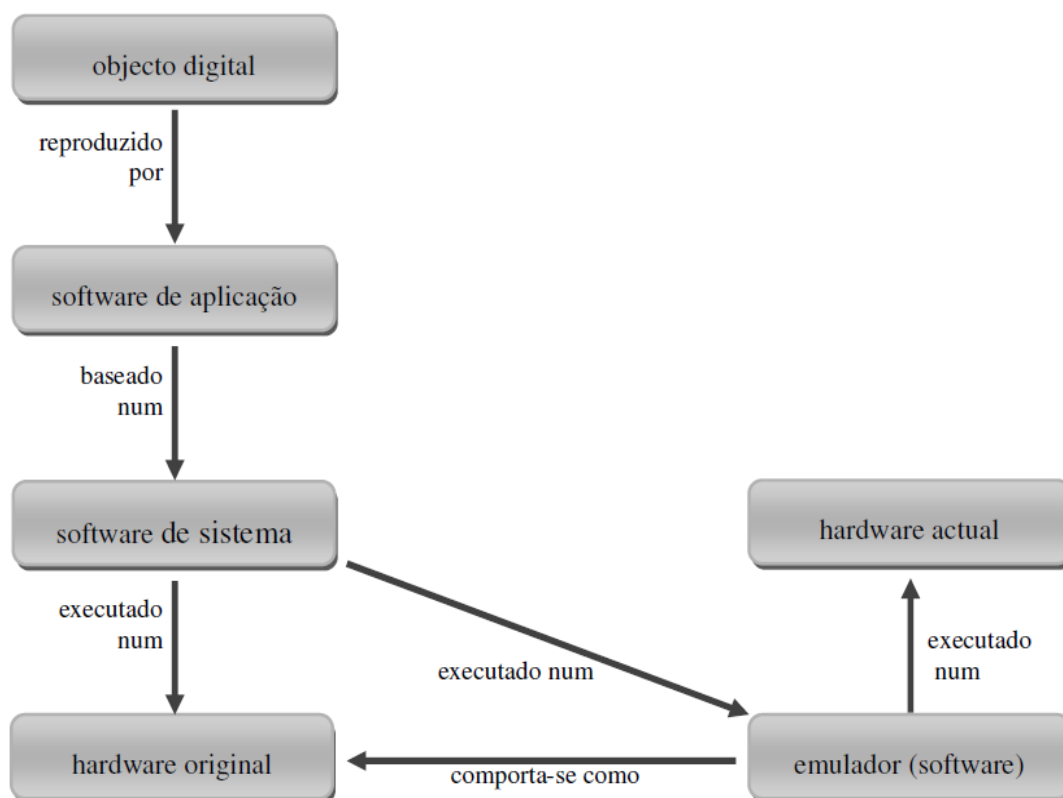
A emulação é uma estratégia de preservação da informação digital que surgiu com o intuito de preservar todo ambiente tecnológico que produziu o documento. A Emulação permite que os usuários possam ter o acesso aos documentos digitais sem necessidade de preservar ou manter em funcionamento o *hardware* e/ou *software* originais. (ARELLANO, 2004)

A Emulação recria o ambiente tecnológico original a partir da virtualização dos *hardwares* e/ou *softwares*, reproduzindo fielmente as funções para a execução e interpretação do documento ou informação digital pelo *software* de virtualização. Esta técnica preserva a dimensão lógica, ou seja, o *software* que produziu o documento. Quanto à parte física, ela é emulada recriando todo ambiente de *hardware* para executar o *software* preservado, possibilitando o acesso à informação ou documento preservado por outra técnica. (THOMAZ; SOARES, 2004)

O e-ARQ Brasil descreve a técnica da emulação de forma conceitual, portanto a emulação é:

A simulação de um determinado hardware ou software através de software. Permite que um computador moderno, possivelmente mais barato e de fácil manutenção, possa executar programas (software) antigos desenvolvidos originalmente para outra plataforma. Para evitar possíveis perdas de informação e funcionalidades, deve ser realizada com bastante rigor. A probabilidade de ocorrência de perdas de informações e funcionalidades aumenta à medida que são utilizadas diversas camadas de emulação, como resultado da aplicação dessa técnica repetidas vezes. (E-ARQ Brasil, 2006, p.102)

**Figura 4 - Recuperação de um objeto digital por meio da emulação.**



Fonte: FERREIRA, 2011.

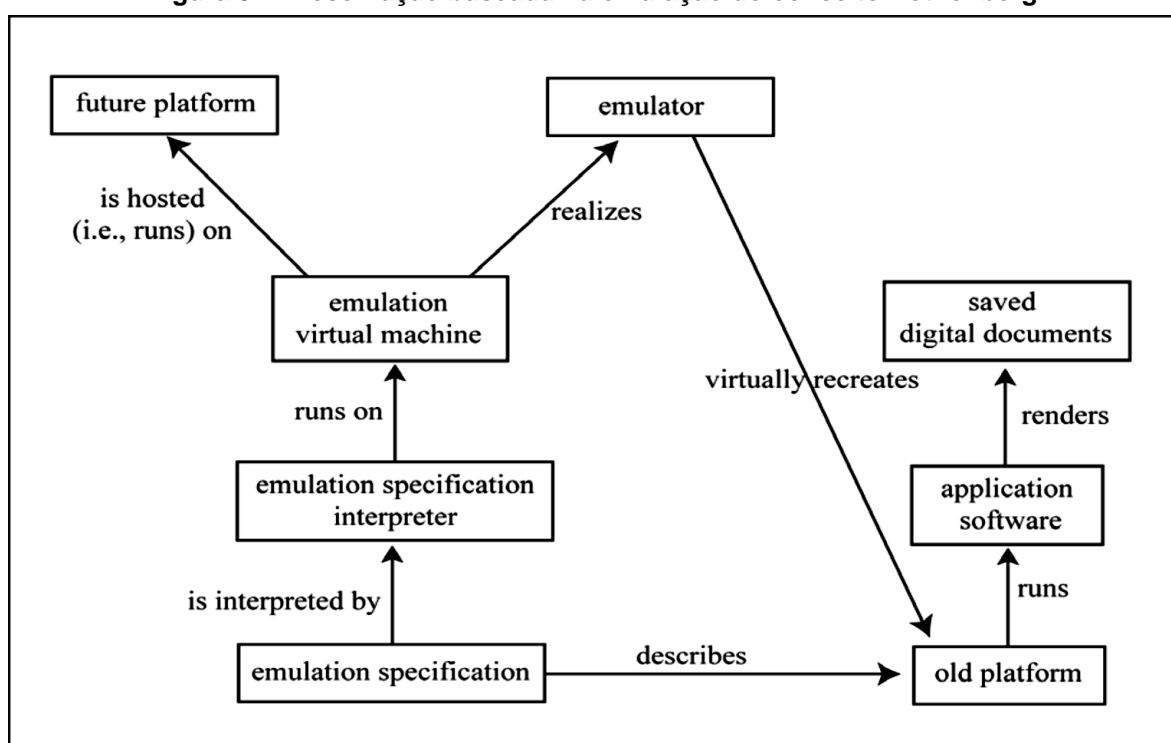
O principal objetivo da emulação é assegurar o formato original do ambiente informático e a funcionalidade do documento digital, permitindo a sua leitura no ambiente tecnológico, originalmente, emulado. Esta técnica sobressai de outras por excluir a possibilidade de obsolescência de *hardwares*, pois estes são emulados pelos *softwares* de virtualização.



Esta técnica é defendida, duramente, por Jeff Rothenberg, como melhor método para se preservar, indefinidamente, um documento digital, seus formatos, padrões e funcionalidades. Segundo Rothenberg (1999), a implementação dessa técnica abarca três métodos padrão para o sucesso da emulação:

- Desenvolver um pacote de informação com o código e técnicas para especificação de emuladores que serão executados em computadores desconhecidos e recriar o ambiente tecnológico do comportamento dos objetos digitais;
- Desenvolvimento de técnicas para a guarda dos metadados necessários para encontrar, acessar e recriar os objetos digitais;
- Desenvolvimento de técnicas de encapsulamento de documentos, seus metadados, *software* e especificações do emulador de forma a assegurar sua coesão e prevenir sua corrupção.

Figura 5 – Preservação baseada na emulação do conceito Rothenberg.



Fonte: LEE, et al. 2002, p. 95

Como podemos ver, a emulação é uma técnica composta por uma série de métodos que preserva não apenas a informação ou documento, mas todo seu ambiente tecnológico pode ser recriado a partir da virtualização para executar a leitura e/ou reprodução do documento ou informação digital preservada.

Embora esta técnica seja uma estratégia promissora, a emulação tem sido utilizada para recriar ambientes de jogos eletrônicos e para emular máquinas virtuais para a obtenção de múltiplos sistemas operacionais. (FERREIRA, 2006)

“Apesar de que existam emuladores para determinados sistemas obsoletos, a emulação para preservar objetos digitais por longo prazo ainda não foi suficientemente testada ou avaliada em termos de custo/preço final”. (THOMAZ; SOARES, 2004). Os emuladores podem auxiliar o profissional da informação não como uma ferramenta para preservação do objeto digital, mas para reconstruir todo ambiente que objeto foi criado e executá-lo nesta plataforma virtualizada.

### 3.3.2.3 Migração de dados

Antes de iniciarmos a análise, é de extrema importância ressaltar a diferença entre a migração de dados da área de arquivologia e a migração de dados da área da informática uma vez que o objeto tratado nas áreas é o dado digital.

A migração de dados, no campo Arquivístico, consiste na transferência periódica dos dados digitais de um determinado sistema de informação, sustentado por *hardware* e *software* específico, para outro sistema de informação com *hardwares* atualizados e *software* compatível, ou seja, transferir os dados de uma geração de tecnologia obsoleta para outra subsequente (FERREIRA, 2009, p. 26; ARELLANO, 2008, p. 63; HEDSTRON et al. 2003; LEE et al., 2002, p. 96; FERREIRA, 2011, p. 52).

Na Informática, a migração de dados é uma técnica utilizada frequentemente por analista de sistemas e administradores de banco de dados, geralmente quando um sistema ou *software* de banco de dados será substituído por outro, mesmo que não seja compatível com a estrutura antiga ou a que produziu o dado. Ou seja, nesta

técnica todos os dados são transferidos de um banco de dados para outro, mesmo que este novo banco de dados esteja *off-line*, *on-line* ou que a plataforma seja incompatível com o sistema de origem. (DATE, 2004, p.18)

Embora as duas áreas relatem de forma similar o processo de migração, existem diferenças que distanciam uma da outra, mas ambas são unidas pelo caráter intrínseco, o dado.

Na Arquivística existem alguns tipos de migração: migração para suportes analógicos, atualização de versões, conversão para formatos concorrentes, normalização, migração a-pedido e migração distribuída. (FERREIRA, 2009)

- **Migração para suportes analógicos:** Consiste em converter um conjunto de dados, objetos ou documentos digitais para um suporte não digital, isto é, analógico. O intuito desta migração é prolongar a longevidade do objeto. Um exemplo desses suportes analógicos, podemos citar o papel, microfilme e outros.
- **Atualização de versões:** Incide em converter um objeto digital de uma versão de *software* anterior para um mais atual. Esta migração é intensamente utilizada pelos analistas de sistema e administradores de bancos de dados, pois as atualizações dos *softwares* proporcionam dinamismo e novas funcionalidades para os usuários.
- **Conversão para formatos concorrentes:** infere na conversão do objeto digital de um determinado formato ou extensão de *software* para outro formato de *software* concorrente. Este tipo de migração previne contra uma possível descontinuação do *software*. Salienciamos que os *softwares* atuais disponibilizam varias opções de formato para gerar os objetos. Como exemplo, podemos citar o *software* Open Office Write da Apache que além de disponibilizar os formatos dos seus concorrentes, oferece o formato o ODF, que é um formato aberto e público que foi aprovado como norma ISO/IEC<sup>6</sup> em 2006, sendo o primeiro formato de documentos editáveis de escritório, como textos, planilhas, bases de dados, desenhos e apresentações.

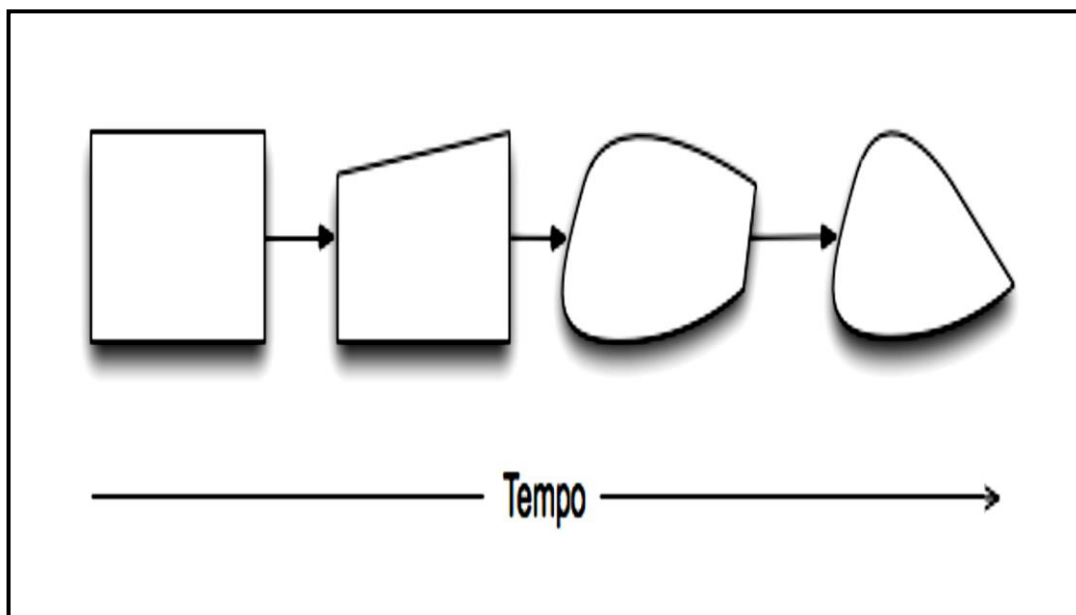
---

<sup>6</sup>Norma internacional que busca estabelecer uma estrutura comum para os processos de ciclo de vida e de desenvolvimento de softwares visando ajudar as organizações a compreenderem todos os componentes presentes na aquisição e fornecimento de software.

- **Normalização:** A normalização, segundo Ferreira (2009, p. 28) tem “como objetivo simplificar o processo de preservação através da redução do número de formatos distintos que se encontram num repositório de objetos digitais”. A normalização ou padronização é uma solução para a migração já esclarecida com a criação do ODF, pois este formato, atualmente, é executado em todos os *softwares* escritórios.

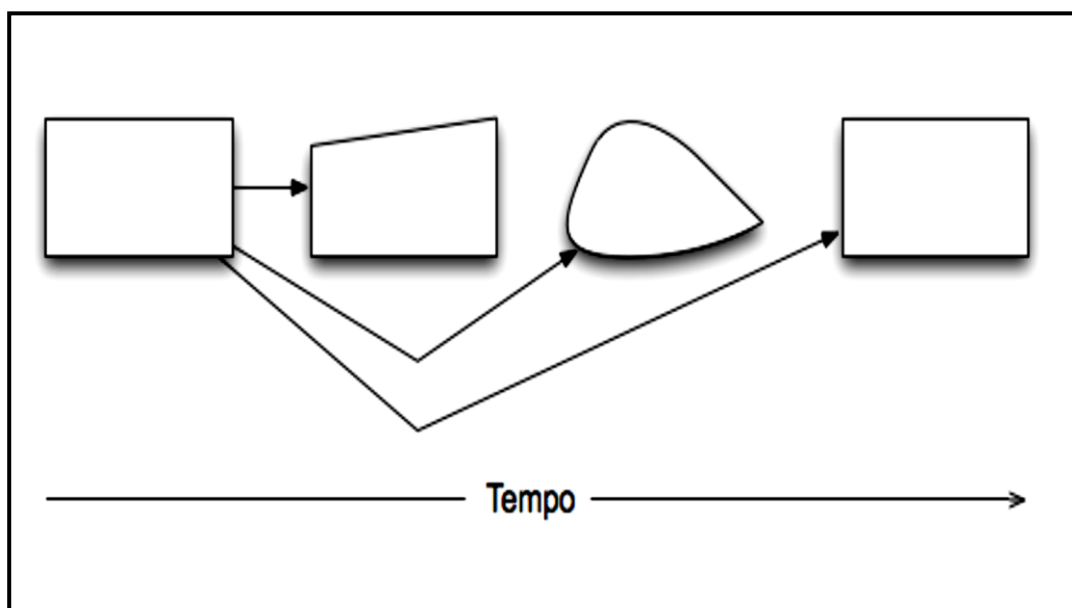
**Migração a-pedido:** Neste tipo de migração, aplica-se o processo de conversão sempre no objeto ou documento digital original e não no documento que já foi convertido. O documento original fica reservado e protegido no repositório para quando houver a necessidade de migrá-lo, utilizarem sempre o original. Segundo Ferreira (2009, p. 30), os documentos podem sofrer alterações, caso for convertido repetidas vezes para outros formatos.

**Figura 6 - Degradação do objeto digital ao longo de sucessivas migrações.**



Fonte: Ferreira, 2009, p.30

Figura 7 – Migração a-pedido.



Fonte: Ferreira, 2009, p.30

- **Migração Distribuída:** Para Ferreira (2009, p. 30), neste tipo de migração, “existe um conjunto de serviços de conversão que se encontram acessíveis através da rede ou da Internet e que poderão ser invocados remotamente recorrendo a um pequeno módulo de software ou aplicação-cliente”. Esse conjunto de serviços pode ser encontrado na Web disponível para download e para o usuário converter na sua própria máquina ou implementá-lo como *web service*<sup>7</sup> que possibilita o usuário realizar o *upload* do documento para o servidor, escolher o formato que será migrado e salvar o documento digital em novo formato. Como exemplo, podemos citar os serviços que estão na Web: docspal.com, convertstandard.com, cometdocs.com, entre outros.

Ainda que existam esses variados tipos, o objetivo da migração é garantir que os dados, documento ou informação sejam compatíveis com tecnologias futuras, a partir de um modelo tecnológico padronizado. Com essa técnica asseguraria o acesso aos documentos sem a necessidade de utilizar outras estratégias auxiliares, como por exemplo, emuladores.

<sup>7</sup> É uma solução utilizada na integração de sistemas e na comunicação entre diferentes aplicativos, permitindo as aplicações enviar e receber dados em formato XML, ou seja, é um método de comunicação entre dois dispositivos eletrônicos em conexão com a internet.

Entretanto, para que essa técnica obtenha sucesso, o administrador deve precaver-se com *backups* de todos os dados que serão migrados, pois neste tipo de estratégias existe uma grande probabilidade de algumas das propriedades que constituem os objetos digitais não serem corretamente transferidas para o formato de destino adotado. Outro fator é a transferência dos dados, pois um grande volume de dados, dependendo da infraestrutura de rede e largura de banda web, poderá acarretar em uma lentidão em toda rede. (FERREIRA, 2006)

Contudo, a transferência dos dados de um sistema, *software* ou plataforma obsoleta para um mais atualizado, provavelmente, resultaria em uma melhor interpretação dos dados, pois os sistemas mais sofisticados realizam reparos quando erros são encontrados nos dados. Este ponto pode ser interpretado como uma vantagem, mas também poderia resultar numa desvantagem oculta, pois ao reparar os dados ou ao permitir aplicações que possam realizar reparos, na ótica arquivística, afetaria diretamente a autenticidade e à sua integridade.

## 4. DO VIRTUAL A VIRTUALIZAÇÃO

Atualmente, vivemos em uma era que eleva a realidade para o plano da virtualidade, atingindo e renovando os padrões da sociedade. A cada dia o mundo observa novos feitos das novas tecnologias para dinamizar e tornar mais acessível os meios de informação, comunicação, empreendimentos, valores democráticos e constrói uma nova axiologia. O desenvolvimento tecnológico potencializa a utilização do espaço real para o virtual. Contudo, as ações praticadas no alusivo plano do espaço virtual são refletidas de modo direto no espaço real. (LÉVY, 1999)

Neste capítulo, exploraremos o significado de virtual e a potencialidade da virtualização, Uma vez que este último nos trará base suficiente para a compreensão da *Cloud Computing* ou Computação em Nuvem. Por fim, tocaremos superficialmente no campo filosófico dos fenômenos, porém aprofundando mais no universo tecnológico.

### 4.1 DESCREVENDO O VIRTUAL

Segundo Lévy (1996, p. 18-19), “a virtualização [...] consiste em uma passagem do atual ao virtual, em uma “elevação à potência” da entidade considerada”. Lévy nos assenta em uma problemática complexa e universal, mas nos traz a ampliação do universo real e cria os multiversos da virtualidade.

A Ciência da Informação explora, exaustivamente, o campo da virtualidade. Entretanto, os estudos da virtualidade são impulsionados pelo desenvolvimento tecnológico que, nos dias atuais, são extremamente voláteis.

As tecnologias desenvolvidas pelo universo da informática atingem diretamente a Ciência da Informação e suas subáreas como: Arquivologia, Biblioteconomia, Sistemas de Informação e entre outras que atêm a informação como objeto de estudo direta ou indiretamente.

A palavra virtual, como é definida por Lévy (1996, p. 15), “vem do latim medieval *virtualis*, derivado por sua vez de *virtus*, força, potência.”. Como podemos perceber a palavra “virtual” não é fruto da modernidade, mas podemos dizer que seu sentido é factível nos dias atuais e porvindouros. Se buscarmos seu significado em dicionários *online* da língua portuguesa, encontraremos algumas seguintes definições:

- Coisa existente apenas em potência ou como faculdade, não como realidade ou com efeito real. (Dicionário InFormal<sup>8</sup>)
- Coisa ou ação que poderá vir a ser, existir, acontecer ou praticar-se. (Dicionário InFormal)
- Equivalente a outro, e capaz de fazer às vezes desse outro; que constitui uma simulação de algo criada por meios eletrônicos. (Dicionário InFormal)
- Que existe potencialmente e não em ação. (web-dicionário<sup>9</sup>)
- Suscetível de se realizar ou de se exercer. (web-dicionário)
- Que é feito ou simulado através de meios eletrônicos. (web-dicionário)

As referentes definições, apesar de serem recursivas, nos proporcionam um significado claro da palavra virtual. Entretanto, algumas dessas definições entram em contradições, quando nos deparamos com a execução das atividades práticas da virtualidade, ou nas discussões e debates teóricos que aguçam as investigações deste fenômeno abordado.

Analisando as questões do plano virtual, uma falsa oposição é vista nitidamente entre o real e o virtual, mas não entre o virtual e o atual. O virtual deve ser considerado quanto algo que existe em potência, ou seja, o que pode vir a ser.

---

<sup>8</sup> <http://www.dicionarioinformal.com.br/>

<sup>9</sup> <http://webdicionario.com/>



Para compreendermos a falsa oposição, Lévy (1996, p. 16) relata de forma brilhante a diferença do plano virtual e do real:

O possível é exatamente como o real: só lhe falta a existência. A realização de um possível não é uma criação, no sentido pleno do termo, pois a criação implica também a produção inovadora de uma ideia ou de uma forma. A diferença entre o possível e real é, portanto, puramente lógica. Já o virtual não se opõe ao real, mas sim o atual. Contrariamente ao possível, estático e já constituído, o virtual é como o complexo problemático, o nó de tendências ou de forças que acompanha uma situação, um acontecimento, um objeto ou uma entidade qualquer, e que chama um processo de resolução: a atualização.

Seguindo a essencial vertente filosófica de Lévy (1996), para que ocorra a realidade de uma imagem ou de um objeto virtual, consideremos o plano virtual como ponto de partida, passando pela reificação e seguindo em rumo à sua potencialização, que conseqüentemente, este se realizará a partir da sua diminuição de potência, passando desta forma para o plano da realidade.

Esta é uma compreensão da passagem do virtual para o real. No entanto, para transformar ou transmutar um objeto ou um elemento do plano real para o virtual, realizaremos o processo inverso da atualização.

Para que suceda a virtualização, partimos deste objeto real em direção a sua objetivação que implica em sua atualização. Neste ponto encontraremos a “solução particular” que Lévy impõe, incidindo em uma escolha da tecnologia a ser executada. Após a escolha, o objeto é reconstruído e passa a existir no plano virtual. (LÉVY, 1996).

Para esclarecer melhor, o esquema abaixo demonstra as atribuições para os processos de virtualidade e realidade. A imagem expõe os dois polos: Manifesto e Latente. O primeiro é representado pelo real e atual. O segundo pelo potencial e virtual. Mas antes, veremos uma rápida explicação de Lévy quanto ao funcionamento da passagem do real para virtual e suas demais relações na seguinte citação.

Realização e potencialização pertencem ambas à ordem da seleção: escolha molar entre os possíveis, para a realização. Triagem molecular e reconstituição de uma forma, para a potencialização. Oponho aqui essa ordem da seleção a outro registro de transformação completamente diferente, o da criação ou do devir, ao qual pertencem a atualização e virtualização. (Levy, 1996, p.139)

Figura 8 – Esquema de passagens entre Real e Virtual



Fonte: LEVY, 1996, p. 145.

A imagem acima traduz, sucintamente, as passagens ou conversões realizadas entre os polos de manifesto e latência. A imagem esboça, claramente, as alusivas oposições relatadas no cabedal teórico e filosófico de Pierre Lévy (1996). A figura ainda expressa a ação do efeito “Moebius”<sup>10</sup> no âmbito da virtualidade, cujo é percebido nos momentos das passagens entre: O Real, Atual, Virtual e Potencial.

A virtualização não é uma desrealização (a transformação de uma realidade num conjunto de possíveis), mas uma mutação de identidade, um deslocamento do centro de gravidade ontológico do objeto considerado: em vez de se definir principalmente por sua atualidade (uma "solução"), a entidade passa a encontrar sua consistência essencial num campo problemático. (LÉVY, 1996, p.17-18)

<sup>10</sup> “Passagem do interior ao exterior e do exterior ao interior”. LÉVY, P. O que é o virtual. São Paulo: 34, 1996. p. 24.

A partir de um ponto de vista prático a virtualidade, desde a explosão tecnológica, faz-se necessário do uso da informática. Este advento se compara a um salto quântico da humanidade para o universo tecnológico, pois com a sua inserção recriou-se o mundo que subentendemos e o chamamos de real, também edificou e expandiu novas dimensões culturais, realizando as interações que unificam as sociedades por todo o globo. (LÉVY, 1999)

A concepção de virtualidade se amplia a cada nova tecnologia apresentada, isto ocorre pelo forte desenvolvimento da computação e da informática que proporciona uma conexão entre o plano real e o plano virtual. O referido vínculo é dito por Lévy, no livro *Cibercultura*: “A palavra virtual pode ser entendida ao menos em três sentidos: o primeiro, técnico, ligado à informática, um segundo corrente e um terceiro filosófico.” (LÉVY, 1999, p. 47).

A potencialização da computação e da informática desencadeou o incógnito mundo da tecnologia e/ou cibernético, proporcionando o que antigamente era exposto como apenas ficção, o universo paralelo ou mundo virtual. Com esse novo mundo à luz, tudo é passivo à atualização, ou seja, tudo pode ser amodernado. Isto nos implica dizer que com o empenho tecnológico adquirido, podemos produzir nossas informações e disseminá-las de modo instantâneo. Por exemplo: A materialização deste trabalho científico foi realizada por meio de um *software* aplicativo de edição de texto e após a conclusão poderá ser exposto em uma rede social ou em um repositório digital de publicações.

Seguindo com outro exemplo: Podemos digitalizar e disseminar um grande exemplar da era medieval, como a Bíblia de Gutemberg, utilizando os recursos apropriados ou outros meios de captura que serão convertidos em código binário e armazenados, potencialmente, em um espaço virtual dando origem a virtualização de dados. Estas são apenas algumas técnicas utilizadas para compreendermos a essência sobre o que é o virtual.

## 4.2 O QUE É VIRTUALIZAÇÃO?

Não há como abordarmos a computação em nuvem sem obtermos uma compreensão sobre virtualização. No momento em que pensamos na palavra virtualização, o que vem em nossas mentes é um processo que está imbuído na passagem do real ao virtual. Contudo, é de hábito forçoso que na maioria dos indivíduos que detêm um conhecimento maior no campo da informática, associe à ideia de múltiplos sistemas operacionais ou máquinas virtuais sendo executados na mesma máquina.

Esses dois pensamentos são válidos, sendo o primeiro estudado pela ciência da informação e o segundo, mais essencialmente, pela informática. No entanto, para o campo teórico a palavra “Virtualização” é um termo para abstração. Enquanto que para o prático é a potencialização dos componentes físicos para melhor desempenho.

Não obstante, a verdade é que este último pensamento é um dos diversos tipos de virtualização aplicado na informática: a de *hardware*. Este tipo de virtualização não é a única a ser utilizada, mas é, seguramente, a mais perceptível.

[...] a virtualização desvincula aplicações e o sistema operacional dos recursos físicos e acaba por agilizar e permitir o surgimento da plataforma digital. [...] A virtualização é a tecnologia utilizada na busca de uma infraestrutura de TI que se autoconfigura de acordo com a demanda das aplicações e dos negócios. [...] A virtualização é o elemento central do novo DATACENTER. (VERAS, 2011, p. 23)

Assim como o virtual, o conceito de virtualização não é recente. Em 1960, a IBM procurava uma solução para que cada usuário pudesse obter seu próprio sistema operacional e suas aplicações, desvinculado dos demais usuários, ou seja, conceito “*Time sharing*”<sup>11</sup>. No decorrer dos experimentos, engenheiros e analistas desenvolveram um sistema operacional matriz para o projeto chamado de: M44/44X. O referente sistema operacional que foi empregue em *Mainframes*, gerando assim o

---

<sup>11</sup> Uso compartilhado do tempo, de forma intercalada, em dispositivo que permite atender a duas ou mais finalidades. CREASY, R. J. The Origin of the VM/370 Time-sharing System. IBM: JOR. RES. DEVELOP. Palo Alto, v. 25, n. 5, p. 483-490, 1981..

Sistema/360, que derivou outros sistemas que interpretavam a parte física como lógica, ou seja, *Hardware* como *Software*, surgindo desta forma a primeira instância de virtualização. (CREASY, 1981)

A virtualização é uma tecnologia que proporciona um conjunto de abstração das soluções de uma máquina, fornecendo *hardwares* virtuais para um respectivo sistema, com a finalidade de ocultar as propriedades físicas e potencializar os sistemas operacionais e aplicações que interatuam com os recursos computacionais. Assim, Veras (2011, p. 85), destaca que:

Os servidores virtuais criados com a VIRTUALIZAÇÃO oferecem um ambiente similar ao de um servidor físico e otimizam o uso de recursos, tornando as aplicações independentes do hardware. Transforma-se assim um ambiente baseado em servidores físicos em um ambiente baseado em servidores virtuais ou máquinas virtuais.

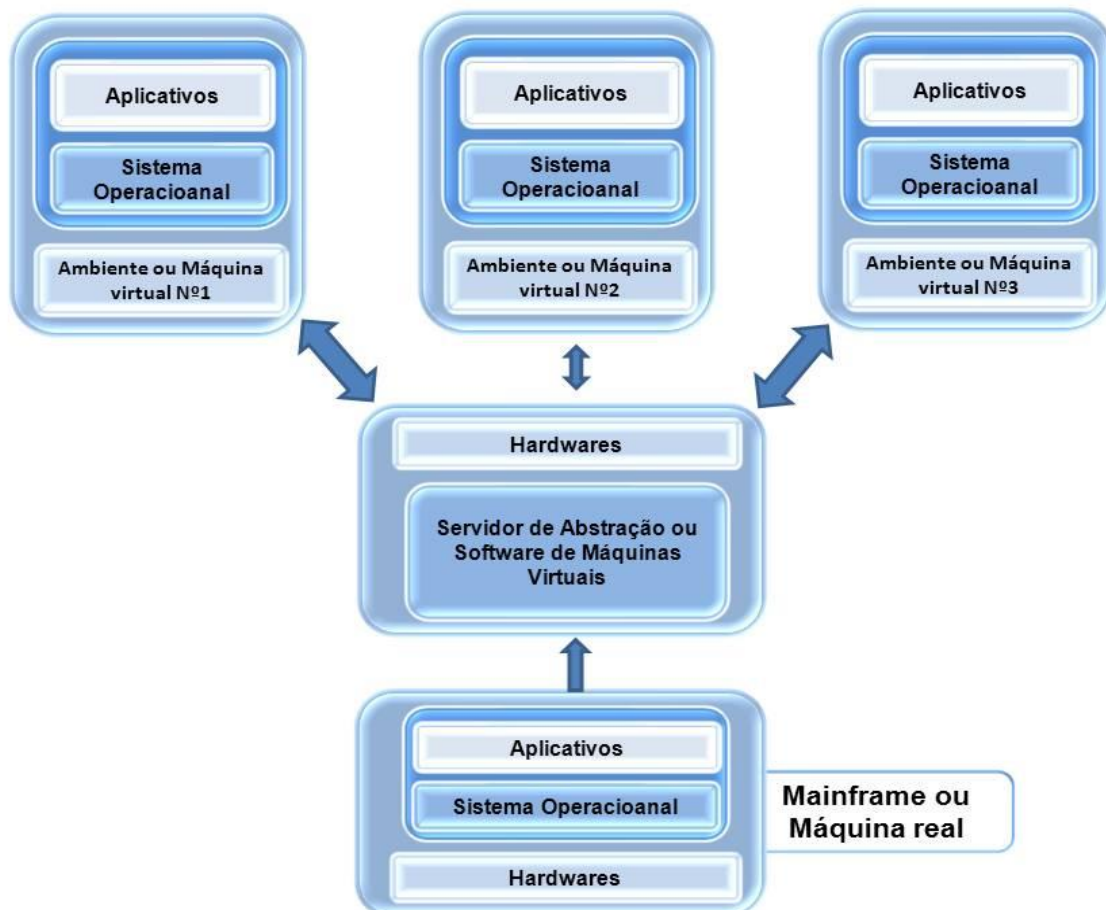
Turban, Volonino e Brodbeck (2013, p. 47), corroboram para uma compreensão melhor sobre a virtualização.

A virtualização é um conceito que possui diversos significados em TI e, portanto, muitas definições. [...] O principal tipo de virtualização é a virtualização de hardwares, que continua popular e é muito utilizada. A virtualização muitas vezes é peça chave no plano de recuperação de uma empresa. Normalmente, a virtualização separa as aplicações e os dados do negócio dos recursos de hardware – em vez de dedicar servidores para as aplicações.

A virtualização é um dos recursos que, atualmente, está sendo mais utilizados por muitas empresas e instituições. Todo o processo prático da virtualização, embora sua parte teórica seja complexa, é de fácil compreensão. A virtualização nada mais é que uma extração dos recursos existenciais e fieis de *hardwares* que serão potencializados por um *software* que recriará os *hardwares* de modo que eles não sejam físicos, mas sim lógico, isto é, *hardware* como *software*. Com a utilização da virtualização uma única máquina física poderá ter, ao invés de um sistema operacional sendo executado, três ou mais máquinas virtuais que gerar ambientes idênticos, ultrapassados ou mais avançados do que a própria máquina física que está encarregada de sustentar essa infraestrutura virtual.

Vejamos a seguir a figura que ilustra bem a virtualização.

**Figura 9 – Representação da criação de Servidores ou Máquinas Virtuais.**



**Fonte: Dados da pesquisa, 2013.**

Para compreender corretamente o conceito da tecnologia abordada, temos que delinear um paralelo entre o que é real e o que é virtual. Simplificando o alusivo paralelo, o real apresentaria características físicas, concretas, tangíveis. Por outro lado, o virtual é pertinente àquilo que é simulado, ou seja, abstrato, lógico, intangível.

Com este sentido, a virtualização pode ser considerada como o pivô para o surgimento dos ambientes virtuais, onde nesse ambiente possui a característica de simular algo físico. A Virtualização origina-se do particionamento de um único servidor físico em múltiplos servidores lógicos. Desta forma, cada servidor lógico pode executar sistemas operacionais e aplicativos de forma independente.

Virtualização simplifica o gerenciamento, permite flexibilizar e ampliar o poder de processamento. Funcionalidades contidas nos softwares de virtualização também permite melhorar a disponibilidade e a recuperação de desastres de ambientes de TI de uma maneira mais simples e com menor custo quando comparado a formas tradicionais. (VERAS, 2012, p. 125)

Veras (2012) demonstra em poucas palavras a relevância da virtualização, denotando a segurança, eficiência, economia e eficácia da virtualização aplicada para os diversos meios da TI. Veras (2012) ainda expõe a virtualização como estratégia para evitar perda da informação em caso de desastres no ambiente tecnológico.

Hoje em dia, a virtualização tornou-se imprescindível para a segurança e interação ao mundo digital, isto se deve ao desenvolvimento das TICs. A sociedade é a grande mola que impulsiona para o desenvolvimento tecnológico, pois o uso massivo das tecnologias faz com que os desenvolvedores proporcionem novas soluções que sejam ajustáveis tanto para o fornecedor quanto para o cliente que utiliza a sua tecnologia. (O'BRIEN; MARACAS, 2010)

Haja vista as pessoas que obtêm acesso a um computador, *smartphone* e *tablet*, possivelmente, já utilizaram um tipo de virtualização e se por porventura conectaram-se a *Internet*, provavelmente, fizeram uso de outro tipo de virtualização. Existem, essencialmente, quatro tipos de virtualização que podem ser implementadas: Virtualização de Hardware, Paravirtualização, Virtualização Completa e Virtualização de Aplicativos. (BERNARD GOLDEN, 2008)

Para facilitar a compreensão, antes de entrar em detalhes sobre as técnicas de implementação de virtualização, será necessário explanar, sucintamente, os termos básico utilizados pela TI para abordarmos o respectivo assunto:

- **Máquina Virtual:** É a nomenclatura dada a uma máquina ou um computador, implementado por um *software* que gerou recursos virtuais para instalar um ou múltiplos sistemas operacionais, mas que não interfira no sistema ou na máquina real. Estes software que implementam recursos virtuais, realizam uma abstração dos *hardwares* reais que são necessários para implantar um sistema operacional ou para realizar a execução de um programa não

suportado pela máquina real. De grosso modo, podemos dizer que são computadores executados dentro de outro computador.

- **Sistema Hospedeiro:** É uma máquina física ou a máquina real. O sistema hospedeiro executa o sistema operacional e o software que irá receber as máquinas virtuais.
- **Sistema Convidado:** É uma máquina virtual executada no sistema hospedeiro. Todo sistema convidado possui a ilusão de conter uma máquina física exclusiva.
- **Monitor de Máquina Virtual:** É um software entre o Sistema hospedeiro e o Sistema convidado. Este *software* faz gerar a ilusão de que cada sistema convidado possui um *hardware* particular, mas na realidade uma única máquina física hospedeira abriga as várias máquinas virtuais convidadas, ou seja, fornece a abstração. O monitor de máquina virtual também pode ser entendido como: Hypervisor. Para compreendermos melhor a sua função, o Hypervisor é um software que executa diretamente no *hardware* do sistema hospedeiro, permitindo a criação de partições isoladas de *hardwares* virtuais para receber os sistemas convidados. Isto é, cria um ou vários ambientes virtuais que possibilita a instalação de sistemas operacionais, gerando assim as máquinas virtuais.

Agora que estamos mais familiarizados com alguns termos básicos da tecnologia da informação, faremos uma explicação sobre os tipos de virtualização. Uma vez entendido os processos de virtualização, tornará mais compreensível à infraestrutura e a implementação da Computação em Nuvem.

A virtualização de *hardware* é realizada quando as partes físicas do sistema hospedeiro providenciam recursos para a virtualização a partir de um software. Para tanto, são geradas instâncias virtuais dos *hardwares* reais, permitindo a execução simultânea de múltiplos sistemas convidados.



A virtualização de *hardware* possibilita que os sistemas convidados sejam executados, isoladamente dos *hardwares* do sistema hospedeiro. Para compreender melhor o processo de virtualização de *hardware*, vejamos a seguir a imagem que representa, relativamente, o processo. (TURBAN; VOLONINO; BRODBECK, 2013)

Virtualização de hardware: é o uso de um software que simula um hardware ou todo um ambiente de computador que não seja aquele software que está rodando. Essa virtualização permite que partes do hardware rodem diversas imagens de sistemas operacionais de uma só vez. Esse tipo de software às vezes é conhecido como máquina virtual. (TURBAN; VOLONINO; BRODBECK, 2013, P. 47)

**Figura 10 – Virtualização de Hardwares**



Fonte: Dados da pesquisa, 2013.

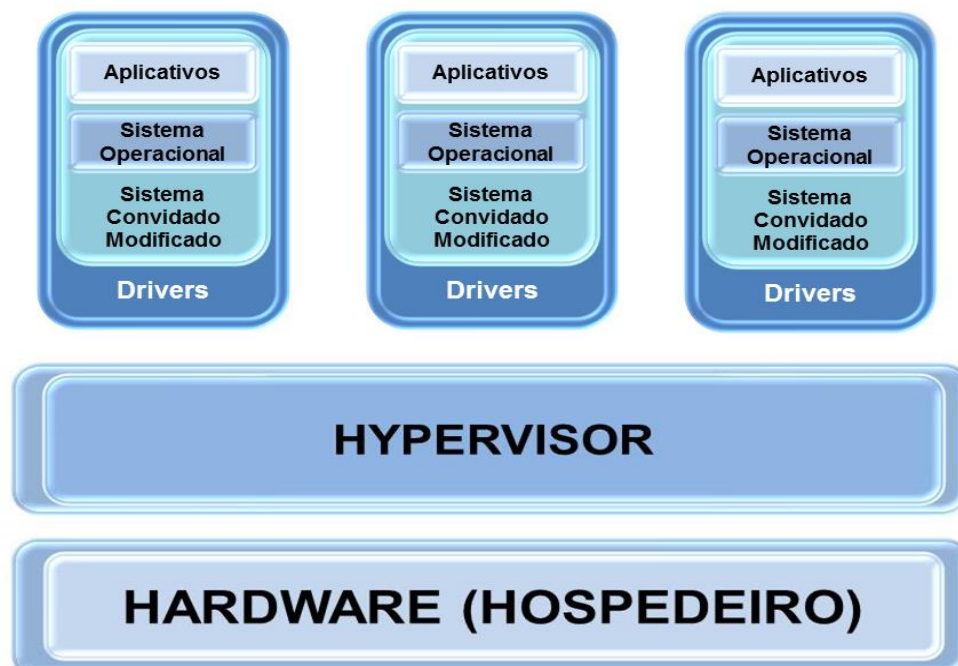
Denota-se a simplicidade da virtualização de *hardwares* quando partimos para uma experiência ilustrativa. O *software* de virtualização é instalado no sistema operacional, criando assim o ambiente de *hardwares* virtualizados, individualmente,

para instalar e executar os múltiplos sistemas convidados e os seus respectivos aplicativos.

A Paravirtualização é um método que utiliza um Hypervisor que organizar e envia para o sistema principal as requisições feitas pelas máquinas virtuais. Simplificando, o Hypervisor na paravirtualização faz o sistema convidado reconhecer os processos mais fáceis de serem executados dos processos mais árduos. Isto significa que o Hypervisor é o mediador. Ele é responsável pela interação entre o sistema convidado e o *hardware* do sistema hospedeiro. Caso o Hypervisor reconheça que o sistema convidado não esteja conseguindo ter um bom desempenho na execução das tarefas, ele enviará o processo para o sistema hospedeiro que os executará.

No entanto, para que este método seja eficaz, haverá de realizar modificações significativas no sistema convidado. “A máquina virtual enxerga a uma abstração do *hardware* que não é idêntico ao *hardware* físico. Os dispositivos de *hardware* são acessados por *drivers*<sup>12</sup> de dispositivos do próprio Hypervisor, pois desta forma otimiza-se o desempenho.” (VERAS, 2011. p. 106). Vejamos a figura que a ilustra perfeitamente.

**Figura 11 – Esquema de Paravirtualização.**



Fonte: Dados da Pesquisa, 2013.

<sup>12</sup> Software que permite ao sistema operacional utilizar os recursos e funcionalidades de um dispositivo. Guia Prático de Redes de Computadores, [Equipe Digirati Books]. – São Paulo: Digirati Books, 2009, p.30

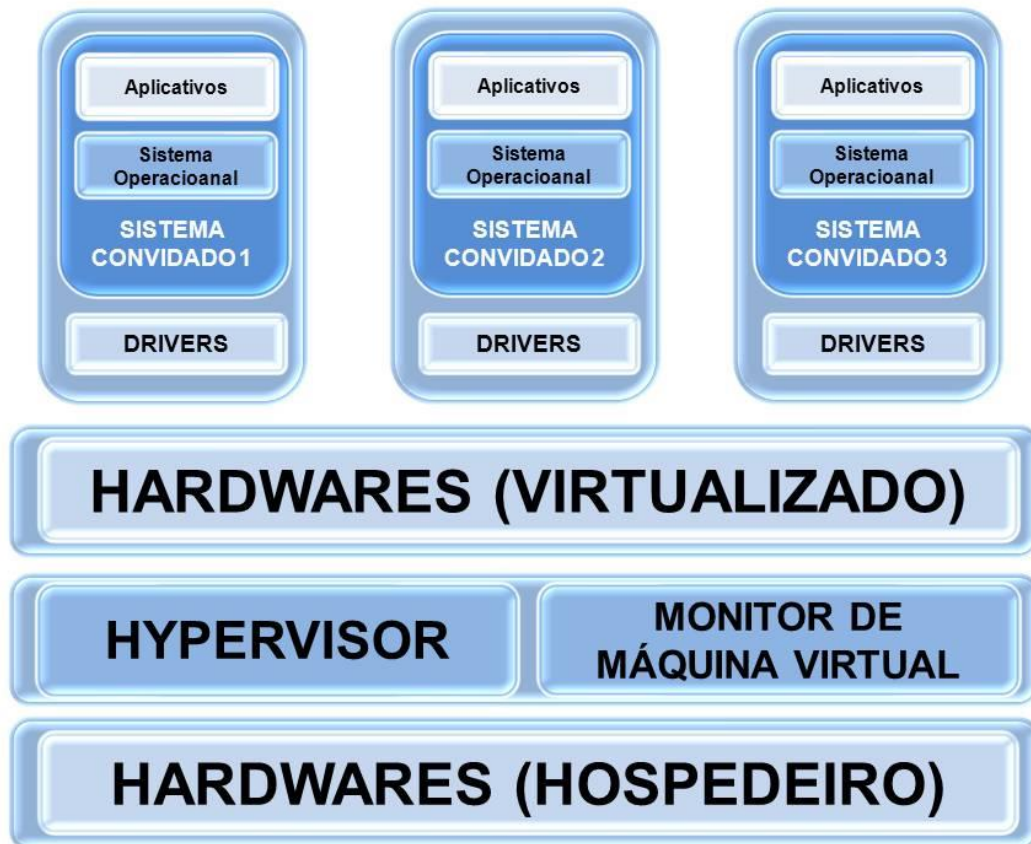
O resultado dessa interação entre *hardware*, Hypervisor e sistema convidado é uma eficiência muito grande no processamento das máquinas. Os sistemas convidados conseguem uma alta performance chegando a atingir, dependendo da configuração, o dobro do desempenho padrão. (VERAS, 2011)

A desvantagem da Paravirtualização é a modificação nos sistemas convidados. Para que o Hypervisor faça a interação é necessário que os sistemas convidados sejam configurados para que não utilize os recursos virtuais em sua totalidade, mas equilibre os processos informáticos com o *hardware* hospedeiro.

Virtualização completa é um método utilizado para permitir que todo software seja executado sem alterações. Para tanto, este método faz uso de uma simulação completa dos *hardwares* do sistema hospedeiro, de modo que o sistema convidado crie a ilusão de estar sendo executado por *hardwares* exclusivos. Na virtualização completa toda infraestrutura de *hardware* é virtualizada, sem a necessidade de alteração nas configurações do sistema convidado. (VERAS, 2011)

Este tipo de VIRTUALIZAÇÃO facilita a migração de máquinas virtuais entre servidores físicos, pois existe total independência das aplicações e dos recursos físicos do servidor. Também a segurança é facilitada pela isolação entre máquinas virtuais. O desempenho neste caso pode ser prejudicado, pois o HYPERVISOR controla todo processo e toda chamada ao hardware é feita sob a sua supervisão. Também a implementação de uma máquina virtual que imite cada dispositivo de hardware é uma tarefa complexa, pois isto é feito baseado em hardwares genéricos, o que inclui no desempenho. (VERAS, 2011, p. 104)

Figura 12 – Esquema da Virtualização Completa.



Fonte: Dados da pesquisa, 2013.

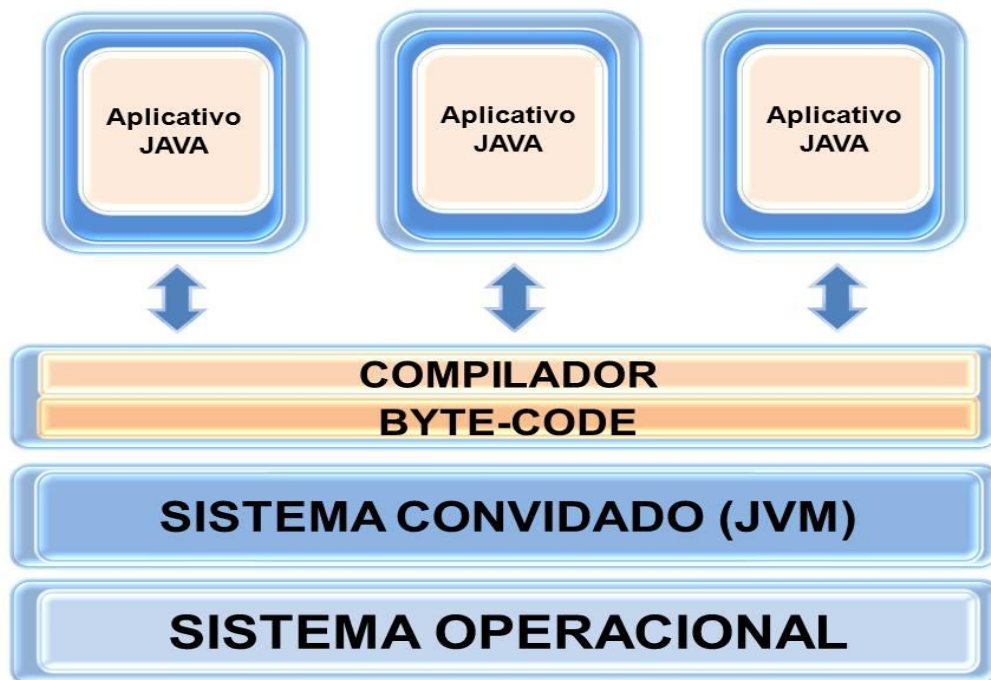
Como ilustra a figura, os *hardwares* virtualizados são simulações fieis dos *hardwares* hospedeiros que interagem com os *drivers* virtuais sem a necessidade de alteração na configuração do sistema convidado.

A atuação do Hypervisor é de gerir a inerente virtualização, como constata Veras (2011). O Hypervisor permite controlar a virtualização para que diferentes sistemas operativos sejam executados sem que haja erros e modificações. Cabe também ao Hypervisor ou monitor de máquina virtual ocultar a virtualização para que o sistema operacional do sistema convidado obtenha o melhor desempenho em sua execução. (VERAS, 2011)

No método de virtualização de aplicativos, para que ocorra a virtualização de um aplicativo, far-se-á uso de um *software* instalado no sistema hospedeiro, mas será executado no sistema convidado. O sistema convidado servirá de intercessor

entre o sistema operacional do hospedeiro e o aplicativo. Em outras palavras, haverá um encapsulamento<sup>13</sup> do aplicativo no sistema convidado que o tornará mais flexível e o fará independente do sistema operacional, podendo, desta forma, ser executado. (VERAS, 2011). Vejamos abaixo a referente ilustração da virtualização de aplicativos.

**Figura 13 – Esquema de funcionamento de virtualização de aplicativo (Java Virtual Machine).**



Fonte: Dados da pesquisa, 2013.

A figura acima demonstra que a máquina virtual JVM é instalada no sistema operacional que irá gerar recursos, possibilitando a leitura da linguagem programada do aplicativo ou imagem virtualizada, acarretando na execução e interpretação do aplicativo para o usuário.

Sem a implementação do referente sistema convidado, o computador não interpretaria o aplicativo, ou seja, o aplicativo não seria executado.

<sup>13</sup> Na programação orientada a objetos, significa separar o programa em partes, o mais isolado possível. A ideia é tornar o software mais flexível, fácil de modificar e de criar novas implementações. Nell B. Dale, Chip Weems, 2007, p. 396

### 4.3 O FUTURO DA VIRTUALIZAÇÃO

Desde o seu surgimento, a virtualização nos oferece os mais variados recursos para tornar flexíveis as atividades do nosso cotidiano. Com o advento da *Internet*, a utilização da virtualização tornou-se essencial para o acesso aos *softwares*, aplicativos, sistemas de informação, *games* e outros componentes que nos proporcionam a interação com o universo digital. (O'BRIEN; MARACAS, 2010)

A virtualização e a *Internet* nos proporcionou a quebra de paradigmas relativos ao espaço-tempo, potencializando a sociedade e recriando-a a partir das ferramentas desenvolvidas para a interação dos atores sócios virtuais ou cibernéticos. Estas duas tecnologias vêm se desenvolvendo desde os seus respectivos surgimentos e com o passar dos anos a sociedade se atualiza com o auxílio das ferramentas geradas pelas indústrias do desenvolvimento tecnológico.

Esta atualização atua na sociedade como uma retroalimentação positiva que impulsiona ainda mais o desenvolvimento tecnológico, realizando, por ventura, a transição da própria tecnologia, como foi o caso da evolução da Web inicial para Web 2.0 e deste para o qual estamos presenciando uma nova atualização, a Web 3.0. (LUZ, 2010)

Para Luz (2010), a partir da Web 2.0 o fluxo de informação na grande rede aumentou, proporcionalmente, de acordo com a inserção de indivíduos que obtiveram acesso ao computador para produção, compartilhamento, disseminação de informação e comunicação instantânea. Este fato fez realçar um problema que emergia no final dos anos 90, a preservação deste grande fluxo de informação digital. (LÉVY, 1996)

Junto a Web, todas as tecnologias se atualizaram e o seu conceito hoje, é fazer com que as informações digitais estejam com o usuário onde ele estiver, ou seja, portar os dados, aplicações, documentos e obter o acesso a partir de qualquer dispositivo com conexão a Web. Este novo meio que utiliza as potências da virtualização para acesso, produção e gerenciamento foi chamado de *Cloud Computing* ou Computação em Nuvem.

## 5. COMPUTAÇÃO EM NUVEM

A Computação em Nuvem está em toda parte acessível na web. Basta lermos algumas revistas eletrônicas sobre tecnologia ou visitar qualquer website de TI, blog ou redes sociais que teremos a certeza de que a Computação em Nuvem está reescrevendo a Web. Para compreendermos mais sobre esta infraestrutura, delinearemos sobre o que, na verdade, é a computação em nuvem e reforçar a nossa definição e compreensão dessa nova infraestrutura que servirá como complemento indispensável para a realização do projeto.

Em 2008, a computação em nuvem surgiu como tendência no mercado mundial. A proposta era de reduzir os custos operacionais e de capital, porém o mais importante era deixar os departamentos de TI se concentrarem em projetos estratégicos em vez de manter todo seu operacional em execução nos centros de dados. (MILLER, 2008)

A computação em nuvem também surgiu com a proposta de um novo conceito de infomática para o usuário doméstico. A proposta seria que o usuário não instalasse os softwares aplicativos no seu computador, deixando assim de ocupar espaço no disco rígido, realizando a execução dos aplicativos a partir de um navegador ou dispositivo que conecte-se a *Internet*. Estes aplicativos executariam tarefas que seriam das mais básicas, como redigir um texto em um editor, as mais complexas, como administrar um sistema de gestão. (TAURION, 2009)

Atualmente, a Computação em Nuvem tornou-se a infraestrutura mais utilizada pelas pequenas, médias e grandes empresas por todo o globo, conforme aponta Lobo (2013, p. 01, grifo nosso)

A Embratel aposta numa infraestrutura latino-americana para marcar seu lugar no negócio de computação na nuvem. A operadora planeja ativar - em outubro - uma oferta única de cloud computing para empresas a partir da interligação dos datacenters do Brasil, México, Argentina e Colômbia.[...] **No mundo da computação na nuvem, as grandes corporações não são o único alvo.** A Embratel também aposta nas PMEs. Tanto que lançou este mês de junho, uma loja virtual para a oferta de soluções de cloud. Por meio do portal, as micro, pequenas e médias empresas podem contratar serviços tecnológicos, de acordo com a demanda. (LOBO, 2013, grifo nosso).

Não obstante, assim como as empresas e instituições, a sociedade também faz uso dessa infraestrutura, por muitas vezes sem perceber. No entanto, todos notaram que as opções para *back-up*, compartilhamento, segurança, acesso a plataformas múltiplas e entre outros foram aumentadas e aperfeiçoadas, no que toca os erros, *bugs*, falhas em *e-mails*, sistemas com negação de serviço, vírus, incompatibilidade de arquivos, dentre outros, diminuíram de forma equivalente a evolução da computação em nuvem. (TAURION, 2009)

Neste capítulo abordaremos, de forma clara, a computação em nuvem de modo que venhamos a compreender o seu conceito, a arquitetura, suas variações, tipos e aplicação. É importante ressaltar que o referente capítulo não focará a computação em nuvem como terceirização da TI e a Arquitetura Empresarial, uma vez que o foco do trabalho é de cunho direcionado ao uso da TI pela Arquivologia a qual se insere na Ciência da Informação. Todavia, versaremos em um arcabouço teórico que servirá de complemento para os capítulos porvindouros.

## 5.1 O CONCEITO DE COMPUTAÇÃO EM NUVEM

Atualmente, no tocante a indústria da Tecnologia da Informação, o que mais se ouve é a repercussão da computação em nuvem. Essa infraestrutura fornece recursos da TI como serviço, possibilitando o usuário acessar os seus aplicativos e executar as suas tarefas onde estiver e quando quiser, ou seja, a computação em nuvem faz uso do acesso remoto para executar tarefas por meio da *Internet*, que é abstrata como uma nuvem. Podemos encontrar muitos conceitos para definir a computação em nuvem, porém poucas trazem uma reflexão prospera como a que veremos a seguir.

What is cloud computing? Cloud computing is the next stage in evolution of the Internet. The *cloud* in cloud computing provides the means through which everything — from computing power to computing infrastructure, applications, business processes to personal collaboration — can be delivered to you as a service wherever and whenever you need. (HURWITZ, et al., 2010, p. 8)



Segundo Hurwitz et al (2010), com este poder de processo computacional, os dados são coletados, processados e enviados para um armazenamento em servidores virtualizados distantes da localização física dos *Datacenters*, mas que se mantêm conectados por uma forte infraestrutura tecnológica. Isto significa que são locais abstratos, apenas visualizados pelos usuários, isto é, os usuários apenas percebem, acessam e executam os dados pelo o computador, *Tablet*, *Smartphone*, ou qualquer sistema operacional seguido de um navegador e conexão com *Internet*, mas os dados não permanecem armazenados em seu dispositivo.

Este e outros processos fazem com que o “estágio evolutivo” da computação em nuvem se consolide a cada dia, pois novas estruturas estão sendo embarcadas, fomentando outros serviços e configurações, tornando a Computação em Nuvem em uma tecnologia padrão para a Web.

O Instituto Nacional de Tecnologia e Padrões – NIST (2011, p. ?), do departamento de comércio dos Estados Unidos da América, define o conceito de computação em nuvem como “um modelo que permite o acesso conveniente à rede sob demanda para utilização de recursos computacionais configuráveis e compartilhados.”

Já Veras (2012, p. 32), ressalta que há inúmeras formas para nós definirmos a computação em nuvem e entre tantas ele destaca que a “Cloud Computing é um conjunto de recursos virtuais facilmente utilizáveis e acessíveis, tais como *hardware*, *software*, plataformas de desenvolvimento e serviços”.

Embora existam inúmeras definições, os conceitos trazem em sua raiz a mesma essência, sendo esta um conjunto de recursos virtuais de *Hardwares* e *Softwares* que permite aos usuários configurar e utilizar esses recursos como aplicações sob demanda que não estão em sua máquina real, mas sim hospedados em servidores virtuais.

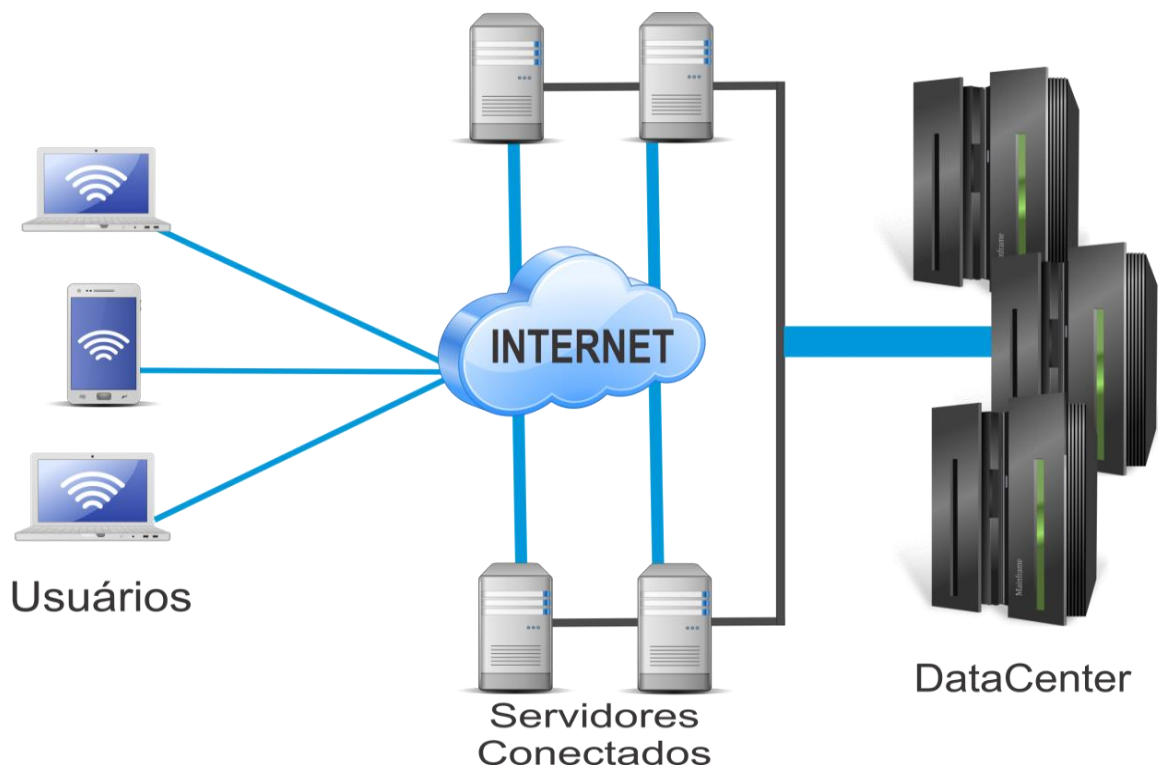
Esses recursos podem ser dinamicamente reconfiguráveis para se ajustarem de acordo com a demanda exigida pelo usuário. Para tanto, o usuário deve-se, a partir da *Internet*, conectar-se a uma rede de servidores virtualizados que ofereça infraestrutura, plataforma e *softwares* como serviços, obtendo assim o acesso à produção, processamento dos dados, gestão e armazenamento nos servidores

físicos em um *Datacenter* remoto. Veras (2012, p. 30), demonstra que captou a essência da computação nuvem e da virtualização ao relatar que:

A arquitetura CLOUD COMPUTING significa mudar fundamentalmente a forma de operar a TI, saindo de um modelo baseado em aquisições de equipamentos para um modelo baseado de aquisições de serviços. A CLOUD COMPUTING, com a VIRTUALIZAÇÃO, teoricamente permite obter melhor dos mundos: otimização do uso de recursos e flexibilidade para o usuário. [...] permite que as organizações escolham o modelo adequado para a arquitetura dos seus aplicativos e onde armazenar seus dados. [...] elemento central do processamento e armazenamento dos dados e das informações na nuvem é o DATACENTER. (VERAS, 2012, p.30)

Abaixo, segue um modelo que ilustra de maneira simplificada a infraestrutura da computação em nuvem.

**Figura 14 – Esquema básico da infraestrutura da computação em nuvem.**



Fonte: Dados da pesquisa, 2013.

## 5.2 INFRAESTRUTURA

Percebendo que a virtualização é o agente principal para a eficiência da computação em nuvem e que os *Datacenters* são responsáveis por fornecer o armazenamento desta infraestrutura, destrincharemos a partir da figura anterior, a base da nuvem e mostraremos a sua arquitetura para compreendermos o seu funcionamento. No entanto, deixaremos claro que os *Datacenters* são locais físicos que possuem máquinas robustas com um alto grau de processamento, armazenamento e desempenho para aplicação da virtualização. Esses *Datacenters* podem estar em um país ou espalhados em vários países, porém conectados pela Internet, gerando desta forma uma nuvem computacional.

Uma tecnologia fundamental ao conceito de nuvem é a virtualização, que é basicamente o uso de software para simular hardware. Quando alugamos os serviços de uma nuvem, na prática não estamos alugando diretamente computadores reais, mas computadores virtuais que existem simulados pelo software, que opera em cima dos computadores reais da infraestrutura do provedor da nuvem. (TAURION, 2009, p. 99)

É uma estrutura de serviço de valor agregado que oferece recursos de processamento e armazenamento de dados em larga escala para que organizações de qualquer porte e mesmo profissionais liberais possam ter ao seu alcance uma estrutura de grande capacidade, flexibilidade, alta segurança e que seja capacitada para gerenciar *hardware* e *software* sem a quebra de serviço. (VERAS, 2012)

Para ser mais objetivo, um *Datacenter* é um conjunto de servidores que fornece os serviços para que o usuário utilize da forma que necessitar. Esses servidores podem estar agrupados em um único lugar ou podem estar distribuídos em vários países e conectados por meio da Internet. Os *DataCenters* englobam toda infraestrutura física e sustentam a estrutura virtual da computação em nuvem.

A seguir veremos as imagens de um dos *Datacenters* do Facebook. Este se localiza no Estado de Oregon nos Estados Unidos.

**Imagem 1 – Visão aérea da Data Center Facebook**



Fonte: Google Imagem

**Imagem 2 – Servidores da Data Center do Facebook.**



Fonte: Google Imagem

Para reforçar a nossa compreensão, Veras (2013, p. 44) explica, sucintamente, o que é um *Datacenter* e como aplica-se na computação em nuvem.

DataCenter é um conjunto integrado de componentes de alta tecnologia que permite fornecer serviços de infraestrutura de TI de valor agregado, tipicamente processamento e armazenamento de dados, em larga escala, para qualquer tipo de organização. [...] Os DataCenters e suas conexões formam a nuvem e podem fazer parte de arranjos de nuvem pública e/ou privados. (VERAS, 2013, p.44)

Nesses *Datacenters* encontra-se os servidores que podem ser físicos ou virtualizados. Em suma, um servidor é basicamente um computador mais robusto do que um computador pessoal. Sua função específica é transmitir dados e prover produtos de software a outros computadores que estiverem conectados a ele por uma rede. (VERAS, 2011)

Uma prática crescente no mundo da TI é a virtualização de servidores, a qual proporciona uma potencialização na infraestrutura dedicada aos serviços disponibilizados na nuvem, ou seja, ao invés de termos um único servidor dedicado à execução de várias atividades simultaneamente, com a virtualização podemos obter múltiplos servidores que serão dedicados a executar tarefas exclusivas, dinamizando o processamento e o envio dos dados para armazenamento no *datacenter*.