



UNIVERSIDADE FEDERAL DE SÃO CARLOS - UFSCar
Centro de Educação e Ciências Humanas - CECH
Programa de Pós-Graduação em Ciência da Informação

JÚLIO CÉSAR SILVEIRA TAUIL

Metadados de preservação em *Cloud Services*

São Carlos
2018

UNIVERSIDADE FEDERAL DE SÃO CARLOS - UFSCar
Centro de Educação e Ciências Humanas - CECH
Programa de Pós-Graduação em Ciência da Informação

JÚLIO CÉSAR SILVEIRA TAUIL

Metadados de preservação em *Cloud Services*

Dissertação apresentada junto ao Programa de Pós-Graduação em Ciência da Informação da Universidade Federal de São Carlos, Campus de São Carlos, como obtenção para o título de Mestre em Ciência da Informação. Linha: Tecnologia, Informação e Representação.

Orientador: Prof. Dr. Fabiano Ferreira de Castro

Pesquisa financiada pela CAPES

São Carlos
2018



UNIVERSIDADE FEDERAL DE SÃO CARLOS

Centro de Educação e Ciências Humanas
Programa de Pós-Graduação em Ciência da Informação

Folha de Aprovação

Assinaturas dos membros da comissão examinadora que avaliou e aprovou a Defesa de Dissertação de Mestrado do candidato Júlio César Silveira Tauil, realizada em 03/06/2018.

Prof. Dr. Fabiano Ferreira de Castro
UFSCar

Profa. Dra. Ana Carolina Simionato
UFSCar

Profa. Dra. Rachel Cristina Vesu Alves
UNESP

Certifico que a defesa realizou-se com a participação à distância do(s) membro(s) Rachel Cristina Vesu Alves e, depois das arguições e deliberações realizadas, o(s) participante(s) à distância está(ão) de acordo com o conteúdo do parecer da banca examinadora redigido neste relatório de defesa.

Prof. Dr. Fabiano Ferreira de Castro

Dedico este trabalho, com muito carinho, afeto e admiração profunda aos meus pais, Ana Maria Silveira Tauil e Elias José Tauil Neto.

AGRADECIMENTOS

Agradeço aos colegas de mestrado, da primeira turma do PPGCI, pelo companheirismo e coletividade.

Agradeço imensamente ao apoio financeiro Capes em apoiar minha pesquisa. Se este não seria possível à realização deste trabalho.

Gostaria de agradecer a todo corpo docente do PPG-CI, em especial as docentes Prof^a Dr^a. Ana Carolina Simionato e Prof^a Dr^a Zaira Regina Zafalon, pelas preciosas contribuições à minha pesquisa.

Meus mais sinceros agradecimentos ao meu orientador, Prof. Dr. Fabiano Ferreira de Castro, pelas explicações, orientação e grande auxílio durante estes dois anos. Particularmente, quero agradecer a confiança deposita em mim e neste trabalho.

Gostaria de agradecer ao meu irmão, Marco Tauil, pelo laço de amizade, o qual sempre procurou inteirar-se da minha caminhada.

Agradeço imensamente aos meus amigos do curso de Biblioteconomia e Ciência da Informação, em especial a Leopoldo Carvalho Júnior, que sempre me ajudou nos momentos mais difíceis, pelo seu bom coração e solidariedade.

Ao colega, Jackson Paul Matsuura Júnior, pelas ricas contribuições, me auxiliando no norte da minha pesquisa, e ao Fábio Santana Barreto, pelas horas de conversas produtivas sobre esta pesquisa.

Também gostaria de agradecer aos amigos que fiz em São Carlos, pelos momentos de descontração, boa companhia e pelos incentivos que me fizeram seguir em frente.

Agradeço a companhia de Fernando Calza, Rafael Formenton, Ricardo Magno e Bruno Cintra. Numa noite qualquer, em Barão Geraldo, eu e Blur conversamos sobre a diferença entre Cloud Computing e Cloud Services. Entender essa diferença foi muito relevante para o andamento da pesquisa.

Quero agradecer ainda a querida Amanda Elisa Ravanini, pelos saborosos quitutes e anos de conversas nessa caminhada.

Não posso deixar de agradecer, também, ao querido casal de amigos Camilo Semensato e Giulia Y. de Campos, os dois testemunharam os momen-

tos mais importantes que tive nessa reta final, muito obrigado por lavarem a louça queridos amigos!

Agradeço, ainda, ao apoio de meus quatro guaxupeanos, parceiros antigos de caminhada, que testemunharam e me ajudaram neste processo, Gustavo Rezende Dias, Alexandre Escarassatti, Eduardo Reda e Rafael Saad.

Por fim, quero agradecer o companheirismo, a parceria, a compreensão, o carinho, a cumplicidade de minha Branquinha, Rafaela Silva Heleno, tudo vai dar certo amor!

Se for tentar,

vá até o fim.

Do contrário, nem comece!

Isso pode significar perder namoradas,
esposas, família, trabalho ... e talvez a cabeça.

Pode significar ficar sem comer por três dias,
pode significar congelar em um banco de parque,
pode significar cadeia,
pode significar zombaria, isolamento.

O isolamento é o presente

O resto é um teste de resistência,
do quanto realmente quer fazer.

E fará, apesar da rejeição.

E será melhor do que qualquer coisa que possa imaginar.

Se for tentar,

vá até o fim.

Não há outro sentimento como esse.

Ficará sozinho com os deuses,
e as noites queimarão como fogo.

Você guiará a vida ao riso perfeito,

É a única luta boa que existe. (Charles Bukowski)

E COM O BUCHO MAIS CHEIO COMECEI A PENSAR
QUE EU ME ORGANIZANDO POSSO DESORGANIZAR
QUE EU DESORGANIZANDO POSSO ME ORGANIZAR
QUE EU ME ORGANIZANDO POSSO DESORGANIZAR
(CHICO SCIENCE)

Um bom lugar se constrói com humildade,
é bom lembrar (SABOTAGE)

HDeus vinte milhões de faces

Rodas de gigas, vidas vividas

Primeira planta no mar, primeira pedra polida (OTTO)

RESUMO

O armazenamento de objetos digitais em *Cloud Services* tem cada vez mais sido adotado para conter a obsolescência tecnológica e preservar as informações digitais. Entretanto, os *Cloud Services* mais tradicionais, ainda não comportam uma estrutura que contempla as aplicações da preservação digital em seu ciclo de vida, tampouco suportam os padrões de metadados de preservação em ambientes virtuais. Nesse sentido, esta investigação busca realizar um estudo das características de interligação dos metadados de preservação digital com *Cloud Services*. Por meio da Revisão Sistemática de Literatura, fez-se o mapeamento das publicações nas áreas de Ciência da Informação e de Ciência da Computação relacionado aos metadados de preservação digital em *Cloud Services*. Os dados obtidos nos documentos recuperados foram analisados com a adoção da análise de conteúdo, destacando-se como resultados as categorias identificadas: conceituação terminológica dos *Cloud Services*; metadados de preservação digital adotados e empregados em *Cloud Services*; e tendências de pesquisa sobre a temática. Foi possível evidenciar as características, as funcionalidades e a relação entre padrões de metadados de preservação digital e os *Cloud Services*. Concluiu-se que os estudos de metadados de preservação em *Cloud Services* ainda se encontram inexpressivos na área de Ciência da Informação na atualidade, tanto no âmbito internacional, quanto nacional. A necessidade de um maior posicionamento da Ciência da Informação no panorama dos metadados de preservação digital em *Cloud Services* é emergente, uma vez que esta ciência detém o *corpus* de conhecimento para responder às questões de armazenamento, de descrição, de representação, de preservação de dados e de informações em ambientes digitais, contribuindo, assim, com a ciência e a inovação.

Palavras-chave: Metadados de Preservação Digital. Cloud Services. Revisão Sistemática de Literatura. Tecnologia, Representação e Informação.

ABSTRACT

The storage of digital objects in Cloud Services has increasingly been adopted to contain technological obsolescence and preserve digital information. However, more traditional Cloud Services do not yet have a framework that encompasses the applications of digital preservation in the life cycle of digital objects, nor do they support preservation metadata standards in their environments. In this sense, this research seeks to perform a study of the interconnection characteristics of the digital preservation metadata with Cloud Services. Through the Systematic Review of Literature, the mapping of the publications in the areas of Information Science and Computer Science that related the digital preservation metadata in Cloud Services was done. The data obtained in the retrieved documents were analyzed with the adoption of content analysis, highlighting as results the identified categories: terminological conceptualization of Cloud Services; digital preservation metadata adopted and employed in Cloud Services; research trends on the subject. It was possible to highlight the characteristics, functionalities and the relationship between digital preservation metadata standards and Cloud Services. It is concluded that the preservation metadata studies in Cloud Services are still inexpressive in the area of Information Science in the current scenario, both internationally and nationally. The need for a greater positioning of Information Science in the landscape of digital preservation metadata in Cloud Services is emerging, since this science holds the corpus of knowledge to answer the questions of storage, description, representation and preservation of data and information in digital environments contributing to science and innovation.

Keywords: Digital Preservation Metadata. *Cloud Services*. Systematic Review of Literature. Technology, Representation and Information.

LISTA DE FIGURAS

FIGURA 1. Formulário para auxiliar a extração de dados nos documentos para sistematização da análise de conteúdo.....	29
FIGURA 2. Número de documentos recuperados.....	30
FIGURA 3. Leitura prioritária dos documentos selecionados para a fase de seleção.....	31
FIGURA 4. Documentos analisados na fase de seleção.	32
FIGURA 5. Prioridade de leitura dos documentos selecionados na fase de extração.....	80
FIGURA 6. Documentos analisados na fase de extração.	81
FIGURA 7. Documentos aceitos na fase de extração.....	82

LISTA DE QUADROS

QUADRO 1. Fases resumidas da Revisão Sistemática de Literatura.	23
QUADRO 2. Formulário do protocolo de revisão sistemática preenchido na ferramenta StArt.	26
QUADRO 3. Critérios de inclusão e exclusão de documentos.....	28
QUADRO 4. Foco de estudo dos documentos pautados na análise de conteúdo.....	83
QUADRO 5. Modelos de <i>Cloud Services</i> e metadados de preservação digital.....	95

LISTA DE SIGLAS

CC - Ciência da Computação
CI - Ciência da Informação
CMS - Content Management System
CS - Cloud Services
DPCMM - Digital Preservation Capability Maturity Model
EC2 - Amazon Elastic Compute Cloud
EDS - Electronic Document Safe
HTML - Hypertext Markup Language
IaaS - Infrastructure as a Service
ISO - Internacional Organization for Standardization
LAPES - Laboratório de Pesquisa em Engenharia de Software
LDPaaS - Long-Term Digital Preservation as a Service
LOTAR - Long Term Archiving and Retrieval
METS - Metadata Encoding and Transmission Standard
MODS - Metadata Object and Description Schema
OAIS - Open Archival Information System
OWL - Web Ontology Language
PaaS - Platform as a Service
PaaST - Preservation as a Service for Trust
PDI - Preservation Description Information
PREMIS - Data Dictionary for Preservation Metadata
RepInfo - Representation Information
RIR - Representation Information Registry
RS - Revisão Sistemática de Literatura
S3 - Amazon Simple Storage Service
SaaS - Software as a Service
SGML - Standard Generalized Markup Language
SHAMAN - Sustaining Heritage Access through Multivalent Archiving
SIRF - Self-contained information retention format
StArt - State of the Art through Review
TI - Tecnologia da Informação
TIC – Tecnologias da Informação e Comunicação

URI - Uniform Resource Identifier

VEO - VERS Encapsulated Object

XML - Extensible Markup Metadata

XFDU - XML Formatted Data Unit

XSLT - eXtensible Stylesheet Language Transformation

SUMÁRIO

1 INTRODUÇÃO	14
1.1 Objetivos de pesquisa	16
1.2 Justificativa.....	17
1.3 Metodologia.....	20
1.3.1 Procedimentos metodológicos da Revisão Sistemática de Literatura	25
1.4 Estrutura do trabalho.....	32
2 REPRESENTAÇÃO, TECNOLOGIA E INOVAÇÃO NA CIÊNCIA DA INFORMAÇÃO.....	34
3 PRESERVAÇÃO DIGITAL DE LONGO PRAZO: ASPECTOS SALUTARES	46
4 ESTRUTURAS, APLICAÇÕES, CARACTERÍSTICAS E DEFINIÇÃO DO CONCEITO DE <i>CLOUD SERVICES</i> A PARTIR DA REVISÃO SISTEMÁTICA DE LITERATURA.....	54
5 PADRÕES DE METADADOS COMO GARANTIA DE PRESERVAÇÃO DIGITAL EM <i>CLOUD SERVICES</i>	69
6 RESULTADOS DA REVISÃO SISTEMÁTICA DA LITERATURA	80
6.1 Enfoques dos documentos analisados na Revisão Sistemática da Literatura	82
6.2 Padrões de metadados de preservação digital adotados e aplicados em <i>Cloud Services</i>	94
6.3 Análise das características, funções e aplicações do padrão de metadados de preservação digital PREMIS e do modelo de referência OAIIS em <i>Cloud Services</i>	110
6.4 Tendências de pesquisas identificadas nos documentos analisados.....	127
CONSIDERAÇÕES FINAIS	130
REFERÊNCIAS.....	134

1 INTRODUÇÃO

A preservação da informação é fundamental para perpetuar o conhecimento, a história, a identidade e os valores humanos. A intenção de preservar os registros a partir do armazenamento organizado, busca proteger os conteúdos intelectuais para serem acessados no futuro ou a qualquer momento, perpetuando, assim, o ciclo virtuoso de disseminação do conhecimento.

De acordo com Arellano (2004, p. 15), “a preservação é um dos grandes desafios do século XXI”. Uma das representações da memória das sociedades é o patrimônio documental, cujo risco de desaparecer é uma ameaça constante. Para Innarelli (2011, p. p. 81), a “memória que está sendo perdida a cada dia em virtude da obsolescência das tecnologias, da deterioração das mídias digitais e principalmente pela falta de políticas de preservação digital” é um fator preocupante na área da Ciência da Informação e na Ciência da Computação.

O dilema que se instala é que a tecnologia digital coloca a humanidade sob o risco de uma amnésia digital, que já está em curso, ao mesmo tempo em que abre oportunidades extraordinárias em todos os campos – da ciência à arte, do trabalho ao lazer (SAYÃO, 2007, p.183).

A preservação cultural, no contexto da sociedade da informação, possui uma relação de interdependência com os conceitos, os princípios e as políticas utilizadas pela Ciência da Informação (CI), pelas três áreas interdisciplinares em seu escopo (Biblioteconomia, Arquivologia, Museologia) e pela Tecnologia da Informação e Comunicação (TIC) (BUCKLAND, 1991).

A quantidade de dados produzidos tem um expressivo aumento a partir da massificação de diversas ferramentas tecnológicas, principalmente de caráter computacional e com o advento da *Internet*. Desse modo, a sociedade moderna está inserida num contexto de explosão informacional, que é denominada como Era Digital.

No caso dos recursos digitais, o principal desafio com relação à preservação, centra-se em torno da obsolescência tecnológica, a qual é permanente, de modo que as ferramentas tecnológicas necessitam de constantes atualizações. O contraste entre recursos impressos e digitais é impressionante. Livros

nas estantes não desaparecem se ninguém usá-los, mas os dados digitais podem desaparecer somente porque ninguém quis acessá-los, dentro de um prazo de um a dois anos após a sua criação (GLUSHKO, 2013).

De acordo com Arellano (2004, p. 16),

com o aumento da produção de informação em formato digital, tem sido questionada cada vez mais a importância de se ter garantida a sua disponibilização e preservação por longos períodos de tempo. Essa preocupação envolve tanto os produtores dos dados quanto os órgãos detentores dessa informação. No início, as práticas relacionadas com a preservação digital estavam baseadas na ideia de garantir a longevidade dos arquivos, mas essa preocupação agora está centralizada na ausência de conhecimento sobre as estratégias de preservação digital e o que isso poderá significar na necessidade de garantir a longevidade dos arquivos digitais.

Portanto, a longevidade do ciclo de vida dos objetos digitais está ameaçada pelo período curto de vida útil das mídias digitais, pelo rápido processo de obsolescência tecnológica de *softwares* e formatos. Tais fatores ameaçam à longevidade dos objetos digitais (SAYÃO, 2007).

Diante desse cenário, questiona-se como a Ciência da Informação tem se posicionado em relação aos estudos sobre os metadados de preservação digital em *Cloud Services*, visando à preservação das informações em longo prazo.

A constante disseminação e popularização dos *Cloud Services* ofertam serviços de armazenamento de informação (entre outros tipos de serviços) em *hardwares* inseridos em estruturas mais complexas conhecidas como *Data Centers*. A partir de então muitas organizações institucionais, empresas e usuários recorrem às estratégias de armazenamento, no intuito de vencer a obsolescência tecnológica.

A terceirização no arquivamento de recursos informacionais na nuvem ocorre por conta dos incessantes avanços tecnológicos. Pode-se afirmar que no futuro, o fato de não dominar as técnicas de preservação em nuvem ou não ter a mínima noção de como funcionam as estruturas dos *Cloud Services*, pode comprometer ou mesmo acarretar a extinção de acervos digitais. Nesse aspecto, é fundamental que os profissionais da CI procurem entender os processos

que envolvam diretamente as questões referentes à preservação de dados em nuvem.

Os Cloud Services procuram atender uma crescente demanda, sendo que, um dos principais aspectos que atraem inúmeros clientes, está na fácil capacidade de recuperar as informações armazenadas em nuvem. Desse modo, esta pesquisa parte da hipótese de que os metadados são considerados elementos essenciais para a garantia de preservação digital em *Cloud Services*, sendo que diversos elementos compõem a relação entre preservação digital e este ambiente, como por exemplo, os objetos digitais, componentes dos objetos digitais, camadas de nuvem, sistemas de serviços das camadas de nuvem, dados, entre outros.

A proposição desta pesquisa consiste em verificar na literatura científica, especificamente nas áreas de Ciência da Informação e da Ciência da Computação, no âmbito internacional e nacional, as publicações sobre metadados de preservação digital em *Cloud Services*.

Nesse sentido, o estudo procurou ressaltar a importância das características que contribuem e interligam os aspectos informacionais que compõe o contexto de preservação no armazenamento em serviços de nuvem, assim, como, a importância dos metadados, principalmente os metadados de preservação digital.

1.1 Objetivos de pesquisa

O objetivo geral dessa pesquisa consiste em mapear os estudos sobre os metadados de preservação digital em *Cloud Services*, por meio de uma Revisão Sistemática da Literatura (RS).

A partir dos desdobramentos deste tema, são objetivos específicos:

- Caracterizar as funcionalidades e as aplicações dos *Cloud Services*;
- Verificar quais padrões de metadados de preservação digital são adotados e aplicados em *Cloud Services*;
- Verificar as tendências de pesquisa que estão sendo desenvolvidas sobre o tema no cenário atual.

1.2 Justificativa

A preservação digital, por longos anos, não consiste num arcabouço de atividades simples, mas, pelo contrário, é norteada por uma gama de problemáticas complexas. Sua manutenção necessita de equipe técnica especializada, a qual deve conter um leque de conhecimento bem específico e atualizado de conhecimentos interdisciplinares.

Podemos destacar, por exemplo, que o profissional do campo da preservação digital precisa dominar diversas abordagens da área da Ciência da Informação e da Ciência da Computação.

A manutenção dos ambientes digitais (repositórios, bases de dados, *Cloud Services* etc.) exige atenção e cuidados constantes, pois rapidamente um objeto digital pode se tornar obsoleto, isto é, o conteúdo intelectual pode se tornar ilegível, podendo acarretar prejuízo social, cultural e econômico de grandes proporções.

No escopo da preservação, são atribuídas diversas técnicas e estratégias conforme o tipo de objeto preservado. Objetos analógicos e objetos digitais necessitam do emprego de distintas técnicas de preservação. No século XX, presenciou-se um repentino avanço tecnológico das ferramentas informacionais e comunicacionais. Essa contínua progressão tecnológica paulatinamente gerou (e continua gerando) um desenfreado fluxo informacional, principalmente, a partir da massificação do consumo dos computadores domésticos, na década de 1980 e, posteriormente, na década de 1990, sobretudo, com a popularização da *Internet*.

No século XXI, os objetos digitais se disseminaram, se massificaram e se institucionalizaram. Podemos afirmar que, em 2018, há uma clara predominância de informações e conhecimentos ancorados no amplo escopo do universo digital, isto é, a maior parte do conteúdo intelectual produzido pela humanidade encontra-se no meio digital.

Usuários, organizações públicas e privadas se comunicam e recuperam informações com muita facilidade em virtude dos recursos tecnológicos da atualidade. Além disso, com o auxílio da internet, podem praticar transações financeiras em qualquer tempo e lugar, entre outras operações, principalmente, a partir da junção de recursos fragmentados (*hardwares*, *softwares*, programas,

arquivos, objetos digitais, dados, metadados, entre outros inúmeros componentes próprios da conjuntura digital), que se complementam através de um conjunto de ambientes, relações e aplicações de tecnologia digital.

Nesse contexto, é possível observar certa despreocupação social com a preservação digital, como se fosse impossível o risco de perda de objetos digitais alocados em armazenamentos digitais. Acredita-se que simplesmente armazenar todos os recursos digitais num HD externo, seja a garantia de que estes sejam preservados por grandes períodos de tempo. Na verdade, existe uma falta de consciência das questões que envolvem todo esse escopo, e existem alguns motivos que explicam tal contexto.

Nas décadas de 1980 e 1990, os avanços tecnológicos possuíam outra dinâmica, pois as inovações ou mesmo atualizações aconteciam num intervalo maior de tempo. Contudo, o processo de obsolescência é completamente interligado aos avanços em questão. A obsolescência dos objetos digitais acontecia de forma mais lenta e menos agressiva em comparação às características do mesmo processo no período atual.

Podemos destacar alguns fatores que influenciam na falta de cuidado com a preservação digital, como por exemplo, a dinâmica do processo de produção de dados, a crença de que não faz sentido um objeto digital se tornar ilegível, a crença de que haverá constantes avanços tecnológicos que vão proporcionar a preservação de maneira fácil e simples, a crença de que é mais fácil preservar um objeto analógico do que um digital e a falta de pesquisas e estudos sobre a preservação digital em longo prazo, estes fatores podem contribuir para a falta de conscientização da população, além de dificultar a implantação de novas soluções para inúmeras instituições que necessitam salvaguardar conteúdos intelectuais e/ou imagens.

Nesse sentido, área de Ciência da Informação deve fomentar e endossar pesquisas e estudos sobre a preservação digital ou até mesmo disseminar conhecimentos sobre a importância de se preservar os conteúdos intelectuais inseridos em ambientes digitais.

Já é consenso na Ciência da Informação que a preservação digital em longo prazo jamais terá uma solução em definitivo ante a obsolescência, porém há soluções de natureza específicas ou mesmo parciais das tecnologias dispo-

níveis no emprego de estratégias de preservação, num sentido coletivo e comunitário e na direção de uma efetiva disseminação de conhecimentos.

Nesse sentido, os *Cloud Services* aparentam ser uma boa solução para a resolução desse problema. Em sintonia com essa ideia emergiu a possibilidade de estudos sobre preservação a partir da aplicação de metadados de preservação, devido à sua crucial importância em salvaguardar os objetos digitais, independente do ambiente digital.

Diante dessa perspectiva, são necessários estudos que busquem um conhecimento aprofundado sobre as características e funções desses ambientes, suas camadas, a relação e a interação entre elas com os objetos digitais, bem como a relação entre *Cloud Services* e os serviços de preservação inseridos, entre outros elementos que fazem parte desse escopo.

Destacamos a importância de se estudar e aprofundar, os metadados de preservação. É possível afirmar que eles se constituem num importante elemento no contexto das estratégias de preservação digital. Contudo, estudar tais elementos pode contribuir para a descoberta de alternativas para solucionar diversos problemas que norteiam a preservação digital de longo prazo nos *Cloud Services*.

As pesquisas sobre metadados de preservação em *Cloud Services* aumentam a dimensão do conhecimento na perspectiva de serem encontradas eficientes estratégias de preservação. Na Ciência da Informação, a preservação digital é destacada pelos desafios de natureza complexa, sendo considerada uma importante área do conhecimento que precisa ser constantemente verificada, estudada e analisada.

Os estudos acerca de metadados de preservação podem contribuir para que seja entendido com mais profundidade as suas relações com as estruturas dos *Cloud Services*; suas aplicações na garantia da acessibilidade, da integridade, da compreensão, da interpretabilidade, da autenticidade, entre outros aspectos, que visam assegurar a preservação digital dos objetos, suas relações com o objeto digital, suas relações e aplicações com outros tipos de metadados de preservação, outras categorias de metadados e dados, entre outros aspectos.

Fomentar a discussão em torno das práticas de preservação de objetos digitais em *Cloud Services* possibilita provocar uma maior consciência da im-

portância de salvaguardar a memória dos registros do conhecimento em ambientes informacionais digitais.

Portanto, esta pesquisa busca contribuir para o aumento de estudos sobre os padrões de metadados de preservação digital em *Cloud Services*, na área da Ciência da Informação no Brasil, de modo a gerar subsídios para novas pesquisas nessa área.

Além disso, acredita-se que esta pesquisa seja relevante por subsidiar discussões contemporâneas na Ciência da Informação, gerando indicadores teóricos e metodológicos para a construção de conhecimento e inovação na temática em questão.

1.3 Metodologia

Trata-se de uma pesquisa exploratória de caráter quali-quantitativa de cunho teórico, de natureza aplicada, na qual se emprega o modelo de pesquisa exploratória e descritiva, pautada na extração de informações, por meio de metodologia bibliográfica.

Configura-se como uma pesquisa quantitativa pela utilização de variáveis por levantar dados sobre a quantidade de publicações encontradas que estudam metadados de preservação digital em *Cloud Services* nas áreas da Ciência da Informação (CI) e Ciência da Computação (CC). Buscar-se-á quantificar o número de publicações sobre a temática, os locais onde os documentos foram publicados, quais os padrões de metadados utilizados em *Cloud Services* e os anos de publicação.

A análise qualitativa foi aplicada, a fim de identificar as relações entre os padrões de metadados de preservação digital com os *Cloud Services*, os padrões de metadados de preservação mais utilizados nesses ambientes, os desafios e soluções envolvendo todo o ciclo de vida dos objetos digitais nos *Cloud Services*, os componentes, os serviços e as categorias de *Cloud Services*, as funcionalidades desses ambientes, os motivos de possíveis variações terminológicas para o mesmo recurso, as questões envolvendo a problemática entre obsolescência tecnológica e preservação digital em longo prazo e as tendências de pesquisas científicas.

A partir de pesquisas teóricas e análises bibliográficas se tornou possível assimilar a importância das estratégias e dos procedimentos que garantam a utilização de metadados de preservação digital em *Cloud Services*, pautados em diferentes abordagens da CI. A abordagem do presente trabalho é de natureza aplicada, pois “objetiva gerar conhecimentos para aplicação prática, dirigidos à solução de problemas específicos” (GERHARDT; SILVEIRA, 2009, p. 35).

A primeira fase da pesquisa consistiu num levantamento bibliográfico, em que foram levantados documentos sobre a preservação digital em longo prazo na Ciência da Informação, em consonância com os denominados *Cloud Services*, na perspectiva de explorar estudos que relacionavam os padrões de preservação digital com *Cloud Services*.

A partir deste levantamento, foi possível constatar poucas abordagens sobre o tema nas áreas de Ciência da Informação. Foram estudados a partir da análise bibliográfica os seguintes pontos: aplicações e funções sobre metadados de preservação digital, características e funções dos *Cloud Services*, conceitos das diferentes terminologias dos *Cloud Services*, os conceitos, as características e as aplicações da preservação digital de longo prazo.

A pesquisa exploratória buscou estabelecer uma maior interligação com as problemáticas da pesquisa e, deste modo, buscou-se explicitá-los ou mesmo tecer hipóteses. No âmbito da pesquisa descritiva, o pesquisador é obrigado a obter um conjunto de informações em torno da temática que almeja pesquisar. Esta modalidade de estudo intenciona buscar, descrever fatos e fenômenos que norteiam determinada realidade (GERHARDT; SILVEIRA, 2009).

A fundamentação dos resultados pauta-se em fontes bibliográficas, as quais fornecem subsídios conceituais para o viés de pesquisa bibliográfica na contextualização teórica. Segundo Gil (2002, p. 44), “a pesquisa bibliográfica é desenvolvida com base em material já elaborado, constituído principalmente de livros e artigos científicos”.

Observou-se e averiguou-se de forma incisiva, no intuito de coletar importantes fragmentos e pistas considerados como forma estratégica de intervir, especificamente, por meio da aplicabilidade de métodos explícitos e sistematizados de busca, apreciação e síntese da informação devidamente filtrada e selecionada. Estratégia de intervenção específica, mediante a aplicação de

métodos explícitos e sistematizados de busca, apreciação e síntese da informação selecionada (GODOY, 2015).

No intuito de contemplar a primeira fase da pesquisa, aplicou-se o método da Revisão Sistemática de Literatura (RS), com a finalidade de mapear os estudos sobre metadados de preservação digital em *Cloud Services* nas principais bases de dados da Ciência da Informação e Ciência da Computação, a fim de identificar os metadados adotados e suas aplicações nos *Cloud Services* e verificar as tendências de pesquisa sobre a temática em questão.

A Revisão Sistemática da Literatura (RS) estabelece de forma reflexiva como a temática sobre metadados de preservação digital em *Cloud Services* está sendo desenvolvida pela área, em âmbito nacional e internacional, pois “[...] utiliza como fonte de dados a literatura sobre determinado tema” (SAMPAIO; MANCINI, 2007, p. 84). “A revisão sistemática é um processo evolutivo e visa a responder algumas questões importantes para pesquisa de um determinado domínio” (NARCISO; NUNES; DELAMARO, 2011, p. 123).

A base da RS é pautada em maneiras de identificar, avaliar e interpretar a totalidade dos trabalhos científicos de importância, a partir de uma questão específica, área temática ou fenômeno de interesse (ANJOS; NUNES; TORI, 2012).

Segundo Kitchenham (2004), a Revisão Sistemática de Literatura pode ser definida como um conjunto de estratégias de busca que objetiva detectar o máximo de literatura relevante possível e visa documentar a estratégia de busca específica para que os leitores possam acessar com rigor e de forma completa os dados disponibilizados sobre a temática. “A revisão sistemática é um processo evolutivo e visa a responder algumas questões importantes para pesquisa de um determinado domínio” (NARCISO; NUNES; DELAMARO, 2011, p.123).

De forma mais precisa e detalhada a RS é a

forma de pesquisa que utiliza como fonte dados a literatura sobre algum tema. Esse tipo de investigação disponibiliza um resumo das evidências relacionadas a uma estratégia de intervenção específica, mediante a aplicação de métodos explícitos e sistematizados de busca, apreciação crítica e sistematização dos resultados (SAMPAIO; MANCINI, 2006, p. 84 apud GODOY, 2015, p.18).

Visando apoiar a construção de todas as fases que contemplam a RS, foi utilizada e aplicada a ferramenta computacional denominada *StArt*¹ (*State of the Art through Review*). Essa ferramenta foi desenvolvida no Departamento de Computação (DC) da Universidade Federal de São Carlos (UFSCar), e elaborada pelo Laboratório em Pesquisa em Engenharia de *Software* (LAPES). Segundo Montebelo et al. (2007), este *software* é uma ferramenta de apoio a RS, o qual objetiva dar suporte ao planejamento, à execução e à análise final da metodologia, englobando qualquer assunto, temática ou área do conhecimento, tornando a pesquisa mais ágil, precisa e replicável.

Desse modo, definidos os aspectos de norteamento do levantamento das pesquisas sobre metadados de preservação digital em *Cloud Services* e as fases da RS foram devidamente aplicadas no *StArt*.

O estudo se estrutura nos métodos pautados em RS no intuito de demonstrar sua importância em captar de forma efetiva o grau de publicação de uma determinada temática, tanto quantitativa, como qualitativamente. Seguindo este viés, visando compreender todas as fases que fazem parte da construção da Revisão Sistemática de Literatura, e conseqüentemente ressaltar sua relevância na Ciência da Informação podem ser visualizadas no quadro 1.

QUADRO 1. Fases resumidas da Revisão Sistemática de Literatura.

Passos	Objetivos	Etapas	Resumo
Planejamento	Planejar o objetivo central da RS	Identificação da necessidade de uma RS; Criação do protocolo da RS.	Objetivos da RS são definidos; Protocolo da RS é definido e validado.
Execução	Executar as etapas validas no passo anterior e coletar material para análise	Identificação da pesquisa; Seleção dos resultados primários; Estudo de avaliação de qualidade.	Identificação de estudos primários; Seleção e avaliação dos estudos primários, de acordo com os critérios de inclusão e exclusão.
Análise dos resulta-	Sintetizar os estu-	Extração de dados;	Dados dos artigos

¹ Disponível em: <http://lapes.dc.ufscar.br/tools/start_tool>. Acesso em: 10 jan. 2018.

dos	dos que atendem o propósito da RS	Síntese de dados.	são extraídos e sintetizados
-----	-----------------------------------	-------------------	------------------------------

Fonte: Montebelo et al. (2007) apud Godoy (2015).

Conforme identificado no quadro 1 e baseado na estrutura utilizada por Godoy (2015), a RS contém várias fases de execução: (I) **planejamento**: definição do protocolo de pesquisa e da estratégia de busca, pautado no levantamento bibliográfico de carácter quantitativo utilizado na análise exploratória; (II) **seleção de documentos**: para seleccionar os documentos relevantes da temática específica da pesquisa foram elaborados critérios de inclusão e exclusão. A partir da confecção dos critérios ocorreu a leitura dos resumos encontrados e posteriormente seleccionados para a próxima fase de acordo com os critérios estipulados; (III) **aplicação do formulário de extração de dados** nos documentos seleccionados anteriormente; (IV) **síntese**: análise dos documentos encontrados.

Após o levantamento quantitativo, foi utilizada a análise de conteúdo nos documentos encontrados na fase de síntese da RS. A análise de conteúdo pode ser definida como o “[...] conjunto de técnicas de análises das comunicações, que utiliza procedimentos sistemáticos e objetivos de descrição do conteúdo das mensagens (BARDIN, 1977, p. 38)”.

Esta metodologia é embasada na pré-análise da bibliografia encontrada, a qual é caracterizada pelo contato inicial dos textos, seleção dos documentos, preparação do material, seleção dos textos, referenciação dos índices e confecção de indicadores, exploração do material, tratamento dos resultados, inferência e interpretação dos documentos (BARDIN, 2007).

Os procedimentos e análises dos documentos foram traçados em várias etapas: consultou o *corpus* obtido, tratamento dos resultados obtidos, leitura e análise final do material estudado, síntese e seleção dos resultados relevantes e utilização dos resultados de análise com fins teóricos. A análise de conteúdo combinou técnicas com características quantitativa e qualitativa para obter detalhes do objeto estudado, possibilitando realizar inferências com mais segurança e obter resultados mais concisos. (VALENTIM, 2005 apud GODOY, 2015).

A análise de conteúdo realizada nos documentos obtidos na Revisão Sistemática de Literatura procedeu à análise dos estudos sobre metadados de preservação, à identificação dos metadados adotados e aplicados em *Cloud Services*, à verificação das características dos *Cloud Services* e às tendências de pesquisas sobre a temática.

1.3.1 Procedimentos metodológicos da Revisão Sistemática de Literatura

Antes da realização da Revisão Sistemática, realizou-se uma pesquisa e análise bibliográfica. A partir de então, verificou-se a ausência de estudos sobre a temática desenvolvida nesta pesquisa.

A RS pode alicerçar uma pesquisa independente se as perspectivas são opostas, pois sua natureza reside em descobrir se determinada temática está gerando publicações ou não. A partir da realidade exposta pelo método, é possível analisar descritivamente os dados apontados e traçar as próximas estratégias de estudo no desenrolar de uma pesquisa, independente do tema.

O intuito de se aplicar a RS na presente pesquisa consiste em demonstrar sua importância em captar, de forma efetiva, o grau de publicação dos metadados de preservação em *Cloud Services*, tanto quantitativa, como qualitativamente.

As fases da RS utilizadas na presente pesquisa são descritas da seguinte maneira:

Planejamento - com o intuito de localizar os documentos, foram utilizadas as seguintes bases de dados: *Scopus*², *Web of Science*³, *IEEE Xplore*⁴, *Computer and Information Systems (ProQuest)*⁵, *Library & Information Science Abstracts (LISA)*⁶, *Information Science & Technology Abstracts (ISTA)*⁷. Das

² Disponível em: <<https://www.scopus.com/freelookup/form/author.uri>>. Acesso em: 09 nov. 2017.

³ Disponível em: <http://apps-webofknowledge.ez31.periodicos.capes.gov.br/WOS_GeneralSearch_input.do?product=WOS&search_mode=GeneralSearch&SID=7Cjk8ECoS9xpFrwnwVQ&preferencesSaved=>>. Acesso em: 09 nov. 2017.

⁴ Disponível em: <<http://ieeexplore.ieee.org.ez31.periodicos.capes.gov.br/Xplore/hom>>. Acesso em: 09 nov. 2017.

⁵ Disponível em: <<https://search-proquest.ez31.periodicos.capes.gov.br/computerinfo/index>>. Acesso em: 09 nov. 2017.

⁶ Disponível em: <<https://search-proquest.ez31.periodicos.capes.gov.br/lisa>>. Acesso em: 09 nov. 2017.

seis bases de dados duas são de escopo geral (*Web of Science* e *Scopus*); duas são da área de Ciência da Informação (ISTA e LISTA); e, por fim, duas são da área da Ciência da Computação (*Computer and Information Systems* [ProQuest]).

O idioma empregado foi o inglês com a mesma *string* padronizada introduzida em todas as bases de dados pesquisadas: ("Cloud Services" OR "Cloud Storage" OR "Cloud Storage Providers" OR "Cloud Operations" OR "Cloud Computing") AND ("Digital Preservation Metadata" OR "Digital Preservation") AND ("Information Science" OR "Computer Science").

No *StArt* há um formulário para preencher o protocolo da pesquisa, que é fundamental ser planejado corretamente, conforme visto no quadro 2.

QUADRO 2. Formulário do protocolo de revisão sistemática preenchido na ferramenta *StArt*.

Propósito da revisão sistemática	Verificar na literatura científica, especificamente nas áreas de Ciência da Informação e da Ciência da Computação, tanto no âmbito nacional e internacional, as publicações sobre metadados de preservação digital em <i>Cloud Services</i> .
Principal questão	Quais publicações interligam diretamente os metadados de preservação digital em <i>Cloud Services</i> visando à preservação das informações em longo prazo?
População	Metadados de preservação digital utilizados em <i>Cloud Services</i> .
Intervenção	Verificar como são interligados os estudos de metadados de preservação digital em <i>Cloud Services</i> e quais são as estratégias utilizadas na preservação em longo prazo.
Controle	Artigos indicados pelo orientador, artigos publicados em periódicos da área, bases de dados interdisciplinares às áreas de Ciência da Informação e Ciência da Computação.

Resultados	Identificar as publicações que estudem os metadados de preservação digital em <i>Cloud Services</i> , focando nas estratégias de preservação em longo prazo no armazenamento dos recursos informacionais. Verificar as tendências de pesquisa que estão sendo desenvolvidas.
Aplicação	Verificar se os estudos de metadados de preservação digital em <i>Cloud Services</i> estão sendo expressivamente pesquisados na área de Ciência da Informação e na área de Ciência da Computação, levando em conta o número de publicações a serem encontradas. Contribuição teórica para os pesquisadores de Ciência da Informação em futuros trabalhos sobre preservação digital em <i>Cloud Services</i> .

Fonte: elaborado pelo autor.

No próprio protocolo desenvolvido pelo *StArt* são definidas as bases de dados utilizadas, os critérios de inclusão e de exclusão; referente à fase de seleção também são devidamente preenchidos no formulário.

Num outro campo, são preenchidas as palavras-chave. Esta etapa é crucial para o bom funcionamento do processo de seleção e extração dos textos publicados numa base de dados e inseridos na ferramenta tecnológica em questão, pois a partir da lista gerada no sistema, as publicações são organizadas das mais importantes para as menos importantes, a partir da pontuação que cada texto alcança.

Após a condução do planejamento respaldado e baseado no protocolo desenvolvido na ferramenta *StArt*, as publicações elencadas nas bases de dados entram na fase de seleção da RS. Dessa forma, a seleção dos textos ocorre a partir dos critérios de inclusão e exclusão.

Na fase de seleção, além das publicações exportadas das bases de dados escolhidas, cada publicação recebe uma classificação, são quatro opções: muito alto, alto, baixo e muito baixo. Tanto a pontuação como a classificação auxiliam de forma categórica a condução da RS.

O quadro 3 apresenta os critérios de inclusão e de exclusão estabelecidos e utilizados no desenvolvimento da pesquisa.

QUADRO 3. Critérios de inclusão e exclusão de documentos.

Critério de inclusão	Critério de exclusão
As principais abordagens do texto focam em características relacionadas com <i>Cloud Services</i> .	As principais abordagens do texto focam em características relacionadas com <i>Cloud Computing</i> .
O documento apresenta relação de metadados de preservação digital em <i>Cloud Services</i> .	O documento não apresenta relação de metadados de preservação digital em <i>Cloud Services</i> .
O documento contempla metadados de preservação digital em <i>Cloud Services</i> nas áreas de Ciência da Informação e Ciência da Computação.	O documento contempla metadados de preservação digital em <i>Cloud Services</i> , mas não são das áreas de Ciência da Informação ou de Ciência da Computação.
O documento apresenta estratégias de preservação digital utilizando metadados de preservação em <i>Cloud Services</i> .	O documento não apresenta estratégias de preservação digital, mesmo apresentando metadados em <i>Cloud Services</i> , porém não relaciona metadados de preservação com <i>Cloud Services</i> .
Somente documentos encontrados em bases de dados <i>online</i>	O texto não está disponível de forma integral.
Qualquer ano.	O documento não está escrito em inglês.

Fonte: Elaborado pelo autor.

Na fase de seleção foram aceitos inúmeros documentos que condiziam com a *string* de busca, além do mais, há um aplicativo de pontuação na ferramenta que também ajuda na decisão de aceitar ou rejeitar o documento. Após os documentos serem selecionados, foram elaboradas questões num formulário relativo à fase de extração, conforme apresentado na figura 1, que auxiliou na sistematização e na categorização dos documentos que foram aceitos na fase de extração. Dos 11 (onze) documentos que restaram para a sistematização descritiva da análise de conteúdo, foram apontadas uma série de questões no âmbito da preservação digital em ambientes de nuvem.

A ferramenta confecciona gráficos que reproduzem o percentual pormenorizado das publicações, há percentuais dos critérios de seleção, de classificação e da própria pontuação, facilitando para o pesquisador trabalhar na última fase, a síntese dos documentos. A extração e a síntese englobam os passos de análises de resultado.

FIGURA 1. Formulário para auxiliar a extração de dados nos documentos para sistematização da análise de conteúdo.

Study Data	Selection Data	⚠ Data Extraction Form	Similar Studies
Quais abordagens foram utilizadas no texto na associação de estudos de metadados de preservação digital com Cloud Services?			
O texto em questão está inserido no escopo de abordagens da Ciência da Informação?			
O texto em questão está inserido no escopo de abordagens da Ciência da Computação?			
O texto em questão conclui que os metadados de preservação digital são necessários ou desnecessários no âmbito dos Cloud Services?			

Fonte: elaborado pelo autor.

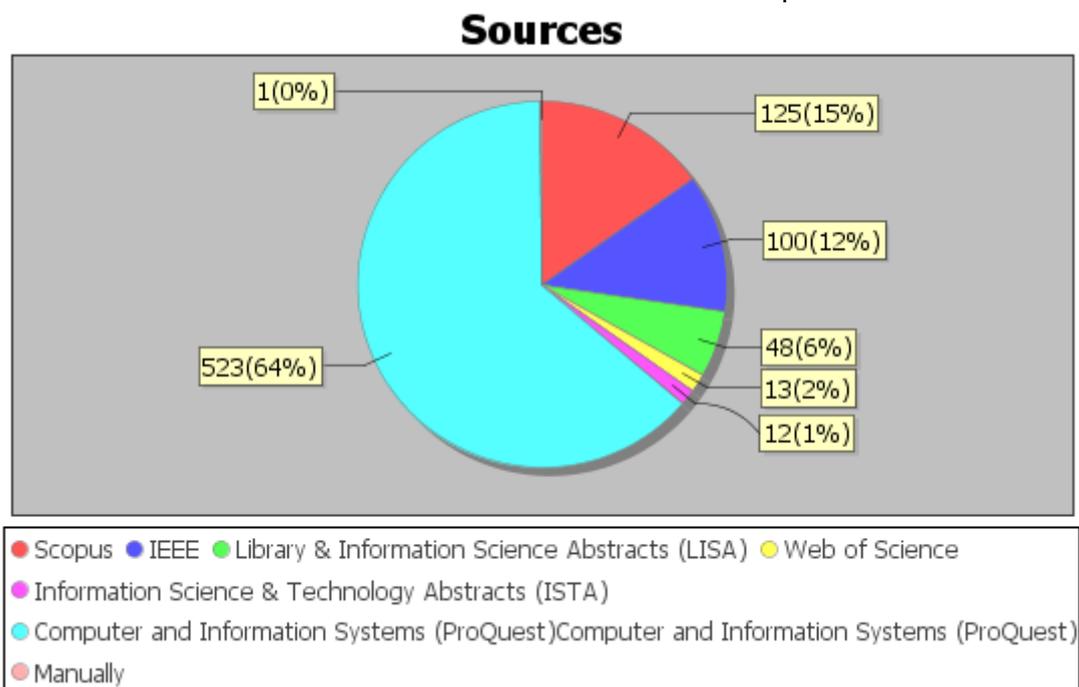
Após a extração das publicações, ocorre a última fase: a síntese dos resultados. Nessa fase, o pesquisador analisará os textos que foram filtrados de acordo com a temática específica. Dessa forma, a pesquisa em questão será conduzida através das análises de resultados obtidas pelo método de RS empregado. A partir dessa análise é possível observar um fluxo baixo de publicações de metadados de preservação digital em *Cloud Services* nas áreas de CI e CC. Concluídas todas as fases da RS, verificam-se as análises dos resultados referentes às publicações consideradas válidas.

A partir da RS, percebe-se que ainda há poucas pesquisas sobre metadados de preservação digital em *Cloud Services*, a princípio foram encontrados

nas seis bases de dados (CI, CC e de escopo geral) 821 documentos, os quais foram diminuindo com o refinamento da busca.

De acordo com os resultados apurados, verificamos que: 523 documentos foram recuperados da *Computer and Information Systems (ProQuest)*, mais da metade dos documentos (64%); na *Web of Science* foram recuperados 13 documentos, apenas 2% do total; na *Information Science & Technology (ISTA)* foram encontrados 12 documentos computando somente 1% dos mesmos; na *Library & Information Science Abstracts (LISA)* 48 documentos contribuíram com 6%; e, na *Scopus* foram encontrados 125 documentos, caracterizando 15% dos documentos, conforme a Figura 2; na *IEEE Xplore* foram encontrados 100 documentos, totalizando 12%. Destacamos que não houve documentos inseridos de forma manual, porém, por questões técnicas, esta vertente aparece na figura 2.

FIGURA 2. Número de documentos recuperados.

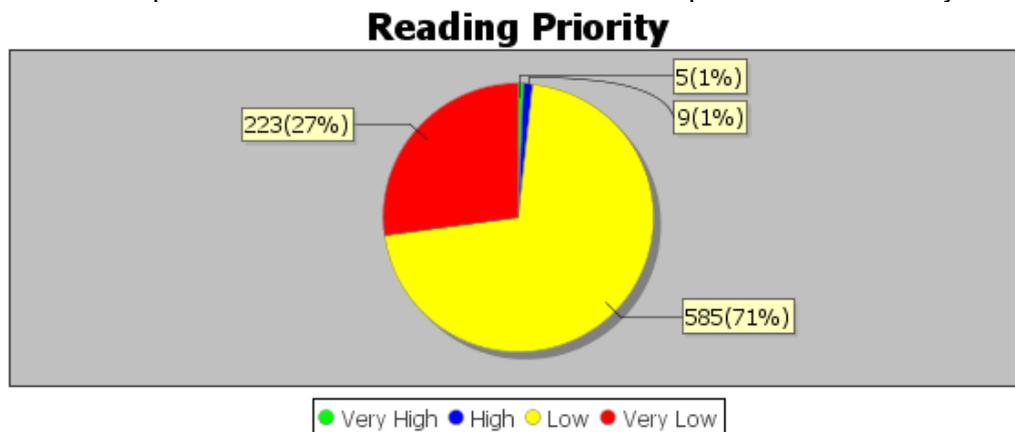


Fonte: elaborado pelo autor.

Essa etapa acontece em conformidade à elaboração das palavras-chave interligadas com o esquema de pontuação decidido a critério do leitor, por exemplo, se um documento contém todas as palavras-chave escolhidas pelo autor no resumo, palavras-chave e título do texto, o documento receberá pon-

tuação máxima e pode ser um forte indício de ser classificado como muito alto na prioridade das leituras. A figura 3 aponta como ficou estabelecida tal divisão.

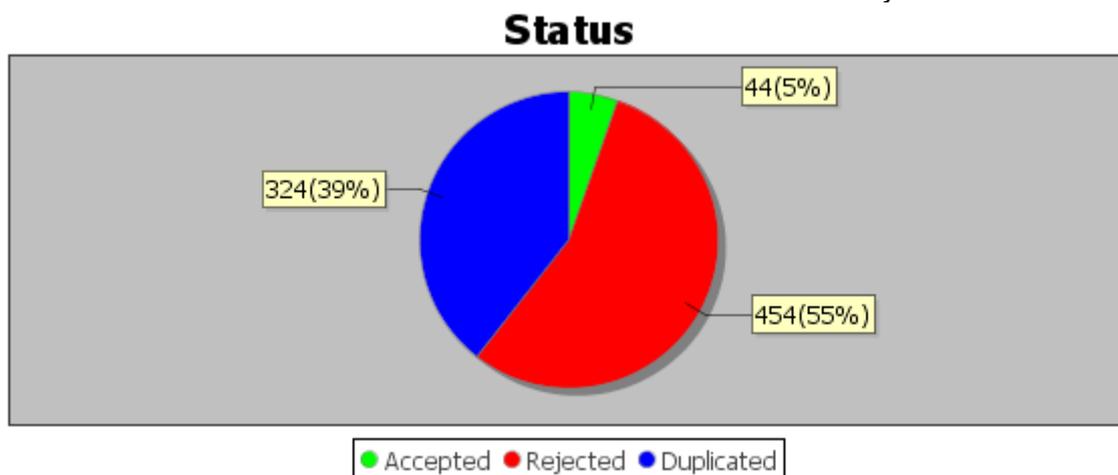
FIGURA 3. Leitura prioritária dos documentos selecionados para a fase de seleção.



Fonte: elaborado pelo autor.

A maioria das prioridades de leitura dos documentos foram cancelados na opção *Low* (baixo), cerca de 585 documentos (71%), 223 (27%) foram enquadrados na opção *Very Low* (muito baixo), somente 9 documentos se enquadraram na opções *High* (alto) e, 5 documentos na modalidade *Very High* (muito alto). Cada uma das fatias foi representada com 1% de documentos considerados com um alto índice de leitura.

Após a ferramenta *StArt* selecionar os documentos, fornecendo dados sobre a prioridade de leitura de cada texto, houve o processo de seleção dos documentos, que foram aceitos para a fase de extração e os documentos rejeitados, os quais, não empregavam conteúdos de preservação digital em *Cloud Services*. Nessa fase, alguns documentos que não citavam a terminologia metadados de preservação digital foram aceitos, pois, utilizando essa terminologia em questão ou não, utilizam esses metadados no escopo dos respectivos textos aceitos. A figura 4 divide os documentos em três seções: aceitos, rejeitados e duplicados (registros que já tinham sido encontradas numa outra base de dados).

FIGURA 4. Documentos analisados na fase de seleção.

Fonte: elaborado pelo autor.

Foram aceitos 44 documentos (5%); enquanto 454 documentos (55%) foram rejeitados e, 324 documentos (39%) foram duplicados.

No intuito de encontrar estudos sobre metadados de preservação digital em *Cloud Services* nas áreas de Ciência da Informação (CI) e Ciência da Computação (CC), a Revisão Sistemática da Literatura (RS) pode ser considerada um método eficaz ao amparar os pesquisadores de CI no emprego de um método que apresenta dados consistentes e auditáveis.

1.4 Estrutura do trabalho

O presente capítulo compreende a Introdução e aborda questões iniciais, bem como o contexto que esta se insere, tais como a questão de pesquisa, a hipótese, a proposição, os objetivos (geral e específico), a metodologia, a justificativa e a relevância para a área de Ciência da Informação.

O capítulo 2 - aborda os aspectos intrínsecos que interligam a relação entre Representação da Informação, avanços tecnológicos e inovação no contexto da Ciência da Informação.

O capítulo 3 – aborda a importância da preservação digital em longo prazo no ciclo de vida dos objetos digitais, abarcando aspectos que interligam obsolescência tecnológica com avanços tecnológicos. São apontados diferenças entre preservação analógica e digital, além de uma série de desafios que fazem parte do arcabouço da preservação digital em longo prazo.

O capítulo 4 – busca explicar as estruturas, as aplicações e as características que abarcam os ambientes do *Cloud Services*, e pelo fato de serem explanadas diversas pluralidades terminológicas encontradas nas publicações, no que se refere aos *Cloud Services*. Neste capítulo buscou-se uma definição conceitual da terminologia.

O capítulo 5 - busca apontar os principais padrões de metadados de preservação digital existentes, endossando a relevância destes padrões em assegurar a preservação digital em longo prazo nos *Cloud Services*.

No capítulo 6 - através da Revisão Sistemática de Literatura, foram extraídas as publicações que tratam sobre a temática do presente estudo, e a partir de uma análise quantitativa, são apontadas diversas características de cada documento. Consequentemente, os resultados construídos, por meio de da análise de conteúdo, identificam nos documentos os tipos de metadados de preservação digital adotados e aplicados, além das especificações e atribuições que norteiam suas aplicações dentro dos ambientes dos *Cloud Services*, assim como no ciclo de vida dos objetos digitais. Por fim, são elencadas tendências de pesquisas mais descritas nos documentos analisados.

2 REPRESENTAÇÃO, TECNOLOGIA E INOVAÇÃO NA CIÊNCIA DA INFORMAÇÃO

Neste capítulo serão delineadas as especificidades, as características e a pertinência da representação da informação no bojo da Ciência da Informação, bem como sua interligação com as tecnologias, em especial, as de informação e de comunicação (TIC). Por último, são ressaltadas características das inovações tecnológicas, a partir de um breve recorte de definições e conceitos. Vale destacar que a inovação permeia a constante relação entre representação da informação e o avanço das ferramentas tecnológicas.

Independente da estrutura de armazenamento dos dados e das informações em formato digital ou analógico, a representação da informação atua de forma decisiva na garantia da preservação da informação e na promoção da inovação, entre outras funções.

Nesse aspecto, o capítulo entra em consonância com o estudo de metadados de preservação digital em *Cloud Services*, pois interliga representação da informação, tecnologias e inovação. O primeiro ponto a ser ressaltado é norteado pela crescente utilização dos *Cloud Services*, a qual pode ser considerada uma inovação tecnológica, que se destaca a partir de meados dos anos 2000. Desse modo, apresenta como uma possível solução no armazenamento de dados, no intuito de conter a obsolescência tecnológica, podendo ser uma alternativa viável para assegurar a preservação digital, a partir da heterogeneidade de recursos e formatos computacionais disponíveis.

Representar é algo inerente ao ser humano, seja no contexto da sua própria sobrevivência como espécie, seja por meio de atividades que o auxiliam na preservação de sua memória e identidade, ou mesmo para organizar de forma mais precisa e adequada o seu próprio cotidiano.

A representação acompanha a história da evolução da humanidade, tendo um papel tão fundamental quanto o desenvolvimento de técnicas de cultivo, de artesanato, de domínio do fogo, entre outros aspectos fundamentais do processo evolutivo.

A partir da representação do pensamento foi possível representar, de forma cada vez mais sofisticada, os processos de comunicação, seja represen-

tada por desenhos, pela linguagem, pelo desenvolvimento da escrita , entre outros fatores de representação.

O desenvolvimento de técnicas de comunicação gerou automaticamente a disseminação do conhecimento, o qual se propagou entre os agentes sociais, possibilitando um contínuo ciclo de assimilação, ensino e aprendizagem.

A representação está interligada com a preservação da memória e identidade de todas as sociedades, além de estar intimamente relacionada com os processos de organização da informação (MARCONDES, 2001).

A representação auxiliou no desenvolvimento da comunicação dos diversos grupos de homínídeos em diversos períodos históricos, seja de uma mesma etnia, seja de uma mesma civilização, ou mesmo de uma 'nação' com os perfis culturais e linguísticos similares, ou mesmo como forma de intercâmbio comunicacional entre povos distintos. Há milhares de anos a representação da informação é utilizada de forma cíclica e constantemente ressignificada. As formas de representar vão se modificando conforme surgem novas ferramentas e tecnologias. No âmbito da representação da informação é possível englobar uma pintura rupestre, uma ficha catalográfica, ou algum padrão de metadados de algum programa computacional.

Os seres humanos usufruíram dos suportes da informação desde sua criação, os quais podem ser considerados como base para sua própria existência. Os suportes de informação possibilitaram a transmissão do conhecimento às gerações futuras, por meio de registros, possibilitando a progressiva evolução da humanidade.

O desenvolvimento dos mecanismos de transmissão de pensamento, através da comunicação verbal, da escrita e pelo registro de símbolos, propiciou a confecção de documentos que expressavam as ideias de uma sociedade. Basicamente o documento representa um item, recurso ou objeto que contém informação, cujo principal intento é produzir conhecimento (CUNHA; CASTRO, 2018).

O conhecimento não se refere ao objeto físico propriamente dito, porém se relaciona diretamente na correlação entre percepção e significação do pensamento, no intuito de atribuir características como fonte e ordem. Não há consenso sobre as abordagens conceituais designadas à representação da informação, ou seja, são trabalhadas de maneira heterogênea nas mais diversas

áreas do conhecimento, como, por exemplo, a filosofia ou a semiótica (MARTINS, 2016).

A Ciência da Informação (CI) apresenta duas tendências no âmbito do tratamento e da gestão da informação. A primeira tendência, no geral, gira em torno da Organização da Informação, com um enfoque voltado aos procedimentos de análise, de síntese, de condensação, de representação e, por último, de recuperação do conteúdo informacional e da reflexão epistemológica e instrumental sobre organização do conhecimento.

A segunda tendência, fundamentalmente inserida nas Tecnologias de Informação e Comunicação, fixa-se nas estruturas e modelos dos sistemas computacionais. Portanto, esta tendência atua nos processos de produção, de armazenamento, de preservação, de representação, de recuperação, de acesso, de uso e de disseminação da informação. Por mais que apresentem diferenças, as duas tendências citadas possuem fundamentações metodológicas e diretrizes bem parecidas e complementares, tanto nos conceitos que norteiam a recuperação, como as que norteiam a representação da informação (SANTOS; VIDOTTI, 2009).

Através da organização das informações armazenadas numa unidade ou sistema de informação, as possibilidades de disseminação da informação são ampliadas. Pode-se constatar que a etapa em questão cumpre uma importante função de intermediar a relação entre o produtor da informação e o usuário, endossando o papel da organização das informações na atuação de mediador da informação (GONZÁLES DE GÓMEZ, 1994).

Os sistemas de organização do conhecimento passaram por diversos tipos de construções lógicas e de categorização, os quais, por sua vez, refletem a estrutura do conhecimento por séculos. Escancarar as ideias de categorização pautadas na construção lógica e organizada pode ser considerado como a melhor tentativa de distribuição de compreensão do conhecimento e, consequentemente, as subdivisões dos seus respectivos elementos (MARTINS, 2016).

Existem duas diferenças no que tange aos processos de organização: uma aplicada às ocorrências individuais de objetos informacionais, isto é, o processo de organização da informação; e, a outra aplicada às unidades do

pensamento (conceitos), intitulado como processo de organização do conhecimento (OC).

Uma das principais características da Organização da Informação (OI) gira em torno da organização de um conjunto de objetos informacionais ordenadas de forma sistemática em coleções. Nesse viés, a organização da informação é aplicada em bibliotecas, museus e arquivos, seja no ambiente tradicional ou digital. Já a Organização do Conhecimento, está direcionada a construção de modelos de mundo caracterizados pelas abstrações da realidade.

Conseqüentemente, estes dois tipos de procedimentos confeccionam dois tipos distintos de representação: a representação da informação, entendida como o grupo de atributos que representa determinado recurso informacional, o qual é obtido pelos processos de descrição física e de conteúdo. O outro tipo de representação é a denominada representação do conhecimento, esta é configurada numa estrutura conceitual com o intuito de representar modelos de mundo, as quais descrevem e fornecem esclarecimentos sobre os fenômenos observados (BRASCHER; CAFÉ, 2008).

As áreas de Filosofia, Linguística, Sociologia e Ciência da Computação enfocam relevantes discussões no cerne da representação do conhecimento. Estas áreas possuem extrema relação de proximidade com a Ciência da Informação devido à interligação com a organização, destacando a importante intersecção com a subárea de Organização e Representação da Informação.

O processo de categorização é o marco que expressa a organização da compreensão do pensamento e, simultaneamente, da representação do conhecimento das coisas. A categorização visa interpretar e compreender o mundo por meio da descrição das coisas/itens/objetos e, conseqüentemente, de suas propriedades. Basicamente as categorias conceituam objetos passíveis de serem conhecidos, organizados e classificados (MARTINS, 2016).

Em relação às principais características relativas ao processo de representação da informação, Novellino (1996, p.38) ressalta:

[...] é a substituição de uma entidade linguística longa e complexa - o texto do documento - por sua descrição abreviada. O uso de tal sumarização não é apenas uma conseqüência de restrições práticas quanto ao volume de material a ser armazenado e recuperado. Essa sumarização é desejável, pois sua função é demonstrar a essência do documento. Ela funciona

então como um artifício para enfatizar o que é essencial no documento considerando sua recuperação, sendo a solução ideal para organização e uso da informação.

A forma efetiva de representação da informação objetiva disseminar de maneira ampla a informação específica endereçada aos usuários finais. Nesse sentido, uma unidade de informação precisa compreender o arcabouço de ferramentas tecnológicas do seu acervo (suportes, recursos), as quais são utilizadas na gestão do ambiente e, conseqüentemente, estruturar os dados no intuito de ter a melhor interpretação possível. A partir da informação estruturada são escolhidos códigos que fazem mais sentido para o usuário e, neste contexto, a semântica é extremamente relevante. Um erro praticado pelo profissional da informação no momento da forma de representação pode acarretar numa informação não disseminada ou mesmo pouco disseminada, isto é, uma informação extremamente valiosa ao usuário final corre o risco de se perder. Por este motivo, o processo da representação da informação precisa ser minucioso, efetivo e prático (MARCONDES, 2001).

A melhor forma de utilização da representação da informação engloba um conjunto de codificações padronizadas direcionadas ao interesse do público-alvo de forma mais abrangente possível e, conseqüentemente, facilitar a recuperação da informação.

No escopo da Ciência da Informação,

pode-se afirmar que os campos de conhecimento trabalhados neste texto são identificados em importantes processos, tais como análise de assunto, interpretação, classificação, descrição, representação e recuperação de conhecimentos constantes de acervos de informações documentais, sejam estes compostos de referências, textos, imagens ou sons, em versão impressa tradicional ou em versão digitalizada. (ALVARENGA, 2003, p. 28).

Cabe salientar que o interesse da presente pesquisa está voltado aos aspectos da descrição que tratam da forma e do conteúdo de recursos informacionais, destacado no processo de Catalogação Descritiva.

No que se refere à utilização intensa dos procedimentos tecnológicos da informação e telecomunicação, a Biblioteconomia e CI, como um campo mais abrangente, são áreas do conhecimento que usufruem intensamente destes

suportes/recursos, pois, conforme as tecnologias do escopo da informação foram sendo criadas, inseridas, disponibilizadas e aperfeiçoadas os sistemas de representação e recuperação testemunharam a uma extrapolação dos limites das convencionais fichas catalográficas e, posteriormente, atingiram as bases de dados online. Esta novidade, em critério de disponibilidade tecnológica, foi responsável, há algumas décadas, por uma enorme transformação no ambiente dos sistemas de representação. (ALVARENGA, 2003).

[...] a Ciência da Informação é complexa e multidimensional e, por esta razão, deve ser intencionalmente abrangente. Sobre o exposto afirma: Ciência da Informação é a disciplina que investiga as propriedades e comportamento da informação [...]. Ela refere-se à origem, disseminação, coleta, organização, recuperação, interpretação, transmissão, transformação e uso da informação. Isso inclui a representação da informação em ambos os sistemas naturais e artificiais, o uso dos códigos para uma transmissão de mensagens eficiente e o estudo de mecanismos de processamento de informação e técnicas, tais como computadores e seus sistemas. (BORKO, 1968, p. 3 apud DANTAS; CÓRDULA; ARAÚJO, 2016, p. 42).

Afinal, a Biblioteconomia é uma disciplina independente ou possui paradigmas que competem à CI? Existem pontos de vista diferentes sobre esta questão. No intuito de evitar uma divisão desnecessária, é possível afirmar que a Biblioteconomia está mais voltada para o exercício profissional, executada tanto em ambientes de bibliotecas, como em serviços de informação. “A Ciência da Informação tem um vínculo muito forte com a Biblioteconomia, seja como complemento, substituição, ratificação ou rompimento com ela” (SOUZA, 2012, p. 51).

As áreas de representação e recuperação da informação são os campos da Ciência da Informação que mais fortemente buscaram estabelecer laços com a biblioteconomia ou, particularmente, com uma parte da biblioteconomia – justamente aquela ligada a instrumentos de descrição e classificação bibliográficas - e que mais fortemente impulsionaram, com isso, a designação da área como *library and information science* (ARAÚJO, 2014, p. 60).

No âmbito da Representação e das Tecnologias, está inserida a Catalogação Descritiva, a qual interliga elementos como editora, título, autoria, e ou-

tras informações acessíveis e identificáveis no documento. A Catalogação Descritiva é uma especialidade intelectual efetuada por profissionais da Biblioteconomia e da Ciência da Informação, pois, são utilizados procedimentos complexos, no intuito de subsidiar a descrição dos elementos presentes no documento.

De acordo com Santos (2008, p. 165), a Catalogação Descritiva é a

[...] disciplina e prática profissional que tem como missão construir as formas de representação para alimentação de catálogos a partir da descrição padronizada de recursos informacionais, contemplando sua forma, seu conteúdo e o seu arranjo em acervos, de modo a tornar a unidade informacional única e multiplicar os pontos de acesso para a sua identificação, localização e recuperação.

A Catalogação Descritiva é responsável pela construção e execução de maneiras de representação de recursos bibliográficos, assim como pelo Tratamento Descritivo da Informação (TDI) nas mais variadas esferas. As formas de catalogação podem ser divididas em três categorias: descrição bibliográfica, pontos de acesso e dados de localização (MEY, 1995).

A Catalogação Descritiva facilita o acesso aos códigos formais de representação simbólica e aos canais de transferência de informação. Está no seu escopo a codificação e a decodificação de regras e códigos utilizados nos sistemas de representação. Descrever os recursos informacionais de forma a representá-los em estruturas passíveis de armazenamento, auxilia e fornece possibilidades à recuperação da informação almejada pelo usuário.

Pode-se dizer que a catalogação descritiva abrange formas de descrição e de representação de recursos informacionais em ambientes informacionais (convencionais ou digitais), sendo que os códigos de catalogação nada mais são que instrumentos utilizados pelos catalogadores no processo de representação, os quais realizam a codificação dos dados bibliográficos do recurso informacional. Os códigos de catalogação fornecem diretrizes e regras para a descrição de recursos informacionais e para a construção de formas de representação que possibilitam o acesso aos documentos e às informações (CUNHA; CASTRO, 2018, p. 151).

No contexto dos ambientes informacionais digitais contemporaneamente, formas diferenciadas de representação para um tratamento mais efetivo das

informações são necessárias, a fim de atender aos requisitos de interoperabilidade.

Uma das formas mais efetivas para tratar as informações, no âmbito digital, gira em torno da representação por metadados⁸. Entre outros aspectos, os metadados são responsáveis por preservar e recuperar de forma adequada os recursos informacionais digitais.

Os metadados estão intimamente concatenados com a subárea de Representação da informação, pois estes fazem parte de todo o processo de codificação de determinado recurso armazenado num sistema de informação. Dentre outras características, os metadados são responsáveis pela recuperação da informação e são elementos do processo de catalogação descritiva. Além disso, os metadados de preservação digital constituem o fio condutor que relaciona tecnologia (no caso em questão os *Cloud Services*), inovações tecnológicas e o escopo da representação da informação.

As recentes ferramentas e meios tecnológicos, incluindo a *Internet*, estão incessantemente disponibilizando inesgotáveis fontes de informações e possibilitando o acesso da maior biblioteca do mundo nos mais diversos locais. A nova noção de tempo na era da *Internet* tem outra dinâmica, na qual, as velocidades das informações e da comunicação estão visivelmente mais aceleradas. Nesse sentido, o novo contexto da sociedade é desencadeado na necessidade em divulgar e receber informações em caráter de urgência (DANTAS; CÓRDULA; ARAÚJO, 2016).

Na década de 1990, a *Internet* foi a grande responsável pelas transformações que aconteceram no fluxo da informação e de canais de comunicação. Esta nova ferramenta tecnológica sofreu algumas resistências por parte de alguns profissionais que não estavam acostumados com essa nova inserção tecnológica, apresentando, assim, certa resistência às mudanças, porém, em curto espaço de tempo já era perceptível o aumento das produções acadêmicas e relatos técnicos, endossando, assim, a sua contribuição ao se aplicar as novas tecnologias do escopo da comunicação e informação no aperfeiçoamento e desenvolvimento do trabalho cotidiano (DANTAS; CÓRDULA; ARAÚJO, 2016).

⁸ Os metadados e os padrões de metadados serão tratados especificamente no capítulo 5.

"Os diferentes aspectos da inovação a tornaram um processo complexo, interativo e não linear. A interação é elemento fundamental para a inovação" (TOMAÉL; ALCARÁ; CHIARA, 2005, p.101).

A pluralidade de agentes contribui para que sejam geradas essas várias oportunidades de comunicação. Do mesmo modo, a pluralidade de distintas, mas complementares, relações cooperativas – mais do que a concentração em um tipo dominante de interação – é uma das principais fontes de inovação (ALBAGLI; MACIEL, 2004, p. 11).

A confecção da informação está sendo vinculada ao potencial tecnológico baseado na amplitude do acesso à informação, ou seja, ao se introduzir novos aportes tecnológicos e investimentos, se faz necessário a atuação constante de um profissional da informação para organizar e gerir a informação.

As inovações técnicas e tecnológicas possibilitaram a transferência de parte da inteligência humana para as máquinas, dotando-as de certa autonomia. A partir desse contexto, é plausível afirmar que a humanidade está redimensionando os domínios de trabalho e de renda e, conseqüentemente, se afastando progressivamente do seu espaço formal de trabalho. É a denominada Sociedade da Informação.

O cenário atual é configurado por uma parcela considerável de transferência dos saberes, da memória, do conhecimento produzido pelos indivíduos para uma rede inteligente de máquinas, alterando significativamente o modo de vida, pois, instantaneamente é processado um conjunto novo de informações que influenciam e definem o comportamento das pessoas, individualmente ou no âmbito coletivo (JESUS, 2012).

As atuais evoluções tecnológicas podem acarretar enormes transformações nos processos de comunicação, escrita, representação primária, secundária e recuperação de informações de conhecimentos. Seguindo a adaptação aos novos meios acessíveis, novas dinâmicas estão sendo criadas. Com a velocidade na qual os conhecimentos e as novas tecnologias avançam, o século XXI, provavelmente, ainda irá desfrutar de incontáveis inovações que vão impactar diretamente nos processos de reconhecimento, codificação, transmissão e recuperação de conhecimentos (ALVARENGA, 2003). "A inovação tecnológica é um processo complexo que envolve várias fases, desde a idéia inicial, ori-

ginada a partir de um problema ou uma oportunidade de negócio, até o desenvolvimento do produto ou processo e seu lançamento no mercado" (TOMAÉL; ALCARÁ; CHIARA, 2005, p. 101).

Um sistema de inovação pode ser definido como um conjunto de instituições distintas que conjuntamente e individualmente contribuem para o desenvolvimento e difusão de tecnologias. Tal noção envolve, portanto, não apenas empresas, mas, principalmente, instituições de ensino e pesquisa, de financiamento, governo, etc. (LASTRES; CASSIOLATO, 2000, p. 247).

No cerne das discussões que caracterizam a natureza da inovação, a teoria Joseph Schumpeter pode ser considerada como um marco na discussão dessa temática. A teoria de Schumpeter é baseada na relevância das inovações e dos avanços tecnológicos no desenvolvimento empresarial. O autor destaca que a inovação promove o ciclo econômico e, conseqüentemente, os elementos crédito, capital, juro e lucro compõem e permeiam todo processo (TOMAÉL; ALCARÁ; CHIARA, 2005).

A versatilidade da inovação é passível de ser interpretada tendo como ponto de partida as redes de comunicações. Entre outros aspectos, as novas concepções de inovação enfatizam as noções de processo e de interatividade. Seguindo esta lógica, os arranjos institucionais (universidades, indústrias e agências governamentais), que norteiam todo contexto das inovações, estão permanentemente se remodelando. É possível salientar que a inovação é resultado de uma série sucessiva de etapas em um *continuum* linear, isto é, o constante fluxo da inovação somente é possível a partir da sólida estrutura comunicacional, entre as entidades institucionais, as quais formam consistentes redes de conhecimento e se modificam, principalmente, conforme os contextos políticos e tecnológicos. Em todo caso, vale ressaltar que o capital físico e humano são núcleos centrais para determinar o desenvolvimento tecnológico e o conceito intitulado de *continuum* linear (CONDE; ARAÚJO-JORGE, 2003).

É importante sublinhar que os *Cloud Services* fazem parte de um contínuo ciclo de inovação tecnológica. Atualmente existem inúmeras empresas que fornecem um número considerável de opções para a preservação, a longo prazo, do armazenamento de dados. Por meio desse caminho, várias ferramentas estão sendo criadas e aperfeiçoadas constantemente. Novas empresas sur-

gem, oferecendo algum diferencial de produtos e serviços para se estabelecer no mercado concorrencial.

Continuamente existem peculiaridades distintas no escopo das ferramentas tecnológicas de cada *Cloud Services*, pois, podem existir várias instituições, públicas ou privadas, fomentando e trabalhando em conjunto para a melhoria da qualidade dos serviços e endossando as inovações tecnológicas, buscando a melhor maneira de aperfeiçoamento dessas tecnologias.

[...] dependendo da área de estudo, o termo inovação apresenta diferentes significados. Na área mercadológica, a inovação é considerada qualquer modificação percebida pelo usuário, mesmo não ocorrendo alteração física do produto. Na área produtiva, a inovação é a introdução de novidades, materializadas em produtos, processos e serviços, novos ou modificados. Outros enfoques relacionam a inovação a [idéias] de alto risco, proporcionando elevados benefícios à organização, que a desenvolve, ou processos que produzem fortes impactos econômicos à sociedade (BARBIERI, 1997 apud TOMAÉL; ALCARÁ; CHIARA, 2005, p. 101).

Há múltiplas abordagens e modelos que surgiram nas últimas duas décadas. Boa parte desses, exibem várias áreas de superposição e implementam concepções de inovação que apontem algum nível de consenso e homogeneidade. Redes de inovação formais e informais no ambiente científico, criaram novas redes mais formalizadas. O que implica a divisão de trabalho entre cientistas em graus e contextos institucionais abundantemente diferentes, os quais conectam o meio universitário com o industrial (CONDE; ARAÚJO-JORGE, 2003).

A procura progressiva pela inovação, através da elaboração e desenvolvimento de novos produtos e processos, diversificação, qualidade e absorção de tecnologias avançadas, pode ser considerada de extrema relevância para certificar excelentes níveis de eficiência, produtividade e competitividade das organizações. Nesse sentido, os fluxos de acumulação capitação tecnológica e de conhecimentos são incessantes e contínuos (TOMAÉL; ALCARÁ; CHIARA, 2005).

Segundo Baracho et al. (2015), o termo computação nas nuvens (*Cloud Computing*) também ganhou impulso nos últimos anos. Por se tratar de uma tecnologia inovadora, a computação em nuvem vem recebendo maior atenção

nos trabalhos acadêmico-científicos desenvolvidos nas instituições de pesquisa. Surgiu a fim de gerir e utilizar enorme quantidade de dados que se tornaram disponíveis com o desenvolvimento da Internet.

O uso da computação em nuvem tem apresentado vantagens significativas, as quais são destacadas por Kamala, Priya e Nandhini (2012) como a redução de custos, a grande capacidade de armazenagem, a escalabilidade, a ausência de instalação e manutenção de *software* e a acessibilidade dos serviços.

Serviços de armazenamento onipresentes e, muitas vezes, livremente disponíveis (*DropBox, Google Drive, iCloud*, por exemplo) são parte da vida de milhares de pessoas e carecem de organização e representação, como proporcionadas pela Ciência da Informação. Dessa forma, estudar os metadados de preservação digital em *Cloud Services* apresenta-se como importância capital para potencializar os aspectos das tecnologias, da representação e da inovação no cenário atual.

Assim, o próximo capítulo apresenta o cenário da preservação digital destacando a importância de salvaguardar os objetos digitais e as estruturas de armazenamento em ambientes informacionais digitais.

3 PRESERVAÇÃO DIGITAL DE LONGO PRAZO: ASPECTOS SALUTARES

Este capítulo visa abordar, de forma geral, aplicações e funções que fundamentam a preservação digital. Principalmente a partir da popularização da *Internet*, o desenvolvimento e o avanço incessante das tecnologias de informação e de comunicação gerou e tem gerado grande impacto informacional para a humanidade. A implantação constante de novas ferramentas tecnológicas interfere diretamente nas ações de preservação dos registros digitais. Diante dessa perspectiva, se torna necessário verificar os desafios que envolvem as atribuições da preservação digital.

No âmbito da preservação digital de longo prazo foram descritos os principais requisitos que os objetos digitais (ou mais especificamente seus componentes) precisam cumprir. Acessibilidade, autenticidade e identificação, são algumas das funções que os objetos digitais precisam dar conta em todo o seu ciclo de vida.

Portanto, neste capítulo, buscaremos fazer alguns apontamentos sobre as problemáticas que desafiam a preservação digital a longo prazo, como também destacar a importância de suas aplicações no ciclo de vida dos objetos digitais. Por último, buscaremos indicar a possibilidade de uma série de desdobramentos das aplicações dos metadados de preservação digital em *Cloud Services*, na esfera da preservação digital a longo prazo.

O emergente crescimento das informações disponibilizadas na *Web*, tendo como ponto de partida sua evolução maciça, passou a promover uma série de questionamentos dos profissionais da Ciência da Informação, os quais passaram a estudar novas alternativas e procedimentos inseridos no âmbito da preservação da informação (ALMEIDA; NASCIMENTO, 2011).

A *web* atual, caracterizada como espaço aberto e de fácil socialização do conhecimento, vem enfrentando mudanças significativas em sua estrutura e na forma de representar e apresentar os recursos informacionais digitais. Isto exige e pressupõe um novo olhar de diversos profissionais, especialmente, os da área de Ciência da Informação (CI) e de Biblioteconomia, que mantêm um papel *sine qua non*, como agentes transformadores, no século XXI. (CASTRO; SANTOS, 2010, p. 156).

A evolução das tecnologias de informação e comunicação desencadeou o aparecimento da denominada era digital, além de introduzir novas formas de organizar a informação e de criar a necessidade de construir novas estruturas que permitam o acesso seguro da informação até o final do seu percurso (ALMEIDA; NASCIMENTO, 2011).

A informação digital pode estar em qualquer lugar, e em qualquer mídia, sendo necessárias novas práticas na gestão do seu ciclo de vida, que se transforma constantemente. A dependência tecnológica das sociedades contemporâneas, junto com a percepção do risco de perder o registro material da produção científica e cultural, leva as instituições e indivíduos a procurarem formas de controle e de garantias de acesso em longo prazo (ARELLANO, 2012, p. 83-84).

A principal finalidade da preservação digital consiste em preservar o conteúdo intelectual de um objeto digital, porém, a compreensão do conteúdo é fragmentada e complexa. Do ponto de vista analógico, os documentos cumprem todo seu percurso de forma monolítica, pois todos os componentes que cerceiam suas atribuições (presença física, *layout*, formatos, ilustrações, conteúdo e suporte), são elementos virtualmente inseparáveis (SAYÃO, 2007).

Segundo Thomaz e Soares (2004), o problema central da preservação digital, tem raiz nos próprios objetos que se busca preservar. Ao contrário dos formatos analógicos, os objetos digitais são acessíveis por meio de combinações exclusivas de componentes *hardwares*, *softwares*, mídia e pessoal técnico.

A preservação de objetos digitais requer atenção contínua e constante, estes são fragmentados em componentes que podem correr o risco de não serem identificados pelas aplicações inseridas nas funções da preservação digital a longo prazo.

De modo incessante, formatos e programas digitais se tornam obsoletos e as aplicações de preservação não conseguem suportar ou reconhecer as atribuições contidas nos componentes, ou seja, o risco de perder o conteúdo de um documento digital é relativamente maior, comparado com um documento analógico.

Assim, a preservação digital combina políticas, estratégias e ações que visam garantir que os objetos digitais permaneçam autênticos e acessíveis aos

usuários e sistemas por um longo período de tempo (WITTEK; DARÁNYI, 2012).

Preservação digital a longo prazo pode ser definida pela capacidade de sustentar a acessibilidade, a compreensão e a usabilidade dos objetos digitais por longo tempo, ou seja, por gerações, independentemente das mudanças tecnológicas e das sucessivas modificações nas configurações das comunidades designadas que usufruem dos recursos desses objetos. Preservar e manter o acesso das informações digitais em longo prazo ainda é difícil, propenso a erros e financeiramente dispendioso (RABINOVICI-COHEN et al., 2011).

Uma das funções primordiais da preservação consiste em garantir a integridade do objeto digital, ou seja, preservar os aspectos que tornam o objeto como único. Ter ciência de como preservar, em termos operacionais, depende da considerável habilidade de segregar as características essenciais que precisam ser preservadas. A sua integridade não depende apenas do objeto em si, mas, sim, está vinculada aos variados tipos de atributos que são incorporados no decurso de seu ciclo de sua vida, como também ao grau de interesse dos indivíduos envolvidos (SAYÃO, 2007).

A dificuldade da preservação digital deve-se principalmente à necessidade de retenção do objeto digital e de seu significado. A carência maior está na definição de técnicas de preservação digital capazes de compreender a forma e a função original do objeto, para garantir sua autenticidade e acessibilidade, pois eles não são apenas objetos físicos. (ARELLANO, 2008, p. 23).

Os objetos digitais de longo prazo podem sofrer inúmeras ameaças, incluindo corrupção do conteúdo digital, ataques virtuais, mudanças organizacionais e obsolescência de *hardware* e *software*. A preservação digital pode ser dividida em preservação de *bits* e preservação lógica. A preservação de bits se define como a capacidade de recuperar *bits* diante da degradação ou obsolescência da mídia física, corrupção e destruição devido a erros e ataques maliciosos e, finalmente, por conta de catástrofes ambientais como incêndios e inundações. Já a preservação lógica está pautada na preservação da compreensão e da usabilidade dos dados, garantindo que o conteúdo em questão seja compreensível no futuro, independentemente das mudanças que podem ocorrer.

rer em servidores, sistemas operacionais, produtos de gerenciamento de dados, aplicativos e, até mesmo, usuários.

A preservação lógica deve manter a proveniência dos dados, em conjunto com sua autenticidade e integridade, no intuito dos sistemas atuais e futuros garantirem apenas o acesso dos usuários legítimos dos dados. É necessário que existam sistemas e processos de preservação que possam habilitar e suportar a preservação lógica e a preservação de *bits* (RABINOVICI-COHEN et al., 2011).

A preservação lógica busca nas ferramentas tecnológicas novos formatos atualizados para a introdução dos dados, novos *softwares* e *hardwares* que mantenham correntes seus bits, no intuito de conservar sua competência de leitura, além de estar coligada com a necessidade de assegurar a conversão de formatos originais obsoletos ou em dispendiosa manutenção (ARELLANO, 2004).

Ainda, no contexto da preservação lógica, quando um livro é preservado todas as suas características são salvaguardadas, em contrapartida, os componentes dos objetos digitais são facilmente fragmentados, divididos em elementos individuais, implicando num esforço complementar e significativo para identificar quais componentes devem ser preservados para assegurar a preservação total do objeto digital (SAYÃO, 2007).

Nas áreas da Ciência da Informação e da Biblioteconomia, a utilização de ferramentas tecnológicas no âmbito digital passa a substituir os tradicionais meios de preservação, levando em conta o grande número de indagações sobre normas que empregam técnicas digitais, além da contínua tarefa que objetiva salvaguardar os objetos digitais. Para trabalhar de forma adequada com materiais em formatos digitais, são constantemente formuladas normas de armazenamento e de compartilhamento, de modo a ter a necessidade de definir políticas institucionais de preservação (ARELLANO; SODRÉ, 2006).

"A preservação da informação em formato digital precisa de um conjunto de práticas técnicas e de gerenciamento que mudam constantemente" (ARELLANO, 2008, p. 24). A preservação digital deve ser vista como um conjunto contínuo e de múltiplos níveis de ferramentas (DECMAN; VINTAR, 2013).

A gestão de informações digitais, devido principalmente à sua transcendência temporal e à sua dependência aos compromissos de longo prazo, envolve ainda questões administrativas, legais, políticas, econômicas e financeiras e, sobretudo, questões referentes à representação das informações via metadados para acesso e gestão da preservação. Tudo isso somado, coloca a preservação digital na classe dos problemas de natureza complexa (SAYÃO, 2007, p. 184).

Os objetos digitais precisam de diversos níveis de compreensão e gestão, exemplificados pelos itens físicos, por uma codificação lógica. O objeto precisa ter significado para os seres humanos e salvaguardar um conjunto de elementos essenciais que ofereça, no futuro, a essência do objeto (ARELLANO, 2008).

Os procedimentos em torno da preservação digital se diversificam amplamente por causa da extensa variedade de tipos de objetos que precisam ser preservados. Por exemplo, textos, imagens e sons são constituídos por componentes diferentes e ao mesmo tempo podem fazer parte de um mesmo objeto digital (SAYÃO, 2007).

O aspecto heterogêneo dos dados é um fator que dificulta a preservação dos objetos digitais. Os dados podem se originar de uma cadeia de domínios especializados, a partir de uma variedade de *softwares* de aplicativos de execução, contidos em diversas plataformas. Entre outros elementos, a representação dos dados exemplifica esta temática, por se alocar em formatos distintos. A necessidade que gira em torno da preservação digital em longo prazo, na maioria das vezes, é acompanhada pela necessidade de prover acessibilidade aos objetos digitais, independente do tipo de *hardware* e *software* que os criaram originalmente (NGUYEN; LAKE, 2011).

A implementação de estruturas que visam à gestão de objetos digitais, no sentido de descrever as necessidades dos responsáveis por esses registros, despertam certas necessidades. A área de preservação exige mais dedicação, em comparação com outras áreas da Biblioteconomia e da CI, principalmente quando são observados os custos dispendidos com a manutenção dos recursos e pela adoção de medidas desempenhadas pelos gestores (ARELLANO; SODRÉ, 2006).

O número expressivo de desafios no centro da preservação digital provoca a necessidade de criar estratégias que mantenham a acessibilidade de

longo prazo da informação digital, respaldadas por métodos e ações que assegurem resolver questões relacionadas à autenticidade e à identidade do conteúdo (ALMEIDA; NASCIMENTO, 2011).

Até o presente momento - e provavelmente até num futuro indefinido - não teremos disponível uma estratégia única capaz de dar conta de todo o espectro de problemas relacionados à preservação digital. O que se apresenta são soluções específicas para casos específicos (SAYÃO, 2007, p. 193).

As tecnologias informacionais e comunicacionais precisam ser usadas como 'meio' e não como 'fim'. A automação foi criada por profissionais das áreas tecnológicas, tendo como foco uma perspectiva que mesclava eficácia e rapidez sem burocracia dos procedimentos sendo que praticamente foram criaram as ferramentas sem contato efetivo com os profissionais da Ciência da Informação e das áreas de documentação. A partir deste contexto, o objeto digital corre uma série de riscos, levando em conta que os próprios criadores das ferramentas, no geral, gerenciam os objetos digitais (INARELLI, 2011).

Conforme aumentam as inovações envolvendo tecnologias de informação e comunicação, aumentam às demandas de investigação na Ciência da Informação, por tal razão, esta área deve ter ou mesmo criar mais espaços que permitam e instiguem a potencialização das competências informacionais e, assim, originar a configuração de arquiteturas computacionais e informacionais no intuito de abarcar uma ampla inclusão na aprendizagem de metalinguagens ou na própria representação informacional (SANTOS; VIDOTTI, 2009).

A partir da massificação da *Internet* e das constantes inovações tecnológicas computacionais, a humanidade produz uma quantidade quase que ilimitada de dados digitais nunca vista antes na história, é uma verdadeira revolução que lança uma série de desafios voltados à preservação da memória, pois atualmente a maior parte dos dados gerados por grande parcela da humanidade é desenvolvida no formato digital. Por outro lado, há um possível risco das tecnologias desencadearem uma espécie amnésia digital coletiva, caso aconteçam falhas estratégicas no âmbito da preservação digital. Da mesma forma, as tecnologias digitais abrem um leque de possibilidades fantásticas (MARCONDES; SAYÃO, 2002).

Gradualmente os *Cloud Services* estão cada vez mais disseminados na sociedade. O crescente uso deste tipo de tecnologia está interligado, dentre outros aspectos, ao contexto de preservação dos diversos tipos de itens em ambientes digitais. Conforme a explosão informacional impulsiona quantitativamente o aumento de dados produzidos, surgem formas cada vez mais expressivas quanto a dependência da utilização dos servidores de nuvem.

Os *Cloud Services* podem oferecer benefícios relativamente atraentes: reduções de custos, eficiência, flexibilidade e escalabilidade, assim como, oportunidades para o desenvolvimento de serviços inovadores. Também traz riscos significativos associados à segurança, à privacidade, à integridade, à autenticidade, à acessibilidade e à continuidade digital de dados. Além disso, estão relacionadas à continuidade comercial e à falta de transparência dos *Cloud Services*, o que afetam a manutenção e o arquivamento de registros (STANCIC; RAJH; BRZICA, 2015).

As características desse ambiente são norteadas por uma série de perspectivas que precisam ser analisadas no âmbito da CI, incluindo o arcabouço de funcionalidades e problemáticas que relacionam preservação digital de longo prazo e *Cloud Services*.

Conforme já salientado, a composição dos objetos digitais é fragmentada em vários formatos de programas, diferentes combinações de *hardwares* e *softwares*, distintos tipos de dados e metadados, entre outros elementos que fazem parte da natureza do objeto. Nesse sentido, as aplicações da preservação digital não conseguem reconhecer os objetos que se tornaram obsoletos. A obsolescência consiste num desafio constante e incessante e pode tornar um objeto digital ineleável num curto período de tempo.

A manutenção dos objetos digitais precisa ser constantemente verificada e atualizada, empregando técnicas especializadas. O rápido e incessante avanço da tecnologia informacional, atualiza ou gera a todo o momento novos suportes, programas, componentes e aplicações, acarretando num infinito esforço em assegurar as funcionalidades do objeto.

Funcionalidades como a usabilidade, a acessibilidade, a autenticidade e a compreensão precisam ser aplicadas e desempenhas corretamente, caso contrário, o objeto digital pode se tornar obsoleto.

Diante desse cenário, há uma série de situações imprevisíveis, produzindo um ciclo interminável de estratégias de preservação. Conforme apontado, a principal finalidade da preservação digital de longo prazo consiste em garantir e salvaguarda os conteúdos intelectuais existentes nos objetos digitais por um longo período de tempo e, assim, permitir o acesso dos conteúdos no futuro .

O compromisso da preservação com o longo prazo é uma árdua e extenuante tarefa, a qual precisa de ininterruptas soluções. Nesse sentido, é necessária a implantação de uma ampla combinação de estratégias, pois, uma solução definitiva provavelmente jamais existirá. Todas essas conjunturas descritas categorizam a natureza complexa dos problemas da preservação digital a longo prazo.

Conforme será apontado nos próximos capítulos, a preservação digital em *Cloud Services* tem sido utilizada como solução ante a obsolescência tecnológica, pois de forma crescente, uma enorme quantidade de objetos digitais está sendo introduzida nesse ambiente.

O emprego dos metadados de preservação digital em *Cloud Services* pode contribuir como um elemento decisivo na aplicação das funcionalidades da preservação dos objetos digitais. Os metadados podem estar presentes nos ambientes de armazenamento, nos suportes, nos objetos digitais, nos componentes dos objetos, no percurso do ciclo de vida dos objetos etc. Dessa forma, se torna necessário verificar o contexto que insere os metadados de preservação digital em *Cloud Services* e, conseqüentemente, seus prováveis desdobramentos que garantam a preservação digital em longo prazo.

Assim, no próximo capítulo serão descritas as estruturas, as aplicações, as características e a definição do conceito de *Cloud Services* com base na Revisão Sistemática de Literatura.

4 ESTRUTURAS, APLICAÇÕES, CARACTERÍSTICAS E DEFINIÇÃO DO CONCEITO DE *CLOUD SERVICES* A PARTIR DA REVISÃO SISTEMÁTICA DE LITERATURA

O presente capítulo procura abordar as questões conceituais a respeito dos *Cloud Services*, identificando os elementos teóricos para o processo de preservação, bem como as características, as funcionalidades, as estruturas e as aplicações tecnológicas disponíveis no ambiente dos *Cloud Services*.

A palavra nuvem faz parte do cotidiano de milhares de usuários com infraestrutura tecnológica em todo o mundo. Existe uma notória tendência para que o número de usuários e empresas venham a utilizar esta expressão por longo tempo.

Entretanto, a palavra nuvem, é um conceito criado por profissionais da Tecnologia da Informação (TI) e que, para a maioria dos usuários, aparenta ser um conceito abstrato. Em muitos casos soa como algo imaterial e impalpável. Para a maioria dos usuários desses serviços, simplesmente basta terceirizar os seus respectivos acervos pessoais em estruturas de armazenamentos ofertados por empresas de *Cloud Services*, limitando-os apenas a reflexões abstratas ou mesmo tendo a simples convicção que ao acessar determinado serviço de *Cloud Services*, os conteúdos informacionais dos respectivos acervos pessoais serão facilmente recuperados assim que o usuário desejar.

A gênese do termo dependeu de múltiplas fontes e não há um consenso estabelecido sobre ele, entretanto, o termo só se popularizou a partir de 2006 quando a empresa *Amazon* lançou o serviço *Elastic Compute Cloud*. É importante salientar que a computação em nuvem é resultado de um conjunto de abordagens, compartilhando muitas de suas características com outros conceitos, como por exemplo: computação em névoa, grade computacional, modelo cliente-servidor, modelo par-a-par⁹, agregados computacionais, entre outros (DUTRA; SANT'ANA; MACEDO, 2016, não paginado).

⁹ Par-a-par ou *peer-to-peer*: contexto onde o usuário, além de consumidor da informação passa a ser o provedor, sinônimo de arquitetura 'descentralizada', como forma de aumentar a escalabilidade. (CECIN, 2005).

No contexto do *Big Data*¹⁰, diante da capacidade espacial das tecnologias de armazenamento passar de *Kilobytes*¹¹ para *Zettabytes*¹², os provedores de nuvem certamente poderão ser importantes aliados para a definição de preservação digital. Porém, se não houver planejamento, acervos inteiros podem desaparecer nos armazenamentos dos *Cloud Services* (GLUSKHO, 2013).

Mesmo que o armazenamento em nuvem traga uma infinidade de possibilidades para preservar os dados, o profissional da informação terá que predefinir, no plano de preservação, mecanismos que auxiliarão na gestão do sistema de preservação em nuvem e que evitará a perda de qualquer objeto digital (PONTES, 2014, p.86).

É possível afirmar que a informação digital pode estar localizada em qualquer lugar, em qualquer tipo de mídia, nesse sentido são necessárias novas práticas na administração do ciclo de vida dos objetos digitais, os quais se transformam constantemente. Há uma grande dependência tecnológica das sociedades contemporâneas, e o risco de perder o registro material da produção científica e cultural leva instituições e indivíduos a procurarem diferentes maneiras de controle e de garantias de acesso em longo prazo. Preservar documentos digitais é uma tarefa dispendiosa, gerando esforço extra na manutenção das características originais para servir como registro da comunicação humana (ARELLANO, 2012).

A utilização da tecnologia em nuvem, por instituições públicas e privadas está notoriamente em ascensão, conseqüentemente há uma grande adoção e aceitação desta infraestrutura nos mercados tecnológico e empresarial, além dos usuários comuns que utilizam esta tecnologia para as mais diversas finalidades. O governo brasileiro já aderiu à nuvem por meio da criação do Serviço

¹⁰ "O barateamento das máquinas computacionais (processamento e memória) e dos dispositivos de captura e armazenagem de dados (sensores, câmeras fotográficas e de vídeo, *celulares*, *pen-drives*, *flash memory*, discos rígidos externos, etc.), criaram inimaginável quantidade de dados, que estão sendo disponibilizados na web, proporcionando a formação da 'Era do *Big Data*'" (LIMA JUNIOR, 2011, p. 49-50).

LIMA JUNIOR, W. T. Jornalismo computacional em função da "Era do Big Data". **LÍBERO (FACASPER)**, n. 28, p. 45-52, 2016.

¹¹ *Kilobytes*: equivalente a 1024 *bytes*. (GERALDI, 2009).

GERALDI, L. M. A. **Elucidando Os Sistemas Operacionais: um estudo sobre conceitos**. Clube de Autores, 2009.

¹² *Zettabytes*: equivalente a 1.000.000.000.000.000.000.000 de *bytes*, isto é, um sextilhão de *bytes*. (SANTANA, 2015).

Federal de Processamento de Dados (SERPRO). Com a criação deste serviço, os servidores federais podem executar suas tarefas de maneira mais dinâmica, pois a interoperabilidade dos serviços será processada sobre o mesmo ambiente (PONTES, 2014).

Conforme já afirmado, uma das consequências dos avanços tecnológicos está diretamente relacionado com a obsolescência tecnológica, tal realidade faz com que o profissional da informação procure contínuas estratégias de preservação digital para garantir, entre outros aspectos, a acessibilidade e recuperação dos objetos digitais alocados nas mais diversas instituições públicas e privadas.

Os conteúdos preservados precisam passar por atividades que garantam a acessibilidade, armazenamento em longo prazo e interpretação de conteúdos, quando necessário. Esse tipo de gestão leva em consideração todos os tipos de conteúdo e de formato de objetos digitais. A criação de repositórios, bibliotecas e arquivos digitais que subsidiam a produtividade intelectual de uma ou várias áreas do conhecimento vem considerando algumas resoluções internacionais como a Carta sobre *la preservación del patrimonio digital* [2003], e As Recomendações sobre *Software Livre para Repositórios e Sistemas de Preservação* [2007] da *Unesco* (ARELLANO, 2012, p. 83).

Entretanto, armazenamento na nuvem não atende a todos os requisitos funcionais necessários para a preservação digital em longo prazo, somente se forem incorporados serviços adicionais; uma vez que uma nuvem de armazenamento adiciona os recursos adicionais, se torna um *Cloud Digital Preservation Service* (FRANKS, 2015), o qual, traduzido para o idioma português, significa serviços de nuvem de preservação digital ou, de forma sucinta e resumida, serviços de preservação.

A questão da preservação digital neste ambiente deve ser criteriosamente analisada, pois já ocorreu a falência, o desaparecimento ou mesmo a venda de provedores de nuvem e, por estas razões registros foram perdidos, danificados de forma irrecuperável, ou utilizados de maneira indevida. Outro ponto importante a ser ressaltado é a impossibilidade de identificar a localização geográfica das ferramentas tecnológicas onde estão armazenados os recursos, pois, vários *Data Centers* não revelam os endereços das instalações, podendo ocasionar amplas discussões jurídicas sobre várias problemáticas, entre elas,

por exemplo, a autenticidade e a garantia de permanente preservação dos recursos digitais armazenados. A utilização da nuvem como um armazenamento incessante de dados não se interromperá e, neste sentido, é necessário criar esquemas para reduzir os riscos ou danos, num patamar aceitável (RIC, 2014).

Em termos práticos, os *Data Centers* são instalações construídas e projetadas, visando a otimização de recursos como: comunicação [conexão e equipamentos de rede], armazenamento [suporte], ambiente [refrigeração, energia, segurança], conhecimento [competências específicas para manutenção e operacionalização], centralizando, assim, custos e recursos (DUTRA; SANT'ANA; MACEDO, 2016, não paginado).

Compreender os requisitos de funcionamento dos *Cloud Services* possibilita nortear os processos que serão necessários desenvolver no âmbito da preservação digital em Ciência da Informação. Conforme verificado pela RS, ainda não existem muitas pesquisas que exploram o tema. Pesquisar sobre preservação digital em provedores de nuvem pode ser considerada uma tarefa complexa, pois estão permeados por conceitos, estudos práticos e teóricos, pautadas na perspectiva de diferentes tipos de ferramentas tecnológicas.

Independentemente dos desafios que norteiam a preservação digital em longo prazo nesse tipo de ambiente, os *Cloud Services* podem consistir numa estratégia viável ante a obsolescência tecnológica dos objetos digitais. Muitas instituições estão migrando os objetos digitais dos seus repositórios institucionais para os *Clouds* (ou alocando de forma mesclada nesses dois ambientes), visando garantir o acesso e recuperação dos conteúdos intelectuais por longos períodos de tempo.

Este contexto pode ser exemplificado através das estratégias de preservação digital de registros de áudio (história oral) da coleção *Jeffco Stories*, pertencente à Biblioteca Pública do Colorado, esta instituição utilizou uma plataforma de acesso aberto e hospedou os registros num *Cloud Services*. Esta medida em questão ofertou um grau básico de preservação digital por um custo baixo. Antes de digitalizar os objetos analógicos e após empregar a migração do repositório para os *Cloud Services*, a instituição criou um projeto no intuito de proporcionar acessibilidade de suas coleções aos membros da comunidade local. Inicialmente a biblioteca não possuía infraestrutura para a dos registros, e

nem mesmo garantiam a preservação em longo prazo, após hospedar os registros digitalizados num repositório institucional, e com o passar dos anos hospedar o acervo num *Cloud Service*¹³, a biblioteca conseguiu assegurar o acesso dos registros de áudio para os membros da comunidade. (MATUSIAK; NEWTON; POLEPEDDI, 2017).

O Escritório Federal de Segurança em Tecnologia da Informação da Alemanha, também aderiu aos *Cloud Services* a partir de um serviço de suporte para plataformas de armazenamento que visam segurança em longo prazo para documentos 'oficiais', o denominado *EDS (Electronic Document Safe)*. Esta instituição alemã armazena documentos com assinatura digital em *Cloud Services* visando assegurar a preservação em longo prazo. Na mesma direção, o primeiro sistema segurança que trabalha com assinaturas eletrônicas da Lituânia, o denominado *e-Servicing System of the Insures (EDAS)*, utiliza um modelo de arquivamento em *Cloud Services* que suportam a preservação em longo prazo, o modelo em questão foi criado para gerar segurança aos registros eletrônicos.¹⁴ (STANCIC; RAJH; BRZICA, 2015).

Além de fornecer softwares e aplicativos, os quais congregam os elementos de processamento, a tecnologia *Cloud* gera uma amplitude dos meios estratégicos para armazenamento de dados. Dessa maneira, os dados podem ser armazenados nos servidores físicos, podendo ainda ser copiados como *Backups*¹⁵ nos *Data Centers*, nos quais apenas os servidores acessam os arquivos. É possível afirmar que após o advento dos *Clouds*, as possibilidades de preservar o objeto digital foram ampliadas devido ao poder computacional ofertado por essa tecnologia (PONTES, 2014).

Além de instituições governamentais, universidades e empresas que necessitam empregar um armazenamento que contemple a preservação em longo prazo, inúmeros usuários utilizam os *Cloud Services* a partir diversos tipos de aparelhos computacionais. Muitos usuários comuns também almejam obter um ambiente que possibilite a preservação em longo prazo de seus objetos

¹³ Mais explicações sobre o contexto em questão estão inseridas no capítulo 6.

¹⁴ Mais explicações sobre o contexto em questão estão inseridas no capítulo 6.

¹⁵ Consiste na cópia de dados em um meio separado do original, com a finalidade de protegê-los de qualquer infortúnio, ou seja, essa cópia objetiva a segurança dos dados. Disponível em: < <https://www.hardware.com.br/termos/backup>>. Acesso em: 11 jul. 2018.

digitais pelos mais diversos motivos. Com o aparecimento das tecnologias móveis, existe uma dependência cada vez maior das pessoas em relação aos *Cloud Services*.

Na segunda metade da primeira da década do século XXI, os usuários dos *Cloud Services* utilizavam os serviços em ferramentas tecnológicas com pouca ou praticamente nenhuma mobilidade, a partir da segunda década do século XXI, surgiram, e posteriormente, houve a massificação dos *smartphones*¹⁶. Devido aos incessantes aprimoramentos do *hardware* e *software* desses aparelhos, e o lançamento de sofisticadas plataformas (*iPhone*¹⁷ e *Android*¹⁸), além da alta qualidade na transmissão de dados via *Internet* 3G e 4G, assim como a onipresença das redes de *Wifi*¹⁹, ocorreu o fornecimento de melhores serviços para os clientes de *Cloud Services*. Agora, os usuários de *smartphones* podem colaborar e compartilhar informações com facilidade (FLORES; SRIRAMA; PANIAGUA, 2011).

Os *Clouds* contêm inúmeros atributos para o campo de Ciência da Informação, pois com esta tecnologia, entre outros aspectos, é possível potencializar o armazenamento de dados, o tráfego de dados em uma rede e, ampliar a segurança ao acesso. Para que os serviços em nuvem seja um trunfo ao ser manipulado pelo profissional da informação, este precisa discernir as estratégias e modelos que podem ser adotados no seu plano de preservação, e dessa forma, utilizar todo o potencial dessa ferramenta tecnológica (PONTES, 2014).

¹⁶ Na tradução literal significa 'telefones inteligentes', esta ferramenta tecnológica é a evolução do celular, as funções desse aparelho permitem uma infinidade de possibilidades, existem diversos tipos de modelos. Os *smartphones* podem ser considerados uma mistura entre celulares e computadores, entretanto não possuem *hardware* potente comparados a um computador doméstico, porém não são tão simples como um telefone. Disponível em: <<https://www.techtudo.com.br/artigos/noticia/2011/12/o-que-e-smartphone-e-para-que-serve.html>>. Acesso em: 11 jul. 2018.

¹⁷ Aparelho desenvolvido pela empresa *Apple*, capaz de entrar na *internet*, de tirar fotos, enviar *e-mails*, reproduzir músicas e vídeos, faz *downloads* de aplicativos, funciona como leitor de livros, entre outras funcionalidades. Pode ser considerado como o símbolo da convergência digital. Disponível em: <<https://www.techtudo.com.br/artigos/noticia/2012/02/o-que-e-iphone.html>>. Acesso: 11 jul. 2018.

¹⁸ Sistema operacional baseado em Linux que opera, entre outros tipos de aparelhos, nos *smartphones*. Disponível em: <<https://www.significados.com.br/android/>>. Acesso em: 11 jul. 2018.

¹⁹ Abreviação de '*Wireless Fidelity*', em português significa fidelidade sem fio. *Wi-fi*, ou *wireless* é uma tecnologia de comunicação que não faz uso de cabos, e geralmente é transmitida através de frequências de rádio, infravermelhos etc. Disponível em: <<https://www.significados.com.br/wi-fi/>>. Acesso em: 13 jul. 2018.

A tecnologia *Cloud* está emergindo como uma infraestrutura adequada para a construção de sistemas grandes e complexos. Os recursos de armazenamento e computação provisionados a partir de infraestrutura convergente apresentam uma alternativa econômica ao tradicional centro de dados interno.

O emergente armazenamento em nuvem é altamente distribuído, escalável e oferece disponibilidade, acessibilidade e compartilhamento de dados com valores monetários relativamente baratos. No entanto, as plataformas de armazenamento em nuvem são projetadas para uso geral e não são especialmente adaptadas às necessidades de preservação (RABINOVICI-COHEN et al., 2011).

Para o *National Institute of Standards and Technology*²⁰ (NIST), tal modelo de ferramenta tecnológica autoriza o fornecimento de muitos recursos computacionais, determinando que estes recursos sejam disponibilizados e liberados com maior rapidez com esforços de administração e de relação do fornecedor relativamente minimizados. (MACHADO, 2013).

Os *Cloud Services* oferecem novos patamares de escalabilidade²¹, elasticidade e disponibilidade, os quais permitem acesso simples aos dados armazenados independente do local ou dispositivo. A nuvem provê um modelo de dados dos objetos armazenados que incluem dados integrados com seus respectivos metadados definidos pelo usuário e pelo sistema como uma única unidade (RABINOVICI-COHEN et al., 2013).

Para que os *Cloud Services* suportem tamanha quantidade de dados armazenados é necessário que exista uma infraestrutura física, a qual se interliga gerando um complexo sistema computacional, denominado *Cloud Computing* (Computação em nuvem). De forma geral, *Cloud Computing* pode ser compreendido como uma infraestrutura em nuvem que coleciona *hardwares* e *softwares*, permitindo o desempenho das características essenciais.

²⁰ Instituto Nacional de Padrões e tecnologia, traduzido para língua portuguesa. Consiste num dos mais antigos laboratórios de ciência física dos Estados Unidos. Centra-se em recomendar padrões para indústrias e agências governamentais, numa ampla variedade de áreas. Disponível em: < <https://www.commerce.gov/doc/national-institute-standards-and-technology#4/37.55/-91.10>>. Acesso em: 01 jul. 2018.

²¹ No escopo que contextualiza telecomunicações, engenharia de *software* e infraestrutura de tecnologia da informação, o conceito de escalabilidade é definido como uma característica de um sistema, de rede, ou de um processo que aponta sua capacidade de manipular uma parcela emergente de trabalho de forma uniforme, sendo organizado e preparado para ascender, crescer. Isso significa que um sistema é escalável partir do momento que a carga total dos recursos (geralmente *hardware*) é requerida. (BONDI, 2000).

Afinal, o que é *Cloud Computing*? Este conceito ainda se aprimora. O objetivo inicial do Cloud Computing visou processar informações e armazenar os dados fora do ambiente corporativo, dentro da grande rede, nas estruturas dos *Data Centers*. A nuvem é um conjunto de grandes pontos de armazenamento e processamento de dados de informação (VERAS, 2012). Verifica-se que o conceito de *Cloud Computing* é abrangente e ao mesmo tempo relativamente confuso devido às inúmeras indefinições que compreendem a pluralidade terminológica empregados em torno do conceito de *Cloud*.

É possível afirmar que na composição do *Cloud Computing*, o *hardware* é a camada física. Os recursos de *hardware* são necessários para se interligar e suportar os *Cloud Services* que estão sendo fornecidos, como por exemplo, componentes de servidor, armazenamento e rede. A camada de abstração consiste no *software* implantado em toda a camada física, isto é, toda infraestrutura que abarcam as características essenciais dos *Cloud Services*. Conceitualmente, a camada de abstração fica acima da camada física (MELL; GRAN-CE, 2011).

A arquitetura dos *Cloud Services* geralmente é representada por camadas compostas por *Infrastructure as a Service*²² (*IaaS*), *Platform as a Service*²³ (*PaaS*) e *Software as a Service*²⁴ (*SaaS*) (MACHADO, 2013). Outra camada menos conhecida é a camada de preservação digital²⁵, esta pode assumir várias categorias e diferentes terminologias. Já existem alguns *Cloud Services* que oferecem este serviço, há poucos estudos sobre as camadas de preservação como serviço, pois ela foi desenvolvida recentemente. A camada de preservação acopla diversas funcionalidades para manipular e acrescentar informações aos objetos digitais, os quais se tornam pacotes de informações. A sua principal função visa interligar sistemas de arquivamento nos sistemas *Interaction Layer* (Interação de Camadas) e *SaaS* (ASKHOJ; SUGIMOTO; NAGAMORI, 2011).

²² Infraestrutura como Serviço.

²³ Plataforma como Serviço.

²⁴ Software como Serviço.

²⁵ A partir do capítulo 6 são explanadas mais especificações sobre esta camada, porém ainda existem poucos sobre esta temática, os poucos documentos encontrados faz parte do arcabouço da RS. Futuras pesquisas sobre a camada de preservação são necessárias e pertinentes.

A variedade de tecnologias empregada nos *Clouds* gera uma confusão conceitual. Além disso as campanhas publicitárias em torno do termo *Cloud* confunde ainda mais a mensagem sobre cada tipo de domínio, ferramenta e ambiente (VAQUERO et al., 2008).

As questões elencadas nos cinco parágrafos acima endossam a imprecisão terminológica sobre *Cloud Computing* e *Cloud Services*, afinal a estrutura de *Cloud Computing* faz parte de um complexo sistema físico aonde os dados são alocados nos denominados *Data Centers*, ofertando suporte e espaço para o desenvolvimento das ações utilizadas a partir das três principais camadas dos *Cloud Services*. No domínio dos *Cloud Services* o usuário e instituições públicas ou privadas tem a possibilidade de interagir e configurar suas ações conforme a necessidade do seu respectivo interesse, ou seja, um usuário somente conseguirá manipular suas ações no ambiente compreendido como *Cloud Services*. As estruturas do *Cloud Computing* são fundamentais para que ocorra a interação entre usuário e *Cloud Services*, porém diretamente não existe interação entre usuário e *Cloud Computing*.

Nesse sentido, é possível entender que há uma correlação intrínseca entre *Cloud Computing* e *Cloud Services*, pois para que os *Cloud Services* suportem tamanha quantidade de dados armazenados é necessário que exista uma infraestrutura física, a qual se interliga e gera um complexo sistema computacional, denominado *Cloud Computing* (MELL; GRANCE, 2011).

Conforme as tecnologias informacionais e comunicacionais avançam a confusão terminológica aumenta, em certo sentido tais tecnologias surgem e, rapidamente são aprimoradas ou substituídas antes de definições consensuais dos conceitos terminológicos, ocasionando uma certa indefinição das diferenças entre, por exemplo, *Cloud Computing* e *Cloud Services*, porém são ambientes distintos, que se inter-relacionam, mas não são necessariamente a mesma coisa. É pertinente afirmar que *Cloud Computing* e *Cloud Services* desempenham distintas funções.

Cloud Services se baseiam na sensação ilusória de uso infinito dos recursos disponíveis sob a demanda. Desse modo, o usuário não precisa projetar outras formas de configuração, isto é, de provisionamento. Os usuários e organizações que se utilizam dos *Cloud Services* podem usar e/ou ocupar o espaço de armazenamento conforme suas respectivas necessidades, ou seja, os com-

promissos financeiros dos usuários giram a partir do que é de fato utilizado, seja por dia, seja por hora, e os recursos considerados sem utilidade podem ser facilmente descartados, sem custos posterior ao descarte (MACHADO, 2013).

Há quatro modelos de serviços em nuvem: a) privada; b) pública; c) comunitária; e, d) híbrida.

Uma nuvem privada se refere a um centro de dados interno de alguma organização, podendo ser empresarial, estatal etc. Significa que não está disponível para o público. Serviços privados de nuvem são capazes de hospedar aplicativos ou máquinas virtuais em um conjunto de *hosts*, de propriedade da instituição. A infraestrutura é da própria organização, ou seja, uma nuvem privada (*Private Cloud*) é muito semelhante a uma rede. As nuvens privadas oferecem o maior controle, uma vez que são provisionadas para o uso de uma única organização. Podem ser de propriedade e gerenciados pela organização ou administrada de forma terceirizada e, por fim, podem estar geograficamente dentro ou fora das instalações (FRANKS, 2015).

O modelo de nuvem pública (*Public Cloud*) disponibiliza uma infraestrutura distribuída de forma que só é pago aquilo que de fato o usuário utilizou. Neste aspecto, um usuário pode fornecer de forma dinâmica qualquer quantidade de recursos de computação aos serviços de nuvem, pois irá pagar apenas pelo serviço demanda de consumo. Geralmente os custos são considerados baixos (WITTEK; DARÁNYI, 2012).

A nuvem comunitária (*Community Cloud*) foi criada para moldar os recursos subutilizados das máquinas de usuários, onde todas as pessoas envolvidas nos papéis de consumidores, de vendedores e principalmente de coordenadores geram um ciclo de sustentabilidade (MARINOS; BRISCOE, 2009).

Por fim, a nuvem híbrida (*Hybrid Cloud*) sugere a combinação da implantação de modelos de nuvem existentes, por exemplo, uma empresa pode mesclar o modelo de nuvem pública com o modelo de nuvem privada; a nuvem híbrida está tendo um crescimento acelerado, pois oferece às organizações opções de gerir dados e aplicativos por meio de ambientes de nuvem privado ou público, conforme a necessidade do usuário envolvido (FRANKS, 2015).

No **quadro 4**, serão exemplificados os principais *Cloud Services* e seus respectivos serviços.

QUADRO 4. Principais provedores de nuvem e seus respectivos serviços.

<i>Provider</i>	<i>IaaS</i>	<i>PaaS</i>	<i>SaaS</i>
<i>Amazon</i>	<i>Amazon EC2 Amazon S3</i>	<i>Amazon Elastic Beans- talk</i>	<i>Available on AWS Market- place</i>
<i>Google</i>	<i>Google Compute</i>	<i>Google App Engine</i>	<i>Google Apps</i>
<i>HP</i>	<i>HP Public Cloud</i>	<i>HP Helion Public Cloud Application Plataform as a Service</i>	<i>Available on HP Software Experience Center</i>
<i>IBM</i>	<i>IBM Cloud Managed Services SoftLayer</i>	<i>IBM Bluemix</i>	<i>IBM Solution Provider - IBM Software as a Service</i>
<i>Oracle</i>	<i>Oracle Compute</i>	<i>Oracle Cloud PaaS</i>	<i>Oracle Applica- tions Cloud; Oracle Analyt- ics Cloud; Ora- cle ERP Cloud etc.</i>
<i>Rackspace</i>	<i>Managed Infrastructu- re</i>		
<i>Salesforce.com</i>		<i>Salesforce1 platforms</i>	<i>Sales force automation and CRM²⁶</i>

Fonte: Doinea e Pocatilu (2014, p.102).

Abaixo serão exemplificados as funções de todos os serviços apontados no quadro 4:

- *Elastic Compute Cloud (EC2)*: sistema responsável pelo gerenciamento da execução de aplicações na infraestrutura da *Amazon*, permitindo controle completo das instâncias de sistemas, podendo acessar e interagir com cada uma destas instâncias de maneira bem semelhante em comparação as máquinas convencionais. (SOUZA; MOREIRA; MACHADO, 2009).

- *Amazon Simple Storage Service*: consiste num armazenamento para a *Internet*, facilita a computação de escala na *web* para os desenvolvedores, possui uma interface que armazena e recupera qualquer quantidade de dados, a qualquer momento, em qualquer lugar da *web* (AMAZON S3, 2018).

²⁶ Esse ambiente abarca colaboração, gestão de contatos, ferramentas de marketing, venda de dados, mobilidade, workflows aprovados, entre outras características. Disponível em <<https://www.salesforce.com/products/sales-cloud/features/>>. Acesso em 10 jul. 2018.

- *Amazon Elastic Beanstalk*: há mais de cem serviços, cada um dos quais expõe uma área de funcionalidade, o qual oferta uma variedade de serviços com flexibilidade para o usuário decidir como gerenciar sua infraestrutura, o qual pode consistir num desafio para o usuário usá-los e provisioná-los. Nesse serviço, é possível implantar e gerenciar (sem restringir as escolhas ou controle) rapidamente os aplicativos sem se preocupar com a infraestrutura que executa esses aplicativos (AWS, 2018).

- *Available on AWS Marketplace*: auxilia o usuário em localizar, comprar e começar imediatamente a usar o *software* e os serviços necessários para criar produtos e administrar seus negócios. (AWS, 2018).

- *Google Compute*: oferta balanceamento de carga de dimensionamento de instâncias individuais para instâncias globais. (GOOGLE COMPUTE ENGINE, 2018).

- *Google App Engine*: facilita o uso de idiomas, bibliotecas e *frameworks*. (GOOGLE SUITE, 2018).

- *Google Apps*: pacote com inúmeros aplicativos ofertados pelo Google. Dentre as principais características estão: armazenamento de dados, gerenciamento de dispositivos, migração segura de dados etc. (GOOGLE APPENGINE DOCS, 2018).

- *HP Public Cloud*: a HP abandonou sua nuvem pública, este *Cloud Service* se tornou inexistente. (NETWORK WORLD, 2015).

- *HP Helion Public Cloud Application Platform as a Service*: integra diversos ambientes de nuvem pública, por ser flexível auxilia o usuário a criarem seus próprios aplicativos portáteis, além de contém funcionalidades de uma nuvem híbrida. (ALPHARETTA, 2018).

- *Available on HP Software Experience Center*: ambiente que permite soluções gráficas para a impressão de selos postais, o qual envolve sua construção, entre outros aspectos. (HP, 2018).

- *IBM Cloud Managed Services SoftLayer*: consiste num conjunto de serviços elaborados para aplicativos empresariais. Serviço de nuvem, prontos para a produção, rico em segurança. (IBM SERVICES, 2018).

- *IBM Bluemix*: contém recursos que possibilitam reunir diversas origens de dados, escalar sistemas e incorporar serviços cognitivos para impul-

sionar o valor de negócios de forma rápida e econômica. Consiste numa implementação híbrida de *Cloud Service*. (BLUEMIX, 2018).

- *IBM Solution Provider - IBM Software as a Service*: o acesso ao software fornece suporte aos negócios dos usuários, no qual o usuário direciona a cotação, contrata serviços e recebe a fatura de seus clientes. É considerada pelos usuários da IBM uma plataforma sustentável, o qual oferta anuidade de elevados valores de negócios, além desse ambiente oferecer flexibilidade para o usuário trabalhar em equipe e com vendas para trabalhar com negócios de forma autônoma. (IBM SOFTWARE AS A SERVICE, 2018).

- *Oracle Compute*: oferecem recursos de computação em muitas configurações, se adaptam a uma variedade de cargas de trabalho e arquiteturas de *software*, suportam armazenamento em blocos remoto, em contrapartida também trabalham com configurações de alta densidade de entradas e saídas no armazenamento, e alta capacidade de armazenamento. (ORACLE CLOUD INFRASTRUCTURE, 2018).

- *Oracle Cloud PaaS*: Consiste numa combinação ampla e padronizada, mesclando as tecnologias da Oracle e de código-fonte aberto. Elaborado para permitir que o usuário desenvolva, implante, migre, proteja e gerencie todos os aplicativos empresariais. (ORACLE CLOUD PLATFORM, 2018).

- *Oracle Applications Cloud*: Este ambiente possibilita ao usuário a criação de produtos e serviços diferenciados, além de possibilitar participar de novos mercados, e conseqüentemente atendendo à demanda global. Simultaneamente contempla operações relativas aos negócios que abarcam as tecnologias da informação a partir de custos e complexidade baixa. (ORACLE CLOUD APPLICATIONS, 2018)

- *Oracle Analytics Cloud*: descobre e exibe os dados, facilitando o trabalho do usuário de acordo com os dados apontados, trabalha com visualizações e cria histórias atraentes, potencializa os recursos de voz para obter respostas, entre outros recursos. Por meio de usuários, dados e sistemas esse ambiente permite que o usuário compreenda e norteie suas ações com base nas informações. (ANALYTICS CLOUD, 2018).

- *Oracle ERP Cloud*: acelera os processos de negócios empresariais com a ferramenta *Enterprise Resource Planning (ERP) Cloud*, criando um serviço de contabilidade de sistemas diferentes, unificando-os para que exista

controle completo e integral da gestão financeira, entre outros recursos e aplicativos. (ORACLE ERP CLOUD, 2018)

- *Managed Infrastructure*: Este ambiente é caracterizado pela assistência para planejar, projetar e arquitetar, e após instalada a infraestrutura os técnicos da *RackSpace* mantém a nuvem constantemente atualizada. Nos seus serviços constam: conselhos de segurança; emissão de relatórios; sistema de integridade e *status* de violação de limites; autoconfiguração, monitoramento e execução de *backups* automáticos; autoconfiguração do escalonamento automático; serviços de aplicativos; monitoramento de desempenho etc. (RACKSPACE OPEN CLOUD, 2018).

- *Salesforce1 platforms*: Consiste numa plataforma que permite o desenvolvimento de aplicativos e a troca de dados através de interfaces de programação de aplicativos (*application programming interfaces - APIs*) e componentes de código de programação pré-construídos. (SALESFORCE1 PLATAFORM, 2018).

- *Sales force automation and CRM*: Esse ambiente abarca colaboração, gestão de contatos, ferramentas de *marketing*, venda de dados, mobilidade, *workflows* aprovados, entre outras características. (SALES CLOUD CRM, 2018).

Na prática uma empresa ou um determinado usuário pagam pelo serviço de armazenamento de dados sem necessidade de preocupar com atualizações ou *backups*. De uma forma geral, embora a nuvem represente certo perigo, por serem cada vez mais utilizadas para guardar dados digitais, o futuro da preservação dos recursos digitais tendem a direcionar-se progressivamente e em consonância com os *Cloud Services*. Para tanto a Ciência da Informação deve participar ativamente desse movimento. "Com essa nova tecnologia o profissional da informação ganha mais possibilidades de elaborar estratégias nos seus planos de preservação" (PONTES. 2014, p.84).

É possível afirmar que não existem muitas pesquisas que inter-relacionam metadados de preservação digital em *Cloud Services*, pois pautado da RS verificou-se, na maioria dos casos, uma fragmentação sobre temática, ora eram encontrados documentos que abordavam aspectos de preservação digital fora do escopo do ambiente dos *Cloud* (independente da terminologia *Cloud Computing*, *Cloud Provider*, *Cloud System*, *Cloud Storage*, *Cloud Servi-*

ces etc.), ora eram encontrados documentos que abordavam *Cloud*, mas sem mencionar a temática da preservação digital, ou quando um documento citava a preservação digital, o conteúdo explanado era de caráter brevíssimo.

Conforme averiguado na RS, não há uma definição de *Cloud Services* que seja ao mesmo tempo, precisa, explícita, sucinta e objetiva nos documentos pesquisados. Há várias definições de ferramentas e sistemas que fazem parte do arcabouço da nuvem, porém não há explicações definitivas dos diversos domínios. Portanto existe a necessidade de construção de um glossário que defina cada terminologia dos diferentes domínios de *Cloud*, o glossário em questão ajudaria na recuperação de documentos, em posteriores RS, assim como , em termos práticos, na delimitação de um objeto de pesquisa sobre temáticas que abordam os *Clouds*.

Pode-se verificar que a pluralidade terminológica respaldada na literatura acerca do tema, tem propiciado uma dificuldade da compreensão conceitual de forma explícita do termo *Cloud Services*. Nesse sentido, entende-se que os *Cloud Services*, aqui, são entendidos como ambientes digitais que oferecem algum tipo de serviço relacionado à armazenagem, terceirizando espaços ao usuário ou permitindo que ele tenha acesso a algum objeto digital, ou mesmo, a disponibilização de serviços de infraestrutura no contexto do armazenamento digital.

Desse modo, como garantia de preservação digital em *Cloud Services*, são apontados os metadados e os padrões de metadados que serão tratados no próximo capítulo.

5 PADRÕES DE METADADOS COMO GARANTIA DE PRESERVAÇÃO DIGITAL EM *CLOUD SERVICES*

Nesse capítulo serão abordados os metadados e os padrões de metadados adotados para a preservação digital em longo prazo, conceitos e especificações. Para a compreensão dos processos que envolvem a representação, a organização e a recuperação de dados armazenados em serviços de nuvem, é necessário conhecer o escopo de metadados que cumprem este importante papel no âmbito da preservação digital, principalmente, os metadados que cumprem importantes funções de preservação digital.

Atualmente, os *Cloud Services* procuram atender uma crescente demanda de usuários. Um dos principais aspectos que atrai inúmeros clientes se encontra na fácil capacidade de acessar e recuperar as informações armazenadas em nuvem. Para tanto, os metadados são referenciados como uma solução para a garantia de preservação no ambiente de nuvem. “A quantidade de informações disponível na rede é tão grande que identificar, localizar, descobrir a existência e acessar informações relevantes torna-se um problema crítico, demandando um tempo proibitivo aos usuários” (MARCONDES; SAYÃO, 2002, p. 47).

É pertinente afirmar que há a constante necessidade de tecnologias de custo acessível e do treinamento de recursos humanos para uso desses sistemas. Os métodos de preservação digital podem dividir-se em dois grupos: os estruturais e os operacionais. Os métodos operacionais consistem de ações sobre os objetos digitais. Os estruturais estão relacionados às decisões institucionais que envolvem a adoção de padrões, a elaboração de normas, a escolha dos metadados, a montagem da infraestrutura e o estabelecimento de consórcios (ARELLANO, 2008)

A informação semântica é caracterizada por adicionar significado à estrutura de dados identificada por meio da informação estrutural. Existe a necessidade de precisão relativa à definição de estruturas lógicas que interligam o conteúdo a ser preservado à grande amplitude de variedade de metadados, os quais darão suporte na administração da sua preservação (SAYÃO, 2010).

Os metadados podem ser considerados agrupamentos mais específicos e complexos de dados. Segundo Jia (2007), não existe um consenso geral sobre quando foi empregado pela primeira vez o termo metadados.

A princípio, o termo metadados foi utilizado em estruturas de banco de dados visando a descrição, o controle de gestão e a utilização de dados. As origens dos metadados estão diretamente ligadas ao processo de catalogação descritiva, realizado principalmente pelas bibliotecas, ou seja, num primeiro momento, a palavra metadados está vinculada ao seu uso no ambiente de bibliotecas, cuja função se refere em como descrever os diferentes tipos de recursos/objeto informacional, tanto analógico quanto digital. A catalogação descritiva tradicional é uma maneira de atribuição de metadados. Os exemplos mais significativos, nesse contexto, estão em torno do padrão de metadados MARC 21²⁷ e o conjunto de regras usadas com ele, tal como o AACR2r²⁸. (SAYÃO, 2010).

Desse modo, os metadados asseguram a padronização dos recursos informacionais, no que diz respeito à forma e ao conteúdo, os quais são baseados em regras, códigos e esquemas internacionais no intuito de simplificar e potencializar a identificação, a busca, a localização, a recuperação, a preservação, o uso e o reuso dos recursos informacionais. A grande diferença a ser destacada ao contextualizar esta forma de representação está vinculada a uma nova abordagem ofertada pelo ambiente tecnológico na qual ela está inserida. (ALVES, 2005 apud CASTRO; SIMIONATO; ZAFALON, 2016).

No campo da Ciência da Informação, os metadados são apontados como uma solução para a descrição e a representação dos dados e, entre outros

²⁷ *MARC 21 Format for Bibliographic Data* (Formato *MARC 21* para Dados Bibliográficos). Padrão de metadados no qual, entre outros aspectos consiste num dos principais padrões de metadados utilizados no intercâmbio de registros bibliográficos (ASSUMPÇÃO; SANTOS, 2013).

ASSUMPÇÃO, F. S.; SANTOS, P. L. V. A. da C. . Metadata Authority Description Schema (MADS): uma alternativa à utilização do formato MARC 21 para dados de autoridade. **Informação & Informação**, v. 18, n. 1, p. 106-126, 2013.

²⁸ AACR2r (*Anglo-American Cataloguing Rules*): em português significa, Código de Catalogação Anglo-Americano. Criado na década de 60 do século XX, publicado pela primeira vez em parceria entre a ALA (*American Library Association*), a *Canadian Library Association* e *Library Association* (FERNANDES; CASTRO, 2013).

FERNANDES, R. G. S. A.; CASTRO, F. F. de . . As relações entre RDA e AACR2: a busca de uma descrição bibliográfica completa. In: Encontro de Iniciação Científica, 23., 2013, São Cristóvão. **Anais do 23º Encontro de Iniciação Científica**. São Cristóvão: Universidade Federal de Sergipe, 2013. v. 23. p. 751-751.

aspectos, preservar os objetos digitais, independentemente do tipo de armazenamento, requerendo algo mais do que o esforço de se preservar o suporte físico, também se faz necessário considerar várias outras dimensões apresentadas pela questão: preservação lógica, física, intelectual, representação formada por metadados, e o constante monitoramento (SAYÃO, 2007).

Para os profissionais da Ciência da Informação, o termo metadados está relacionado com o tratamento da informação, mais especificamente com as formas de representação de um recurso informacional para fins de descrição, identificação, localização, busca e recuperação, ou seja, dados bibliográficos e catalográficos que servem para organizar, representar e tornar a informação identificável, localizável e acessível. (GILLILAND-SWETLAND, 1999 apud CASTRO; SANTOS, 2010, p.16).

Para se compreender determinada categoria de descrição de metadados, os esquemas e os elementos de metadados são considerados de suma importância: “os esquemas de metadados são conjuntos de elementos desenhados para encontrar as necessidades de comunidades particulares”. (CASTRO; SIMIONATO; ZAFALON, 2016, não paginado).

Ao trabalhar com elementos, esquemas e padrões de metadados que almejam a preservação da informação armazenadas em *Cloud Services* se faz necessário compreender os conceitos, as teorias, as metodologias e as abordagens epistemológicas que estão intimamente ligadas aos metadados.

Os metadados garantem a identidade e a preservação dos documentos digitais armazenados nas nuvens. “A organização dos metadados dá-se em estruturas formais chamadas de esquemas de metadados, que consistem de conjuntos de elementos concebidos para um fim específico, como a descrição de um certo tipo de recurso de informação” (FORMENTON et al., 2017, não paginado). Segundo Sayão (2010, p. 6), os metadados:

[...] são agrupados em estruturas abstratas conhecidas como esquemas ou formatos de metadados, que são conjuntos de elementos criados com fins específicos, por exemplo: descrever um tipo particular de recurso de informação. Muitos e diferentes esquemas de metadados têm sido continuamente desenvolvidos tendo como perspectiva uma grande variedade de usos em contextos variados, porém cada qual é limitado por suas especificidades e pelos seus domínios de aplicação próprios. Os poucos exemplos a seguir nos mostram um pouco dessa diversidade: *MODS (Metadata Object Description Sche-*

ma) esquema bibliográfico derivado do *MARC 21*; *EAD* (*Encoded Archival Description*) voltado para a área de Arquivologia; *LOM* (*Learning Object Metadata*) para gerenciar, avaliar e localizar objetos de aprendizagem; *MPEG Multimedia Metadata* para representação de objetos multimídiaicos.

De acordo com a literatura científica existe uma pluralidade conceitual acerca da compreensão dos metadados. Para entender a sua relevância para a preservação digital, é necessário conhecer o conceito terminológico da palavra metadados e respectivamente suas características gerais e básicas, que explicam sua importância no escopo da Ciência da Informação.

Neste estudo, adota-se o conceito de Alves (2010, p. 47-48), que define metadados como,

[...] atributos que representam uma entidade (objeto do mundo real) em um sistema de informação. Em outras palavras, são elementos descritivos ou atributos referenciais codificados que representam características próprias ou atribuídas às entidades; são ainda dados que descrevem outros dados em um sistema de informação, com o intuito de identificar de forma única uma entidade (recurso informacional) para posterior recuperação.

Os padrões de metadados são estruturas de descrição constituídas por um conjunto predeterminado de metadados (atributos codificados ou identificadores de uma entidade) metodologicamente construídos e padronizados. O objetivo do padrão de metadados é descrever uma entidade gerando uma representação unívoca e padronizada que possa ser utilizada para recuperação da mesma.

Metadados de preservação em longo prazo define o espaço de tempo determinado pelo acesso continuado aos recursos digitais ou pelo menos à informação neles contida, indefinidamente (SARAMAGO, 2004). Conforme afirma Arellano (2004, p. 19),

Os metadados de preservação são uma forma especializada de administrar metadados que podem ser usados como um meio de estocar a informação técnica que suporta a preservação dos objetos digitais.

É de fundamental importância sublinhar que a interligação entre metadados de preservação com iniciativas, recursos e seus respectivos componentes digitais se deve à preservação de estruturas próprias, isto é, características

específicas ou atribuídas ao recurso, as quais estão registradas em outras categorias de metadados.

Não apenas os metadados para preservação são importantes, mas também os metadados descritivos, estruturais, administrativos e técnicos auxiliam o alcance da preservação por longo prazo. Desta maneira, não há como afirmar qual é o único esquema de metadados existente que assegure totalmente a preservação digital e, sim, que diferentes padrões podem trabalhar juntos, de modo a se auto complementarem para o registro eficaz das informações requeridas, na gestão da preservação e do acesso utilizável ao longo do tempo de objetos/recursos digitais (FORMENTON et al. 2017, p. 91).

Saramago (2004) subdivide os metadados de preservação em três categorias: os descritivos, os administrativos e os estruturais, salientando que os metadados de preservação necessitam percorrer todo processo de existência de item/objeto informacional, estes ciclos são divididos em seis partes: criação, seleção, identificação persistente, descrição e acesso, armazenamento e preservação.

[...] os metadados de preservação registram informações do conteúdo de recursos digitais e dados administrativos, estruturais e técnicos para preservação digital. Esses metadados propiciam a preservação por longo prazo e o acesso contínuo aos objetos digitais, com garantias de autenticidade, de integridade e de confiabilidade. Neste cenário, a determinação e, possivelmente, adaptação de padrões de metadados torna-se uma ação indispensável e complexa, devendo-se incorporar os vários tipos específicos de processos e de recursos digitais tratados e preservados, o rol abundante de informações a serem registradas e fornecidas e, ainda, as decisões tomadas diante de um futuro de imprevisões intrínsecas à preservação em ambientes digitais [...] (FORMENTON et al., 2017, p. 85).

Os fluxos de conteúdos são incessantes. Nessa direção, novas situações e problemáticas surgem a partir da necessidade de se utilizar as tecnologias disponíveis ou mesmo os novos formatos digitais. A implantação da preservação em objetos digitais, as quais envolvem a definição de estruturas para representar e registrar as informações relacionadas ao objeto digital, isto é, os metadados. Metadados são dados estruturados que descrevem, permitem pesquisar, gerenciar e preservar documentos por longos períodos de tempo, e são fundamentais no âmbito do universo digital. Tais metadados em questão

podem ser classificados como descritivos, técnicos, estruturais e de preservação (PAVÃO; CAREGNATO; ROCHA, 2016).

Serão brevemente descritos e conceituados alguns dos principais padrões de metadados que descrevem, administram e/ou preservam os objetos digitais. Seguindo o pensamento de Riley (2010), os padrões metadados a seguir desempenham importantes funções de preservação digital na interligação do ciclo entre ambientes digitais e objetos digitais, são eles: *SGML (Standard Generalized Markup Language)*, *MODS (Metadata Object and Description Schema)*, *XML (Extensible Markup Metadata)*, *METS (Metadata Encoding and Transmission Standard)*, *PREMIS (Data Dictionary for Preservation Metadata)*, *AES Process History* e *MIX (NISO Metadata for Images in XML Schema)*

Os metadados de preservação digital representam a informação necessária para manter a viabilidade (permanência e integridade), representatividade (capacidade de ser representado por aplicações e visualizado) e compreensibilidade (que seja interpretado e compreendido pelos usuários) dos objetos digitais a longo prazo. Ou seja, os metadados de preservação são metadados descritivos, estruturais e administrativos que permitem a preservação a longo prazo de materiais digitais (PAVÃO; CAREGNATO; ROCHA, 2016, p.419-420)

O *SGML* é o padrão de metadados considerado o precursor, além de ser a principal e a atual metalinguagem utilizada pelo padrão *XML*, não é tão rigoroso em sua estrutura se comparado ao *XML*, incluindo a capacidade de não exigir *tags* de fechamento. Muitos padrões de metadados de interesse para a comunidade de patrimônio cultural iniciaram como linguagens *SGML* e, posteriormente, migraram para o *XML*. Atualmente, é mais vantajoso utilizar o padrão *XML* para o desenvolvimento de novas linguagens de marcação, em grande parte devido à estrutura mais rígida do *XML* (RILEY, 2010).

O padrão *SGML* é baseado na [idéia] de que documentos contêm estrutura e outros elementos semânticos que podem ser descritos sem que se faça referência à forma com que estes elementos serão exibidos. O conjunto de todas as *tags* – marcações sintáticas que descrevem os dados e comandos para manipulação de um documento – passíveis de serem utilizadas por uma linguagem derivada do *SGML* é chamado de *DTD*, ou *Document Type Definition*. (SOUZA; ALVARENGA, 2004, p. 134).

O *MODS* destina-se à complementar outros formatos de metadados e a fornecer uma alternativa entre um formato de metadados simples com um mínimo de campos e pouca ou nenhuma subestrutura, seu formato é muito detalhado com muitos elementos de dados, com várias complexidades estruturais. Utilizam *tags*²⁹ baseadas em linguagem que podem ser facilmente compreendidas (GUENTHER; MCCALLUM, 2003). O padrão de metadados *MODS* está classificado na categoria dos metadados descritivos, os quais abordam aspectos intelectuais ou artísticos relacionados ao objeto (PAVÃO; CAREGNATO; ROCHA, 2016).

O *XML* é ideal na representação das estruturas lógicas dos documentos (títulos, seções, capítulos, parágrafos, fluxo de leitura etc.), podendo ser descrito de forma semelhante ao que é utilizado nos documentos, codificando explicitamente os vários elementos que os compõem, tais como autores, datas, organização ou nomes de produtos, dados financeiros, entre outros. Portanto, fornece um ponteiro natural entre bancos de dados e conteúdo. Algumas propostas em estudo contemplam a adoção da linguagem de marcação *XML* em conjunto com procedimentos complementares (como, por exemplo, padrões de metadados em formato eletrônico) que permitam conferir elementos semânticos à Internet. (WITTEK et al., 2011).

O *XML* é uma arquitetura que não possui elementos e marcas predefinidas. Não especifica como os autores vão utilizar metadados, sendo que existe total liberdade para utilizar qualquer método disponível, desde simples atributos, até a implementação de padrões mais complexos. (ALMEIDA, 2002, p. 6)

METS é um padrão aberto, o esquema estrutural deste padrão de metadados é considerado flexível e relativamente simples. Concentra seis módulos que contém diferentes elementos de metadados necessários para um recurso digital, a maioria dos módulos não é definida pelo *METS*, os elementos de metadados, nem as *tags* que serão utilizadas. Para vários módulos, permite que

²⁹Forma de organizar e classificar informações, a qual utiliza palavras-chave para interligar informações semelhantes, muito comum na *internet Web 2.0*. Tais palavras-chave em questão são conhecidas como *Tags* ou metadados. A palavra *Tag*, em português, significa etiqueta, na *internet* é utilizada justamente como uma etiqueta e auxilia na organização das informações, agrupando aquelas que receberem a mesma marcação, e dessa forma facilitando recuperar outras relacionadas. Atualmente as *tags* são relacionadas a diferentes tipos de conteúdo, como páginas de *sites*, fotos, programas para *downloads* etc. Disponível em: <<https://www.tecmundo.com.br/navegador/2051-o-que-e-tag-.htm>>. Acesso em: 10 jul. 2018.

os metadados residam fora do pacote, isto é, os pacotes apontados para dentro do documento *METS*. É um padrão com características bem flexíveis (GUENTHER; MCCALLUM, 2003). Categorizado na classe dos metadados estruturais, os quais descrevem as representações dos objetos digitais. Os metadados estruturais identificam as relações entre os objetos digitais. Baseado num modelo que incluem metadados descritivos, administrativos e estruturais, os metadados estruturais do *METS* podem ser utilizados para reconstruir um documento a partir de múltiplos artigos de imagem, ou também são utilizados na identificação da relação entre elementos de uma coleção de documentos. (PAVÃO; CAREGNATO; ROCHA, 2016).

O principal padrão de metadados de preservação digital é o dicionário de dados *PREMIS* (*Data Dictionary for Preservation Metadata*), o qual desenvolveu os parâmetros de infraestrutura conceitual, baseados nas definições através do modelo de referência *OAIS* (SAYÃO, 2010).

O Dicionário de Dados *PREMIS* para Metadados de Preservação é o padrão internacional de metadados para apoiar a preservação de objetos digitais e garantir sua usabilidade em longo prazo. Desenvolvido por uma equipe internacional de especialistas, o *PREMIS* é implementado em projetos de preservação digital em todo o mundo, e o suporte para *PREMIS* é incorporado em uma série de ferramentas e sistemas de preservação digital comercial e de código aberto. O Comitê Editorial da *PREMIS* coordena as revisões e implementação do padrão, que consiste no Dicionário de Dados, um esquema *XML* e documentação de suporte (*PREMIS*, 2016, tradução nossa).

A composição do *PREMIS* inclui entidades, unidades semânticas e relações, além de contemplar a descrição das entidades: objeto de preservação (incluindo entidade intelectual, representações e *bitsreams*). (PAVÃO; CAREGNATO; ROCHA, 2016).

O metadados *AES Process History*, além de ser um dicionário de dados, também é um esquema *XML*. A principal característica deste padrão reside na capacidade de registrar informações sobre processos que foram executados num recurso de áudio ao longo do tempo, o qual não está limitado à transferência de áudios entre formatos físicos e digitais ou de um formato físico para digital. Fornece elementos para rastrear detalhes sobre configurações do dis-

positivo, cadeias de sinal e números de série do equipamento (RILEY, 2010). O padrão de metadados *AES Process History* são:

Metadados administrativos para objetos sonoros - Esquema de processo histórico - É um dicionário de dados e Esquema *XML* para gravar informações sobre os processos que têm sido realizados em arquivos de áudio ao longo do tempo. Inclui estruturas de campo para descrever e fornecer acesso ao conteúdo de áudio contidos em arquivos digitais. Permite a transferência, a preservação e restauração da informação. (SOUSA; SIMIONATO, 2016, p. 15).

O padrão de metadados *MIX* foi desenvolvido pela *Library of Congress*, sendo considerada uma representação *XML* para dados e metadados técnicos para imagens fixas digitais. É utilizado na descrição de imagens natodigitais, e imagens que foram reformatadas a partir de originais analógicos como digitalizações de fotografia ou páginas de texto. Inclui quatro áreas básicas de metadados: informações básicas de objetos digitais; informações básicas de imagem; metadados de captações de imagem e metadados de avaliação de imagens (RILEY, 2010). O *MIX* está categorizado na classe dos metadados técnicos.

Os metadados técnicos basicamente descrevem o objeto digital e seus arquivos, considerando o entorno tecnológico no qual o objeto digital foi produzido. O uso principal é dar suporte para que o objeto possa ser visualizado, migrado e processado. A grande maioria dos metadados técnicos é gerada de forma automática, no momento da ingestão (PAVÃO; CAREGNATO; ROCHA, 2016 p. 419).

Quando se trata de padrões de metadados de preservação, é plausível afirmar que o modelo de referência *OAIS* (*Open Archival Information System*) é o principal. Ele concentra-se na preservação lógica e especifica os termos, os conceitos, os pacotes de informação e os padrões de metadados a serem utilizados em um sistema dedicado à preservação de recursos digitais para um grupo de usuários dedicados, as denominadas 'comunidades designadas', a qual precisa acessar e entender as informações preservadas (RABINOVICI-COHEN et al., 2011).

O modelo de referência *OAIS* pode ser aplicado por variadas instituições em parceria. Desenvolvido pelo *Consultative Committee for Space Data Systems* (*CCSDS*), descreve um enquadramento conceitual para o repositório digi-

tal genérico, aberto a todas as comunidades com garantias de confiabilidade, além de conter um vocabulário próprio que viabiliza a comunicação entre as comunidades e os repositórios (SARAMAGO, 2004).

A taxonomia de classes de objeto de informação do OAIS especifica a informação de descrição de preservação e subdivide os dados em quatro: 1) Informação de referência - cujo objetivo foca em identificar e localizar um recurso ao longo do tempo visando sua integridade; 2) Informações de contexto - a qual está ligada ao quadro de uma enorme quantidade de recurso que não são interpretados corretamente sem a compreensão do seu contexto; 3) Informação de proveniência que está associada ao princípio de que a parcela de integridade de um recurso é dependente de sua trajetória histórica, ou seja, documenta a história do objeto armazenado; 4) Informação de fixidade - documenta mecanismos particulares de autenticação usados para assegurar que o item/objeto armazenado em qualquer documentação, isto é, não sofre nenhuma alteração não documentada e, desta forma, sua integridade acaba não sendo comprometida. Estes conjuntos de informação são as bases das principais estruturas dos metadados de preservação, os quais são norteados com fundamentos a partir das categorias estabelecidas no relatório *Task Force on Archiving of Digital Information* comissionados pela *Commission on Preservation and Access - CPA* e pela *Research Library Group - RLG* (SAYÃO, 2010).

No que diz respeito aos *Cloud Services*, o modelo de referência OAIS oferece serviços avançados, como verificações de fidelidade (integridade), proveniência e auditoria que complementam as capacidades de nuvens genéricas. Além disso, suporta objetos inter-relacionados complexos na nuvem e gerencia relacionamentos e *links*, mantendo a integridade referencial (RABINOVIC-COHEN et al., 2013).

O conjunto de requisitos necessários para a preservação de recursos de informação em formato digital por longo período, segundo o modelo OAIS, basicamente consiste num instrumento importante de orientação a ser analisado pelas organizações na elaboração de um planejamento adequado, assim como eficiente para a implantação do projeto ou da política de preservação digital (FORMENTON; GRACIOSO; CASTRO, 2015, p.189).

O modelo de referência OAIS se destaca por fornecer informações na elaboração de planos de preservação, os quais interagem com consumidores e produtores ao rastrear modificações em requisitos de serviços e produtos tecnológicos disponíveis, além de rastrear tecnologias, padrões e plataformas computacionais emergentes. Outra característica salutar do modelo de referência OAIS consiste nas especificações de recursos úteis, como guias e exemplos para a implantação da preservação digital (PAVÃO; CAREGNATO; ROCHA, 2016). O modelo de referência OAIS é pautado norma ISO: 16363: 2012³⁰: / ISO: 14721: 2012³¹ (FRANKS, 2015).

Estudar metadados de preservação, ou mesmo categorias de metadados que contemplam a preservação digital em *Cloud Services* caracteriza-se num importante desafio no âmbito da CI. Promover debates, pesquisas e estudos que objetivem a preservação de recursos informacionais em *Cloud Services* podem gerar resultados frutíferos, entre os quais assegurar e potencializar as melhores estratégias de preservação dos objetos digitais.

Todas as formas de preservação digital, exceto as mais simples, podem se beneficiar pela criação, manutenção e evolução de metadados detalhados para o apoio aos processos de preservação. Por exemplo, metadados podem documentar o processo técnico associado com a preservação, especificar informações de direitos autorais e estabelecer a autenticidade do conteúdo digital. Eles podem registrar a cadeia de custódia de um objeto digital e identifica-lo individualmente tanto interna como externamente em relação ao arquivo em que reside. Em resumo, a criação e instalação de metadados de preservação parece ser um componente chave para as estratégias de preservação (OCLC/RGL, 2001, p.2).

O próximo capítulo abordará os resultados alcançados na Revisão Sistemática da Literatura destacando a análise dos dados e a discussão dos metadados de preservação em *Cloud Services*.

³⁰ Define as práticas de recomendações que consiste em avaliar a confiabilidade dos repositórios digitais, a qual pode ser aplicável em todo escopo de repositórios digitais, além da possibilidade de ser utilizada como base para certificação. Disponível em: <<https://www.iso.org/standard/56510.html>>. Acesso em 11 jul. 2018.

³¹ Consiste na norma que define o modelo de referência OAIS. Disponível em <<https://www.iso.org/standard/57284.html>>. Acesso em: 11 jul. 2018.

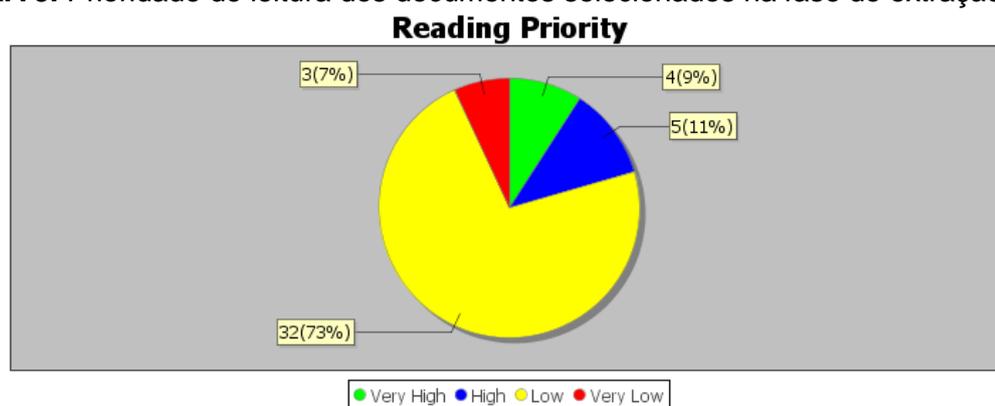
6 RESULTADOS DA REVISÃO SISTEMÁTICA DA LITERATURA

A presente pesquisa recuperou 821 resultados usando seis bases de dados, entre as fases de seleção e de extração, poucos documentos foram aceitos, restando apenas 11 artigos.

Nesse capítulo, são descritos os processos da fase extração dos documentos aceitos a partir dos procedimentos metodológicos pautados na RS, destacando de forma pormenorizada os seguintes pontos: leitura prioritária dos documentos selecionados, documentos analisados e documentos aceitos.

Todos os documentos considerados com prioridade alta e muito alta apresentavam características que relacionavam metadados de preservação digital com serviços de nuvem, entretanto, alguns pontuados com prioridade baixa de leitura foram aceitos, pois no caso dessa pesquisa, os diversos termos que englobam o universo dos *Cloud Services* se misturam, e, em muitos casos, os conceitos não entram no escopo do consenso geral, e ao mesmo tempo, padrões de metadados de preservação são empregados nos textos, mas a terminologia metadados de preservação digital não é explicitada. Conforme destacado na figura 5, foram incluídos 44 documentos para a fase de extração, a prioridade de leitura nesse momento foi categorizada como: muito baixo - 3 (7%), baixo - 32 (73%), alto - 5 (11%) e muito alto - 4 (9%).

FIGURA 5. Prioridade de leitura dos documentos selecionados na fase de extração.

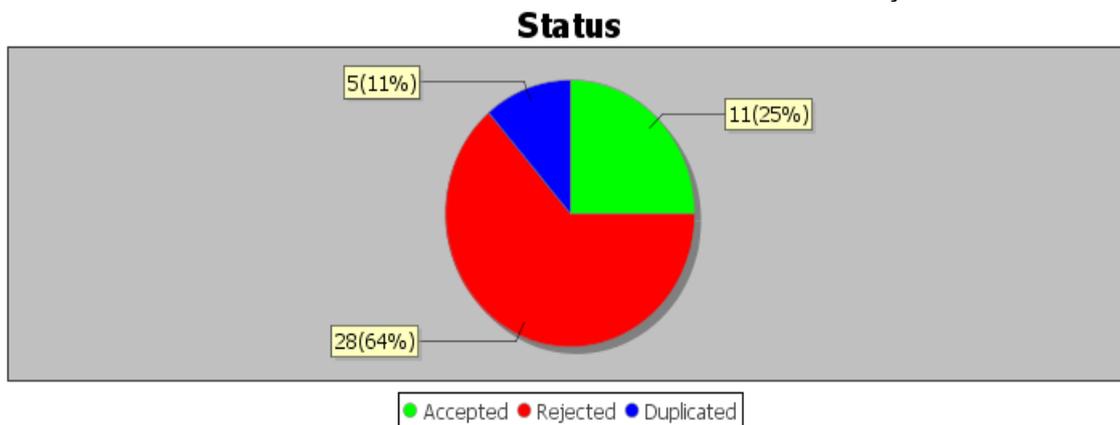


Fonte: Elaborado pelo autor.

A **figura 6** indica quantitativamente quais documentos foram aceitos e quais rejeitados. Houve documentos que não foram duplicados automaticamente na fase de seleção, porém foram identificados e categorizados como dupli-

cados, exatamente 5 documentos (11%). Foram aceitos 11 documentos, 25% do total e foram rejeitados 28 documentos, cerca de 64% do total. Nessa fase ocorreu leitura pormenorizada e na íntegra dos textos, a partir das análises das leituras, pautadas pelo formulário de extração.

FIGURA 6. Documentos analisados na fase de extração.

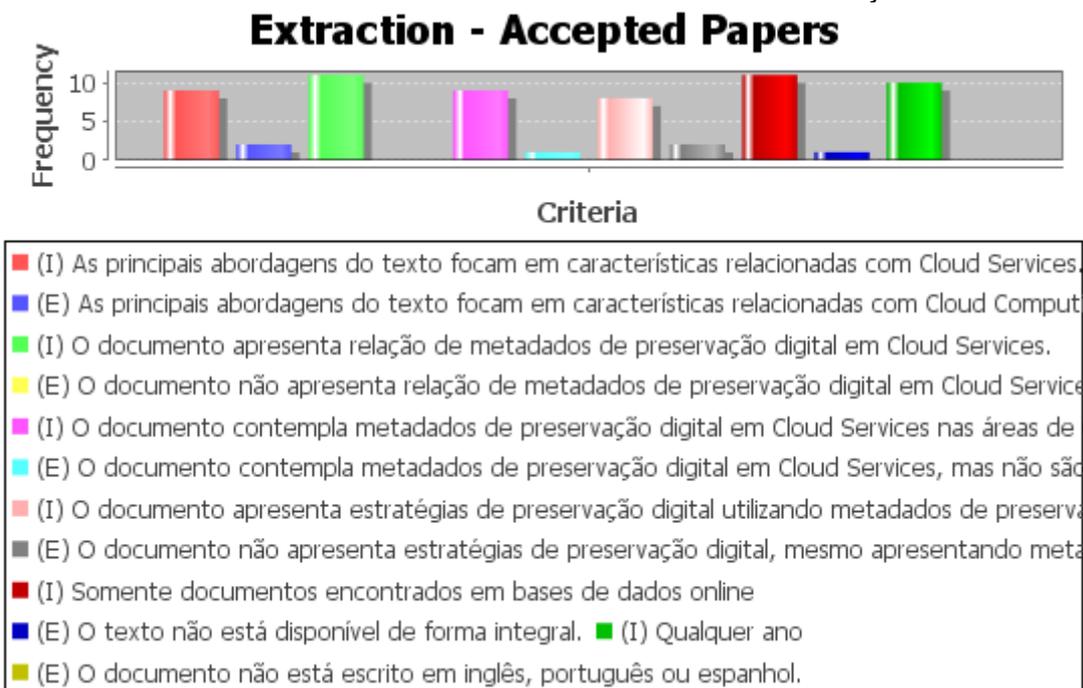


Fonte: elaborado pelo autor.

A maioria dos artigos rejeitados foca em abordagens relativamente próximas ao escopo da preservação digital, mas não refletem o objetivo central da temática pesquisada.

Muitos textos focam na questão da autenticidade dos serviços de nuvem; há alguns textos que priorizam abordar as questões de confiança nas políticas entre usuários e serviços de nuvem; existem pesquisas que apontam a preservação digital sem explicar o funcionamento em *Cloud Services*.

FIGURA 7. Documentos aceitos na fase de extração.



Fonte: elaborado pelo autor.

Por outro lado, há estudos sobre a infraestrutura dos serviços de *Cloud Services*, suas camadas, programas e *softwares* que possibilitam utilizar repositórios fora do ambiente de nuvem como um fator intermediário entre usuário e nuvem; também há textos que focam na combinação preservação digital e nuvem, mas não citam padrões, esquemas e metadados de preservação digital.

Com base na leitura e na análise de conteúdo dos onze documentos aceitos foram extraídos os dados necessários para entender os metadados de preservação digital em *Cloud Services*, as aplicações, as características, as tipologias e as tecnologias utilizadas que serão descritos e detalhados nos subcapítulos subsequentes.

6.1 Enfoques dos documentos analisados na Revisão Sistemática da Literatura

Após todas as fases da Revisão Sistemática de Literatura, onze documentos se enquadraram nos aspectos temáticos da presente pesquisa, os documentos em questão se relacionam diretamente com as estratégias de preservação digital em *Cloud Services*, seja na área da Ciência da Informação ou na área da Ciência da Computação. Por meio da análise de conteúdo se verifi-

cou os aspectos de semelhança e de discrepância em relação aos seguintes focos: os tipos de padrões e metadados de preservação digital encontrados nos serviços de nuvens pesquisados em cada documento; as possíveis comparações entre *Cloud Services*, incluindo as plataformas de hospedagens, os tipos de objetos digitais (recursos de áudio, documento de texto, documento de imagem etc.), os programas utilizados, os formatos, os *softwares* etc.

Outra característica extraída nos documentos gira em torno da camada de nuvem de preservação digital, poucas pesquisas abordam essa temática no escopo das áreas que tratam de comunicação, informação e tecnologia. Alguns dos documentos que chegaram à última fase da sumarização abordam a questão apontando soluções criativas, pautadas em metodologias da área de exatas, ao explicar o funcionamento da infraestrutura de uma camada de preservação digital em nuvem, incluindo abordagens temáticas que interagem diretamente com os aspectos de preservação, como interoperabilidade, migração, autenticidade, segurança de dados, retenção, persistência etc.

Em todos os documentos que trataram da camada de preservação digital em nuvem, padrões e metadados de preservação eram vistos como um atributo importante na conjuntura da camada. O foco de estudo de cada documento, fruto da análise de conteúdo, está explicitado no quadro 5.

QUADRO 4. Foco de estudo dos documentos pautados na análise de conteúdo.

Documento 1	
Autores	Jan Askhoj, Shigeo Sugimoto; Mitsuharu Nagamori
Título	<i>Preserving records in the cloud</i>
Ano	2011
Área	Ciência da Informação
Tipo	Artigo
País(autor/publicação)	Japão
Instituições	Universidade de Tsukuba
Foco de estudo	Compara características de gerenciamento de registros dentro do ambiente de nuvem com os modelos de arquivamento existentes.
Documento 2	
Autores	Simona Rabinovici-Cohen; Mary G. Baker; Roger Cummings; Sam Fineberg; John Marberg
Título	<i>Towards SIRF: Self-contained information retention format</i>
Ano	2011
Área	Ciência da Computação
Tipo	Artigo
País(autor/publicação)	Israel
Instituições	IBM; HP; Symantec Research Labs

Foco de estudo	Verifica sistemas de preservação e os processos para suportar o armazenamento de dados em longo prazo por meio de estratégias de preservação. Descreve os atributos do <i>SIRF (self-contained information retention format)</i> , especializado em retenção de longo prazo.
Documento 3	
Autores	Hrvoje Stancic; Arian Rajh; Hrvoje Brzica
Título	<i>Archival Cloud Services: Portability, Continuity, and Sustainability Aspects of Long-term Preservation of Electronically Signed Records</i>
Ano	2015
Área	Ciência da Informação
Tipo	Artigo
País(autor/publicação)	Croácia; Canadá
Instituições	Universidade de Toronto; Universidade de Zagreb; <i>Finacial Agency (Croácia)</i>
Foco de estudo	Examina o conceito de documento eletrônico no contexto da segurança e confiança da nuvem.
Documento 4	
Autores	Peter Wittek; Thierry Jacquin; Hervé D'éjean; Jean-Pierre Chanod; Sándor Da ranyi
Título	<i>XML Processing in the Cloud: Large-Scale Digital Preservation in Small Institutions</i>
Ano	2011
Área	Ciência da Informação
Tipo	Artigo
País(autor/publicação)	Suécia; França
Instituições	<i>Xerox Research Center Europe</i> ; Universidade de Boras
Foco de estudo	O texto destaca os fluxos de trabalhos (<i>workflows</i>) nos ambientes de nuvem, propondo novas estratégias de preservação digital nestes ambientes, possibilitando migração direta de uma infraestrutura operacional para outra.
Documento 5	
Autores	Quyen L. Nguyen; Alla Lake
Título	<i>Content Server System Architecture for Providing Differentiated Levels of Service in a Digital Preservation Cloud</i>
Ano	2011
Área	Ciência da Informação
Tipo	Artigo de conferência
País(autor/publicação)	Estados Unidos
Instituições	<i>College Park</i>
Foco de estudo	Fornece pontos de vista que analisam os desafios de sistemas específicos no âmbito da preservação digital em longo prazo, focando na infraestrutura da camada de preservação em nuvem.
Documento 6	
Autores	Jan Askhoj; Shigeo Sugimoto; Mitsuharu Nagamori
Título	<i>Developing an ontology for cloud-based archive systems</i>
Ano	2015
Área	Ciência da Informação
Tipo	Artigo
País(autor/publicação)	Japão
Instituições	Universidade de Tsukuba; <i>Nippon Records Management</i>

Foco de estudo	Descreve e analisa o contexto da ontologia ³² em arquivos de nuvem a partir da interação entre as camadas baixas, destacando a relação semântica neste tipo de armazenamento, e relaciona os esquemas de metadados de preservação com a camada de preservação.
Documento 7	
Autores	Krystyna Matusiak; Allison Tyler Catie Newton; Padma Polepeddi.
Título	<i>Finding access and digital preservation solutions for a digitized oral history project: A case study</i>
Ano	2017
Área	Ciência da Informação
Tipo	Artigo
País(autor/publicação)	Estados Unidos
Instituições	Universidade de Denver; Universidade de Michigan; Faculdade de Wooster; Biblioteca Pública de <i>Jefferson County</i>
Foco de estudo	Percorre as estratégias de preservação digital de registros de áudio (história oral) da coleção <i>Jeffco Stories</i> , a qual pertence à biblioteca pública do Colorado, e para preservar o serviço utilizou uma plataforma de acesso aberto e hospedou os registros num serviço de nuvem. Por meio desta experiência os autores abordam os pontos positivos e os percalços ao utilizar esta estratégia de preservação digital.
Documento 8	
Autores	Ingmar Pappel; Ingrid Pappel
Título	<i>Integral and secure cloud architecture based system for backup and retention of public sector information</i>
Ano	2012
Área	Ciência da Computação
Tipo	Artigo de conferência
País(autor/publicação)	Estônia
Instituições	Universidade de Tecnologia de Tallinn
Foco de estudo	Visa utilizar a funcionalidade de armazenamento na nuvem, baseado na retenção de objetos digitais através da aplicação de um modelo de retenção de dados, por meio do conceito de <i>MetaClouds</i> armazenamento em nuvem.
Documento 9	
Autores	Peter Wittek; Sándor Darányi
Título	<i>Digital Preservation in Grids and Clouds: A Middleware³³ Approach</i>
Ano	2012
Área	Ciência da Informação

³² No escopo da Ciência da Computação, o termo ontologia é utilizado para designar as chaves de indexação ou temáticas referentes a objetos digitais constantes na *Web*. A ontologia permite não apenas representações secundárias a partir dos conhecimentos sobre os seres, mas também se constitui de representações primárias ou de excertos de conteúdos das próprias coisas ou objetos virtuais. (ALVARENGA, 2003).

³³ O *middleware* é o *software* que se encontra entre o sistema operacional e os aplicativos nele executados. Funcionando de forma essencial como uma camada oculta de tradução, o *middleware* permite a comunicação e o gerenciamento de dados para aplicativos distribuídos. Disponível em: <<https://azure.microsoft.com/pt-br/overview/what-is-middleware/>>. Acesso: 06 jul. 2018.

Tipo	Artigo
País(autor/publicação)	Suécia
Instituições	Universidade de Boras
Foco de estudo	Examina as características de gerenciamento de registros em <i>Cloud Services</i> e compara a existência de modelos de arquivo, exemplificados pela OAI/S como modelo de referência.
Documento 10	
Autores	Patrícia Franks
Título	<i>Government Use of Cloud-based Long Term Digital Preservation as a Service: An Exploratory Study</i>
Ano	2015
Área	Ciência da Informação
Tipo	Artigo
País(autor/publicação)	Estados Unidos
Instituições	Universidade do Estado de San José
Foco de estudo	Conceitua e explica a camada de nuvem refere à preservação digital, por meio da ação de metadados de preservação digital, organização dos requerimentos de preservação na norma ISO 14721, visando identificação na otimização da preservação, além de explanar sobre a captura e criação de registros em longo prazo. Compara serviços de hospedagem em nuvem como <i>DuraCloud</i> e o <i>Preservica</i> .
Documento 11	
Autores	Simona Rabinovici-Cohen; John Marberg–Israel Kenneth Nagin IBM Research – David Pease
Título	<i>PDS Cloud: Long Term Digital Preservation in the Cloud</i>
Ano	2013
Área	Ciência da Computação
Tipo	Artigo
País(autor/publicação)	Estados Unidos/Israel
Instituições	IBM Research
Foco de estudo	O objetivo deste estudo foi avaliar as capacidades das plataformas de serviços de nuvem concorrentes, e identificar diferenças importantes entre elas, no sentido de compreender as funcionalidades que podem ser exploradas no contexto das necessidades de preservação dos objetos digitais.

Fonte: elaborado pelo autor.

O estudo da literatura, embasada pela análise de conteúdo, proporcionou averiguar os principais elementos que relacionam metadados de preservação digital, destacando-se a preservação digital em longo prazo, com os serviços de nuvem. Baseados nessa temática os documentos revelam estratégias adotadas utilizando os metadados de preservação digital, enfocando nas necessidades, nos tipos de armazenamento que estão alocados os objetos digitais, nos custos da preservação, nos pontos de vista quanto aos outros aspectos

tos que norteiam a preservação nesse contexto, nas trajetórias dos objetos digitais, nos modelos de servidores de nuvem existentes etc.

Nem todos os documentos selecionados na fase de extração utilizam a terminologia metadados de preservação digital, embora todos os textos foquem em algum padrão de metadados de preservação como estratégia na preservação de objetos digitais em nuvem. Porém, vale destacar a omissão da terminologia, independente do documento englobar a temática da pesquisa. A omissão em si, principalmente em documentos da área da Ciência da Informação pode revelar nuances a serem desvendados, por exemplo, a compreensão conceitual do tema.

Não foram encontrados documentos sobre a temática na América Latina, incluindo o Brasil. Além de ser possível afirmar que os metadados de preservação digital em nuvem são relativamente pouco pesquisados no mundo.

Os documentos aceitos na fase de extração foram confeccionados por autores e/ou publicados nos seguintes países: Estônia, Japão, Estados Unidos, Canadá, Croácia, Suécia, Israel e França. Vale observar que a maioria dos documentos é proveniente de países denominados como desenvolvidos. Embora a temática ainda seja pouco explorada mundialmente, os estudos sobre preservação digital em longo prazo em serviços de nuvem estão vagarosamente aumentando. Todos os documentos são da década X do século XXI. Dos 11 documentos selecionados, 4 são do ano de 2011, apenas 1 documento do ano de 2013, 3 documentos são de 2015, 2 documentos foram publicados em 2012, e apenas um documento de 2017. Os aspectos tecnológicos também influenciaram no contingente de estudos sobre a temática, pois os serviços de nuvem se popularizaram em meados dos anos 2000 e, no desenrolar dos anos surgiram novos tipos de serviços em nuvens, como novos aplicativos e novas camadas.

O conceito de camada de preservação é uma tecnologia relativamente nova, em meados da década passada, provavelmente pouquíssimos usuários e profissionais se preocupavam com a preservação em longo prazo. As discussões sobre estratégias de preservação em *Cloud Computing* possivelmente não eram recorrentes na Ciência da Informação.

Para tanto, uma análise mais aprofundada acerca do detalhamento das características dos documentos aceitos na fase de extração que contemplam

estratégias de preservação digital em *Cloud Services*, a partir dos metadados de preservação é realizada, com base no quadro 5.

- Documento 1: Utiliza um modelo que permite o compartilhamento de funcionalidades e objetos de informação disponibilizando com outras camadas de nuvem, norteando a discussão em conceitos do modelo de referência *OAIS*. Apesar de concluir que existem várias áreas que o modelo *OAIS* não se integra bem com os sistemas de *Cloud Computing*, o documento se baseia em conceitos desse modelo, focando na facilidade compartilhamento. O texto também aponta características de interação entre camada de serviços de nuvem. Afirma que existe facilidade de compartilhamento, e utiliza um modelo de serviço de nuvem que contém os seguintes aspectos: provisão automática de metadados, repositório confiável, pacotes de informação abrangentes e padronizados. O documento destaca a camada de preservação no sentido de ser mais do que um armazenamento seguro, pois fornece tipos de informações necessárias para a preservação por longo prazo e cria pacotes de informações para sistema de arquivo. O texto não compara *Cloud Services*.

- Documento 2: Não cita de forma explícita uma estratégia de preservação utilizando metadados de preservação, porém menciona a questão do uso de padrão de metadados. A principal estratégia de preservação é norteada pelo *SIRF*, um formato que utiliza um *container* de retenção, e uma das principais características é o fornecimento expressivo da quantidade de metadados no processo de encapsulação em nível de armazenamento. O trabalho explana sobre a camada de preservação e cita a plataforma *DuraCloud*, a qual gera um serviço integrado aonde os dados podem ser gerenciados por vários serviços de nuvem. O suporte do *container* pode ser interpretado por vários serviços de *Cloud*.

- Documento 3: O texto foca na portabilidade, na continuidade e na sustentabilidade dos aspectos da preservação em longo prazo pautados na assinatura de registros eletrônicos. O artigo descreve conceitos envolvendo soluções de infraestrutura em *Cloud*. O documento conceitua importantes elementos de infraestrutura de chave pública (*PKI*³⁴), os quais são bem relevantes

³⁴ Sua importância consiste em verificar a identidade de um usuário com mais confiança, o foco dos ambientes da *PKI* giram em torno da segurança na comunicação. É caracterizado pelo seu investimento em larga escala, ou seja, o suporte controle de acesso pode ser considerado

no domínio da preservação, além de examinar o conceito de documento eletrônico no contexto de confiança e segurança deste ambiente. Analisa modelos de arquivos digitais baseados em *Cloud Services* e ainda faz uma proposta de modelo de nuvem que suporta a preservação em longo prazo. Os provedores de nuvem deveriam estar mais sintonizados com os padrões arquivísticos. O texto não utiliza a terminologia metadados de preservação digital e tampouco cita exemplos de camada de preservação.

- Documento 4: O texto foca em fluxo de trabalho (*workflows*) em *Cloud*. O padrão de metadados mais utilizado é o *XML*, porém não usa o termo metadados de preservação. O estudo sobre preservação é focado na migração, na transformação, na escalabilidade. Usa o *Xeproc*, programa de domínio específico de linguagem. A pesquisa propõe uma nova estratégia usando nuvens. Do ponto de vista da preservação digital permite a migração direta de uma infraestrutura operacional para outra. Cita o padrão de metadados *METS*. Apresenta os fluxos de trabalho (*workflows*) nos serviços de nuvem. Faz comparações relativas ao armazenamento entre *Cloud Services* com outros ambientes. O estudo não faz comparações entre serviços de nuvem, além de não citar plataformas de preservação.

- Documento 5: Utiliza camada de preservação como estratégia, fornece pontos de vista que analisa desafios de sistemas específicos para preservação digital em longo prazo; descreve como a arquitetura padrão consiste num serviço de conteúdo, usando núcleo de uma camada de preservação em nuvem, podendo fornecer flexibilidade e diferentes níveis de serviços. Cita padrões de metadados (mas não cita a terminologia) e também afirma que uma camada de preservação pode ter mais de uma categoria. Vale salientar que o texto foca na infraestrutura da camada de preservação.

- Documento 6: Descreve uma ontologia para arquivos em nuvem. O principal padrão de metadados utilizado é o *PREMIS*, o qual é de suma importância na construção de vocabulários para ambientes em *Cloud* (criação, transferência, incluindo registros de conteúdo de criação de sistemas em nuvem). A ontologia descreve os componentes de escolha de forma sucessiva, além de fornecer interoperabilidade entre os conteúdos, criando aplicações e

serviços fornecidos por metadados de preservação. O texto cita que o modelo *OAIS* não endereça de forma específica as características de *Cloud Computing*. Aborda a interação entre camadas, além de destacar a importância da camada de preservação. O estudo, no contexto de modelo de camadas, afirma que os objetos digitais precisam se transformar em pacotes de informação, ou seja, precisam de metadados anexos, pois estes metadados são necessários para garantir a preservação em longo prazo. A camada de preservação gerencia aplicativos de criação onde seu armazenamento está alocado. Na ontologia em nuvem a descrição semântica é importante para descrever o arquivo digital e a relação entre os esquemas de metadados usados na criação do aplicativo dos metadados de preservação, usados na camada de preservação. Utiliza vários padrões de metadados, mas não especifica a terminologia (padrão de metadados).

- Documento 7: O texto foca na preservação da coleção de áudios de história oral (*Jeffco Stories*) da Biblioteca pública do Colorado. Compara o *software Omeka* e a plataforma de hospedagem em nuvem *DuraCloud*. O padrão de metadados citados é o *MODS* e o *OAIS*. Porém, não explicita a terminologia metadados de preservação. Segundo a conclusão do documento os objetos digitais não podem ficar armazenados eternamente numa plataforma, e sendo assim, as estratégias usando metadados se torna importante no processo de migração. O *DuraCloud* não é compatível em vários aspectos com o modelo *OAIS*, porém, fornece alguns serviços que se estendem além do armazenamento básico em nuvem. O estudo finaliza afirmando que o *Omeka* tem as maiores taxas de soluções em termos de gestão de conteúdo, porém *Omeka* e *DuraCloud* apresentam limitações. O texto não foca na camada de preservação. Tanto *Omeka* como *Duracloud* são plataformas intermediárias que usam serviço em nuvem, mas não são nuvens por si só.

- Documento 8: O texto utiliza o termo *Metacloud*, um conceito que trabalha com a unificação de diferentes interfaces de armazenamento na nuvem para um arquivo digital baseado na retenção de objetos digitais. O documento trabalha com o *software X-Road*³⁵, o qual foca no intercâmbio de do-

³⁵ Utilizado para integrar diversas bases de dados (públicas e privadas) da Estônia, permite que vários setores se conectem e funcionem em harmonia. Consiste numa grande solução eletrônica no país em questão, pois contempla uma gama de completa de serviços para o público em

cumentos. Um dos principais objetivos do texto gira em torno de soluções baseadas em esquema de nuvem. O estudo sugere uma nova base de dados com novo sistema de informações que usa *backup* e arquivamento, baseado nos serviços de nuvem que possam ter a capacidade de se comunicar com *Metacloud*. Baseado no aplicativo de hospedagem em nuvem, só será possível armazenar os objetos digitais se o *Metacloud* conseguir uma instalação de armazenagem especializada, ou seja, os provedores de nuvem necessitam garantir as especificidades no âmbito da preservação e da flexibilidade. O estudo não usa a terminologia metadados de preservação, também não compara serviços de provedores de nuvem.

- Documento 9: O texto examina as características de gerenciamento em *Cloud Computing* e compara a existência de modelos de arquivo, exemplificados pelo *OAIS* como modelo de referência; outro fator sublinhado no documento é o sistema de *middleware* que compara escolhas entre redes (*Grid*³⁶) ou nuvem. A vantagem da nuvem gira em torno dos baixos custos. No contexto da preservação, os estudos dizem respeito à transformação, à escalabilidade e à reutilização. Os padrões de metadados mais citados foram o METS, o XML. Outros modelos são testados por conta das características das plataformas de nuvem que não combinam com algumas características do pacote *OAIS*. O XML é idealmente adequado para representar a estrutura lógica do documento (títulos, sessões, capítulos, parágrafos e fluxo de leitura, também é de grande utilidade no ciclo de vida do objeto digital, no gerenciamento, no reuso de conteúdo e reposição, além de assegurar qualidade e segurança). Porém a conversão de objetos digitais para XML é uma tarefa bem complexa e delicada, precisando utilizar o formato XSLT. O estudo não cita camada de

geral, além de garantir transferências seguras, todos os dados que saem do *X-Road* são assinados e criptografados, e todos os dados que são recebidos pelo *software* são autenticados e registrados. Disponível em: <<https://e-estonia.com/solutions/interoperability-services/x-road/>>. Acesso em: 11 jul. 2018.

³⁶ Pode ser compreendida por uma malha tecida pela intersecção de um conjunto de linhas horizontais e outro de linhas verticais, um dos conjuntos definem as colunas e o outro conjunto definem as linhas. Dentro de um *Grid*, a partir da construção de suas linhas, é possível inserir elementos de marcação. Um *Grid* pode constituir-se das seguintes finalidades: dimensões fixas ou flexíveis; posicionamento dos itens; criação de itens adicionais; alinhamento e; controle sobre conteúdos sobrepostos. *Grids* também podem ser usados para *design* de *layouts* de grandes sessões de uma *webpage*, assim como de pequenos elementos de interface. Disponível em: <https://developer.mozilla.org/pt-BR/docs/Web/CSS/CSS_Grid_Layout/Basic_Concepts_of_Grid_Layout>. Acesso em: 12 jul. 2018.

preservação. A preservação de dados no fluxo de trabalho usando nuvem funciona sem problemas, pois a nuvem tem por características flexibilidade, preço razoável e recursos fixos. O trabalho não usa o termo metadados de preservação e não compara serviços de nuvem.

- Documento 10: O texto explica os quatro modelos de nuvem existentes. Pauta-se no modelo *OAIS* amparado pelas normas da *ISO 14721; 16363*. Apresenta modelo de preservação para organizar os requerimentos de preservação digital conforme os procedimentos da norma *ISO*, por meio do modelo *Digital Preservation Capability Maturity Model (DPCMM)*, amparados por três domínios: infraestrutura, um ou mais repositórios digitais e *Cloud Services*. Objetivos do modelo: identificação (otimização da capacidade de preservação); continuidade, isto é, captura ou criação de registros em longo prazo. Compara dois serviços que habilitam fluxos de trabalho na nuvem, o *Preservica* e o *DuraCloud*, ambos trabalham em parceria para fornecer soluções. O trabalho cita o termo metadados de preservação. Apresenta revisão atual dos serviços de preservação digital em serviços de nuvem, e considera a camada de preservação em nuvem uma opção viável. O estudo afirma que há literatura escassa sobre preservação digital em serviços de nuvem, cita o projeto *PaaST*³⁷ do programa *Inter Trust*³⁸. As normas da *ISO* citadas acima podem ser usadas para desenvolver capacidade de maturidade dos modelos de preservação digital.

- Documento 11: O texto cita uma camada de nuvem, *Preservation Digital Services (PDS)*, a qual é um serviço de armazenamento com preservação baseada em *OAIS* que emprega vários provedores de nuvens heterogêneos. Foram examinadas várias plataformas de nuvem (*Cloud Platforms*), visando à usabilidade no emprego da preservação digital, levando em consideração os atributos de armazenagem da *PDS Cloud*. O objetivo deste estudo foi avaliar as capacidades das plataformas candidatas, identificar as diferenças importantes entre elas e entender quais funcionalidades podem ser exploradas

³⁷ Preservação como um Serviço de Confiança ou *Preservation as a Service for Trust (PaaST)*: consiste num projeto cujo o principal objetivo abarca desenvolver especificações que sejam neutras em relação à parte ou partes que prestam o serviço. (FRANKS, 2015).

³⁸ Instituição criada com o objetivo dar suporte para empresas nos quesitos: configuração, manutenção e gerenciamento. A equipe é formada por especialistas que gerenciam diariamente as empresas que contratam os serviços dessa instituição. Disponível em: <<https://www.intertrustgroup.com/our-services/corporate-services>>. Acesso em: 12 jul. 2018.

para necessidades de preservação. O principal enfoque da pesquisa contempla as capacidades de armazenamento em nuvem. Outros requerimentos também são incluídos na *PDS*: funcionalidade e especificações para publicar e empregar aplicações que são usados para interpretação da preservação em objetos digitais. A relevância desta camada gira em torno das características similares de muitos serviços de nuvem que são similares ou mesmo idênticos, como por exemplo, o emprego do uso de metadados. Somente lançar dados na nuvem não é uma solução adequada para repositórios de preservação digital em longo prazo, pois são necessárias mecanismos de gerenciamento e confiabilidade. A *PDS Cloud* suporta a preservação lógica e materializa o objeto lógico para o armazenamento de informação-objeto no dispositivo de armazenamento físico. Isso é motivado pela ideia de que os sistemas de preservação digital serão mais robustos e têm menor probabilidade de corrupção ou perda de dados se a funcionalidade relacionada à preservação for enviada ao armazenamento. Trabalha com o conceito de *Multi-Clouds*: conjunto heterogêneo de plataformas de armazenamento que objetiva integrar diferentes *Cloud Services*, pois a arquitetura suporta a implantação de múltiplas nuvens de diferentes fornecedores.

É possível identificar a partir da análise sistemática da literatura que os estudos de metadados de preservação em *Cloud Services* ainda se encontram inexpressivos na área de Ciência da Informação no cenário atual, tanto no âmbito internacional quanto nacionalmente. A necessidade de um maior posicionamento da Ciência da Informação no panorama da preservação digital em *Cloud Services* é emergente, uma vez que essa ciência detém o *corpus* de conhecimento para responder às questões de armazenamento, de descrição, de representação e de preservação de dados e de informações em ambientes digitais.

O próximo subcapítulo apresenta os padrões de metadados de preservação digital adotados e aplicados em *Cloud Services*, a partir dos onze documentos analisados.

6.2 Padrões de metadados de preservação digital adotados e aplicados em *Cloud Services*

Com base na metodologia de Análise de Conteúdo foram verificados os padrões de metadados de preservação digital identificados a partir dos onze documentos estudados. Somente não será destacado os padrões de metadados de preservação digital *PREMIS*, as características, as funcionalidades e as aplicações deste padrão estão destacadas no próximo subcapítulo. Este padrão foi destacado por causa dos detalhamentos e especificações de suas atribuições em *Cloud Services*.

Antes de explicar sobre os padrões de metadados encontrados nos documentos, vale salientar que nem todos os padrões de preservação digital foram empregados na finalidade de exercer funções e aplicações no âmbito da preservação, em muitos casos tais padrões são utilizados para outras funcionalidades e aplicabilidades, porém mesmo nestes casos eles serão descritos. Em segundo lugar, em oito documentos analisados, os padrões de metadados de preservação digital adotados e aplicados empregados com a finalidade de exercer suas atribuições no âmbito da preservação não são minuciosamente explicados, suas funções são brevemente explanadas e suas aplicabilidades são descritas superficialmente. Nesses oito documentos, a ênfase que relaciona preservação digital de longo prazo em *Cloud Services* consistem de outros elementos e abordagens, por essa razão os padrões de metadados de preservação não se destacam como protagonistas das descrições textuais destes documentos.

No quadro 5 foram apontados modelos de *Cloud Services* criados pelos autores além dos existentes, modelos e serviços de preservação (elaborado pelos autores e modelos existentes, como por exemplo, os serviços de hospedagem), as camadas que empregam serviços de preservação digital e padrões de metadados de preservação digital encontrados nos documentos, além dos padrões de metadados classificados em outras categorias, mas que estão diretamente ligados com contextos, assim como a relação da preservação digital com o modelo de referência *OAIS*.

QUADRO 5. Modelos de *Cloud Services* e metadados de preservação digital.

Documento 1- <i>Preserving records in the cloud</i>	
Serviços de preservação ³⁹ ou Camadas de <i>Cloud Services</i>	Modelo de camada simples (<i>Simple layered model</i>); Modelo em camadas para serviços de arquivamento (<i>Layered model for archiving services</i>): contém quatro camadas (<i>PaaS, SaaS, Camada de Preservação e Camada de Interação</i>), a Camada de preservação fornece tipos de informação necessárias para a preservação em longo prazo e criam pacotes de informação para arquivo, manipula e adiciona informações dos objetos digitais transformando-os em pacotes de informação. A Camada de Interação contém sistemas voltados para o usuário, usada diretamente para manipular objetos digitais.
<i>Cloud Services</i> identificados no texto (modelos ou existentes)	Amazon S3
Metadados de preservação; metadados classificados em outras categorias que estão relacionados com a preservação; e modelo de referência OAIS	OAIS; PREMIS; XML
Documento 2 - <i>Towards SIRF: Self-contained information retention format</i>	
Serviços de preservação ou Camadas de <i>Cloud Services</i>	<i>SIRF (Self-contained information retention format)</i> : é um formato de contêiner de armazenamento para um conjunto de objetos (digitais) de preservação que ao mesmo tempo fornece um catálogo com metadados relacionados a todo o conteúdo do contêiner. <i>DuraCloud</i> : é um sistema de gerenciamento de conteúdo de código aberto, o qual é utilizado como um sistema de hospedagem em <i>Cloud Services</i> cuja finalidade consiste em fornecer serviços para preservação digital.
<i>Cloud Services</i> identificados no texto (modelos ou existentes)	<i>First Cloud; Second Cloud</i> : modelos de <i>Cloud Services</i> criados pelos autores. Fornecem serviços de preservação.
Metadados de preservação; metadados classificados em outras categorias que estão relacionados com a preservação; e modelo de referência OAIS	OAIS; PREMIS; XML; VERS; METS; LOTAR
Documento 3 - <i>Archival Cloud Services: Portability, Continuity, and Sustainability Aspects of Long-term Preservation of Electronically Signed Records</i>	
Serviços de preservação ou Camadas de	<i>Electronic Document Safe (EDS)</i> : consis-

³⁹ Podem estar embutidos na arquitetura dos *Cloud Services* ou acoplados no ambiente através de um serviço de hospedagem. Para melhor compreensão do conceito ler a página 71 deste presente texto, no qual Franks (2015), explica a finalidade de um *Cloud Digital Preservation Service*.

<p><i>Cloud Services</i> (modelos criados ou existentes)</p>	<p>te numa serviço de suporte para plataformas de armazenamento em nuvem visando segurança em longo prazo para documentos 'oficiais'; Camadas <i>PaaS</i> e <i>SaaS</i>; <i>Escritório Federal de Segurança em Tecnologia da Informação (Bundesamt für Sicherheit in der Informationstechnik - BSI)</i>: modelo de armazenamento de arquivo em <i>Cloud Services</i> caracterizado por suportar preservação em longo prazo para documentos com assinatura digital;</p> <p>Sistema de Segurança <i>e-Servicing (e-Servicing System of the Insurers - EDAS)</i>: sistema armazenamento que em <i>Cloud Services</i>, consiste em preservar documentos com assinatura eletrônica; <i>Model of archival cloud services supporting long-term preservation</i> (Modelo de arquivamento em <i>Cloud Services</i> que suportam a preservação de longo prazo): criado para gerar segurança aos documentos eletrônicos, levando em conta as especificidades dos registros assinados eletronicamente, os quais são arquivados no ambiente de nuvem como um processo de negócio.</p>
<p><i>Cloud Services</i> identificados no texto (modelos ou existentes)</p>	<p>-</p>
<p>Metadados de preservação; metadados classificados em outras categorias que estão relacionados com a preservação; e modelo de referência <i>OAIS</i></p>	<p><i>XML</i>; <i>OAIS</i></p>
<p>Documento 4 - <i>XML Processing in the Cloud: Large-Scale Digital Preservation in Small Institutions</i></p>	
<p>Serviços de preservação ou Camadas de <i>Cloud Services</i></p>	<p><i>SHAMAN</i>⁴⁰: investiga os processos de preservação em longo prazo de grandes volumes de objetos digitais desenvolvendo uma estrutura de preservação.</p>
<p><i>Cloud Services</i> identificados no texto (modelos criados ou existentes)</p>	<p><i>Amazon Elastic Compute Cloud (EC2)</i>; <i>Amazon Simple Storage Service (S3)</i></p>
<p>Metadados de preservação; metadados classificados em outras categorias que estão relacionados com a preservação; e modelo de referência <i>OAIS</i></p>	<p><i>XML</i>; <i>METS</i></p>
<p>Documento 5 - <i>Content Server System Architecture for Providing Differentiated Levels of Service in a Digital Preservation Cloud</i></p>	

⁴⁰ *Sustaining Heritage Access through Multivalent ArchiviNg (SHAMAN)*. Maiores informações consulte o site: Disponível em: <https://cordis.europa.eu/project/rcn/85468_en.html>. Acesso em: 28 jun. 2018.

Serviços de preservação ou Camadas de <i>Cloud Services</i>	SOA ⁴¹ (<i>Service-Oriented Architecture</i>) -: Arquitetura de Serviços Orientados: sistema arquitetado para trabalhar em camadas de nuvem, o servidor de conteúdo apenas ganha funcionalidade com seu respaldo; <u>Servidores de Conteúdo</u> : construção de arquitetura lógica que abrange componentes de gestão de dados, o armazenamento de componentes do modelo OAIS, considerado o núcleo central da LDPaaS. Sua importância reside na implementação interna da abstração de dados e metadados, permitindo um grande de variabilidade. Suas atribuições são divididas em quatro camadas (Camada de dados físicos, Camada de Serviço, Camada de Processo e Camada de Interface).
<i>Cloud Services</i> identificados no texto (modelos ou existentes)	Preservação Digital em longo Prazo como serviço (<i>Long-Term Digital Preservation as a Service - LDPaaS</i>): modelo de <i>Cloud Service</i> planejado para fornecer armazenamento e preservação de registros digitais. Pode ser uma alternativa viável na provisão independente de registros dos usuários.
Metadados de preservação; metadados classificados em outras categorias que estão relacionados com a preservação; e modelo de referência OAIS	OAIS; XML
<i>Documento 6 - Developing an ontology for cloud-based archive systems</i>	
Serviços de preservação ou Camadas de <i>Cloud Services</i>	Modelo em camadas para serviços de arquivamento (funcionalidades descritas no documento 1); Serviço de Preservação (<i>Preservation Service</i>): utilizado como uma de abstração entre os aplicativos de criação de conteúdo e o arquivo;
<i>Cloud Services</i> identificados no texto (modelos criados ou existentes)	Amazon EC2; Amazon S3
Metadados de preservação; metadados classificados em outras categorias que estão relacionados com a preservação; e modelo de referência OAIS	OAIS; PREMIS; Metadados de preservação dinâmica; Metadados de preservação estatístico; XML; Metadados registrados pelo produtor; Metadados voltados ao produtor; Metadados de Preservação de Registros Automáticos; Metadados de informações de arquivo automático.
<i>Documento 7 - Finding access and digital preservation solutions for a digitized oral history project: A case study</i>	
Serviços de preservação ou Camadas de <i>Cloud Services</i>	Omeka: sistema de gerenciamento de código aberto que oferece opções de

⁴¹ Disponível em: < <https://tqrwrvtjb03.storage.googleapis.com/MDEzMTg1ODU4MA==03.pdf>>. Acesso em: 01 jul. 2018.

	hospedagem para coleções pequenas e médias; <i>DuraCloud</i> : (Documento 2).
<i>Cloud Services</i> identificados no texto (modelos ou existentes)	-
Metadados de preservação; metadados classificados em outras categorias que estão relacionados com a preservação; e modelo de referência <i>OAIS</i>	<i>OAIS</i> ; <i>MODS</i>
Documento 8 - <i>Integral and secure cloud architecture based system for backup and retention of public sector information</i>	
Serviços de preservação ou Camadas de <i>Cloud Services</i>	<i>Metacloud</i> : infraestrutura conjunta de elementos que unifica diferentes interfaces de armazenamento em nuvem;
<i>Cloud Services</i> identificados no texto (modelos ou existentes)	-
Metadados de preservação; metadados classificados em outras categorias que estão relacionados com a preservação; e modelo de referência <i>OAIS</i>	<i>OAIS</i> ; <i>XML</i>
Documento 9 - <i>Digital Preservation in Grids and Clouds: A Middleware Approach</i>	
Serviços de preservação ou Camadas de <i>Cloud Services</i>	<i>SHAMAN</i> (documento 4);
<i>Cloud Services</i> identificados no texto (modelos ou existentes)	<i>Amazon Elastic Compute Cloud (EC2)</i> ; <i>Amazon Simple Storage Service (S3)</i>
Metadados de preservação; metadados classificados em outras categorias que estão relacionados com a preservação; e modelo de referência <i>OAIS</i>	<i>OAIS</i> ; <i>XML</i> ; <i>METS</i>
Documento 10 - <i>Government Use of Cloud-based Long Term Digital Preservation as a Service: An Exploratory Study</i>	
Serviços de preservação ou Camadas de <i>Cloud Services</i>	<i>Preservica</i> : aplicativo de hospedagem de preservação em longo prazo - pode ser hospedado em nuvem; armazena duas cópias de conteúdo na nuvem, uma no <i>Amazon S3</i> , que oferta alta disponibilidade e rápida velocidade de transferência e uma no <i>Amazon Glacier</i> para armazenamento e recuperação em longo prazo com baixos custos.
<i>Cloud Services</i> identificados no texto (modelos ou existentes)	<i>Amazon Glacier</i> ; <i>Amazon S3</i> ; <i>Amazon RDS</i> .
Metadados de preservação; metadados classificados em outras categorias que estão relacionados com a preservação; e modelo de referência <i>OAIS</i>	<i>OAIS</i>
Documento 11 - <i>PDS Cloud: Long Term Digital Preservation in the Cloud</i>	
Serviços de preservação ou Camadas de <i>Cloud Services</i>	<i>PDS Cloud</i> : serviço de armazenamento especializado em preservação digital em longo prazo em ambientes de nuvem, o serviço é baseado no modelo de referência <i>OAIS</i> . Oferece preservação para obje-

	tos armazenados em nuvem. Pode ser dividido em duas camadas descritas a seguir: <i>Multi-Cloud Service</i> : trabalha no acesso de distintas plataformas de armazenamento; <i>Preservation Engine</i> : fornece funcionalidade aos pacotes de informação;
Cloud Services identificados no texto (modelos ou existentes)	<i>Amazon Storage (S3)</i> ; <i>Amazon Compute (EC2)</i> , categorizada como nuvem pública; <i>Openstack Swift Storage</i> ; <i>Nova compute</i> (nuvem privada).
Metadados de preservação; metadados classificados em outras categorias que estão relacionados com a preservação; e modelo de referência <i>OAIS</i>	<i>OAIS</i>

Fonte: elaborado pelo autor.

- *Preserving records in the cloud* (1): nesse documento, o *XML* está embutido num *software* empresarial acoplado na camada *Software as a Service*; de forma sucinta, consiste num *Software* de escritório *Online*. Nesse contexto, é preciso que as informações dos elementos inseridos nessa ferramenta precisam ser registradas antecipadamente, usando um formato legível por máquina, no caso este formato é o *XML*.

- *Towards SIRF: Self-contained information retention format* (2): as funcionalidades do padrão de metadados de preservação digital são aplicadas no *SIRF* (*Self-contained information retention format*), o qual adaptado para a tradução em português significa: Formato autônomo de informações retidas. O *Self-contained information retention format* é um formato de contêiner de armazenamento lógico especializado em conservação por longo prazo. O pacote de informação de arquivo do modelo *OAIS* é aplicado na preservação dos objetos digitais, entre outras características, descreve as informações de preservação do conteúdo de informação dos objetos digitais. Além do mais, o modelo *OAIS* descreve os elementos que devem ser incluídas dentro do pacote de informações de arquivo, porém não especifica os formatos ou como eles são empacotados juntos.

São indicados os seguintes padrões de metadados de preservação ou padrões de metadados que fazem parte do conjunto de ações de preservação, os quais fornecem especificações para o formato e o empacotamento de um objeto de preservação: *XML Formatted Data Unit (XFDU)* - registra todos os

tipos de metadados no seu pacote de informações, permitindo que os metadados registrem cada item separado no objeto digital; também permite que o *designer* de pacotes defina qualquer modelo de metadados, fornecendo atributos para as categorias de metadados e um esquema de classificação mais precisa dentro das categorias; outra característica consiste no seu fornecimento de categorias e classes de metadados predefinidas pelas atribuições do modelo de referência *OAIS* (CCSDS, 2008); *VERS Encapsulated Object (VEO)* - assim como o *XFDU*, este padrão é utilizado para dados espaciais; o padrão de metadados *METS* é aplicado em registros eletrônicos; as aplicações do padrão de preservação digital *PREMIS* em bibliotecas digitais; *Long Term Archiving and Retrieval (LOTAR)* - são aplicados em dados aeroespaciais.

- *Archival Cloud Services: Portability, Continuity, and Sustainability Aspects of Long-term Preservation of Electronically Signed Records* (3): Nesse documento a terminologia metadados de preservação não é utilizada, porém foi encontrado o padrão de metadados *XML*. O modelo de referência *OAIS* somente é citado no texto. O *XML* é citado por conta de formato legível, permitindo a realização de assinaturas eletrônicas. No contexto do *EDS*, o *XML* consiste num arquivo que trabalha com arquivamento de pacotes de informações, o qual contém dados e metadados correspondentes. Os autores do texto criaram um modelo de arquivamento em *Cloud Services* que suportam a preservação de longo prazo, a arquitetura é dividida em três camadas: *Archiving Layer* (camada de arquivamento), *Middleware Layer* (camada Middleware) e *Application Layer* (camada de aplicação); nesse contexto, os aplicativos alocados na camada de aplicação são usados para criar documentos que serão arquivados a partir da compatibilidade interligada com a padronização da preservação digital em longo prazo, e o *XML* é aplicado um dos formatos compatíveis entre a relação dos documentos e a preservação.

- *XML Processing in the Cloud: Large-Scale Digital Preservation in Small Institutions* (4): os autores do presente texto empregaram no *SHAMAN* um programa específico de domínio de linguagem, o *Xeproc*; este programa foi usado especificamente para modelar *pipelines* (ferramenta de gestão usada para acompanhar as etapas de vendas) de processamento *XML* e pontos de verificação de validação *XML*. Os *pipelines* processados em formato *XML* se concentram na identificação de metadados estruturais que descrevem a orga-

nização do documento. Embora o *XML* seja usado em consonância com atividades que norteiam a preservação em longo prazo, e auxilia na preservação, pois sua arquitetura de processamento é descrita para suportar a preservação digital em *Cloud Services*, mas não no sentido de ser empregado como padrão de metadados. A principal função do *XML* nesse documento, consiste na ligação entre *Cloud Services* e *workflows*. Os *pipelines* produzem metadados dentro do padrão de metadados *METS*, os quais endereçam vários aspectos das coleções dos documentos: extração de elementos lógicos dos documentos (tabelas de conteúdo, capítulos); extração do número de páginas dos livros; extração de ilustrações e captações associadas; geração de um arquivo em *XML* com a finalidade de indexar refletindo a organização lógica dos livros. Os dois padrões citados no texto se correlacionam, numa outra situação o *pipeline* de processamento *XML* produziu uma coleção de documentos *XML* no formato *METS*, esta aplicação conjunta faz parte do contexto da preservação digital e também possibilita a extração automática dos documentos.

- *Content Server System Architecture for Providing Differentiated Levels of Service in a Digital Preservation* (5): o documento em questão cita uma categoria de *Cloud Services* denominada *Cloud Preservation*, tal categoria de ambiente de nuvem é anunciada comercialmente oferecendo serviços de registros, armazenamento e arquivamento dos registros digitais, porém tais serviços não empregam as aplicabilidades do modelo de referência *OAIS*; mesmo assim tal serviço enfatiza sua capacidade de usar um registro por muito tempo desde o momento de sua criação. Segundo os autores do texto, preservar registros digitais conforme as exigências do modelo *OAIS* em *Cloud Services*, exige um significativo esforço no planejamento de engenharia e de *design*, além de um conjunto de recursos de computação. O texto apresenta seis características do modelo de referência *OAIS*: inserção de dados com metadados associados e preparação dos dados para o arquivamento em longo prazo; gerenciamento de dados que administram os metadados e objetos digitais - essa função é necessária para monitorar a os dados inseridos e o processo de arquivamento; planejamento de preservação para o gerenciamento da preservação digital em longo prazo - incluem regras e estratégias que regem o processo de preservação; o armazenamento arquivístico é representado por um confiável repositório de objetos digitais; a acessibilidade permite que os usuários

pesquisem, descubram e recuperem os objetos digitais no arquivo; o contexto da administração é aplicado para monitorar os serviços e os processos de gestão do sistema. *Long-Term Digital Preservation as a Service (LDPaaS)* é um modelo de *Cloud Service* especializado em preservação digital - este modelo fornece flexíveis e diferenciados níveis de preservação. A arquitetura do modelo é baseada no modelo de referência *OAIS*. Os autores também examinaram a maioria dos componentes empregados na *LDPaaS*, excluindo apenas os componentes de administração, e tais componentes foram contextualizados a partir das principais funcionalidades do *OAIS*: inserção de serviços - este componente transfere um objeto digital para um repositório em longo prazo; serviços de acesso - a principal finalidade consiste em acessar a área de armazenamento temporário em sistema de grande escala, dessa forma ocorre o aumento da eficiência das explorações que acessam o sistema, o serviço em questão é considerado um desafio no âmbito das ações de provisionamento em *Cloud Services*; os serviços de preservação - consistem no principal núcleo dos componentes elencados, nesse sentido, dois desafios são levantados -, em primeiro lugar, a preservação deve assegurar que os objetos digitais recuperados não sejam corrompidos em datas posteriores; em segundo lugar, uma estratégia de preservação contundente consiste nas transformações dos arquivos sempre que os formatos de arquivo correrem o risco de serem ilegíveis pelo *software* atual ou pelas plataformas de computação; o serviço de armazenamento pode ser consumido ou oferecido como armazenamento em nuvem, nesse contexto, a entrada e a saída nesse ambiente consiste num desafio significativo para grande arquivos.

Vale destacar que este documento descreve de forma abrangente a relação das funcionalidades entre o *Cloud Service LDPaaS* e o modelo de referência *OAIS*. No entanto, as aplicações desse não são desenvolvidas em comparação com os três documentos selecionados; as descrições do modelo em questão terminam nas funcionalidades, ou seja, as aplicações do *OAIS* perdem o protagonismo para uma série de elementos que destacam a arquitetura do *Cloud Service*.

No caso do *XML*, este padrão é empregado como serviço na arquitetura da nuvem, o qual foi empregado com a finalidade de decodificar arquivos, isto

é, no contexto do documento não se enquadra como um padrão de metadados que visa dar suporte nas ações de preservação.

É possível notar uma enorme dificuldade dos textos desenvolverem explicações sobre as aplicações dos metadados de preservação digital tanto nos componentes das camadas de *Cloud Services*.

- *Developing an ontology for cloud-based archive systems* (6): As descrições de classes da ontologia confeccionada pelos autores deste documento, o *Preservation Service* (Serviço de Preservação) é responsável por assegurar a validade e a integridade dos metadados de preservação, no intuito de criar pacotes de informação. E como o domínio de descrição de linguagem (OWL) empregado na ontologia não especifica nenhuma restrição sintática, o serviço de preservação fornece um modelo de registro do esquema XML, o qual deve ser preenchido pela organização proprietária do Aplicativo de Criação (*Drupal*)⁴². Seguindo por outro viés, as atribuições do XML são mescladas com os pacotes de informação criados pelo *PREMIS*, tal junção entre o XML e os pacotes de informação incluem entidades específicas *dos Cloud Services*, como por exemplo, as descrições de plataforma (*platform descriptions*).

- *Finding access and digital preservation solutions for a digitized oral history project: A case study* (7): o texto inicia a explicação sobre o modelo *OAIS*, ressaltando sua importância para a integridade dos objetos digitais inseridos em serviços de hospedagem, pois há um certo nível de diferença entre os serviços de hospedagem existentes, porém todos tentam cumprir as normas estabelecidas pelo modelo. Uma das funcionalidades do *OAIS*, destacada pelos autores, consiste na oferta de uma estrutura conceitual e de uma terminologia consistente para o desenvolvimento e a manutenção dos sistemas de preservação da informação. Os objetos digitais da coleção *Jeffco Stories* foram preservadas no serviço de hospedagem *DuraCloud*, algumas características do serviço de hospedagem em questão com o modelo de referência *OAIS* não são totalmente compatíveis, porém o *DuraCloud* fornece alguns serviços que não são encontrados no armazenamento de *Cloud Services* tradicionais. A

⁴² O *Drupal CMS* é um componente que pertence à classe dos Aplicativos de Criação (*Creating Application*). As funções do *CMS* são descritas nas anotações das classes, as quais também interligam as classes da camada *Software as a Service*. *CMS* (*Content Management System*), cuja tradução para o português significa Sistema Gerenciador de Conteúdo (ASKHOJ; SUGIMOTO; NAGAMORI, 2015).

unidade de informação responsável pela coleção não empregou as funcionalidades exigidas pelo modelo de referência *OAIS* no contexto da preservação digital em longo prazo, os objetos digitais foram armazenados a partir das funcionalidades empregadas pelo *DuraCloud*, as quais não eram totalmente normeadas pelas normas *OAIS*; o modelo de referência em questão estabelece exigências para a criação de pacotes de envio, de arquivamento e de disseminação, e tais pacotes conteriam significativamente mais metadados que auxiliariam, entre outros aspectos, a proveniência, visando à preservação digital em longo prazo. Os autores do presente texto salientam que os recursos empregados pelo *OAIS* são preferíveis no contexto da preservação dos objetos digitais, quando comparados às funcionalidades das aplicações do *DuraCloud*. Um fator decisivo na escolha desse serviço de hospedagem reside na questão dos custos, dentro do contexto da Unidade de Informação (que armazena a coleção), não havia necessidade de um significativo investimento financeiro em infraestrutura técnica. Em contrapartida, o *DuraCloud* oferta um nível básico de preservação digital por baixo custo.

Antes de digitalizar seus objetos analógicos e posteriormente empregar a imigração do repositório para os serviços de hospedagem, a Biblioteca *Jeffco* elaborou um projeto pautado em fornecer acesso de suas coleções aos membros de sua comunidade, porém inicialmente não continham infraestrutura para digitalização dos objetos ou mesmo asseguravam a preservação em longo prazo dos objetos digitais; o repositório que hospedou os objetos digitalizados empregou a plataforma *Islandora/Fedora*, a qual utilizava o *MODS* como o principal padrão de metadados. A principal função do *MODS* nesse ambiente consistia no fornecimento de uma interface para o usuário. A migração da coleção *Jeffco Stories* do repositório para os serviços de hospedagem necessitou empenhar inúmeras técnicas e desafios relacionados aos metadados.

Outro serviço de hospedagem empregado no contexto de migração das coleções *Jeffco Omeka* e o principal padrão de metadados que compõe o escopo de aplicabilidades do *Omeka* é o padrão de metadados *Dublin Core*, porém, no repositório, o padrão de metadados original era o *MODS*; para economizar tempo na reconstrução da coleção no ambiente *Omeka*, os dois padrões de metadados foram cruzados no repositório. Porém, a mescla dos dois padrões implicou numa série de problemáticas, por exemplo, houve falhas na i-

identificação dos objetos digitais e duplicação de itens, desencadeando na inability de corrigir e salvar os vários componentes dos arquivos de áudio. Cruzar diferentes tipos de metadados aumenta o tamanho do objeto digital, não só por inserir mais componentes, mas pelos conflitos que tal relação pode gerar (ao menos no contexto empregado neste documento) na identificação, na acessibilidade, na integridade, na autenticidade, entre outros aspectos que norteiam a preservação digital em longo prazo em *Cloud Services*. Isso significa que os denominados metadados robustos não fornecem uma coleção digital útil, a mixagem do *MODS* com o *Dublin Core* provou a incapacidade do *Omeka* exercer suas funcionalidades. Nesse caso o *MODS* se aplica na categoria de padrão de metadados que trabalham um conjunto de ações objetivando a preservação digital. O texto não utiliza a terminologia metadados de preservação.

- *Integral and secure cloud architecture based system for backup and retention of public sector information* (8): O conceito de *metacloud* descrito pelos autores de forma concisa consiste na junção das diversas interfaces dos *Cloud Services*, dessa forma os conteúdos dos objetos digitais seriam legíveis em qualquer ambiente. O *metacloud* deve se estruturar através do *software X-Road*, o qual intercambia os objetos digitais e, conseqüentemente, permite trocar os objetos inseridos num padrão de contêiner *XML* sem a necessidade de empregar a estrutura destes.

Segundo os autores do presente texto todos os sistemas de informações são capazes de transferir dados para o formato *XML*, e dessa forma as estruturas de dados divergentes são definidos. Uma estrutura criada e unificada, baseada no armazenamento em *Cloud Services* organiza a comunicação mútua e organizada entre os componentes, e conseqüentemente transfere os dados com segurança e conforme a intenção.

Os autores somente constatarem que o modelo de *metacloud* descrito segue os apontamentos estabelecidos pelo modelo de referência *OAIS*.

- *Digital Preservation in Grids and Clouds: A Middleware Approach* (9): As descrições do presente documento iniciam-se de forma fragmentada a partir de uma afirmação bem taxativa, os autores compararam o *SHAMAN* e o modelo de referência *OAIS*.

Para entender amplamente o contexto dos objetos digitais que precisam ser preservados, *SHAMAN* define características que não estão presentes no

OAIS, tais como os pacotes de informação em camadas, onde cada camada é endereçada por uma atividade particular de preservação; tais atividades precedem a fase de inserção, que sucedem a fase de acesso e por último, o refinamento dos pacotes de informação, os quais asseguram informações necessárias na inclusão da preservação digital em longo prazo (WITTEK; DARANYI, 2012).

Os autores conceituam praticamente todas as funções exercidas pelo *OAIS*, verificam as aplicações dos pacotes de informação e as características dos sistemas de informações de arquivo; a arquitetura é devidamente detalhada a partir de um modelo desenvolvido e apontado numa ilustração do texto; tal arquitetura foi elaborada para elucidar o mecanismo de funcionamento do planejamento de preservação, nesse ambiente vários componentes são destacados (gerenciamento de dados, armazenamento de arquivos etc.), todo o ciclo de vida percorrido pelo objeto digital é destacado, do produtor ao consumidor, do envio ao acesso, entre tantos outros componentes e funções que explicam de forma compreensível o modelo funcional *OAIS*.

Ainda, segundo os autores, o modelo de referência *OAIS* foca numa visão completamente arquivística, tendo um escopo bem limitado quando se trata do ciclo de vida completo de um objeto digital; na contramão desse contexto, o *SHAMAN* investiga a preservação em longo prazo de grandes volumes de objetos digitais, no qual é desenvolvido um *framework* de preservação, aberto verificável e extensível.

Com relação ao *XML*, quase todas as descrições, as funcionalidades, o contexto das temáticas (por exemplo, a relação entre este padrão e os *pipelines*), foram tecidos por muitos dos autores do documento 4 (*XML Processing in the Cloud: Large-Scale Digital Preservation in Small Institutions*). Etapas de fluxo de trabalho de preservação são relativamente caras, principalmente no âmbito computacional, e nesse contexto migrar um formato *XML* para ser transformado num formato *XSLT* consiste numa atividade complexa, afinal os processadores *XSLT* são constantemente atualizados, empregando técnicas de otimização em linguagens de programação e em consultas de banco de dados.

Os autores implementaram um modelo de objeto de documento, cuja utilização consistia na aplicação de plataforma cruzada não otimizada, combinando uma linguagem independente para representar e interagir com objetos

num documento *XML*, porém tal processo demanda muito tempo, além do mais, os recursos computacionais ofertam recursos suficientes para tal procedimento. As transformações em *XML* podem ser consideradas particularmente onerosas, porém metadados de extração automática é norteado por um conjunto de alto teor de complexidade nas atribuições relativas ao processamento de linguagem, no contexto dos aspectos computacionais da preservação digital. Nesse escopo, estes metadados processam tarefas de análises profundas ao reconhecer detalhadamente cada entidade nomeada. É possível observar a importância das aplicações do *XML* no sistema de *workflows* de preservação, os quais foram processados para trabalhar em *Cloud Services*. Entretanto, o *XML* não é utilizado no contexto da das funcionalidades e das aplicabilidades que envolvem a preservação digital nos *Cloud Services*. A função mais adequada do *XML* é norteada pela estrutura lógica do documento (título, seção, capítulo, parágrafo etc.), em vários textos, inclusive nas descrições deste. Sua funcionalidade na preservação dos *workflows* (*Preservation Workflows*) é construída a partir das ações do *Xeproc*, este domínio específico de linguagem preserva as representações abstratas dos objetos digitais, e posteriormente emprega o *XML*, os quais podem ser vistos como especificações, independente das especificações serem instanciadas ou implantadas por longos períodos de tempo.

O *METS* aborda vários aspectos das coleções de documentos, incluindo a organização lógica, a extração de documentos, as legendas, a extração de ilustrações etc. Também emprega casualmente *pipelines* empacotados, os quais podem ser implantados numa plataforma de produção, vale destacar que a saída de um *pipeline* processado por *XML* consiste numa coleção de documentos *XML* no formato *METS*.

Este documento contém todos os elementos que contemplam a temática da presente pesquisa: metadados de preservação (mesmo sem explicitação do termo); serviços de preservação projetados para executar funções em *Cloud Services*; diversos componentes, suportes e recursos; *frameworks*; *pipelines*; domínios específicos de linguagem; novas abordagens que contestam e contrastam com o modelo de referência *OAIS*, a partir dos modelos do *SHAMAN*; arquivamento em rede; funcionalidades do *XML* aplicadas na preservação dos *workflows* etc. Porém, os aspectos específicos que envolvem os padrões de

metadados de preservação digital em *Cloud Services* foram descritas de forma bem esclarecedora, elucidando pontos importantes, relacionando metadados de preservação; no entanto, o foco do texto é concentrado em outros recursos e ferramentas do escopo da preservação digital, sendo assim não foram desenvolvidas as aplicabilidades dos padrões de preservação, estruturais, técnicos e descritivos. Foram elucidados o funcionamento de sistema de engenharia, *framework* etc.

- *Government Use of Cloud-based Long Term Digital Preservation as a Service: An Exploratory Study* (10): Segundo a afirmação da autora, em 2015 existiam poucos *Cloud Services* de preservação digital em longo prazo. As características, as funções, as regras, as aplicabilidades e as normas *ISO* foram contextualizadas nos outros documentos analisados. Foram elencados dois modelos de serviços de hospedagem que conseguem suportar os pacotes de informação do modelo de referência *OAIS* na preservação digital em longo prazo em *Cloud Services*.

O *Preservica* pode incluir dois sistemas num pacote, um compatível com *OAIS*, suas funções consistem em preservar o conteúdo, na acessibilidade, no compartilhamento e apresentação de conteúdos e coleções; o outro pacote tem a função de importação automatizada, isto é, importa o conteúdo exportado do *DSpace*⁴³, *SharePoint*⁴⁴, entre outros ambientes, dentre outras aplicações é possível destacar, a coleta avançada de *sites*, inserção em massa do conteúdo digitalizado. Visando o armazenamento durável dos metadados em *Cloud Services*, o denominado *Cloud Edition* da *Preservica*⁴⁵ utiliza estes três ambientes

⁴³ Consiste num *software* livre, cujo principal objetivo reside na transferência de responsabilidades e custos com outras com atividades de arquivamento e publicação da produção institucional das organizações que adotaram esse *software*. O *DSpace* contém natureza operacional específica de preservar objetos digitais. Também possibilita a criação de repositórios digitais com funções de armazenamento, gerenciamento, preservação e visibilidade da produção intelectual, além de permitir o gerenciamento da produção científica em qualquer tipo de material digital, garantindo sua acessibilidade por longos períodos de tempo. Disponível em: <<http://www.ibict.br/pesquisa-desenvolvimento-tecnologico-e-inovacao/Sistema-para-Construcao-de-Repositorios-Institucionais-Digitais>>. Acesso em: 11 jul. 2018.

⁴⁴ Consiste numa plataforma de colaboração que interliga internet, extranet e intranet, esta plataforma foi criada pela Microsoft. Atualmente *SharePoint* teve seus serviços desenvolvidos, criando e implantando uma ampla gama de opções. *SharePoint* pode consistir em um ou mais produtos, incluindo: *SharePoint Online*; *SharePoint Server*; *SharePoint Foundation*; *SharePoint Designer 2013* etc. Disponível em: <<https://support.office.com/pt-br/article/o-que-%C3%A9-o-sharepoint-97b915e6-651b-43b2-827d-fb25777f446f>>. Acesso em: 11 jul. 2018.

⁴⁵ Consiste num serviço de hospedagem em *Cloud Services* que visa solucionar questões que norteiam a preservação em longo prazo e a acessibilidade dos objetos digitais, geralmente utilizado por organizações e por consórcio de pequeno e médio porte. Caracterizada por otimi-

em nuvem: *Amazon RDS* para armazenamento durável de metadados, assim como o *Amazon S3* e/ou *Glacier*⁴⁶. Já o *Archives Direct* consiste num *software* de código aberto nos quais desenvolveram vários serviços, o *DuraCloud*, preservação de *workflows*, armazenamento arquivístico e serviços de preservação para armazenamento.

Existe certa homogeneidade na descrição da maioria dos documentos analisados quando se trata do modelo de referência *OAIS*.

Basicamente a estrutura é a mesma: os documentos analisados citam que *OAIS* é uma norma *ISO: 16363: 2012: / ISO: 14721: 2012*, em seguida são destacados as finalidades do modelo, suas características, as especificações dos pacotes de informação, entre outros aspectos elencados repetidamente na maioria dos onze documentos. Por fim, depois de salientados os recursos e as características, a maioria dos textos não menciona de forma detalhada o contexto das aplicações e das atribuições do *OAIS* nos *Cloud Services*; tais aplicações geralmente privilegiam programas, *softwares*, aplicativos, sistemas, pacotes de informação etc., mas geralmente é explanado um recorte. A maioria dos onze documentos analisados finaliza as construções textuais elencando os pontos positivos do modelo e resumindo as problemáticas nas questões em torno do custo e do tamanho dos metadados nos arquivos, os quais não são suportados nos pacotes de informações do modelo de referência *OAIS* que trabalham com prese.

- *PDS Cloud: Long Term Digital Preservation in the Cloud* (11): As especificações, as características, as funcionalidades e as aplicações do modelo *OAIS* estão devidamente detalhadas no subcapítulo (6.3).

Diante do cenário apresentado é possível destacar e novamente ressaltar, que os padrões de metadados de preservação digital adotados e aplicados na preservação em longo prazo (assim como os padrões de metadados que

zar necessidades e custos de armazenamento do usuário, o qual pode decidir armazenar diferentes tipos de conteúdo na *Amazon S3* ou na *Amazon Glacier Preservica*. Disponível em: <<https://preservica.com/digital-archive-software/cloud-edition>>. Acesso em: 11 jul. 2018.

⁴⁶Consiste num serviço de armazenamento em nuvem, seguro, durável e de custo relativamente baixo para arquivamento e backup de dados de longo prazo. Foi projetado para fornecer resiliência de 99,999999999%, além de oferecer recursos que abrangem a segurança e conformidade, cujo objetivo reside em poder auxiliar no cumprimento a requisitos normativos considerados mais rigorosos. Esse ambiente oferta funcionalidade de consultas no local, permitindo a execução de análises avançadas diretamente em dados ociosos arquivados. Disponível em: <<https://aws.amazon.com/pt/glacier/>>. Acesso em: 12 jul. 2018.

agem de forma decisiva nas estratégias de preservação digital) em *Cloud Services* são pouco explorados nos documentos analisados. Em oito dos onze documentos selecionados foi possível verificar que as funções e a aplicabilidade não são pormenorizadas nos estudos, isto é, dentro do contexto analisado em oito documentos, os padrões de metadados de preservação não são os objetos de estudo protagonistas na relação que envolve preservação digital de longo prazo e *Cloud Services*.

6.3 Análise das características, funções e aplicações do padrão de metadados de preservação digital PREMIS e do modelo de referência OAIS em *Cloud Services*

Conforme destacado anteriormente no presente estudo, há uma ínfima quantidade de pesquisas que interligam metadados de preservação digital em *Cloud Services*. Independente de ser uma tecnologia que se popularizou recentemente, independente das incessantes modificações nos diversos tipos de serviços disponíveis e independente das novas ferramentas tecnológicas que continuamente acoplam o escopo de funções dos *Cloud Services*, a preservação em longo prazo necessita de estudos mais aprofundados, com a finalidade de serem desenvolvidas implementações para possíveis alternativas na resolução dos problemas que envolvem a obsolescência dos objetos digitais.

Vale salientar novamente que em vários documentos verificados na literatura, a terminologia metadados de preservação digital em *Cloud Services* não foi mencionada, apesar da citação e até mesmo o emprego dos padrões de metadados de preservação digital. Nessa direção, tal omissão do termo por outras áreas do conhecimento dificulta o entendimento e a importância de todo o contexto envolvendo os metadados, incluindo os metadados de preservação digital. Por exemplo, entre outros efeitos negativos acarretados pela indiferença com o termo, várias publicações que poderiam ser endereçadas ao cientista da informação correm o risco de não serem recuperadas.

Os metadados são fundamentais em todo contexto do ciclo de vida de um objeto digital, sendo assim é preciso conscientizar outras áreas do conhecimento sobre o valor dos metadados e, por conseguinte endossar a dimensão

de sua relevância na preservação dos conteúdos digitais para a posteridade. A impactante omissão da terminologia revela pistas do quão a preservação em longo prazo está destoada dos incessantes avanços tecnológicos, principalmente em *Cloud Services*.

Outra característica salutar observada nos documentos analisados reside no fato dos padrões de metadados de preservação serem minimamente explorados nos estudos que interligam funções e aplicações dos objetos digitais relacionadas aos *Cloud Services* existentes, ou mesmo em alguns modelos projetados.

Porém todos os onze documentos analisados se complementam quando se tratam das seguintes características: construção de indagações sobre problemáticas sem soluções, e sobre alternativas estratégicas que norteiam a preservação digital em longo prazo nesse tipo de ambiente; indiferentemente do foco dos estudos se pautar exclusivamente nos padrões de metadados de preservação digital, a importância implícita ou explícita destes padrões estão contempladas em todos os documentos.

Estudar minuciosamente os padrões metadados de preservação digital inseridos num ambiente de constante mobilidade é uma tarefa dispendiosa e complexa. Todo o processo que envolve a preservação digital em ambientes de nuvem é caracterizado por muitos elementos, e mesmo com as várias semelhanças temáticas agrupadas em todos os textos analisados, não há como especificar todas as interligações entres os elementos que compõem o ciclo de vida dos objetos digitais inseridos em *Cloud Services*, mesmo se tratando apenas de preservação digital em longo prazo, tampouco é possível detalhar num único texto todas as relações entre as entidades envolvidas no processo.

Da mesma forma a maioria dos textos não contemplam de forma completa as especificações do modelo de referência OAIS, tal modelo em questão contém características explicitamente relevantes em priorizar a preservação dos objetos digitais nos mais diferentes tipos de ambientes de armazenamento, incluindo claramente os metadados que exercem funções e ações que abrangem a preservação digital.

Entretanto, três documentos descreveram e detalharam as funções e as aplicações do padrão de metadados de preservação digital PREMIS e o modelo de referência OAIS em *Cloud Services: Preserving records in the cloud* (1);

Developing an ontology for cloud-based archive systems (6); *PDS Cloud: Long Term Digital Preservation in the Cloud* (11). Diversos atributos foram especificados minuciosamente, em todo percurso percorrido foram referenciadas as relações com camadas, sistemas, arquivos, pacotes de informação etc.

Neste subcapítulo, apenas foram destacados os três documentos citados acima, os quais detalharam de forma específica, características, funções e aplicações dos padrões de metadados de preservação digital alocados nas camadas dos *Cloud Services*, e conseqüentemente suas particularidades que visam garantir à preservação do conteúdo intelectual dos objetos digitais.

Nos oito documentos restantes, características, funções e aplicações deste tipo de padrão tiveram uma contribuição delimitada na conjuntura dos assuntos explanados sobre a temática, isto é, as explicações de suas ações tiveram uma atuação coadjuvante, afinal o enfoque abarcava comentários sobre aplicações de outras questões que envolvem os processos de preservação em *Cloud Services*.

- *Preserving records in the cloud* (1): o estudo examinou características de gerenciamento de registro em ambientes de nuvem ao comparar modelos de arquivamento existentes, exemplificados pelo modelo de referência *OAIS*, e buscou correlacionar entidades funcionais em *OAIS* com um modelo de *Cloud Services* criado pelos autores do documento em questão, nesse sentido, os autores verificaram que há diversas áreas onde o modelo *OAIS* não se integra bem às camadas dos *Cloud Services*. Entretanto, um novo modelo de camadas para sistemas de arquivos em nuvem foi desenvolvido usando conceitos e tipos de informações do modelo de referência *OAIS*.

O modelo de *Cloud Services* proposto pelos autores aspira demonstrar aos produtores de registro os benefícios da provisão de metadados de preservação no gerenciamento de sistemas de arquivo. A *Preservation Layer* (camada de preservação) deve fornecer metadados de preservação para sistemas de arquivamento. Se isso não puder ser feito, os arquivos ficarão sobrecarregados na tarefa de criar ou complementar os metadados ausentes.

Para exemplificar a construção do modelo proposto, os autores utilizaram o metadados de preservação *PREMIS*. Apesar do *PREMIS* não cobrir todos os aspectos dos metadados de preservação, este padrão define as informações que a maioria dos repositórios de preservação precisa para garan-

tir a preservação digital em longo prazo. Os padrões de metadados de preservação *PREMIS* foram divididos em três subcategorias de metadados, conforme Askhoj, Sugimoto e Nagamori (2011):

- Metadados gerados para sistemas de negócio no momento exato da criação ou uma declaração de registro: são no mínimo metadados descritivos para preservação que só podem ser fornecidos pelos sistemas de negócio;
- Metadados de Informações pré-registradas: as informações estatísticas devem ser fornecidas aos sistemas avançados no modelo de nuvem, isto é, devem ser fornecidas informações sobre entidades registradas no sistema;
- Metadados de informações relacionadas aos eventos: informações que descrevem alterações em objetos digitais e em metadados durante o processo de preservação.

As principais funcionalidades dos metadados de preservação *PREMIS* no modelo de *Cloud Services* proposto pelos autores podem ser divididas da seguinte maneira: devem ser fornecidos pela *Preservation Layer*, por meio de categorias caracterizadas, no intuito de gerenciar os objetos digitais para sistemas de negócios e para metadados; suas atribuições giram em torno de registrar as informações sobre os sistemas de negócios (representação de informações simples, informações sobre esquemas de metadados, informações sobre sistemas técnicos etc.); e por fim, documentar mudanças ocorridas durante a criação do pacote em objetos digitais e em outros tipos de metadados. Dessa forma é possível definir os requisitos que possibilitam fornecer metadados de preservação para os *Cloud Services* existentes.

Os sistemas de arquivamento na camada de interação registram informações básicas na camada de preservação, garantindo que apenas os sistemas autorizados possam acessar os pacotes. Os sistemas de arquivamento recebem pacotes completos de informações virtuais diretamente da Camada de Preservação, tendo a opção de gerar e salvar metadados adicionais na camada *Platform as a Service*. No sistema de arquivamento, um típico fluxo de

informações segundo Askhoj, Sugimoto e Nagamori (2011) seria da seguinte forma:

- Um e-mail é criado pelo usuário, o documento é anexado em *HTML*⁴⁷ e enviado para um destinatário;
- O e-mail é registrado no sistema de negócios, e está bloqueado na *Platform as a Service*, porém uma notificação no formato *XML* é enviada para o Empacotador⁴⁸;
- Baseado na notificação, o Empacotador recupera os dados relevantes da camada *Platform as a Service*. Os dados são validados e convertidos. Na conversão dos dados um arquivo *XML* é criado através dos metadados de preservação *PREMIS*, o qual se baseia nos metadados de sistemas de negócios, informações pré-registradas e informações relacionadas ao evento;
- Essas informações e o objeto de dados são salvos na camada *Platform as a Service* como um pacote virtual. Uma notificação indicando as principais cadeias de *bits* é enviada para camada de interação.

- *Developing an ontology for cloud-based archive systems* (6): Um arquivo necessita que o objeto digital seja acessível por muito tempo, mesmo se a organização que os criaram corra o risco de desaparecer. Arquivos digitais usados ou fornecidos por diversas organizações em *Cloud Services* são armazenados de forma inadequada. Num ambiente de nuvem, a capacidade de abstrair serviços é importante, em particular quando se trata de interoperabilidade, porque permite que uma organização concentre nas partes que os serviços

⁴⁷ O *HTML* pode ser definido como uma linguagem base da internet, na qual é utilizada para desenvolver *websites*, foi criada para ser de fácil entendimento pelos usuários e também pelas máquinas. Sua sigla se origina do inglês e significa *Hypertext Markup Language*, ou em português - Linguagem de Marcação de Hipertexto. Disponível em: <<https://tableless.com.br/o-que-html-basico/>>. Acesso em: 11 jul. 2018.

⁴⁸ Empacotador ou *Packager*. são definidos como arquivos que contém instruções pormenorizadas em etapas, direcionadas a um computador, para que o computador siga corretamente as instruções e realize as funções. Disponível em: <<https://www.solvusoft.com/pt-br/files/erro-remo%C3%A7%C3%A3o-do-v%C3%ADrus/exe/windows/microsoft/msdn-disc-1665/packager-exe/>>. Acesso em 11 jul. 2018.

precisam interoperar, tendo a segurança que outras peças desse sistema apenas funcionem conforme os parâmetros predefinidos.

Nos *Cloud Services* existentes e analisados no texto, as camadas de *Infrastructure as a Service*, *Platform as a Service* e *Software as a Service* podem estar fora do controle da organização do sistema de arquivamento (já devidamente explicado nas descrições do documento 1). Nesse sentido, se faz necessário descrever os tipos de dados produzidos e recebidos por cada camada. Sem essa informação, torna impossível abstrair a funcionalidade, pois sem esta aplicação não há garantias que os dados importantes para esta finalidade serão produzidos no formato certo.

Askhoj, Sugimoto e Nagamori (2015) desenvolveram uma ontologia que não apenas descreve valores de metadados, mas também dependências, entradas e resultados. O domínio de descrição de linguagem escolhido foi o *OWL* (*Web Ontology Language*⁴⁹) por oferecer melhor expressão semântica interpretada pela máquina, isto é, ocorreu o desenvolvimento de uma ontologia como uma ferramenta de aprovação, a qual aponta as possibilidades de utilizar a validação automática dos objetos digitais através de diferentes variáveis. O objetivo da ontologia consiste em definir um vocabulário comum em arquivos de nuvem, atribuindo responsabilidades para criação e transferências de dados, incluindo o registro de sistemas de aplicativos de criação de conteúdo. A descrição dos componentes possibilita melhor interoperabilidade de metadados entre os aplicativos de criação de conteúdo, e também entre os serviços que fornecem metadados de preservação.

Os metadados de preservação digital *PREMIS* foram adotados para trabalhar em conjunto com a ontologia descrita anteriormente, devido sua notoriedade em assegurar a preservação digital em longo prazo nos processos que envolvem o ciclo de vida dos objetos digitais em *Cloud Services*, e em conjunto com as descrições da ontologia projetada. A criação de um modelo de *Cloud Services* visa preencher o elo entre produtores e arquivos que geram pacotes de informação de submissão, por meio de um processo automatizado os obje-

⁴⁹ Foi elaborada para representar um conhecimento rico e complexo sobre coisas, grupos de coisas e relações entre as coisas. Sua linguagem é baseada em lógica computacional, podendo ser explorado por programas de computador, entre outros aspectos, para verificar a consistência desse conhecimento ou para tornar explícito o conhecimento implícito. Disponível em: <<https://www.w3.org/OWL/>>. Acesso em 11 jul. 2018.

tos digitais são enviados junto com os metadados *PREMIS*, permitindo a interoperabilidade entre aplicativos de criação de conteúdo e o serviço de preservação.

Há uma sólida inclusão na pesquisa do dicionário de dados *PREMIS*. Para melhor aplicação prática do modelo de *Cloud Services* elaborado pelos autores, todos os componentes do modelo foram hospedados na plataforma *EC2* da *Amazon*.

Este padrão de metadados de preservação digital cria um pacote de informação mesclado com atribuições de *XML* genérico, o qual inclui entidades específicas dos *Cloud Services*, como a *Preservation Description Information* (Informações sobre Descrição da Preservação). As características do padrão de metadados de preservação digital *PREMIS* são importantes por garantir um certo grau de interoperabilidade entre aplicativos de criação de conteúdo e o serviço de preservação.

Askhoj, Sugimoto e Nagamori (2015) confeccionaram um modelo de camadas para *Cloud Services*, as quais se familiarizam, em alguns aspectos, com modelos de *Cloud Services* existentes, porém o modelo planejado se diferencia dos modelos de camadas existentes por possuir quatro camadas e ao mesmo tempo, por subtrair a *Infrastructure as a Service*, esta camada foi mesclada e inculida nas camadas de *Platform as a Service* e de *Software as a Service*:

- Primeira camada (*Platform as a Service*): ambiente onde estão armazenados os objetos digitais. Esta camada interage com a segunda camada por meio de *bits-streams*⁵⁰ e arquivos. Inclui fornecimento de processamento, armazenamento, redes/recursos básicos de computação, capacidade de controle sobre os sistemas operacionais e garante a integridade dos objetos;

⁵⁰ "[...] refere-se à maneira como o sistema interpreta os arquivos neste processo, *byte a byte*, isto é, independentemente da entrada (qualquer que seja a semântica da mesma), tudo é uma sequência de bits". (ROCHA et al., 2004, p. 5).

ROCHA, A. et al. . Camaleao: um software de esteganografia para protec ao e privacidade digital. In: Simpósio Segurança em Informática (SSI), 2004, São José dos Campos. **Anais do 6º Simpósio de Segurança em Informática**, 2004. p. 1-9.

- Segunda camada (*Software as a Service*): os aplicativos de criação de conteúdo são desenvolvidos nesse ambiente, e interage com a terceira camada (*Preservation Layer*, traduzido para língua portuguesa significa camada de preservação) através da representação dos objetos digitais. A principal função dos aplicativos de conteúdo consiste em possibilitar a criação de um documento com vários tipos de dados diferentes (imagens, textos, formatação de dados etc.). Os aplicativos de criação de conteúdo devem ter a capacidade de exportar os objetos digitais num formato aceitável, para que tais objetos sejam entendidos;
- Terceira Camada (*Preservation Layer*): basicamente nessa camada se desenvolve o *Preservation Service*, e assim como as outras camadas também interage com a camada acima, a quarta camada. Os objetos digitais precisam acompanhar as aplicações dos metadados antes de serem inseridos no arquivo. Estes metadados são necessários para certificar a preservação em longo prazo (por exemplo, informações sobre proveniência e infraestrutura), e se diferem dos metadados adicionados na camada *Software as a Service*, os quais são específicos para aplicativos de criação, tais como metadados de descrição (*Dublin Core*, *MODS* ou similares). Diferente do contexto aplicado pelo modelo *OAIS*, nos quais os pacotes de informação são fragmentados em vários para fornecer conteúdos de informação e preservação, a maior parte das informações sobre aplicativos de criação de conteúdo, objetos digitais e metadados é produzida para serviços de preservação avançados. Porém, é possível criar pacotes de informação de submissão no intuito de completar a *Preservation Description Information*, e da Representação da Informação (*Representation Information - ReplInfo*). A *Preservation Description Information* e a *ReplInfo* fazem parte do escopo de metadados necessários

para preservação digital em longo prazo. No modelo proposto a *ReplInfo* deve tomar forma de uma Compartilhada Representação da Informação registrada (*Representation Information Registry - RIR*). A *Preservation Description Information* pode ser dividida em quatro partes: informação de referência (em ambiente de nuvem utiliza-se o formato *URI*); informações de proveniência (documentam o histórico de um objeto digital e as alterações durante seu ciclo de vida, incluindo o arquivamento); descrição das informações que relacionam objetos digitais com seu ambiente; novas informações são fixadas e adicionadas nos objetos digitais para certificar que estes não sofram mudanças por conta de erros no sistema ou por adulteração;

- Quarta camada (*Interaction Layer*): nessa camada são elaborados os sistemas de arquivo. Os agentes (usuários ou sistemas) criam, gerenciam ou arquivam um objeto digital através um *browser*⁵¹, ou por meio de algum sistema específico. Na camada de interação, os usuários podem acessar e ler informações recebidas, mas não podem fazer alterações ou sobrescrevê-las.

A *Preservation Layer* também se responsabiliza por alocar e referenciar o armazenamento na *Platform as a Service*, dessa forma, arquivos podem ter acesso aos conteúdos através de identificadores únicos. O armazenamento dos objetos digitais se aloca nessa camada. Sistemas de arquivo e aplicações de conteúdo compartilham em comum uma mesma plataforma de armazenamento. Como o armazenamento precisa ser confiável e de longa duração ele é alocado no *Preservation Service*, a qual serve como uma camada de abstração entre o aplicativo de criação de conteúdo e o arquivo. Nesse cenário, usar o

⁵¹ Consiste num programa de computador responsável pela navegação na *Internet*, é tradicionalmente conhecido pelo termo navegadores. Disponível em: <<http://bloginformaticamicrocamp.com.br/computadores/o-que-e-browser/>>. Acesso em: 12 jul. 2018.

registro do *template*⁵² originário do *Preservation Service* se constitui num aspecto decisivo, pois permite e define o formato de quaisquer metadados enviados. A atribuição do registro do *template* registra informações sobre os objetos digitais produzidos, incluindo de qualquer esquema de metadados e das aplicações de criação de conteúdo.

O *Preservation Service* desenvolve um pacote de informação de submissão para o sistema do arquivo contendo *URI* destinado aos objetos digitais, por essa razão são empregadas às aplicações que subdividem os metadados de preservação em dois tipos. Os Metadados de Preservação Estatísticos consistem em informações que caracterizam o tipo de inibidor (por exemplo, o *Blowfish*⁵³) e são empregados nas aplicações de criação de conteúdo. Já os metadados responsáveis pelo registro são diferentes em cada objeto digital, denominados Metadados de Preservação Dinâmica, possuem diferentes tamanhos em cada objeto digital e cobrem grande parte de informações necessárias para criar pacotes de informação de submissão.

A ontologia pode ser usada para descrever metadados de preservação nos pacotes de preservação, além de descrever entidades inclusas num arquivo de nuvem. O sistema baseado na ontologia em *Cloud Services* contém um alto grau de complexidade. Os metadados de preservação digital se originam de mais de uma fonte, ou seja, precisam passar por uma série de etapas: submissão, agregação e cruzamento potencial antes de se incluir num pacote de envio. Para cada objeto digital alocado na camada de preservação há quatro fontes diferentes de metadados de preservação apontadas por Askhoj, Sugimoto e Nagamori (2015):

- Metadados registrados pelo produtor: consiste nos metadados que foram pré-registrados pelo produtor que utilizou o modelo de registro;

⁵² De forma sucinta, *templates* podem ser definidos como páginas pré-configuradas, as quais permitem serem editadas por meio de um formulário. Além do mais são úteis para a publicação de conteúdo de forma consideravelmente rápida. Disponível em: <<https://screencorp.zendesk.com/hc/pt-br/articles/201227124-O-que-s%C3%A3o-templates->>>. Acesso em: 12 jul. 2018.

⁵³ O *Blowfish* foi elaborado como uma alternativa gratuita, considerado mais veloz para os algoritmos, além de ser considerado um algoritmo forte. O *Blowfish* não é patenteado, tem sua licença grátis e está à disposição para todos. Disponível em: <<https://www.schneier.com/academic/blowfish/>>>. Acesso em: 12 jul. 2018.

- Metadados voltados ao produtor: são fornecidos ao produtor no momento da exportação;
- Metadados de Preservação de Registros Automáticos: são criados pelo Serviço de Preservação no momento da importação;
- Metadados de informações de arquivo automático: relativo ao tratamento de metadados do sistema com propriedades de arquivo, como tamanho, data de criação e extensão.

O núcleo central da ontologia desenvolvida para *Cloud Services* gira em torno do conceito de níveis de entrega em diferenciados serviços de preservação, localizados na camada de preservação. A descrição sobre os componentes de *Cloud Services* existentes é mais rudimentar, não indo além de diferentes descrições que tratam as entidades dos *Cloud Services* com funções atribuídas a diferentes camadas e, portanto, não utilizam, entre outros aspectos, a interação entre as camadas.

A ontologia proposta descreve os metadados de preservação nos pacotes de informação, além de descrever entidades inclusas num arquivo inserido no modelo de *Cloud Services* (elaborado pelos autores deste artigo), e contém um alto grau de complexidade. Pelo fato dos metadados de preservação se originar de mais de uma fonte é necessário que eles passem pelas etapas de submissão, agregação e cruzamento em potencial e, posteriormente, sua inclusão num pacote de envio.

- *PDS Cloud: Long Term Digital Preservation in the Cloud* (11): Rabinovici-Cohen et al. (2013) desenvolveram um modelo de serviço de armazenamento: *Preservation Data Stores in the cloud (PDS Cloud)*, em português o termo significa preservação de dados no armazenamento em nuvem. O objetivo principal do *Preservation Data Stores in the cloud* gira em torno de manter a capacidade de compreensão do conteúdo digital (preservação lógica) em longo prazo, aderindo às mudanças dinâmicas dos requisitos de preservação e se adaptando ao incessante panorama de evolução da tecnologia.

Este serviço foi construído com base no modelo de referência *OAIS*, interconectando o padrão de metadados de preservação digital *OAIS* aos diversos *Cloud Services* existentes. O modelo de referência *OAIS* é reconhecido por utilizar de forma extensiva os metadados, acoplados em dados brutos, como

parte integrante dos pacotes de informação de arquivos, conforme já salientados nos documentos 1 e 6.

Os atuais serviços de armazenamento nesses ambientes têm suporte bastante limitado de metadados. O espaço permitido para metadados (por objeto) é muito pequeno para o tamanho extenso dos metadados de preservação, pois faltam recursos para empregá-los. Além disso, ocorre a possibilidade dos metadados se transformarem, crescendo de tamanho durante o ciclo de existência dos pacotes de informação.

Os *Cloud Services* existentes não suportam a atualização de metadados sozinhos, assim como o custo de transferência de dados de seus dados. A *Preservation Data Stores in the cloud* consegue trabalhar suportando a preservação lógica, se caracterizando como um agente que interconecta com o modelo de referência *OAIS* com diversos *Cloud Services* existentes. No presente documento, foram examinados diversos *Cloud Services*, utilizando como fio condutor da análise, os mecanismos de armazenamento norteados pela atuação da *Preservation Data Stores in the cloud*.

Foram mapeadas diferentes descrições de estruturas dos pacotes de informação de arquivo no *Amazon S3*. Os pacotes foram divididos na seguinte ordem: estrutura lógica, mapeamento, representação da informação e versões de pacotes. No contexto de estrutura de descrição dos pacotes, a representação da informação consiste numa importante classe de metadados de preservação, pois descreve como interpretar os dados do conteúdo.

Os pacotes de informação do *OAIS* contêm um compartimento de Informações sobre descrições de preservação (*Preservation Description Information* - *PDI*) - no *Preservation Description Information* são encontrados os Conteúdos de Informação, como o *Representation Information (RepInfo)*. A *Representation Information* consiste numa importante classe de metadados de preservação, pois descreve como interpretar o conteúdo dos dados. Na *Preservation Data Stores in the cloud*, a *Representation Information* é mantida separado dos pacotes de informação com seu próprio e exclusivo identificador.

A camada de preservação garante o agrupamento em longo prazo de metadados e dados, auxiliando na automatização dos processos de preservação. A partir do emprego e aplicação da *Preservation Data Stores in the cloud* como uma camada de preservação nas múltiplas camadas de *Cloud Services*

existentes, foram identificados vários recursos relevantes para assegurar a preservação por longos períodos. Em muitos casos, os recursos eram parecidos, mas não semelhantes, por outro lado, as camadas dos *Cloud Services* existentes também evidenciaram falhas encontradas nas múltiplas plataformas, as quais se destacam: confiabilidade do *bit*; bloqueio de dados; certificação e confiança; preservação lógica; acompanhamento de eventos; armazenamento e cálculo de sinergia. Da mesma forma, também foram constatadas falhas em torno dos metadados, mais especificamente nos metadados de preservação.

A *Preservation Data Stores in the cloud* foi um serviço arquitetado para trabalhar como uma camada intermediária, pois se interconecta entre *OAIS* e vários *Cloud Services*, expondo um conjunto de serviços de preservação baseados nos pacotes de informação *OAIS*, exemplificadas por ações de inclusão, de acesso, de exclusão e de preservação. Nesse contexto, a *Preservation Data Stores in the cloud* se divide em duas camadas: a *Multi-Cloud Service* (múltiplos serviços em nuvem) e a *Preservation Engine* (mecanismo de preservação). A *Multi-Cloud Service* acessa um conjunto heterogêneo de plataformas de armazenamento dos *Cloud Services*. A *Preservation Engine* fornece funcionalidades de preservação para os pacotes de informação.

Vale salientar que a *Preservation Data Stores in the cloud* serve como uma estrutura conceitual de referência para arquivos e provedores que almejam armazenar e preservar os registros nos *Cloud Services*. Dentro do âmbito de atribuições do modelo elaborado, um provedor só precisa fornecer um tipo de metadados que atenda aos requisitos do serviço de preservação. A camada *Software as a Service* necessita ter a capacidade de exportar o conteúdo e os metadados em formatos utilizáveis ao ser submetido ao processo de preservação em longo prazo.

A partir da análise feita na literatura, é possível reafirmar que ainda há poucos estudos sobre preservação digital em longo prazo em *Cloud Services*. Quando se trata de metadados de preservação digital nesse ambiente os resultados são ainda mais escassos. Na maioria dos textos encontrados não se discorrem explicações profundas, minuciosas e detalhadas, das atribuições específicas referentes ao escopo de funções, de características e de aplicações dos metadados de preservação digital em *Cloud Services*, ou seja, o número de publicações que abordam os metadados de preservação digital como prota-

gonista nas ações de preservação digital em longo prazo em *Cloud Services* diminui drasticamente em comparação ao número de publicações sobre outras temáticas envolvendo preservação digital de longo prazo em *Cloud Services*.

Também foram apontadas nestes três documentos, as dificuldades dos *Cloud Services* existentes em suportar as ações destes tipos de metadados embutidos em objetos digitais. Simultaneamente, os três documentos elaboraram e discutiram soluções viáveis, criando modelos de *Cloud Services*, incluindo a camada de preservação e a camada interação; é importante sublinhar que a estrutura dos principais *Cloud Services* existentes não emprega estas duas camadas.

O documento 11 (*Preservation Data Stores in the cloud*) buscou evidenciar soluções através de modelos de sistemas de armazenamento, os quais podem ser hospedados em *Cloud Services* existentes, funcionando como uma camada de preservação, ou por meio de mecanismos atribuídos aos denominados serviços de preservação.

Tanto a camada de preservação, como os sistemas de armazenamento são amparados por aplicações de metadados de preservação digital, e assim sendo, são descritos os desdobramentos percorridos pelo objeto digital.

A partir das descrições observadas em dois documentos (*Preserving records in the cloud* e *Developing an ontology for cloud-based archive systems*), verificou-se que os padrões de metadados de preservação *PREMIS* contém um melhor rendimento ao ser aplicado nos modelos de *Cloud Services* projetados com camadas de preservação e interação. Outro fator constatado sobre as funcionalidades dos padrões de metadados *PREMIS*, são norteadas pelas divisões sofridas por este padrão, se transformando em metadados de preservação com atribuições bem peculiares e específicas; estes metadados de preservação oriundos da divisão do padrão penetram nas camadas de preservação e de interação, evidenciado a complexidade de suas atribuições para assegurar a preservação digital em longo prazo.

Por mais que *OAIS* seja considerado um modelo de referência de alta qualidade, seus respectivos pacotes de informação não possuem o rendimento adequado no ciclo de vida de um objeto digital inserido em *Cloud Services*, (segundo os documentos: *Preserving records in the cloud* e *Developing an on-*

ology for cloud-based) pelo fato desses ambientes não suportarem as informações dos seus metadados embutidas no objeto digital.

Entretanto, o documento 11 (*PDS Cloud: Long Term Digital Preservation in the Cloud*) descreve de forma vasta a importância da aplicação deste modelo de referência, a partir de seus pacotes de informações, e por mais que tenham sido destacadas no texto as dificuldades enfrentadas de se inserir um objeto digital em *Cloud Services* que abarcam o modelo em questão. Rabinovici-Cohen et al. (2013) concluem que novas medidas estratégicas precisam ser implantadas na resolução dessa problemática, pois consideram que o emprego das ações dos pacotes de informação *OAIS* consiste na melhor forma de assegurar a preservação digital em longo prazo nos *Cloud Services*.

Outro aspecto identificado no documento 11 (*PDS Cloud: Long Term Digital Preservation in the Cloud*) se refere ao protagonismo do *OAIS* (Rabinovici-Cohen et al. 2013), são elencados uma série de atribuições específicas que interligam as aplicações e as funcionalidades deste modelo, inclusive ressaltam a importância da padronização dos objetos digitais de acordo com as regras estabelecidas pelo modelo. Entretanto, muitos aspectos das atribuições do percurso do objeto digital usando *OAIS* não são totalmente explorados em *Cloud Services*, pois a temática principal gira em torno de outras funcionalidades da *Preservation Data Stores in the cloud*.

Nesse sentido, *PDS Cloud: Long Term Digital Preservation in the Cloud* (11), não é tão minucioso quando se trata das atribuições dos metadados de preservação digital no ciclo de vida dos objetos digitais alocados nesse ambiente, em determinado momento o foco pauta-se nas características da arquitetura dos *Cloud Services*.

Outra diferença entre o último documento analisado, em comparação com os outros dois documentos, reside no fato dos autores trabalharem com um modelo de serviço de armazenamento, isto é, seu sistema foi elaborado para ser hospedado num *Cloud Service* existente, podendo ser empregado como um serviço de preservação, ou como uma camada de preservação que pode se subdividir em duas, além de outras funcionalidades. Construir um modelo de *Cloud Services* com camadas de preservação e interação, e elaborar um modelo de sistema de armazenamento com funções de hospedagem (para ser empregado num *Cloud Service* existente) são considerados projetos distin-

tos, embora o intuito seja o mesmo: solucionar os desafios que norteiam a preservação de longo prazo por meio de estratégias concisas.

Há uma distinção que pode ser determinante para se compreender o contexto que levam os dois primeiros documentos analisados, *Preserving records in the cloud* e *Developing an ontology for cloud-based*, em explicar aplicações, características e funções dos metadados de preservação digital de forma mais detalhada, em comparação com o último texto analisado neste subcapítulo, *PDS Cloud: Long Term Digital Preservation in the Cloud*, afinal os dois documentos são da área da Ciência da Informação, enquanto o último pertence à área de Ciência da Computação.

A área de Ciência da Informação, historicamente é mais ligada aos contextos que norteiam questões sobre preservação, antes mesmo do surgimento de elementos digitais, [...] "nas subáreas da Ciência da Informação, a preservação é um aspecto central de preocupação e ocupação dos profissionais que atuam nos espaços destinados à memória" (MONTEIRO; CARELLI; PICKLER, 2008, p.1). Existe uma gama de razões que relaciona o campo da preservação (seja analógica ou digital) com uma subárea da CI em especial, a denominada Representação da Informação. Nessa subárea os padrões de metadados são fundamentais para o desenvolvimento de pesquisas, elaboração de teorias, debates etc. Inúmeros estudos nessa subárea são completamente norteados pelos padrões de metadados.

Em contrapartida, as pesquisas que permeiam a área da Ciência da Computação não concentram o foco de estudos de maneira enfática e específica no escopo da preservação. Por exemplo, uma das principais linhas de pesquisa que norteia essa área é pautada em estudos sobre Inteligência Artificial e Engenharia de *Software*. Questões em torno da preservação digital em longo prazo, utilizando metadados de preservação digital em *Cloud Services* emergiram recentemente nessa área (baseado na literatura encontrada e explorada, a qual verificou as datas das primeiras publicações sobre essa temática, iniciando-se a partir de 2011).

Preserving records in the cloud (2011) e *Developing an ontology for cloud-based archive systems* (2015) são documentos pertencentes aos mesmos autores. Embora os dois textos contenham temáticas que se distinguem em alguns aspectos, ao mesmo tempo se assemelham nos seguintes pontos:

nas características relacionadas ao modelo de *Cloud Services* (elaborado pelos autores); nas abordagens sobre preservação digital em longo prazo; e nas aplicabilidades e nas funcionalidades dos metadados de preservação digital. Conforme pôde ser observado, os dois textos foram publicados numa diferença de cinco anos, e segundo a análise da presente pesquisa, foram os documentos que mais descreveram detalhadamente abordagens sobre aplicações de metadados de preservação digital em *Cloud Services*.

O estudo empregado no presente subcapítulo buscou verificar as aplicações dos padrões de metadados de preservação em *Cloud Services*, focando todo ciclo de vida dos objetos digitais. Alguns padrões de metadados foram utilizados para explicar alguma temática envolvendo preservação em longo prazo em armazenamentos de nuvem. Porém, alguns padrões de metadados de preservação foram adotados para outras finalidades de aplicação em *Cloud Services*, que não foram objeto dessa pesquisa.

Por meio do estudo dos três documentos analisados, objetivando identificar principalmente as funções e as aplicações detalhadas e específicas dos padrões de metadados de preservação digital em comunhão com as atribuições dos *Cloud Services* no contexto dos diversos processos que envolvem o percurso dos objetos digitais, e com a finalidade de garantir a preservação em longo prazo, apenas um padrão de metadados de preservação se enquadraram nas descrições elencadas acima: o *PREMIS*.

A partir de 821 documentos recuperados de seis bases de dados, por meio dos procedimentos metodológicos da RS, apenas três documentos descrevem todo o processo envolvendo a temática, e apenas dois padrões de metadados de preservação são detalhados e especificados de forma pormenorizada, no contexto das aplicações e das funções exercidas em *Cloud Services*. Isso significa que existe um vasto conteúdo que ainda foi inexplorado, o que possibilita o aprofundamento de estudos de metadados de preservação digital, e do modelo de referência *OAIS* em *Cloud Services* principalmente pela Ciência da Informação.

6.4 Tendências de pesquisas identificadas nos documentos analisados

Dentro do arcabouço temático envolvendo a presente pesquisa, foi possível verificar uma importante contribuição dos padrões de metadados de preservação digital, diante do desafio contínuo e incessante das problemáticas em torno da obsolescência em objetos digitais armazenados em *Cloud Services*.

Dentre os onze documentos analisados apontaram a necessidade de se implementar outros tipos de camadas, as quais se distinguem da tradicional estrutura empregadas em *Cloud Services* existentes. Em alguns textos, os autores destacavam a importância dos sistemas de armazenamento hospedados entre as camadas dos *Cloud Services*; os também denominados serviços de hospedagem são introduzidos entre as camadas dos *Cloud Services* existentes, exercendo funções semelhantes das camadas de preservação, ou também podem trabalhar como serviços de preservação (serviços de preservação ficam alocados na camada de preservação, constituindo num importante elemento para o funcionamento desta camada). É possível encontrar modelos de serviços de hospedagem elaborados pelos autores e serviços de hospedagem que atuam com a finalidade de empregar suas aplicações, no intuito de garantir à preservação digital em longo prazo em diversos tipos de ambientes, nesse contexto se destacam o *Preservica* e o *DuraCloud*.

Estudos envolvendo a elaboração de novos modelos de *Cloud Services* (em destaque os modelos que utilizam camadas de preservação) e serviços de hospedagem em nuvem estão intrinsecamente relacionados aos metadados de preservação, pois se torna inviável assegurar a preservação digital em longo prazo dos objetos digitais inseridos em *Cloud Services*, sem as aplicações dos padrões de metadados de preservação (as quais implicam em dois momentos: aplicações neste tipo de ambiente e nos próprios objetos digitais).

Em algumas pesquisas há uma latente preocupação em solucionar disfunções que são determinantes para a aplicação correta e efetiva do modelo de referência *OAIS*, novas alternativas na aplicação do *OAIS* podem facilitar consideravelmente os percursos e as funções que compõe o ciclo de vida do objeto digital em *Cloud Services*. O *OAIS* é o modelo de referência mais utilizado nos documentos consultados, e considerado pela maioria dos autores como o principal no contexto da preservação digital, inclusive em ambientes de nuvem,

mesmo sendo salientadas as dificuldades de se inserir seus pacotes de informação nos *Cloud Services*. É possível observar a necessidade de novas pesquisas que objetivem a correção das falhas de sua implantação em tais ambientes, ou seja, abordar o modelo de referência *OAIS* pode gerar inúmeros estudos, abordagens, discussões, debates e assim se concretizar numa relevante tendência de pesquisa sobre a temática em questão.

Quando se trata de pesquisas pautadas na preservação digital em longo prazo em *Cloud Services*, as temáticas que norteiam as camadas de preservação se destacam em vários documentos analisados. O protagonismo dos estudos sobre camadas de preservação pode alavancar pesquisas sobre metadados de preservação nesses ambientes, a fim de garantir a consolidação de aplicações e funcionalidades que visam assegurar a preservação digital.

As possíveis combinações entre diversos metadados de preservação existentes precisam ser analisadas, podendo beneficiar as questões que norteiam o escopo da preservação em *Cloud Services*; a relação de combinações entre metadados de preservação com outras categorias de metadados (principalmente os padrões de metadados técnicos, descritivos e estruturais) também precisam ser verificadas, tendo potencial de possuir uma boa perspectiva.

As opções envolvendo novos estudos sobre metadados de preservação com outras entidades são extensas e podem ser destacadas as tendências: interação com as camadas; serviços de preservação; objetos digitais; serviços de hospedagem; extensões; diferentes tipos de dados; domínios de linguagem; pacotes de informação; perspectivas em diferentes contextos de suas aplicações durante o percurso e ciclo de vida do item digital em ambientes de nuvem; eficiência e produtividade de um mesmo metadado em diferentes *Cloud Services* etc.

É possível que a camada de preservação se torne uma forte tendência de pesquisa na Ciência da Informação, pois há uma gama de recursos que podem ser introduzidos, além de diversas aplicações existentes capazes de serem aprimoradas, incluindo funcionalidades e aplicações em torno dos padrões de metadados de preservação. A possibilidade de criação ou descoberta de novos elementos e mecanismos pode se materializar numa realidade palpável a partir do interesse científico, social, cultural e até mesmo econômico, em questões que giram em torno da preservação digital de longo prazo em *Cloud*

Services. Devido a sua importância, a camada de preservação está sendo apontada como um ponto de partida em fomentar o aumento da quantidade de pesquisas e publicações científicas.

A conexão entre estudos sobre camadas de preservação e metadados de preservação pode concretizar o aumento de pesquisas referentes aos metadados de preservação em *Cloud Services*. O crescimento do número de pesquisas sobre metadados de preservação fomentaria novas pesquisas em torno da camada de preservação, a partir da análise cada vez mais aprofundada sobre os metadados desencadearia novos panoramas nas camadas de preservação, seja no sentido de aperfeiçoar a camada, seja pelos incrementos de novas ferramentas.

Dessa maneira, o próximo capítulo apresenta as conclusões, as observações e as ponderações alcançadas nessa pesquisa.

CONSIDERAÇÕES FINAIS

São diversos os desafios em torno da preservação digital em longo prazo, os quais começam paulatinamente a despertar na comunidade acadêmica, ainda que bastante tímida, na procura por tecer análises, fomentar pesquisas e soluções que visam nortear todo o processo do ciclo de vida dos objetos digitais armazenados em *Cloud Services*.

Em primeiro lugar, o objeto digital é dividido em diversos componentes, envolvendo suportes, programas, arquivos, dados, metadados etc. Em segundo lugar, os avanços das tecnologias de informação e comunicação são incessantes, facilitando a rápida obsolescência dos objetos digitais. Outra questão que complementa esse contexto gira consiste na dependência cada vez mais maior dos *Cloud Services*; o presente estudo procurou evidenciar a emergente dependência por parte de usuários e de instituições em alocar seus objetos digitais nesse tipo de ambiente.

No entanto, foi possível observar que a maioria dos *Cloud Services* ainda não trabalham adequadamente com as questões relacionadas com a preservação digital de longo prazo, isto é, em breve pode ocorrer certas dificuldades na recuperação de objetos digitais, caso os *Cloud Services* existentes não foquem em sistemas e ferramentas tecnológicas que atendam as questões dessa problemática.

Por fim, essa pesquisa ressalta o baixo número de publicações envolvendo padrões de metadados de preservação digital nas áreas de Ciência da Informação e de Ciência da Computação, além de sublinhar a omissão da terminologia metadados de preservação digital nos documentos.

Na área de Ciência da Informação, os metadados de preservação digital são fundamentais para assegurar a preservação digital em longo prazo. O baixo número de publicações sobre essa categoria de metadados atuando em *Cloud Services* precisa ser revisto com atenção pelos profissionais da área, pois os poucos estudos analisados sinalizaram que as camadas da maioria dos *Cloud Services* não conseguem suportar os metadados de preservação, principalmente devido ao tamanho ocupado por objetos digitais nas aplicações em ambientes de nuvem, ou mesmo pelo alto custo gerado para serem suportados, principalmente no processo de entrada (*input*) e saída (*output*), as descri-

ções sobre a relação dos padrões de metadados de preservação com os pacotes de informação do modelo de referência *OAIS* é um exemplo desse contexto.

De 821 documentos foram aceitos apenas onze na fase de extração da Revisão Sistemática de Literatura, o que reforça o caráter inovador do tema, sobretudo na área de Ciência da Informação, destacando a pesquisa como uma fonte de informação em língua portuguesa sobre os metadados de preservação digital em *Cloud Services*.

Destes onze documentos, apenas três publicações especificam os percursos do ciclo de vida dos objetos digitais pautados na interligação dos padrões de metadados de preservação com esse tipo de ambiente (ASKHOJ; SUGIMOTO; NAGAMORI, 2011; 2015; RABINOVICI-COHEN et al., 2013). Destaca-se o padrão de preservação *PREMIS*, assim como o modelo de referência *OAIS* nas principais ações referentes à preservação digital em longo prazo em *Cloud Services*.

Os onze documentos recuperados através da Revisão Sistemática de Literatura, e posteriormente pautados no método de análise de conteúdo, contribuíram para a temática desenvolvida no presente estudo nos seguintes aspectos: identificação das tendências de pesquisa em Ciência da Informação que estão sendo desenvolvidas com base no estado da arte; há necessidade de mais estudos envolvendo o modelo de referência *OAIS*, padrões de metadados que desempenham ações na conjuntura da preservação digital (padrões de metadados: de preservação, administrativos, técnicos, estruturais e descritivos) e, *Cloud Services*; os documentos evidenciaram a necessidade de fomentar pesquisas, no campo da Ciência da Informação, sobre as camadas de preservação, no Brasil ainda não foram desenvolvidas pesquisas sobre essa temática no campo em questão; a omissão terminológica metadados de preservação digital obriga o profissional da informação em criar mais eventos sobre este metadados em questão, e conseqüentemente sobre outros padrões de metadados que estão diretamente interligados com a preservação digital; outra questão fundamental pauta-se na utilização da Revisão Sistemática de Literatura, sabe-se que ainda não existem muitas pesquisas na área de Ciência da Informação que utilizam essa metodologia, a RS pode ser considerada uma grande aliada do cientista da informação, inclusive em novas pesquisas envol-

vendo preservação digital e *Cloud Services*; ocorreu a necessidade de definição do conceito de *Cloud Services*; por fim, evidenciou-se a necessidade de construção de um glossário definindo as diversas terminologias que abarcam o universo da tecnologia *Cloud*, pois a pluralidade terminológica dos conceitos gera um obstáculo no desenvolvimento de estudos sobre estes ambientes.

Um dado salutar verificado reside na diversificação dos países que estudam a temática. Nações com realidades culturais e sociais diferentes (apesar da maioria das pesquisas terem sido realizadas ou mesmo publicadas em países considerados desenvolvidos) e, geograficamente esparsos em três continentes: América do Norte, Europa e Ásia. Não foi encontrada nenhuma pesquisa sobre a temática em questão no Brasil ou na América Latina. Outros dados apurados constataram a repetição do nome de autores em documentos diferentes, isto é, são poucos autores que de fato estudam padrões de metadados de preservação digital em *Cloud Services*. Vale salientar que o documento recuperado mais antigo é da data do ano de 2011, ou seja, trata-se de uma temática recentemente nova.

Houve necessidade de conceituar de forma padronizada os significados das variantes das funcionalidades dos *Cloud*, pois há uma pluralidade terminológica para os conceitos, como por exemplo, *Cloud Storage Services*, *Cloud Services*, *Cloud Providers Services*, *Cloud Providers* etc. e essas nomenclaturas, dificultam a compreensão precisa da temática específica. Todos os contextos encontrados nos documentos, discutiram descrições de características, de funções e de aplicações que contemplavam os denominados serviços em nuvem, e sendo assim, buscou-se empregar a padronização do conceito terminológico das palavras *Cloud Services*.

Como pesquisas futuras destaca-se o estudo da Revisão Sistemática da Literatura identificando a relação dos metadados de preservação digital na camada de preservação dos *Cloud Services*, suas características, funcionalidades e componentes desses ambientes digitais. A principal tendência de pesquisa constatada abarca a implantação das denominadas camadas de preservação. As camadas de preservação, os serviços de hospedagens e os sistemas de serviços acoplados (tanto nas camadas de preservação, como nos serviços de hospedagem) são destacados como foco entre as publicações. É possível afirmar que há uma indissociável intersecção entre os padrões de metadados

de preservação digital e a camada de preservação e uma tendência de pesquisa, principalmente no campo da Ciência da Informação, justificada pelo fato de trazer um conceito tecnológico relativamente novo nesse tipo de ambiente, e por ainda não ser utilizada com frequência pela maioria dos tradicionais *Cloud Services*.

A análise dos documentos identificou uma repetição constante de padrões como o *PREMIS*, *XML*, *MODS* e *METS*. Porém, em poucos casos foram localizados padrões de metadados de preservação digital que trivialmente não são exemplificados, como por exemplo, o *LOTAR*.

Baseado na literatura específica do campo da preservação sabe-se que padrões de metadados de preservação digital são componentes fundamentais na garantia da autenticidade, da integridade, da compreensão, da acessibilidade, da segurança, da descrição, da representação, da gestão, da recuperação e da preservação dos objetos digitais, em qualquer tipo de armazenamento em longo prazo.

Pode-se presumir que muitos estudos no campo da preservação digital na área de Ciência da Informação e provavelmente na área de Ciência da Computação ascenderão, pois as práticas e as técnicas de manutenção, de conservação e de preservação precisam ser constantes e englobam todos os componentes de um objeto digital, suportes, programas, armazenamentos, arquivos etc.

Diante desse cenário, os *Cloud Services* podem consistir numa alternativa estratégica viável para a resolução dos constantes desafios da preservação digital, porém esses ambientes necessitam adotar e aplicar os padrões de metadados de preservação digital de forma efetiva e adequada. A implantação contínua das tecnologias em torno dos *Cloud Services* é recente, os questionamentos que levaram às abordagens sobre preservação digital nesses ambientes são ainda mais recentes, porém é fundamental que a Ciência da Informação tome a frente nos processos de investigação dessa natureza.

Espera-se que essa pesquisa possa contribuir com elementos teóricos e metodológicos para o estudo mais aprofundado das questões dos metadados de preservação digital em *Cloud Services* promovendo a discussão interdisciplinar e a construção de conhecimento para a promoção da inovação.

REFERÊNCIAS

ALBAGLI, S.; MACIEL, M. L. Informação e conhecimento na inovação e no desenvolvimento local. In: **Ciências da Informação**, v. 33, n. 3, p. 9-16, 2004.

ALMEIDA, M. B. Uma introdução ao XML, sua utilização na Internet e alguns conceitos complementares. **Ciência da informação**, v. 31, n. 2, p. 5-13, 2002.

ALMEIDA, A. C. L.; NASCIMENTO, G. B. Considerações sobre a preservação de documentos em formato digital. **Biblionline**, João Pessoa, v. 7, n. 2, p. 22-27, jul./dez. 2011.

ALPHARETTA. **The HP Graphics Experience Center**. 2018. Disponível em: <<https://www8.hp.com/us/en/commercial-printers/floater/demo-center-atlanta.html>>. Acesso em: 10 jul. 2018.

IBM SERVICES. **IBM Cloud**. 2018. Disponível em: <<https://www.ibm.com/cloud/managed-services>>. Acesso em: 10 jul. 2018.

ALVARENGA, L. Representação do conhecimento na perspectiva da ciência da informação em tempo e espaço digitais. **Encontros Bibli: revista eletrônica de biblioteconomia e ciência da informação**, Florianópolis, v.8 n. 15, p.18-40, jan. 2003.

ALVES, R. C. V. Metadados e padrões de metadados para preservação digital no domínio arquivístico. In: ENCONTRO NACIONAL DE PESQUISA EM PÓS-GRADUAÇÃO EM CIÊNCIA DA INFORMAÇÃO, 17., 2016, Salvador, BA. **Anais...** Salvador: ENANCIB, 2016.

ALVES, R. C. V. **Web semântica: uma análise focada no uso de metadados**. 180f. 2005. Dissertação (Mestrado em Ciência da Informação) – Faculdade de Filosofia e Ciências, Universidade Estadual Paulista, Marília, SP, 2005.

ALVES, R. C. V. **Metadados como elementos do processo de catalogação**. 2010. 132f. Tese (Doutorado em Ciência da Informação) – Faculdade de Filosofia e Ciências, Universidade Estadual Paulista, Marília, 2010.

ALVES, R. C. V.; SANTOS, P. L. V. A. da C. **Metadados no domínio bibliográfico**. Rio de Janeiro: Intertexto, 2013.

AMAZON S3. **Amazon Simple Storage Service**. 2006. Disponível em: <https://docs.aws.amazon.com/pt_br/AmazonS3/latest/dev/Welcome.html>. Acesso em: 10 jul. 2018.

ANALYTICS CLOUD. **Oracle Cloud**. 2018. Disponível em: <https://cloud.oracle.com/pt_BR/oac>. Acesso em: 10 jul. 2018.

ANJOS, A. M.; NUNES, F. L. S; TORI, R. Avaliação de habilidades sensório-motoras em ambientes de realidade virtual para treinamento médico: uma revisão sistemática. **Journal of Health Informatics**, v. 4, n. 1, p. 28-34, 2012.

ARELLANO, M. A. M. Cariniana: uma rede nacional de preservação digital. **Ciência da Informação**, v. 41, n. 1, 2012.

ARELLANO, M. A. M. **Critérios para a preservação digital da informação científica**. 2008. 354 f. Tese (Doutorado em Ciência da Informação) – Universidade de Brasília, Departamento de Ciência da Informação, 2008.

ARELLANO, M. A. M; SODRÉ, R. S. Preservação Digital e os profissionais da informação. **Datagramazero: Revista Ciência da Informação**, v.7, n.5, out. 2006.

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

ASKHOJ, J.; SUGIMOTO, S.; NAGAMORI, M. Preserving records in the cloud. **Records Management Journal**, v. 21, n. 3, p. 175-187, 2011.

AWS. **AWS Elastic Beanstalk**. 2018. Disponível em: <https://docs.aws.amazon.com/pt_br/elasticbeanstalk/latest/dg/Welcome.html>. Acesso em: 10 jul. 2018.

AWS. **AWS MARKETPLACE**. 2018. Disponível em: <https://docs.aws.amazon.com/pt_br/marketplace/latest/userguide/what-is-marketplace.html>. Acesso em: 10 jul. 2018. Disponível em: < >. Acesso em: 10 jul. 2018.

BAKER, M. Topics should merit their metadata: metadata matters, objects vs. chunks. In: BAKER, M. **Every page is page one**: topic-based writing for technical communication and the WebEvery. Ottawa, Canada: XML Press, 2011.

BARACHO, R. M. A. et al. Ciência da informação: sinalizações para o presente e o futuro. In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 16., 2015, João Pessoa, PB. **Anais...** João Pessoa, PB: ENAN-CIB Pessoa, 2015.

BLUEMIX. **IBM Cloud**. 2018 <<https://www.ibm.com/cloud-computing/bluemix/pt/what-is-bluemix>>. Acesso em: 10 jul. 2018.

BONDI, A. B. Characteristics of scalability and their impact on performance. In: INTERNATIONAL WORKSHOP ON SOFTWARE AND PERFORMANCE, 2., 2000, San Francisco, USA. **Proceedings...** San Francisco, USA: ACM, 2000. p. 195-203.

BRASCHER, M.; CAFE, L. Organização da informação ou organização do conhecimento? In: ENCONTRO NACIONAL DE PESQUISA EM PÓS-

GRADUAÇÃO EM CIÊNCIA DA INFORMAÇÃO, 9., 2008, São Paulo. **Anais...** Brasília: ENANCIB, 2008.

BUCKLAND, M. K. What is a document? **Journal of the American Society for Information Science (JASIS)**, v. 48, n. 9, p. 804-809, 1997.

CASTRO, F. F. de; SANTOS, P. L. V. A. da C. Os metadados como instrumentos tecnológicos na padronização e potencialização dos recursos informacionais no âmbito das bibliotecas digitais na era da web semântica. **Informação & Sociedade**, v. 17, n. 2, p. 13-19, 2007.

CASTRO, F. F. de; SANTOS, P. L. V. A. da C. Representação e descrição de recursos informacionais: aspectos estruturantes no delineamento de ambientes informacionais digitais. **Tendências da Pesquisa Brasileira em Ciência da Informação**, v. 3, n. 1, p. 155-168, 2010.

CASTRO, F. F. de; SIMIONATO, A. C.; ZAFALON, Z. R. Aspectos relacionais entre ontologia e metadados: considerações interdisciplinares. In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 17., 2016, Salvador. **Anais...** Salvador: ENANCIB, 2016.

CECIN, F. R. **FreeMMG: uma arquitetura cliente-servidor e par-a-par de suporte a jogos maciçamente distribuídos**. 2005. 101 f. Dissertação (Mestrado em Computação) - Programa de Pós-Graduação em Computação, Universidade Federal do Rio Grande do Sul, Porto Alegre, RS, 2005.

GOOGLE COMPUTE ENGINE. **Google Cloud**. 2018. Disponível em: <<https://cloud.google.com/compute/>>. Acesso em: 10 jul. 2018.

CONDE, M. V. F.; ARAÚJO-JORGE, T.C. de. Modelos e concepções de inovação: a transição de paradigmas, a reforma da C&T brasileira e as concepções de gestores de uma instituição pública de pesquisa em saúde. **Ciência & Saúde Coletiva**, v. 8, p. 727-741, 2003.

CONSULTATIVE COMMITTEE FOR SPACE DATA SYSTEMS - CCSDS. **XML Formatted Data Unit (XFDU) structure and construction rules**. Washington: [s. n.], 2008. 76 p. Disponível em: <<https://public.ccsds.org/Pubs/661x0b1.pdf>>. Acesso em: 05 jul. 2018.

CONSULTATIVE COMMITTEE FOR SPACE DATA SYSTEMS - CCSDS. **Reference model for an Open Archival Information System (OAIS)**. Washington: [s. n.], 2002. 139 p. Disponível em: <<http://ddp.nist.gov/refs/oais.pdf>>. Acesso em: 01 dez. 2016.

CUNHA, G. A.; CASTRO, F. F. O estudo e a identificação dos padrões de metadados para a representação e a recuperação da imagem digital na perspectiva da web. **Em Questão**, v. 24, p. 145- 173, 2018.

DANTAS, C. M.; CÓRDULA, F.R.; ARAÚJO, W. J. Análise da representação da informação em modelos entidade relacionamento com base em metadados. **Archeion Online**, v. 4, n. 1, p. 40-63, 2016.

DECMAN, M.; VINTAR, M. A possible solution for digital preservation of e-government: A centralised repository within a cloud computing framework. In: **Aslib Proceedings: new information perspectives**. Emerald Group Publishing Limited, 2013. p. 406-424.

DOINEA, M.; POCATILU, P. Security of heterogeneous content in cloud based library information systems using an ontology based approach. **Informatica Economica**, v. 18, n. 4, p. 101-110, 2014.

DUTRA, M. L.; SANT'ANA, R. C. G.; MACEDO, D. D. J. Sublimação de dados: dos objetos físicos às nuvens. In: ENCONTRO NACIONAL DE PESQUISA EM EM CIÊNCIA DA INFORMAÇÃO, 17., 2016, Salvador. **Anais...** Salvador: E-NANCIB, 2016.

FERNANDES, R. G. S. A.; CASTRO, F. F. de. As relações entre RDA e AACR2: a busca de uma descrição bibliográfica completa. In: Encontro de Iniciação Científica, 23., 2013, São Cristóvão. In: **Anais do 23º Encontro de Iniciação Científica**. São Cristóvão: Universidade Federal de Sergipe, 2013. v. 23. p. 751-751.

FLORES, H.; SRIRAMA, S. N.; PANIAGUA, C. A generic middleware framework for handling process intensive hybrid cloud services from mobiles. In: **Proceedings of the 9th International Conference on Advances in Mobile Computing and Multimedia**. ACM, 2011. p. 87-94.

FORMENTON, D. et al. Os padrões de metadados como recursos tecnológicos para a garantia da preservação digital. **Biblios: Revista electrónica de bibliotecología, archivología y museología** **Biblios**, n. 68, p. 82-95, 2017.

FORMENTON, D.; GRACIOSO, L. de S.; CASTRO, F. F. de. Revisitando a preservação digital na perspectiva da ciência da informação: aproximações conceituais. **Revista Digital de Biblioteconomia e Ciência da Informação**, v. 13, n. 1, p. 170-191, 2015.

FRANKS, P. C. **Government use of cloud-based long term digital preservation as a service: an exploratory study**. Granada, Spain: [s. n.], 2015. p. 371-374.

GIL, A. C. **Como Elaborar Projetos de pesquisa**. 3. ed. Belo Horizonte: Atlas, 1991.

GLUSHKO, R. J. (Ed.). **The discipline of organizing**. Massachusetts, Londres: MIT Press, 2013.

GODOY, K. V. C. de. **Identificação de diretrizes para elaboração de política de acesso aberto aos conteúdos educacionais disponibilizados em repo-**

sitórios digitais no contexto da educação a distância. 2015. 101 f. Dissertação (Mestrado em Multidisciplinar) - Programa de Pós-Graduação em Ciência, Tecnologia e Sociedade, Universidade Federal de São Carlos, São Carlos, SP, 2015.

GONZÁLES DE GÓMEZ, M. N. Organização e representação do conhecimento visando a recuperação da informação: linha de pesquisa. In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 1., 1994, Belo Horizonte. **Anais...** Belo Horizonte: UFMG, 1994. p.13-14.

GOOGLE APPENGINE DOCS. **Google Cloud.** 2018. Disponível em: <<https://cloud.google.com/appengine/docs/>>. Acesso em: 10 jul. 2018.

GOOGLE SUITE. **Google Cloud.** 2018. Disponível em: <<https://gsuite.google.com.br/intl/pt-BR/>>. Acesso em: 10 jul. 2018.

GUENTHER, R.; MCCALLUM, S. New metadata standards for digital resources: MODS and METS. **Bulletin of the Association for Information Science and Technology**, v. 29, n. 2, p. 12-15, 2003.

HP. **New Model Public Cloud.** 2018. Disponível em: <<https://community.hpe.com/t5/Shifting-to-Software-Defined/A-new-model-to-deliver-public-cloud/ba-p/6804409#.W0S3HdVKjIU>>. Acesso em: 10 jul. 2018.

IBM SOFTWARE AS A SERVICE. **IBM Partner World.** 2018. Disponível em: <https://www-356.ibm.com/partnerworld/wps/servlet/ContentHandler/swg_com_sfw_ibm-solution-provider-ibm-software-as-a-service-overview>. Acesso em: 10 jul. 2018.

INNARELLI, H. C. Preservação digital: a influência da gestão dos documentos digitais na preservação da informação e da cultura. **Revista Digital de Biblioteconomia e Ciência da Informação**, Campinas, v. 8, n. 2, p.72-87, 2011.

JESUS, M. S. de. Os analistas de informação e a sua relação com a Ciência da Informação: histórico da integração desses profissionais. **Revista Ibero-Americana de Ciência da Informação**, v. 4, n. 1, 2012.

JIA, L. **Metadata and its applications in the digital library:** approaches and practices. Westport: Libraries Unlimited, 2007.192 p.

KAMALA, B.; PRIYA, B.; NANDHINI, J. M. Platform autonomous custom scalable service using service oriented cloud computing architecture. **International Journal of Engineering Research and Applications (IJERA)**, v. 2, n. 2, p.1467-1471, mar./apr. 2012.

KITCHENHAM, B. **Procedures for performing systematic reviews.**Staffordshire: Keele University, 2004. 33 p.

LASTRES, H. M. M.; CASSIOLATO, J. E. Sistemas de inovação: políticas e perspectivas. 2000. **Parcerias Estratégicas**, Brasília, p. 237-255, 2000.

MACHADO, M. A. S. **Uma abordagem para indexação e buscas Full-Text baseadas em conteúdo em sistemas de armazenamento em nuvem**. 2013. 61 f. Dissertação (Mestrado em Ciência da Computação) - Universidade Federal de Pernambuco, Recife, 2013.

MARCONDES, C. H. Representação e economia da informação. **Ciência da Informação**, Brasília, v. 30, n. 1, p. 61-70, jan./abr. 2001.

MARCONDES, C. H.; SAYAO, L. F. Documentos e novas formas de cooperação entre sistemas de informação em C&T. **Ciência da Informação**, v.31, n.3, p.42-54, 2002.

MARINOS, A.; BRISCOE, G. Community cloud computing. In: IEEE INTERNATIONAL CONFERENCE ON CLOUD COMPUTING, 1., Beijing, China. **Proceedings**...Heidelberg: Springer, 2009. p. 472-484.

MARTINS, G. K. A Representação do conhecimento em uma perspectiva fenomenológica. In: ENCONTRO NACIONAL DE PESQUISA EM PÓS-GRADUAÇÃO EM CIÊNCIA DA INFORMAÇÃO, 17., 2016, Salvador, BA. **Anais...** Salvador: ENANCIB, 2016.

MELL, P.; GRANCE, T. **The NIST definition of cloud computing**: recommendations of the National Institute of Standards and Technology. Gaithersburg, Maryland: NIST, 2011.

MONTEIRO, S. D.; CARELLI, A. E.; PICKLER, M. E. V. Ciência da Informação, memória e esquecimento. **Datagramazero: Revista de Ciência da informação**, Rio de Janeiro, v. 17, p. 1-17, 2008.

MEY, E. S. A. **Introdução à catalogação**. Brasília: Briquet de Lemos, 1995.

MONTEBELO, R. et al. START (SystematicReviewAutomatic Tool): uma ferramenta computacional de apoio à revisão sistemática. In: EXPERIMENTAL SOFTWARE ENGINEERING LATIN AMERICAN WORKSHOP, 5., São Carlos, SP. **Proceedings**... São Carlos, SP: ICMS, 2007.

NARCISO, E. N.; NUNES, F. L.S.; DELAMARO, M. E. Seleção de casos de teste utilizando conceitos de variabilidade: uma revisão sistemática. In: SIMPÓSIO BRASIDELIRO DE SISTEMAS DE INFORMAÇÃO, 8., 2011, São Paulo. **Anais...** São Paulo: SBSI, 2011. p. 115-125.

NETWORK WORLD. **HP Public Cloud**. 2015. Disponível em: <<https://www.networkworld.com/article/2996536/cloud-computing/hp-just-dropped-out-of-the-public-cloud-now-what.html>>. Acesso em: 10 jul. 2018.

NGUYEN, Q. L.; LAKE, A. Content server system architecture for providing differentiated levels of service in a digital preservation cloud. In: **Cloud compu-**

ting (CLOUD), 2011 IEEE international conference on. IEEE, 2011. p. 557-564.

NOVELLINO, M. S. F. Instrumentos e metodologias de representação da informação. **Informação & Informação**, v. 1, n. 2, p. 37-45, dez. 1996.

OCLC - Online Computer Library Center. **Preservation Metadata for digital objects**: a review of the state of the art. 2001. Disponível em: <http://www.oclc.org/content/dam/research/activities/pmwg/presmeta_wp.pdf>. Acesso em: 02 jan. 2017.

OCLC - Online Computer Library Center. **Preservation metadata and the OAIS Information Model**: a metadata framework to support the preservation of digital object. 2002. Disponível em: <http://www.oclc.org/content/dam/research/activities/pmwg/pm_framework.pdf>. Acesso em: 02 dez. 2016.

ORACLE CLOUD APPLICATIONS. **Oracle Cloud**. 2018. Disponível em: <https://cloud.oracle.com/pt_BR/oac>. Acesso em: 10 jul. 2018.

ORACLE CLOUD INFRASTRUCTURE. **Oracle Cloud**. 2018. Disponível em: <https://cloud.oracle.com/pt_BR/compute>. Acesso em: 10 jul. 2018.

ORACLE CLOUD PLATAFORM. **Oracle Cloud**. 2018. Disponível em: <https://cloud.oracle.com/pt_BR/paas>. Acesso em 10 jul. 2018.

ORACLE ERP CLOUD. **Oracle Cloud**. 2018. Disponível em: <https://cloud.oracle.com/pt_BR/erp-cloud>. Acesso em: 10 jul. 2018.

PAVÃO, C. M. G.; CAREGNATO, S. E.; ROCHA, R. P. da. Implementação da preservação digital em repositórios: conhecimento e práticas. **Revista Digital de Biblioteconomia e Ciência da Informação**. Campinas, SP. Campinas. Vol. 14, n. 3 (set./dez. 2016), p. 407-425, 2016.

PONTES, Glaucio Ranniere de Souza. **Arquivando nas nuvens: um recurso estratégico para a preservação de documentos arquivísticos digitais**. 2014. 126f. Trabalho de Conclusão de Curso (Graduação em Arquivologia)- Universidade Estadual da Paraíba, João Pessoa, 2014.

PREMIS. **Data dictionary for preservation metadata**. Disponível em: <<http://www.loc.gov/standards/premis/>>. Acesso em: 02 dez. 2016.

RABINOVICI-COHEN, S. et al. PDS cloud: long term digital preservation in the cloud. In: IEEE INTERNATIONAL CONFERENCE ON CLOUD ENGINEERING, 6., 2013, Santa Clara, California. **Proceedings...** Santa Clara, California: IEEE, 2013. p. 38-45.

RABINOVICI-COHEN, S. et al. Towards SIRF: self-contained information retention format. In: ANNUAL INTERNATIONAL CONFERENCE ON SYSTEMS AND STORAGE, 4., 2011, Haifa, Israel. **Proceedings...** Haifa, Israel: ACM, 2011. p. 15.

RACKSPACE OPEN CLOUD. **Rackspace**. 2018. Disponível em: <<https://www.rackspace.com/openstack/public/service-levels>>. Acesso em: 10 jul. 2018.

RIC. **Records in the Cloud**. 2014. Disponível em: <<http://www.recordsinthecloud.org>>. Acesso em: 21 dez. 2017.

RILEY, J. **Glossary of metadata standards**. 2010. Disponível em: <http://jennriley.com/metadatamap/seeingstandards_glossary_pamphlet.pdf >. Acesso em: 21 nov. 2017.

ROCHA, A. et al. Camaleão: um software de esteganografia para proteção e privacidade digital. In: Simpósio Segurança em Informática (SSI), 2004, São José dos Campos. **Anais do 6º Simpósio de Segurança em Informática**, 2004. p. 1-9.

SALES CLOUD CRM. **Salesforce**. 2018. Disponível em <<https://www.salesforce.com/products/sales-cloud/features/>>. Acesso em 10 jul. 2018.

SALESFORCE1 PLATAFORM. **Salesforce**. 2018. Disponível em: <<https://searchsalesforce.techtarget.com/definition/Salesforce1-Platform>>. Acesso em: 10 jul. 2018.

SAMPAIO, R. F.; MANCINI, M. C. Estudos de revisão sistemática: um guia para síntese criteriosa da evidência científica. **Revista Brasileira de Fisioterapia**, São Carlos, v. 1, n. 11, p. 83-89, 2007.

SANTANA, W. R. de. **Introdução à programação**. 2015. Disponível em: <<http://www.neutronica.com.br/wp-content/uploads/IAP-Material-v1.5.pdf>>. Acesso em: 10 jul. 2018.

SANTOS, P. L. V. A. da C. Redes informacionais como ambientes colaborativos e de empoderamento: a catalogação em foco. In: GUIMARÃES, J. A. C.; FUJITA, M. S. L. (Orgs.). **Ensino e pesquisa em biblioteconomia no Brasil: a emergência de um novo olhar**. Marília: Cultura acadêmica, 2008, p. 155-171.

SANTOS, P. da C.; VIDOTTI, S. A. B. G. Perspectivismo e tecnologias de informação e comunicação: acréscimos à Ciência da Informação. **DataGramaZero: Revista de Ciência da Informação**, Rio de Janeiro, v. 10, n. 3, 2009.

SARAMAGO, M. de L. Metadados para a preservação digital e aplicação do Modelo OAIS. In: CONGRESSO NACIONAL DE BIBLIOTECÁRIOS, ARQUIVISTAS E DOCUMENTALISTAS, 8., 2004, Estoril. **Anais...** Lisboa: BAD, 2004.

SAYÃO, L. F. Conservação de documentos eletrônicos. In: GRANATO, M.; ROCHA, C. R. A. da; SANTOS, C. P. (Org.). **MastColloquia**, v. 9: conservação de acervos. Rio de Janeiro: MAST, 2007.

SAYÃO, L. F. Uma outra face dos metadados: informações para a gestão da preservação digital. **Encontros Bibli: Revista Eletrônica de Biblioteconomia e Ciência da Informação**, v. 15, n. 30, p.1-31, 2010.

SAYÃO, L. F. Padrões para bibliotecas digitais abertas e interoperáveis. **Encontros Bibli: Revista Eletrônica de Biblioteconomia e Ciência da Informação**, v. 12, p. 18-47

SOUSA, F. R. C.; MOREIRA, L. O.; MACHADO, J. C. Computação em nuvem: conceitos, tecnologias, aplicações e desafios. In: ESCOLA REGIONAL DE COMPUTAÇÃO CEARÁ, MARANHÃO E PIAUÍ, 3., Sobral. **Anais...** Sobral: SBC, 2009. cap. 7, p. 150-175.

SOUSA, V. E. B.; SIMIONATO, A C. Podcast: uma proposta de descrição por meio de padrões de metadados. In: XI Seminário de Pesquisa em Ciências Humanas, 2016, Londrina - PR. **Anais do XI Seminário de Pesquisa em Ciências Humanas**, 2016. v. 2. p. 10-21.

SOUZA, Renato Rocha; ALVARENGA, Lídia. A Web Semântica e suas contribuições para a ciência da informação. **Ciência da Informação**, v. 33, n. 1, 2004.

SOUZA, E. D. de. A institucionalização da ciência da informação no Brasil. **Informação & sociedade: estudos**, v. 22, p. 49-64, 2012.

STANCIC, H.; RAJH, A.; BRZICA, H. Archival cloud services: portability, continuity, and sustainability aspects of long-term preservation of electronically signed records. **Canadian Journal of Information & Library Sciences**, v. 39, n. 2, p. 210-227, 2015.

THOMAZ, K. P.; SOARES, A. J. A preservação digital e o modelo de referência Open Archival Information system (OAIS). **Datagrama: Revista de Ciência da Informação**, Rio de Janeiro, v. 5, n. 1, p. 1-17, 2004.

TOMAÉL, M. I.; ALCARÁ, A. R.; CHIARA, I. G. D. Das redes sociais à inovação. **Ciência da Informação**. Brasília, v.34, n.2, p.93-104, 2005.

VAQUERO, L. M. et al. A break in the clouds: towards a cloud definition. **ACM SIGCOMM Computer Communication Review**, v. 39, n. 1, p. 50-55, 2008.

VERAS, M. **Cloud Computing: nova arquitetura da TI**. Brasport, 2012.

WITTEK, P. et al. XML processing in the cloud: large-scale digital preservation in small institutions. In: INTERNATIONAL SYMPOSIUM ON PARALLEL AND DISTRIBUTED PROCESSING WORKSHOPS AND PHD FORUM,1., 2011, Shanghai. **Proceedings...** Shanghai: ISPD, 2011.p. 1072-1081.

WITTEK, P.; DARANYI, S. Digital preservation in grids and clouds: a middle-ware approach. **Journal of Grid Computing**, v. 10, n. 1, p.133-149, 2012.