XIX encontro nacional de pesquisa em ciência da informação
ENANCIB
// SUJEITO INFORMACIONAL E AS PERSPECTIVAS ATUAIS EM CIÊNCIA DA INFORMAÇÃO. //
22-26 OUTUBRO 2018
LONDRINA/PR
Universidade Estadual de Londrina

XIX ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO – ENANCIB 2018

GT-8 – Informação e Tecnologia

MAPEAMENTO DE PADRÕES DE METADADOS DE PRESERVAÇÃO DIGITAL EM CLOUD SERVICES

Fabiano Ferreira de Castro (UFSCar)

Júlio César Tauil Silveira (UFSCar)

MAPPING METADATA PATTERS OF DIGITAL PRESERVATION IN CLOUD SERVICES

Modalidade da Apresentação: Comunicação Oral

Resumo: O armazenamento de objetos digitais em *Cloud Services* tem sido adotado de forma cada vez mais frequente no intuito de conter a obsolescência tecnológica e assim preservar as informações digitais. Objetivou-se nessa pesquisa o mapeamento dos estudos sobre os metadados de preservação digital em Cloud Services, no intuito de identificar os padrões de metadados adotados e aplicados nesse ambiente digital. Como matriz metodológica, a Revisão Sistemática da Literatura e a análise de conteúdo foram definidas, a fim de identificar especificamente nas áreas de Ciência da Informação e da Ciência da Computação, no âmbito internacional e nacional, as publicações que relacionam os metadados de preservação digital em *Cloud Services*. Como resultados evidenciam-se as características, as funcionalidades, a relação entre padrões de metadados de preservação digital e os *Cloud Services* e as tendências de pesquisa sobre o assunto. Conclui-se que os estudos de metadados de preservação em *Cloud Services* ainda se encontram pouco explorados na área de Ciência da Informação no cenário atual, nacional e internacionalmente. O olhar da Ciência da Informação no contexto dos metadados de 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 contribuindo com a ciência e a inovação.

Palavras-Chave: Padrões de metadados de preservação digital; *Cloud Services*; Revisão Sistemática de Literatura. Informação e Tecnologia.

Abstract: The storage of digital objects in Cloud Services has been adopted more and more frequently in order to contain technological obsolescence and thus preserve digital information. The objective of this research was to map the studies on the digital preservation metadata in Cloud Services, in order to identify metadata standards adopted and applied in this digital environment. As a methodological matrix, the Systematic Review of Literature and content analysis were defined in

order to identify, specifically in the areas of Information Science and Computer Science, at the international and national levels, the publications that relate the digital preservation metadata in Cloud Services. As results, the characteristics, the functionalities, the relationship between the digital preservation metadata standards and the Cloud Services, and the research trends on the subject are evidenced. It is concluded that the preservation metadata studies in Cloud Services are still little explored in the area of Information Science in the current scenario, nationally and internationally. The view of Information Science in the context of the 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 patters; Cloud Services; Systematic Review of Literature; Information and Technology.

1 INTRODUÇÃO

A preservação da informação é fundamental para perpetuar o conhecimento de história de identidade e de valores humanos. A intenção de preservar os registros a partir do armazenamento organizado visa proteger os conteúdos intelectuais para serem acessados num futuro distante perpetuando o ciclo virtuoso de disseminação do conhecimento, onde preservar e salvaguardar são considerados um dos grandes desafios do século XXI.

Estudar padrões de metadados de preservação digital em *Cloud Services* num cenário onde a obsolescência constantemente ameaça a preservação digital em longo prazo implica numa série de eventos envolvendo vários autores e processos. E todas as complexas relações ocorrem com a única finalidade: assegurar a preservação do conteúdo intelectual, garantindo o acesso para o uso das informações. (FORMENTON; GRACIOSO; CASTRO, 2015; FORMENTON *et al.*, 2017).

Ainda se conhece pouco das funções e aplicações dos metadados de preservação em *Cloud Services*, porém é possível afirmar que eles se constituem num importante elemento no contexto das estratégias de preservação digital, e podem contribuir ao serem utilizados como alternativa para solucionar diversos problemas que norteiam a preservação digital de longo prazo nesses ambientes. A preservação digital deve ser vista como um conjunto contínuo e de múltiplos níveis de ferramentas (DECMAN; VINTAR, 2013).

No que se refere ao arcabouço de percursos, situações e panoramas, envolvendo preservação em longo prazo, metadados de preservação digital e *Cloud Services* se faz necessário analisar as estruturas de composição do principal ator deste cenário: os objetos digitais, pois eles precisam de diversos níveis de compreensão e gestão, exemplificados pelos itens físicos, por uma codificação lógica (DECMAN; VINTAR, 2013). 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, texto, imagem e som são constituídos por componentes diferentes e ao mesmo tempo podem fazer parte de um mesmo objeto digital (YAMANE; CASTRO, 2018).

É objetivo nessa pesquisa o mapeamento dos estudos acerca dos metadados e padrões de metadados de preservação digital em *Cloud Services*, por meio de uma Revisão Sistemática da Literatura (RS).

As pesquisas sobre metadados de preservação em *Cloud Services* possibilitam aumentar a dimensão do conhecimento sobre suas ações e implicações diante da perspectiva de se viabilizar novas alternativas de preservação, sobretudo no campo da Ciência da Informação.

Nesse contexto, 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.

Como aporte metodológico utilizou-se a Revisão Sistemática da Literatura e a análise de conteúdo onde foram mapeadas as principais publicações que tratam dos metadados de preservação digital em *Cloud Services*, bem como as tendências de pesquisa na área de Ciência da Informação.

2 TECITURAS CONCEITUAIS DOS *CLOUD SERVICES*

Os *Cloud Services* foram desenvolvidos com a vantagem de fornecer serviços de fácil acesso, baixo custo e com garantias de disponibilidade e escalabilidade (GLUSHKO, 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* ou 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. Nesse sentido, a infraestrutura em *Cloud Computing* contém uma camada física e uma camada de abstração. A camada física consiste nos recursos de *hardware* necessários para suportar os *Cloud Services* que estão sendo fornecidos e normalmente inclui componentes de servidor, armazenamento e rede (MELL; GRANCE, 2011). 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; GRANCE, 2011).

O armazenamento em *Cloud Services* é altamente distribuído, escalável e oferece disponibilidade, acessibilidade e compartilhamento de dados a baixo custo. 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). Diante da capacidade espacial das tecnologias de armazenamento passar de *Kilobytes* para *Zettabytes*, a utilização de *Cloud Services* certamente poderá ser importante aliada para a definição e o processo de preservação digital. Porém, se não houver planejamento, acervos inteiros podem desaparecer nos seus armazenamentos (GLUSKHO, 2013).

Há quatro modelos de *Cloud Services*: a denominada nuvem privada (*Private Cloud*) - é referente a um centro de dados interno de alguma organização, pode ser empresarial, estatal etc. significa que não está disponível para o público; já a nuvem pública (*Public Cloud*) - disponibiliza uma infraestrutura distribuída de forma que só é pago aquilo que de fato o usuário utilizou; a nuvem comunitária (*Community Cloud*) - foi criada para moldar os recursos subutilizados das máquinas de usuários; por fim, a nuvem híbrida (*Hybrid Cloud*) - sugere a combinação da implantação de modelos de nuvem existentes, mesclar o modelo de nuvem pública com o modelo de nuvem privada (MARINOS; BRISCOE, 2009; WITTEK; DARÁNYI, 2012; FRANKS, 2015).

Cloud Services não possuem suporte limitado de metadados, pois o espaço para metadados por objeto digital é muito escasso para o tamanho extensivo dos metadados de preservação. A maioria dos serviços em nuvem não é capaz de pesquisar metadados, ou seja, não conseguem filtrar objetos digitais por meio de *tags*. Outro aspecto relevante consiste no fato de que esses ambientes não suportam a atualização de metadados sozinhos, além de questões pautadas nos custos de transferência de dados dos seus próprios dados (RABINOVICI-COHEN et al. 2013).

A arquitetura dos *Cloud Services* geralmente é representada por camadas compostas por *Infrastructure as a Service* (IaaS), *Platform as a Service* (PaaS) e *Software as a Service* (SaaS). Outra camada menos conhecida, a qual foi verificada nos documentos encontrados a partir da revisão sistemática, é a camada de preservação digital, esta camada pode assumir várias categorias e até mesmo diferentes terminologias. Já existem alguns *Cloud Services* que oferecem este serviço, há poucos estudos sobre as camadas de preservação como serviço,

pois esta camada foi desenvolvida recentemente. A principal função desta camada é pautada em interligar sistemas de arquivamento nos sistemas *Interaction Layer* (Interação de Camadas) e SaaS (ASKHOJ; SUGIMOTO; NAGAMORI, 2011).

No quadro 1**,** são descritos os principais *Cloud Services* e suas respectivas camadas. Vale salientar que nesses ambientes não há camada de preservação, prevalecendo a arquitetura tradicional: Infraestrutura como Serviço, Plataforma como Serviço, *Software* como Serviço.

QUADRO 1. Principais *Cloud Services*.

Provider	***IaaS***	***PaaS***	***SaaS***
Amazon	Amazon EC Amazon S3	Amazon Elastic Beanstalk	Available on AWS Marketplace
Google	Google Compute	Google App Engine	Google Apps
HP	HP Public Cloud	HP Helion Public Cloud Application Plataform as a Service	Available on HP Software Experience Center
IBM	IBM Cloud Managed Services SoftLayer	IBM Bluemix	IBM Solution Provider - IBM Software as a Services
Oracle	Oracle Compute	Oracle Cloud PaaS	Oracle Applications Cloud Oracle Analytics Cloud Oracle ERP Cloud etc.
Rackspace	Managed Infrastructure		
Salesforce.com		Salesforce1 platforms	Sales force automation and CRM

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

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, a tecitura conceitual construída nessa pesquisa parte do entendimento que os *Cloud Services* 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.

3 METODOLOGIA

No intuito 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, esta pesquisa se respaldou na Revisão Sistemática da Literatura (RSL) e na análise de conteúdo dos documentos, a fim de identificar as tendências de pesquisas sobre a temática.

A Revisão Sistemática da 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 (KITCHENHAM, 2004).

Para tanto, para apoiar a construção as fases que contemplam a RSL, é aplicada a ferramenta computacional denominada *StArt* (*State of the Art through Review*). Essa ferramenta foi desenvolvida dentro do 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 revisão sistemática, a qual objetiva dar suporte ao planejamento, execução e análise final da metodologia, englobando qualquer assunto, temática ou área do conhecimento, tornando a pesquisa mais ágil, precisa e replicável. Definido os aspectos de norteamento do levantamento das pesquisas sobre metadados de preservação digital em *Cloud Services*, as fases da RSL foram devidamente aplicadas no *StArt*.

Na fase de planejamento, com o intuito de localizar os documentos, foram utilizadas as seguintes bases de dados: Scopus, *Web of Scienc*, IEEE Xplore, *Computer and Information Systems* (ProQuest), *Library* & *Information Science Abstracts* (LISA), *Information Science* & *Technology Abstracts* (ISTA). Das 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 duas, pertencem à área da Ciência da Computação (*Computer and Information Systems* [ProQuest]).

No *StArt* há um formulário para preencher o protocolo da pesquisa. É fundamental que o protocolo seja planejado corretamente, é possível que ocorram modificações nos campos de pesquisa conforme as estratégias de pesquisa vão se afunilando e consequentemente se desenhando de forma mais concisa. O quadro 2 apresenta o formulário preenchido conforme os padrões da Revisão Sistemática.

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

Título:	Metadados de preservação em Cloud Services
Descrição:	Verificação na literatura em duas bases de dados da Ciência da Informação, duas de escopo geral e duas Ciência Computação, quais documentos pesquisam metadados de preservação digital em *Cloud Services*.
Objetivos:	Mapear nas principais base de dados, utilizando a análise quantitativa pautada na Revisão Sistemática de Literatura, com respaldo do *software StArt*. Posteriormente através do método de análise de conteúdo aplicou-se a análise qualitativa do intuito de analisar o grau de relação dos metadados de preservação digital em *Cloud Services* no contexto da preservação digital em longo prazo.
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 Cloud Services.
Intervenção:	Verificar como são interligados os estudos de metadados de preservação digital em Cloud Services e quais são as estratégias utilizadas na preservação em longo prazo.
População:	Metadados de preservação digital utilizados em Cloud Services.
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 Cloud Services, 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 Cloud Services 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 Cloud Services.
Palavras-chave:	*Cloud Services; Cloud Storage; Cloud Storage Providers; Cloud Operations; Cloud Computing; Digital Preservation; Metadata Digital Preservation; Information Science Computer Science.*
Fontes do critério de seleção:	Artigos, periódicos, capítulos de livro, teses e dissertações recuperadas nas bases de dados selecionadas.
Estudo de idiomas:	Inglês.

Fonte dos métodos de pesquisa:	Busca sistemática utilizando os booleanos AND e OR.
Ferramenta de busca:	*Computer and Information Systems* (*ProQuest*), *Web of Science* (WoS); *Information Science & Technology* (ISTA); *Library & Information Science Abstracts* (LISA) *Scopus* e; *IEEE Xplore.*
Critérios de inclusão e exclusão:	As principais abordagens do texto focam em características relacionadas com *Cloud Services* (I); As principais abordagens do texto focam em características relacionadas com *Cloud Computing* (E); O documento apresenta relação de metadados de preservação digital em *Cloud Services* (I); O documento não apresenta relação de metadados de preservação digital em *Cloud Services* (E); O documento apresenta estratégias de preservação digital utilizando metadados de preservação em *Cloud Services* (I); O documento contempla metadados de preservação digital em *Cloud Services*, mas não são das áreas de Ciência da Informação ou de Ciência da Computação (E); O documento apresenta estratégias de preservação digital utilizando metadados de preservação em *Cloud Services (I)*; O documento não apresenta estratégias de preservação digital, mesmo apresentando metadados em *Cloud Services*, porém não relaciona metadados de preservação com *Cloud Services (E);* Somente documentos encontrados em bases de dados *online* (I); O texto não está disponível de forma integral (E); Qualquer ano (I); o documento não está escrito em inglês (E)
Definição dos tipos de estudo:	Artigos científicos.
Informações do campo de extração:	Abordagens que foram utilizadas no texto na associação de estudos de metadados de preservação digital com Cloud Services; as abordagens do texto em questão que fazem parte do escopo da Ciência da Informação; as abordagens do texto em questão fazem parte do escopo da Ciência da Computação; definições do texto sobre as necessidades de se utilizar metadados de preservação digital em Cloud Services.
Resultados da sumarização:	1 - Mapeamento dos documentos que pesquisam metadados de preservação digital em Cloud Services. A partir do levantamento bibliográfico, da Revisão Sistemática de Literatura e Análise de Conteúdo, foram recuperados 11 documentos nas seis bases de dados elencadas. Com o aporte da ferramenta StArt, tais documentos foram selecionados de 821 documentos recuperados. Foi possível identificar um escasso número de pesquisas sobre a temática em questão. Não foram encontrados documentos no Brasil, nem na América Latina. Muitos textos são de autores repetidos. 2 - Definição conceitual de Cloud Services. 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. 3- Enfoques dos documentos analisados na Revisão Sistemática de Literatura. Por meio da análise de conteúdo se verificou 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. 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 que norteiam a preservação nesse contexto, nas trajetórias dos objetos digitais, nos modelos de servidores de nuvem existentes etc. 4 - Principal tendência de pesquisa verificada nos 11 documentos. A principal tendência de pesquisa verificada 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. 5 - Omissão da terminologia metadados de preservação digital. 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. 6 - Textos que documentam todo o percurso dos metadados de preservação digital do ciclo de vida dos objetos digitais nos 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*. 7 - Identificação dos padrões de metadados de preservação digital mais citados nos onze documentos. 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*. 8 - Criação de estratégias ante a imprecisão terminológica sobre o termo Cloud. 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*.

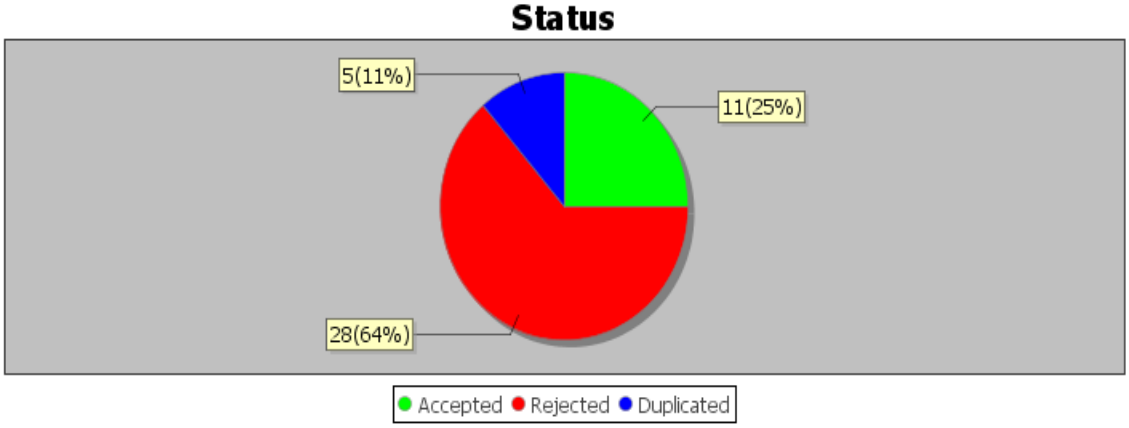
Fonte: elaborado pelos autores.

Após os documentos serem selecionados foram elaboradas questões num formulário relativo à fase de extração, que auxiliou na sistematização e na categorização dos documentos que foram aceitos na fase de extração. Após a extração das publicações, ocorre a última fase: a síntese dos resultados. Nesta 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 revisão empregado.

A figura 1 indica quantitativamente quais documentos foram aceitos e quais rejeitados. A partir do protocolo criado na fase de planejamento da Revisão Sistemática foram encontrados 821 documentos, na fase de seleção foram aceitos 44 documentos, 454 foram rejeitados e 324 foram duplicados. Na fase de extração somente 11 documentos

foram selecionados (25% do total), enquanto 28 documentos foram rejeitados (64% do total), e 5 documentos foram duplicados (11%).

FIGURA 1. Documentos analisados na fase de extração.



Fonte: elaborado pelos autores.

Muitos documentos focam na questão da autenticidade dos serviços de nuvem; há alguns documentos que priorizam abordar as questões de confiança nas políticas entre usuários e serviços de nuvem; existem pesquisas que apontam para a questão da preservação digital sem entrar na questão da preservação em *Cloud*; por outro lado, há estudos sobre a infraestrutura dos serviços de nuvens, suas camadas, programas e *softwares* que possibilitam utilizar repositórios fora do ambiente de nuvem como um fator intermediário entre usuário e nuvem; também há documentos que focam na combinação preservação digital e nuvem, mas não citam padrões, esquemas e metadados de preservação digital.

3 RESULTADOS ALCANÇADOS A PARTIR DA REVISÃO SISTEMÁTICA DA LITERATURA

Após todas as fases da Revisão Sistemática da Literatura, onze documentos se enquadraram nos aspectos temáticos da presente pesquisa, os quais 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 verificou 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, *softwares* etc.

Em todos os documentos que trataram da camada de preservação digital em nuvem, padrões e metadados de preservação eram quistos como um atributo importante na conjuntura da camada.

O estudo da literatura sobre metadados de preservação digital em *Cloud Services*, embasada pela análise de conteúdo, proporcionou averiguar os tipos de metadados de preservação adotados e aplicados nos *Cloud Services*; as características; as funções; as aplicações nas camadas de nuvens e nas interações das camadas; as aplicações em sistemas de preservação já inseridos na nuvem ou hospedados por algum serviço terceirizado; as aplicações em objetos digitais inseridos nos *Cloud Services*, incluindo programas, componentes, dados, outras categorias de metadados e outros tipos de metadados de preservação etc.

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

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. Característica essa encontrada nos documentos da área de Ciência da Informação e que pode ser considerado um problema na compreensão conceitual do tema.

O conceito de camada de preservação é uma tecnologia relativamente nova, conforme pôde ser verificado na literatura científica, as discussões sobre estratégias de preservação em *Cloud Services* iniciam-se a partir de 2011 nas áreas de Ciência da Informação e na Ciência da Computação.

O quadro 4 apresenta os tipos de modelos de *Cloud Services* criados ou serviços utilizados em consonância com algum ambiente em nuvem e os metadados de preservação

encontrados, a partir dos 11 documentos que foram aceitos e analisados na fase de extração da RS aplicada.

QUADRO 4: Sistematização e análise de conteúdo dos documentos recuperados.[1]

Documento 1	
Título	*Preserving records in the cloud*
Autores	Jan Askhoj, Shigeo Sugimoto; Mitsuharu Nagamori
Serviços de preservação[2] ou Camadas de *Cloud Services*	Modelo de camada simples (*Simple layered model*); Modelo em camadas para serviços de arquivamento (*Layered model for archiving services*): contém quatro camadas (PaaS, SaaS, Camada de Preservação e Camada de Interação), 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.
Cloud Services identificados no texto	Amazon S3
Metadados de preservação encontrados	OAIS; PREMIS; XML
Documento 2	
Título	*Towards SIRF: Self-contained information retention format*
Autores	Simona Rabinovici-Cohen; Mary G. Baker; Roger Cummings; Sam Fineberg; John Marberg
Serviços de preservação ou Camadas de *Cloud Services*	SIRF: é um formato de contêiner de armazenamento para um conjunto de objetos (digitais) de preservação que ao mesmo tempo

[1] A análise minuciosa e detalhada dos documentos e dos padrões de metadados em *Cloud Services* encontra-se na íntegra na dissertação de mestrado intitulada Metadados de preservação em *Cloud Services,* e está disponível e acessível no repositório institucional da universidade.

[2] Podem estar embutidos na arquitetura dos *Cloud Services* ou acoplados no ambiente através de um serviço de hospedagem.

	fornece um catálogo com metadados relacionados a todo o conteúdo do contêiner. *DuraCloud*: é um sistema de gerenciamento de conteúdo de código aberto, o qual é utilizado como um sistema de hospedagem em Cloud Services cuja finalidade consiste em fornecer para preservação digital.
Cloud Services identificados no texto	*First Cloud*; *Second Cloud*: modelos de *Cloud Services* criados pelos autores. Fornecem serviços de preservação
Metadados de preservação encontrados no texto	OAIS; PREMIS; XML; VERS; METS; LOTAR
Documento 3	
Título	*Archival Cloud Services: Portability, Continuity, and Sustainability Aspects of Long-term Preservation of Electronically Signed Records*
Autores	Hrvoje Stancic; Arian Rajh; Hrvoje Brzica
Serviços de preservação ou Camadas de *Cloud Services*	*Electronic Document Safe* (EDS): consiste num serviço de suporte para plataformas de armazenamento em nuvem visando segurança em longo prazo para documentos 'oficiais'; Camadas PaaS e SaaS; *Escritório Federal de Segurança em Tecnologia da Informação (Bundesamt fur Sicherheit in der Informationstechnik* - BSI*):* modelo armazenamento de arquivo em Cloud Services caracterizado por suportar preservação em longo prazo para documentos com assinatura digital; Sistema de Segurança e-Servicing (*e-Servicing System of the Insurers* - EDAS): sistema armazenamento que em *Cloud Services*, consiste em preservar documentos com assinatura eletrônica.
Cloud Services identificados no texto	-

Metadados de preservação encontrados no texto	XML; OAIS
Documento 4	
Título	*XML Processing in the Cloud: Large-Scale Digital Preservation in Small Institutions*
Autores	Peter Wittek; Thierry Jacquin; Herv´eD´ejean; Jean-Pierre Chanod; S´andorDar´anyi
Serviços de preservação ou Camadas de *Cloud Services*	SHAMAN[3]: investiga os processos de preservação em longo prazo de grandes volumes de objetos digitais desenvolvendo uma estrutura de preservação.
Cloud Services identificados no texto	Amazon *Elastic Compute Cloud* (EC2); Amazon *Simple Storage Service* (S3)
Metadados de preservação encontrados no texto	XML; METS
Documento 5	
Título	*Content Server System Architecture for Providing Differentiated Levels of Service in a Digital Preservation Cloud*
Autores	Quyen L. Nguyen; Alla Lake
Serviços de preservação ou Camadas de *Cloud Services*	SOA: sistema arquitetado para trabalhar em camadas de nuvem, o servidor de conteúdo apenas ganha funcionalidade com seu respaldo; Servidores de Conteúdo: é uma construção de arquitetura lógica que abrange componentes de gestão de dados, englobando a gestão de dados e o armazenamento de componentes do modelo OAIS, considerado o núcleo central da *Long-Term Digital Preservation as a Service* (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).

[3] *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.

Cloud Services identificados no texto	Preservação Digital em longo Prazo como serviço (*Long-Term Digital Preservation as a Service* - LDPaaS): modelo de *Cloud Service* planejado para fornecer armazenamento e preservação de registros digitais. É caracterizada pelo serviço de preservação e de confidencialidade dos dados. Pode ser uma alternativa viável na provisão independente de registros dos usuários.
Metadados de preservação encontrados no texto	OAIS; XML
Documento 6	
Título	*Developing an ontology for cloud-based archive systems*
Autores	Jan Askhoj; Shigeo Sugimoto; Mitsuharu Nagamori.
Serviços de preservação ou Camadas de *Cloud Services*	Modelo em camadas para serviços de arquivamento (Documento 1); Serviço de Preservação (Preservation Service): utilizado como uma de abstração entre os Aplicativos de criação de conteúdo e o arquivo;
Cloud Services identificados no texto	Amazon EC2; Amazon S3
Metadados de preservação encontrados no texto	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.
Documento 7	
Título	*Finding access and digital preservation solutions for a digitized oral history project: A case study*
Autores	Krystyna Matusiak; Allison Tyler Catie Newton; PadmaPolepeddi.
Serviços de preservação ou Camadas de *Cloud Services*	Omeka: sistema de gerenciamento de código aberto que oferece opções de hospedagem para coleções pequenas e médias;

	DuraCloud: (Documento 2).
Cloud Services identificados no texto	-
Metadados de preservação encontrados no texto	OAIS; MODS
Documento 8	
Título	*Integral and secure cloud architecture based system for backup and retention of public sector information*
Autores	Ingmar Pappel; Ingrid Pappel
Serviços de preservação ou Camadas de *Cloud Services*	Metacloud: infraestrutura conjunta de elementos que unifica diferentes interfaces de armazenamento em nuvem;
Cloud Services identificados no texto	-
Metadados de preservação encontrados no texto	OAIS; XML
Documento 9	
Título	*Digital Preservation in Grids and Clouds: A Middleware Approach*
Autores	Peter Wittek; S´andorDar´anyi
Serviços de preservação ou Camadas de *Cloud Services*	SHAMAN (documento 4);
Cloud Services identificados no texto	Amazon Elastic Compute Cloud (EC2); Amazon Simple Storage Service (S3)
Metadados de preservação encontrados no texto	OAIS; XML; METS
Documento 10	
Título	*Government Use of Cloud-based Long Term Digital Preservation as a Service: An Exploratory Study*
Autores	Patrícia Franks
Serviços de preservação ou Camadas de *Cloud*	Preservica: aplicativo de hospedagem

Services	preservação em longo prazo, pode ser hospedado em nuvem; ---: aplicativo de hospedagem preservação em longo prazo. Armazena duas cópias de conteúdo na nuvem, uma no Amazon S3, que oferta alta disponibilidade e rápida velocidade de transferência, e uma no Amazon Glacier para armazenamento e recuperação em longo prazo com baixos custos.
Cloud Services empregados	Amazon Glacier; Amazon S3; Amazon RDS
Metadados de preservação encontrados no texto	OAIS
Documento 11	
Título	*PDS Cloud: Long Term Digital Preservation in the Cloud*
Autores	Simona Rabinovici-Cohen; John Marberg–Israel Kenneth Nagin IBM Research – David Pease
Serviços de preservação ou Camadas de *Cloud Services*	PDS *Cloud*: serviço de armazenamento especializado em preservação digital em longo prazo em ambientes de nuvem, o serviço é baseado em OAIS. Oferece suporte no âmbito da preservação lógica e preservação para objetos armazenados em nuvem. Pode ser divido em duas camadas descritas logo abaixo; *Multi-Cloud Service*: trabalha no acesso de distintas plataformas de armazenamento; *Preservation Engine*: fornece funcionalidade aos pacotes de informação;
Cloud Services identificados no texto	*Amazon Storage* (S3); *Amazon Compute* (EC2), categorizada como nuvem pública; *Openstack Swift storage*; Nova compute (nuvem privada).
Metadados de preservação encontrados no texto.	OAIS; RepInfo

Fonte: Elaborado pelos autores.

As informações apontadas no quadro 4 evidenciam a necessidade de novos modelos de *Cloud Services* que possam assegurar a preservação em longo prazo. Em algumas pesquisas há uma latente preocupação em solucionar disfunções que são determinantes para a aplicação correta e efetiva dos pacotes de informação elaborados pelo modelo de

referência OAIS (ou padrão de metadados de preservação), novas alternativas na aplicação desses pacotes de informação podem facilitar consideravelmente nos percursos do ciclo de vida de um objeto digital em *Cloud Services* existentes, ou novos modelos que futuramente serão implantados.

4 CONSIDERAÇÕES FINAIS

O presente estudo procurou evidenciar a emergente dependência por parte de usuários e de instituições em alocar seus objetos digitais no ambiente de *Cloud Services*.

Nessa pesquisa ressalta-se o baixo número de publicações envolvendo padrões de metadados de preservação digital em *Cloud Services* nas áreas de Ciência da Informação e de Ciência da Computação, além de sublinhar a omissão da terminologia metadados de preservação digital nos documentos analisados. De 821 documentos, foram aceitos apenas onze na fase de extração da Revisão Sistemática da Literatura; destes onze, apenas três publicações especificam os percursos do ciclo de vida dos objetos digitais pautados na interligação dos padrões de metadados de preservação em *Cloud Services*.

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, excetuando a terminologia *Cloud Computing*, discutiam as descrições de características, de funções e de aplicações que contemplavam os denominados serviços em *Cloud Services*.

A análise dos documentos identificou os padrões *PREMIS*, *OAIS*, *XML*, *MODS* e *METS,* como os mais citados nesse contexto. Porém, em poucos casos foram localizados padrões de metadados de preservação digital que trivialmente não são exemplificados, como por exemplo, o *LOTAR*.

Somente foi possível verificar as atribuições das características, das funções e das aplicações dos padrões de metadados de preservação digital, de forma pormenorizada, específica e detalhada, englobando todo o percurso dos objetos digitais nos *Cloud Services*, em apenas três documentos (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. Nos outros oito textos analisados, os padrões de metadados de preservação são especificados, mas apenas para contextualizar o cenário dos *Cloud Services*.

Baseado na literatura específica do campo da preservação sabe-se que os 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. Nessa direção 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*.

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 das tecnologias em torno dos *Cloud Services* é bem recente, os questionamentos que levaram às abordagens sobre preservação digital nesses ambientes, é ainda mais recente, 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

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

DECMAN, M.; VINTAR, M. A possible solution for digital preservation of e-government: A centralised repository within a cloud computing framework. In: ASSOCIATION FOR INFORMATION MANAGEMENT. **Aslib Proceedings:** new information perspectives. London: Emerald Group, 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.

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, 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.

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

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

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

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

MONTEBELO, R. *et al.* START (Systematic Review Automatic 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.

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.

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, 2007.

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

YAMANE, G. A. C.; 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.

Agradecemos à CAPES pelo apoio financeiro ao desenvolvimento dessa pesquisa.