

Metadados de preservação digital: uma abordagem através da revisão sistemática de literatura

Júlio César Silveira Tauil

Universidade Federal de São Carlos – UFSCar, Email: jtauil86@gmail.com

Fabiano Ferreira de Castro

Universidade Federal de São Carlos – UFSCar, Email: fabianocastro.ufscar@gmail.com

Resumo

A pesquisa em andamento aborda os metadados de preservação digital em *Cloud Services*, a partir dos dados que estão sendo levantados utilizando o método de Revisão Sistemática de Literatura. Serão explicitados os procedimentos necessários para a condução do método em questão, incluindo as fases e a importância de sua utilização, através do *software* criado para facilitar o desenvolvimento dos estudos através da Revisão Sistemática de literatura: o StArt. Destaca-se a relevância da adoção da Revisão Sistemática da Literatura e seu potencial ao levantar as publicações referentes aos metadados de preservação digital em *Cloud Services*. Na análise preliminar dos resultados são relacionados os percentuais de documentos encontrados sobre a temática e a prioridade de leitura nas respectivas bases de dados utilizadas, obtidos por meio de dois gráficos gerados pelo StArt. Neste cenário é possível observar uma tendência baixa de publicações de metadados de preservação em *Cloud Services* nas áreas de Ciência da Informação e Ciência da Computação.

Palavras-chave: Revisão Sistemática de Literatura. *Cloud Services*. Metadados de Preservação Digital.

1 INTRODUÇÃO

Gradualmente os *Cloud Services* estão cada vez mais disseminados na sociedade, e o crescente uso deste tipo de tecnologia está interligado, entre 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, vão surgindo de forma cada vez mais expressiva, a dependência dos servidores de nuvem.

A computação em nuvem oferece benefícios atraentes, como 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 e registros na nuvem, além do mais estão relacionadas à continuidade comercial e à falta de transparência dos serviços na nuvem que afetam a manutenção e o arquivamento de registros (STANCIC; RAJH; BRZICA, 2015). As possíveis vantagens e desvantagens dos serviços de nuvem são norteadas por uma série de perspectivas que precisam ser analisadas no âmbito da CI, incluindo as problemáticas que envolvem preservação digital nestes ambientes.

São variados as complexidades e os desafios que norteiam a preservação digital, e conforme as mudanças tecnológicas vão acarretando em obstáculos na manutenção dos acervos digitais alocados num determinado ambiente virtual, novas estratégias precisam ser pensadas e colocadas em práticas, pois a integridade e a usabilidade de um documento podem ser comprometidas principalmente pela obsolescência digital, processo que ocorre rapidamente.

Há inúmeras problemáticas no que tange à preservação digital, há vários tipos de armazenamentos com características similares e opostas, há uma verdadeira diversificação dos tipos de objetos digitais, de ferramentas tecnológicas (*hardwares* e *softwares*), de programas, de formatos de programas, de dados, de metadados, e assim por diante.

Uma das principais estratégias para se efetuar a preservação digital com êxito é baseada nos serviços de nuvem, afinal a nuvem tem se mostrado como uma alternativa viável no combate contra a obsolescência tecnológica, acarretando na usabilidade de uma quantidade cada vez mais crescente, ávida para reduzir os custos para armazenar dados e visando não ter mais problemas sérios em consequência da obsolescência. A preservação utilizando a tecnologia pautada em *Cloud Services* pode ser considerada uma alternativa viável, e metadados, mais especificamente metadados de preservação digital são fundamentais na garantia eficaz desse processo.

Para se compreender os processos que envolvem o funcionamento das estratégias de representação, de organização e de 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 denominados metadados de preservação.

Nesse estudo adota-se o conceito de Alves (2010, p. 47-8) 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.

No âmbito da (CI), o intuito de manter os objetos digitais acessíveis para uso, é requerido algo mais do que o esforço de se preservar o artefato físico. Para solucionar o problema é necessário considerar também várias outras dimensões que a questão apresenta: como a preservação lógica, física, intelectual, a representação formada por metadados, e o constante monitoramento (SAYÃO, 2007). Estas ações se consolidam a partir das estruturas dos metadados, englobando nessa conjectura os metadados de preservação. “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 nessa perspectiva que apontam-se os metadados como elementos que garantirão efetivamente a preservação dos dados e das informações no ambiente digital.

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).

No contexto da preservação digital em longo prazo, mais precisamente em ambientes de serviços em nuvem, os metadados exercem funções primordiais, especificamente os metadados de preservação digital, há diversos padrões e esquema de metadados de preservação. Dentro do escopo do universo dos *Cloud Services* há vários tipos de serviços de armazenagem, diferentes infraestruturas e camadas, entre outras características que variam de acordo com as estratégias do próprio serviço de nuvem em consonância com as atividades de organizações e de usuários.

No escopo da preservação digital são destacados segundo Riley e Becker (2010): OAIIS (*Open Archival Information System*), METS (*Metadata Encodin and Transmission Standard*), PREMIS (*Data Dictionary for Preservation*), XML (*Extensible Markup Metadata*), MODS (*Metadata Object and Description Schema*), SGML (*Standard Generalized Markup Language*) e AES *Process History*.

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 de Literatura (RS) pode ser considerada um método eficaz ao amparar os pesquisadores de CI ao serem refinadas buscas em bases de dados visando encontrar

publicações sobre esta temática específica, e conseqüentemente levantar dados a partir das publicações sobre a temática.

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. Os métodos aplicados em RS são considerados explícitos e rigorosos, os quais objetivam identificar, de forma pormenorizada, o conhecimento científico em uma determinada área, através da coleta, combinação e avaliação crítica de descobertas de diversas abordagens já realizadas (ANJOS; NUNES; TORI, 2012).

As avaliações sistemáticas requerem mais esforços do que as análises tradicionais. Sua principal vantagem consiste no fornecimento de informações sobre os efeitos de algum fenômeno em uma ampla gama de configurações e métodos empíricos. Se os estudos fornecem resultados consistentes, as revisões sistemáticas fornecem evidências de que o fenômeno é robusto e transferível. Se os estudos fornecem resultados inconsistentes, as fontes de variação podem ser estudadas (KITCHENHAM, 2004, p. 2, tradução nossa).

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, pautado na extração de informações por meio de metodologia bibliográfica.

A fundamentação dos resultados pauta-se em fontes bibliográficas, os quais fornecem subsídios conceituais para o viés de pesquisa bibliográfica na contextualização teórica. A partir de pesquisas teóricas e análises bibliográficas se torna viável assimilar a importância das estratégias e procedimentos que garantam a utilização de metadados de preservação digital em *Cloud Services* pautados em diferentes abordagens da CI. A natureza da abordagem do presente trabalho é de natureza aplicada, pois “objetiva gerar conhecimentos para aplicação prática, dirigidos à solução de problemas específicos” (GERHARDT; SILVEIRA, 200, p. 35).

Outra metodologia que será empregada na presente pesquisa será a análise de conteúdo, “[...] conjunto de técnicas de análises das comunicações, que utiliza procedimentos sistemáticos e objetivos de descrição do conteúdo das mensagens (BARDIN, 1977, p.38).

A RS estabelece de forma reflexiva como a temática sobre metadados de preservação digital 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).

O foco principal deste 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, o objetivo geral do presente estudo visa compreender todas as fases que fazem parte da construção da Revisão Sistemática de Literatura, e conseqüentemente ressaltar o grande poder de utilidade ao empregar este método na Ciência da Informação.

A partir dos desdobramentos do objetivo geral, serão apresentados os seguintes objetivos específicos:

- Analisar o prévio desenvolvimento da temática pautada nas publicações sobre metadados de preservação digital em *Cloud Services* nas áreas de CI e CC utilizando a RS.
- Verificar as funcionalidades da ferramenta StArt¹ (*State of the Art through Review*) na construção de todas as fases que contemplam a RS.

Para buscar com precisão documentos de uma temática específica numa determinada área a RS afunila os resultados de forma precisa, auxiliando de forma prática e decisiva na temática sobre metadados de preservação em *Cloud Services*, pois as etapas são criteriosamente construídas, e por mais que possa ser considerada mais enfadonha e trabalhosa que uma análise exploratória, na realidade a RS poupa tempo ao pesquisador, além proporcionar resultados consistentes de temáticas específicas.

A princípio ocorre à identificação da pesquisa, a seguir são desenvolvidos os critérios de seleção dos estudos, posteriormente ocorre a avaliação dos estudos considerados importantes, a próxima etapa é baseada na extração dos dados e monitoramento do progresso, e a síntese dos dados consiste na última etapa da RS. (GODOY, 2015).

Os estudos ancorados em RS esquadriham um panorama peculiar dos estudos acadêmicos criteriosamente delimitados, e realizados por meio de publicações numa área específica: artigos em revistas e jornais, teses, dissertações, monografias e livros.

1 Disponível em: <http://lapes.dc.ufscar.br/tools/start_tool>. Acesso em: 20 fev. 2018.

2 A IMPORTÂNCIA DA REVISÃO SISTEMÁTICA DE LITERATURA

A metodologia pautada em Revisão Sistemática de Literatura originou-se na área da Saúde e Medicina no final da década de oitenta do século passado. Esta metodologia passou a ser aplicada por conta da baixa qualidade das publicações da área devida à falta de métodos científicos nas revisões bibliográficas. A RS integra as evidências disponíveis até certo momento, dando base de sustentação à combinação de resultados de vários estudos, geralmente isolados, possibilitando o aumento do poder estatístico na avaliação entre resultados. Independente da sua origem a RS pode ser utilizada em diversos tipos de estudo, em qualquer área. Mesmo com diferenças cruciais na formulação da pergunta primária, os formulários e a condução da pergunta primária ocorrem da mesma maneira (MONTEBELO et. al., 2007).

O processo metodológico, de caráter formal e sistemático, torna possível que outros pesquisadores, em posteriores pesquisas, possam reproduzir o mesmo protocolo, e desta forma, possam conseguir avaliar os padrões escolhidos para o caso em questão. Uma RS tem a capacidade de fornecer informações sobre a temática escolhida através de um leque de opções e métodos empíricos, além de possibilitar uma análise conjunta dos estudos primários e resultados, e desta forma, encontrando meios concretos de detectar resultados que estudos menos planejados não conseguem chegar (GONÇALVES; NUNES; DELAMARO, 2011).

Vale ressaltar que os estudos desenvolvidos através da RS são estudos secundários, isto é, se faz necessário promover um eficiente e minucioso levantamento bibliográfico de estudos primários de uma temática específica de determinada área objetivando listar assuntos abordados a partir da comparação de critérios de inclusão e exclusão estabelecidos pelo pesquisador.

Algumas características diferenciam uma RS de uma revisão convencional, como por exemplo: a RS começa a ser definida por um protocolo que especifica as questões de pesquisa sendo adicionados os métodos que serão utilizados para efetuar a revisão; a RS é baseada na definição de estratégias de busca que objetiva detectar o máximo de literatura relevante possível; a RS 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; na RS é requerida de forma explícita a formulação de critérios de inclusão e exclusão para acessar cada estudo primário em potencial (KITCHENHAM, 2004).

O quadro 1, exemplifica os passos, objetivos, etapas, além de resumir a finalidade de cada fase da RS.

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 resultados	Sintetizar os estudos que atendem o propósito da RS	Extração de dados; Síntese de dados.	Dados dos artigos são extraídos e sintetizados.

Fonte: Montebelo et al (2007) apud Godoy (2015, p. 19).

O método da RS é considerado eficiente e consistente, quando uma pesquisa utilizando RS demonstra graves falhas significa que alguma fase, ou mesmo em todas fases da RS não foram desenvolvidas corretamente pelo pesquisador, ou mesmo na fase que antecede a RS, conhecida como RS análise exploratória.

3 PROCEDIMENTOS METODOLÓGICOS

Há várias fases que definem a Revisão Sistemática de Literatura. Inicia-se a RS com a fase de planejamento, compreende a definição do protocolo, ou seja, uma pesquisa piloto e o projeto de uma versão preliminar do formulário de extração de dados. Esta fase é crucial para o bom andamento das seguintes fases, pois nesta fase que são definidas questões bem

pontuais da temática a ser trabalhada na pesquisa. A fase de seleção inicial é responsável pelas atividades estabelecidas com a pré-seleção de estudos primários, como a padronização sequencial das palavras-chave em cada base de dados, e assim verificar a ocorrência de outras palavras-chave. Posteriormente será realizada a fase de seleção e extração dos documentos, as quais agregam novas palavras-chave que podem surgir. A síntese é a última fase do processo, a qual é responsável em analisar os dados extraídos e conseqüentemente relatar os resultados (GODOY, 2015).

Para efetuar todas as fases da Revisão Sistemática de Literatura será utilizado o software StArt (*State of the Art through Review*), esta 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 RS, 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.

Baseado na estrutura utilizada por Godoy (2015) a realização da RS na presente pesquisa está sendo elaborada de acordo com as seguintes fases da metodologia: (I) planejamento: definição do protocolo de pesquisa e da estratégia de busca, pautado no levantamento bibliográfico de caráter quantitativo utilizado na análise exploratória; (II) seleção de documentos: para selecionar 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 selecionados para a próxima fase de acordo com os critérios estipulados; (III) aplicação do formulário de extração de dados nos documentos selecionados anteriormente; (IV) síntese: análise dos documentos encontrados.

Na fase de planejamento, com o intuito de localizar os documentos, estão sendo utilizadas as seguintes bases de dados: Biblioteca Digital Brasileira de Teses e Dissertações (BDTD)², Base de Dados Referenciais de Artigos de Periódicos em Ciência da Informação (BRAPCI), Portal de Periódicos da Capes, *Scientific Electronic Library Online* (SciELO)³, Scopus⁴, *Web of Science*⁵, *IEEE Xplore*⁶, *Computer and Information Systems* (ProQuest)⁷,

2 Disponível em: <<http://bdttd.ibict.br/vufind>>. Acesso em: 20 fev. 2018.

3 Disponível em: <<http://www.scielo.org/php/index.php>>. Acesso em: 20 fev. 2018.

4 Disponível em: <<https://www.scopus.com/freelookup/form/author.uri>>. Acesso em: 20 fev. 2018.

*Academic Search Premier - ASP (EBSCO)*⁸, *Library & Information Science Abstracts (LISA)*⁹, *Information Science & Technology Abstracts (ISTA)*¹⁰. As palavras-chave que encabeçam as *strings* nas bases de dados que estão sendo pesquisadas são: *Cloud Services*; *Digital Preservation Metadata*; *Digital Preservation*; e *Cloud Computing*. Exemplo: (“*Cloud Services*”) AND (“*Cloud Computing*”) AND (“*Digital Preservation Metadata*”) AND (“*Digital Preservation*”).

Cada palavra-chave que encabeçava uma *string* carrega uma variação de palavras-chave equivalentes que seguem uma ordem hierárquica, conforme exemplificada nos esquemas abaixo:

("Cloud Services" OR "Serviços em Nuvem" OR "Servicio en la Nube" OR "Cloud Storage Services" OR "Serviço de Armazenamento em Nuvem" OR "Servicio de Almacenamiento en la Nube" OR "Cloud Storage" OR "Armazenamento em Nuvem" OR "Almacenamiento en la Nube" OR "Cloud Providers" OR "Provedores de Nube" OR "Cloud Storage Providers" OR "Provedores de Armazenamento de Nuvem" OR "Provedores de Almacenamiento en Nube" OR "Cloud Operations" OR "Operações na Nuvem" OR "Opereciones en la Nube" OR "Cloud System" OR "Sistema de Nuvem" OR "Sistema de Nube").

("Cloud Computing" OR "Computação em Nuvem" OR "Computación en la Nube").

("Digital Preservation Metadata" OR "Metadados de Preservação Digital" OR "Metadatos de Preservación Digital").

("Digital Preservation" OR "Preservação Digital" OR "Preservación Digital").

É possível observar que o termo *Cloud* gera uma quantidade enorme de variantes. Por utilizar as principais bases de dados da CI e da CC, além das principais bases de dados que

5Disponível em: <[http://apps-
webofknowledge.ez31.periodicos.capes.gov.br/WOS_GeneralSearch_input.do?product=WOS&search_mode=G
eneralSearch&SID=7Cjk8ECoS9xpFrwnwVQ&preferencesSaved=>](http://apps-
webofknowledge.ez31.periodicos.capes.gov.br/WOS_GeneralSearch_input.do?product=WOS&search_mode=G
eneralSearch&SID=7Cjk8ECoS9xpFrwnwVQ&preferencesSaved=>)>. Acesso em: 20 fev. 2018.

6Disponível em: <<http://ieeexplore.ieee.org.ez31.periodicos.capes.gov.br/Xplore/hom>>. Acesso em: 20 fev. 2018.

7 Disponível em: <<https://search-proquest.ez31.periodicos.capes.gov.br/computerinfo/index>>

8 Disponível em: <[http://web.b-ebsohost-
com.ez31.periodicos.capes.gov.br/ehost/search/basic?vid=0&sid=5b7ecd13-a8e2-4b41-8335-
f5c7bf78c0b4%40sessionmgr103](http://web.b-ebsohost-
com.ez31.periodicos.capes.gov.br/ehost/search/basic?vid=0&sid=5b7ecd13-a8e2-4b41-8335-
f5c7bf78c0b4%40sessionmgr103)>. Acesso em: 20 fev. 2018.

9 Disponível em: <<https://search-proquest.ez31.periodicos.capes.gov.br/lisa>>. Acesso em: 20 fev. 2018.

10Disponível em: <[http://web.a-ebsohost-
com.ez31.periodicos.capes.gov.br/ehost/search/basic?vid=0&sid=eab0f04c-19f8-444e-a930-
e976cc7e0310%40sessionmgr4008](http://web.a-ebsohost-
com.ez31.periodicos.capes.gov.br/ehost/search/basic?vid=0&sid=eab0f04c-19f8-444e-a930-
e976cc7e0310%40sessionmgr4008)>. Acesso em: 20 fev. 2018.

abrangem inúmeras áreas possuírem filtros por áreas somente em casos excepcionais serão utilizadas as palavras-chave: *Information Science* e *Computer Science*, assim como as respectivas palavras-chave equivalentes nas línguas portuguesa e espanhola. Entretanto serão utilizadas várias *strings* de buscas em cada base de dados pesquisada, no intuito de obter a maior quantidade de publicações sobre a temática pesquisada, contudo a principal *string* de busca forma foi confeccionada da seguinte forma:

("Cloud Services" OR "Serviços em Nuvem" OR "Servicio en la Nube" OR "Cloud Storage Services" OR "Serviço de Armazenamento em Nuvem" OR "Servicio de Almacenamiento en la Nube" OR "Cloud Storage" OR "Armazenamento em Nuvem" OR "Almacenamiento en la Nube" OR "Cloud Providers" OR "Provedores de Nube" OR "Cloud Storage Providers" OR "Provedores de Armazenamento de Nuvem" OR "Provedores de Almacenamiento en Nube" OR "Cloud Operations" OR "Operações na Nuvem" OR "Opereciones en la Nube" OR "Cloud System" OR "Sistema de Nuvem" OR "Sistema de Nube") AND ("Cloud Computing" OR "Computação em Nuvem" OR "Computación en la Nube") AND ("Digital Preservation Metadata" OR "Metadados de Preservação Digital" OR "Metadatos de Preservación Digital") AND ("Digital Preservation" OR "Preservação Digital" OR "Preservación Digital").

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

Quadro 2 – Formulário de preenchido de acordo com o 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 que serão utilizadas, os critérios de inclusão e exclusão referente a 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 que foram inseridos na ferramenta tecnológica em questão, pois a partir da lista gerada no sistema, as publicações são encabeçadas das mais importantes para as menos importantes a partir da pontuação que cada texto alcança.

Após a condução da fase de planejamento respaldado baseado no protocolo desenvolvido na ferramenta StARt as publicações elencadas nas bases de dados entram na fase de seleção da RS, a seleção dos textos vão ocorrer 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. Tanto a pontuação como a classificação auxilia de forma categórica a condução da RS.

Seguem, no Quadro 3, os critérios de inclusão e de exclusão estabelecidos para o desenvolvimento da presente 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 online.	O texto não está disponível de forma integral.
Qualquer ano	O documento não está escrito em inglês, português ou espanhol.

Fonte: Elaborado pelo autor.

Na fase de seleção são aceitas as publicações que condizem 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 seleção dos documentos são confeccionadas questões num formulário relativo à fase de extração, apresentado na figura 2, que auxiliam na sistematização e na categorização dos documentos que serão aceitos na fase de extração.

FIGURA 2 - Formulário para auxiliar a extração de dados nos documentos para sistematização da análise de conteúdo.

The image shows a web-based 'Data Extraction Form' with a navigation bar at the top containing 'Study Data', 'Selection Data', 'Data Extraction Form' (highlighted with a warning icon), and 'Similar Studies'. The main content area contains four questions, each followed by a large empty text box for the answer:

- 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.

Na ferramenta se encontram gráficos com o percentual das publicações que interessam ao pesquisador e as publicações que não interessam ao pesquisador por não se enquadrar na temática. O StArt confecciona gráficos que reproduzem o percentual pormenorizado das publicações, há percentuais dos critérios de seleção, da classificação e da própria pontuação, facilitando para o pesquisador trabalhar na última fase: a síntese dos documentos. A extração

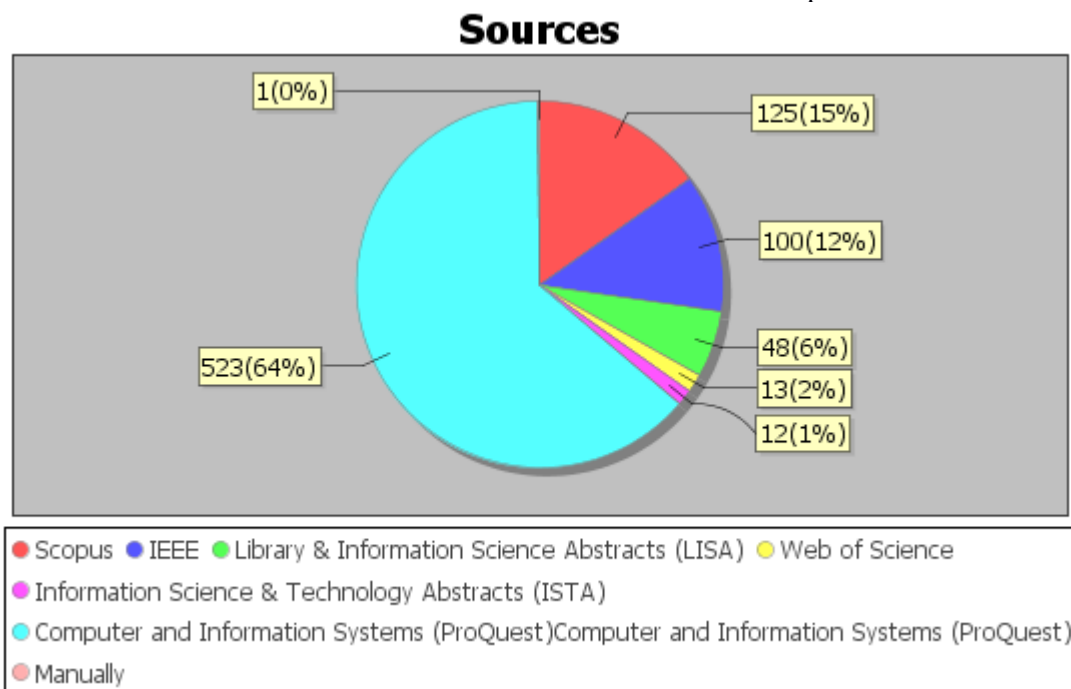
e síntese englobam os passos de análises de resultado.

Após a fase de 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, desta forma a pesquisa em questão será conduzida através das análises de resultados obtidas pelo método de RS empregado.

4 RESULTADOS: APRESENTAÇÃO E DISCUSSÃO

A presente pesquisa em questão se encontra em fase de desenvolvimento. A princípio foram encontrados em seis bases de dados (da CI, CC e de escopo geral) 821 documentos, conforme apontados na figura 3.

FIGURA 3 - Número de documentos recuperados.

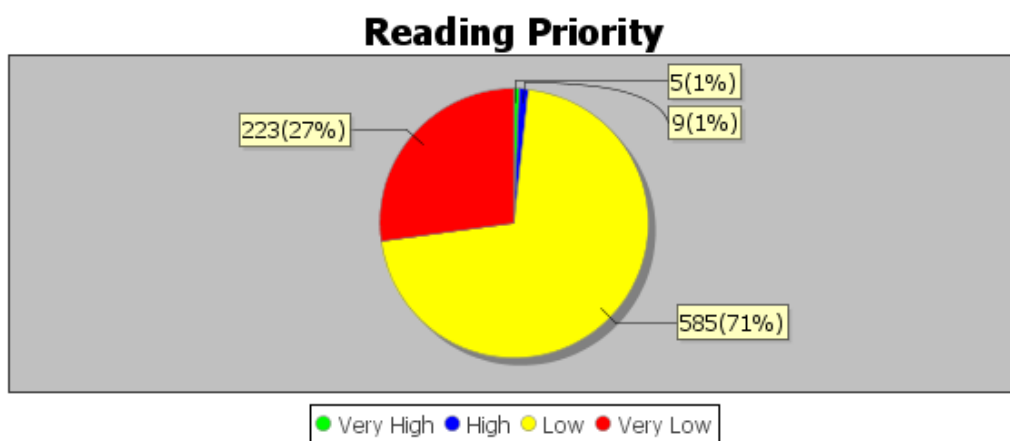


Fonte: elaborado pelo autor.

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)* 125 documentos contribuíram com 15% conforme Figura 1; na *IEEE Xplore* foram encontrados 100 documentos, totalizando 12%.

No desenrolar das etapas referentes à fase de seleção, a partir dos critérios de inclusão e exclusão, os números de documentos foram diminuindo drasticamente. A prioridade de leitura dos documentos em consonância com a temática, ainda na fase de seleção, é dividida em quatro modalidades: muito baixo, baixo, alto, muito alto. 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 pontuação máxima e pode ser um forte indício de ser classificado como muito alto na prioridade das leituras. A figura 4 aponta como ficou estabelecida tal divisão.

FIGURA 4. Prioridade de leitura 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 *VeryLow* (muito baixo), somente 13 documentos se enquadraram nas opções *High* (alto) e *Very High* (muito alto), cada uma das fatias foi representada com 1% de documentos considerados com um alto índice de leitura.

A partir da análise exploratória, da elaboração do protocolo na fase de planejamento e do uso do StArt na fase de seleção é possível observar uma tendência baixa de publicações de metadados de preservação digital em *Cloud Services* nas áreas de CI e CC.

5 RESULTADOS PARCIAIS

A pesquisa em andamento visa contribuir na construção de futuras pesquisas na área

de Ciência da Informação a partir da temática em questão. O estudo em andamento objetiva demonstrar que o futuro da preservação dos recursos digitais, tão amplamente discutido na Ciência da Informação, apresenta consonância com os metadados de preservação em *Cloud Services*.

Dessa forma, as próximas etapas consistem em: avaliar e analisar de forma quantitativa e descritiva os dados dos documentos publicados nas bases de dados, a partir da metodologia de análise de conteúdo e as tendências de pesquisa na temática.

A pertinência da realização de uma Revisão Sistemática se dá pela capacidade desta de promover uma visão geral do desenvolvimento de iniciativas nacionais e internacionais acerca dos metadados de preservação digital em *Cloud Services*.

Metadata of digital preservation: an approach through the Systematic Literature Review

ABSTRACT

The current research addresses the digital preservation metadata in Cloud Services, from the data being collected using the Systematic Literature Review method. The procedures necessary to conduct the method in question, including the phases and the importance of its use, will be explained through the software created to facilitate the development of the studies through the Systematic Literature Review: StArt. The relevance of the adoption of the Systematic Review of Literature and its potential in raising the publications regarding the digital preservation metadata in Cloud Services is highlighted. In the preliminary analysis of the results are related the percentage of documents found on the theme and the priority of the reading in the respective databases, obtained through two graphs generated by StArt. In this scenario it is possible to observe a low tendency of preservation metadata publications in Cloud Services in the areas of Information Science and Computer Science.

Keywords: *Systematic Literature Review. Cloud Services. Digital Preservation Metadata.*

REFERÊNCIAS

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.

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, 2012. Disponível em: <<http://www.jhi-sbis.saude.ws/ojs-jhi/index.php/jhi-sbis/article/viewFile/179/111>>. Acesso em: 09 nov . 2017.

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.

GONÇALVES, V. M.; NUNES, F. L.S; DELAMARO, M.E. Avaliação de funções de similaridade em sistemas de CBIR: uma revisão sistemática. **Anais do VI Workshop de Visão Computacional**, Presidente Prudente, p.199-204. 2011. Disponível em: <<http://www.lbd.dcc.ufmg.br/colecoes/wvc/2010/0035.pdf>>. Acesso em: 07 ago. 2017.

GODOY, K. V. C. de et al. **Identificação de diretrizes para elaboração de política de acesso aberto aos conteúdos educacionais disponibilizados em repositórios digitais no contexto da educação a distância**. 2015. 101 f. Dissertação (Mestrado em Multidisciplinar) - Universidade Federal de São Carlos, São Carlos, SP, 2015.

KITCHENHAM, B. **Procedures for performing systematic reviews**, 2004. Disponível em: <http://www.ifs.tuwien.ac.at/~weipl/systemicReviewsSoftwareEngineering.pdf>>. Acesso 09 nov. 2017.

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.

MONTEBELO, R. et. al. **START (Systematic Review Automatic Tool): Uma ferramenta Computacional de Apoio à Revisão Sistemática**. Departamento de Computação – Universidade Federal de São Carlos (UFSCar), 2007. Disponível em: <http://www.lbd.dcc.ufmg.br/colecoes/eselaw/2007/002.pdf>>. Acesso em: 05 nov. 2017.

NARCISO, E. N.; NUNES, F. L.S; DELAMARO, M. E. Seleção de Casos de Teste Utilizando Conceitos de Variabilidade: Uma Revisão Sistemática. **Anais do VIII Simp. Bras. de Sistemas de Informação**, São Paulo, p. 115-125. 2011. Disponível em: <https://www.researchgate.net/profile/Fatima_Nunes3/publication/267708133_Selecao_de_Casos_de_Testes_Utilizando_Conceitos_de_Variabilidade_Uma_Revisao_Sistemica/links/563c00c508ae34e98c47da74.pdf>. Acesso em: 07 nov. 2017.

RILEY, J.; BECKER, D. **Seeing standards: a visualization of the metadata universe**. 2009. Disponível em: <<http://jennriley.com/metadatamap/>>. Acesso em: 27 nov. 2017.

SAMPAIO, R. F.; MANCINI, M. C. Estudos de revisão sistemática: um guia para síntese criteriosa da evidência científica. **Rev. Bras. Fisioter.**, São Carlos, v.1, nº 11, p.83 -89. 2007. Disponível em: <<http://www.scielo.br/pdf/%0D/rbfig/v11n1/12.pdf>>. Acesso em: 30 nov. 2017.

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.

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.