

## Publicação Ampliada no Contexto de Teses e Dissertações

### **Fernando de Assis Rodrigues**

*Universidade Estadual Paulista Júlio Mesquita Filho - UNESP, Email:  
fernando@rodrigues.pro.br*

### **Ricardo César Gonçalves Sant'Ana**

*Universidade Estadual Paulista Júlio Mesquita Filho - UNESP, Email:  
ricardosantana@marilia.unesp.br*

### **RESUMO**

O objetivo desta pesquisa é estudar aspectos envolvidos no contexto de Publicação Ampliada, em especial para os casos de Teses e Dissertações e os diversos documentos que passam a compor o novo espectro de elementos destes resultados, propiciando subsídios para o desenvolvimento de uma base conceitual que sustente propostas de estruturas de coleta, armazenamento e recuperação deste novo conjunto de documentos que as compõem, a partir de conceitos já estabelecidos na Ciência da Informação. A metodologia utilizada foi a triangulação metodológica, de caráter exploratório e descritivo, composta por: (i) identificação de referencial teórico, através do levantamento bibliográfico sobre Publicação Ampliada em língua portuguesa, de estudos disponíveis nas bases de dados *Google Scholar* e *SciELO* e de resultados do mecanismo de busca *Google Search*; (ii) análise de requisitos, com o uso de fases e objetivos do Ciclo de Vida dos Dados para a Ciência da Informação (CVD), e; (iii) estudo de caso, a partir dos conjuntos de documentos de uma Tese em processo de defesa do Programa de Pós-Graduação em Ciência da Informação da Universidade Estadual Paulista. Os resultados desta pesquisa estão demonstrados na forma de análise de requisitos, a partir do uso de grafos para a elaboração das relações entre os documentos da Publicação Ampliada; e de requisitos nas fases de coleta, armazenamento e recuperação para a implementação das funcionalidades. Conclui-se que, com a explicitação da análise de requisitos das fases de coleta, armazenamento e recuperação, começam a surgir mais preocupações na forma em que as Publicações Ampliadas serão apresentadas, no momento da implementação de funcionalidades, como a necessidade de explicitação de aspectos de privacidade, além de um maior detalhamento e explicitação de ações nesta temática, e delimitação elementos necessários para a instanciação de Publicações Ampliadas.

**Palavras-chave:** Publicação ampliada. Dados. Tese. Dissertação. Ciência da Informação.

---

### **1 INTRODUÇÃO**

Em tempos de revoluções em inúmeras áreas do conhecimento, vivemos no chamado Quarto Paradigma Científico ou A Nova Ciência (HEY, 2009, 2010), permeado pelo crescente uso de dados; do acesso aberto a artigos científicos, bases de dados, índices, tesouros, repositórios e outros documentos; do barateamento do custo de armazenamento de dados; da popularização de dispositivos com alta capacidade de processamento de dados, e; do uso da internet como forma de infraestrutura de comunicação, de compartilhamento e de interoperabilidade (BULLINGER et

al., 2003; CASTELLS, 2008; HEY, 2012; RODRIGUES; SANT'ANA, 2012; SAYÃO; SALES, 2014).

Na comunicação científica as publicações se configuram como o capital para a ciência, sendo o principal critério para o estabelecimento de descobertas científicas e parte importante na resolução de litígios de propriedade, e as Teses e as Dissertações fazem parte integrante deste cenário (SALES; SOUZA; SAYÃO, 2014). O processo de publicação destes resultados demanda não só a disponibilidade do texto – com já comprovada relevância – mas também de outros conteúdos, tais como: tabelas, algoritmos, imagens, áudios e vídeos. Estes conteúdos complementares apresentam importância não só para o entendimento dos resultados nos quais estão vinculados, mas também podem ser aproveitados em outros contextos, como na colaboração entre pesquisadores, de onde emerge a necessidade de ambientes que propiciem a ampliação do espectro de documentos disponíveis destas publicações.

Entretanto, conforme estudado por Sales, Sayão e Souza (2013), Araya (2014), Sayão e Sales (2014) e Mucheroni, Silva e Paletta (2015), ainda há uma carência de ambientes informacionais digitais que propiciem a Publicação Ampliada e neste contexto também estão inseridos as Teses e Dissertações, em que estes conteúdos também são parte integrante – e não apenas complementar – dos resultados.

## **2 OBJETIVOS E PROCEDIMENTOS METODOLÓGICOS**

O objetivo desta pesquisa é estudar aspectos envolvidos no contexto de Publicação Ampliada, em especial para os casos de Teses e de Dissertações e os diversos documentos que passam a compor o novo espectro de elementos destes resultados, propiciando subsídios para o desenvolvimento de uma base conceitual que sustente propostas de estruturas de coleta, armazenamento e recuperação deste novo conjunto de documentos que as compõem, a partir dos conceitos já estabelecidos na Ciência da Informação.

A metodologia utilizada foi a triangulação metodológica (DUARTE, 2009), de caráter exploratório e descritivo, composta por: identificação de referencial teórico, através do levantamento bibliográfico sobre Publicação Ampliada em língua portuguesa; análise de requisitos, com o uso de fases e objetivos do Ciclo de Vida dos Dados para a Ciência da Informação (CVD), e; estudo de caso, a partir dos conjuntos de documentos de uma tese em

processo de defesa do Programa de Pós-Graduação em Ciência da Informação (PPGCI) da Universidade Estadual Paulista (UNESP).

No levantamento bibliográfico, foram recuperados estudos disponíveis nas bases de dados *Google Scholar*<sup>1</sup> e SciELO; e no mecanismo de busca *Google Search*<sup>2</sup> (com corte da amostra na quarta página), através da aplicação do termo “publicação ampliada”, disponíveis no mês de julho de 2016. Foram descartadas apresentações de slides, editais licitatórios, editoriais, citações, patentes, notícias, páginas de apresentação contendo perfis de pesquisadores, postagens de blog, verbetes de enciclopédias, e índices, bem como estudos recuperados que não apresentam relação com o tema.

Foram identificados 74 (setenta e quatro) resultados: 25 (vinte e cinco) no *Google Scholar*, 4 (quatro) no SciELO e 44 (quarenta e quatro) no *Google Search*. Foram descartados 34 (trinta e três) resultados: 15 (quinze) notícias, 3 (três) apresentações de slides, 3 (três) editais licitatórios, 3 (três) perfis de pesquisador, 2 (dois) editoriais, 2 (duas) postagens de blog, 2 (dois) verbetes de enciclopédias, 2 (duas) citações, 1 (um) índice de periódico e 1 (um) item que não estava acessível no momento da pesquisa.

Dos 40 (quarenta) resultados restantes, foram considerados fora do escopo desta pesquisa 24 (vinte e quatro): 14 (catorze) artigos em periódicos científicos, 6 (seis) Dissertações, 3 (três) artigos em congresso científico e 1 (uma) tese – por não tratarem do tema Publicação Ampliada e sim documentos que citam o termo mas abordam outro assunto.

**Quadro 1** – Documentos sobre Publicação Ampliada

<b>Título</b>	<b>Ano</b>	<b>Autores</b>	<b>Tipo</b>	<b>URL</b>
Modelo triádico de relações: um protótipo de modelagem conceitual para a área nuclear	2012	Luana Farias Sales e Luís Fernando Sayão	Artigo em Congresso	<a href="http://carpedien.ien.gov.br/bitstream/ien/650/1/artigo_ena_ncib2012_sales_say%C3%A3o.pdf">http://carpedien.ien.gov.br/bitstream/ien/650/1/artigo_ena_ncib2012_sales_say%C3%A3o.pdf</a>
Publicações ampliadas: um novo modelo de publicação acadêmica para o ambiente de e-science	2013	Luana Farias Sales, Luís Fernando Sayão e Rosali Fernandez de Souza	Artigo em Congresso	<a href="http://eprints.rclis.org/20665/1/publica%C3%A7%C3%B5es_ampliadas_2013.pdf">http://eprints.rclis.org/20665/1/publica%C3%A7%C3%B5es_ampliadas_2013.pdf</a>

1 *Google Scholar* é uma marca registrada de Google, LLC.

2 *Google Search* é uma marca registrada de Google, LLC.

<b>Título</b>	<b>Ano</b>	<b>Autores</b>	<b>Tipo</b>	<b>URL</b>
Publicação ampliada: um novo modelo de publicação científica voltada para os desafios de uma ciência orientada por dados.	2014	Luana Farias Sales, Rosali Fernandez de Souza e Luís Fernando Sayão	Artigo em Congresso	<a href="http://ridi.ibict.br/handle/123456789/824">http://ridi.ibict.br/handle/123456789/824</a>
Dados abertos de pesquisa: ampliando os conceitos de acesso livre	2014	Luís Fernando Sayão e Luana Farias Sales	Artigo em Periódico	<a href="http://www.reciis.icict.fiocruz.br/index.php/reciis/article/download/611/1252">www.reciis.icict.fiocruz.br/index.php/reciis/article/download/611/1252</a>
Comunicação científica: agregação, compartilhamento e reúso de elementos informacionais	2014	Elizabeth Roxana Mass Araya	Tese	<a href="http://repositorio.unesp.br/bitstream/handle/11449/121981/000812425.pdf?sequence=1&amp;isAllowed=y">http://repositorio.unesp.br/bitstream/handle/11449/121981/000812425.pdf?sequence=1&amp;isAllowed=y</a>
Integração semântica de publicações científicas e dados de pesquisa: proposta de modelo de publicação ampliada para a área de Ciências Nucleares	2014	Luana Farias Sales	Tese	<a href="http://repositorio.ibict.br/bitstream/123456789/806/1/LUANA%20SALES%20D.pdf">http://repositorio.ibict.br/bitstream/123456789/806/1/LUANA%20SALES%20D.pdf</a>
Entre a publicação ampliada e a multimodalidade	2015	Marcos L. Mucheroni, Fernando José Modesto da Silva e Carlos Francisco Paletta	Artigo em Congresso	<a href="http://www.ufpb.br/evento/lti/ocs/index.php/enancib2015/enancib2015/paper/viewFile/2873/1178">http://www.ufpb.br/evento/lti/ocs/index.php/enancib2015/enancib2015/paper/viewFile/2873/1178</a>
<i>Enhanced publication: a new model of scientific publication for the nuclear área</i>	2015	Luana Farias Sales e Luís Fernando Sayão	Artigo em Periódico	<a href="http://revistas.iien.gov.br/index.php/iienprogressreport/article/view/131">http://revistas.iien.gov.br/index.php/iienprogressreport/article/view/131</a>
Dados de pesquisa em repositório institucional: o caso do <i>Edinburgh DataShare</i>	2015	Denise Ramires Machado	Dissertação	<a href="http://www.lume.ufrgs.br/bitstream/handle/10183/119157/000970298.pdf?sequence=1">http://www.lume.ufrgs.br/bitstream/handle/10183/119157/000970298.pdf?sequence=1</a>

Fonte: Elaborado pelos autores.

Da recuperação dos resultados através das bases de dados foram obtidos 5 (cinco) documentos idênticos, ou seja, dos 19 (dezenove) resultados foram identificados 9 (nove) documentos que tratam sobre Publicação Ampliada: 4 (quatro) artigos em congressos científicos, 2 (dois) artigos em periódicos, 2 (duas) Teses e 1 (uma dissertação) – apresentadas no Quadro 1.

Na análise de requisitos, utilizou-se o conceito de CVD, proposto por Sant’Ana (2013), para a explicitação dos elementos vinculados com a Publicação Ampliada para as fases: de Coleta, fase inicial de planejamento e outras atividades relacionadas a obtenção e processamento dos dados; de Armazenamento, com atividades relacionadas a estruturação, a inserção, a

manipulação, a migração e a transmissão na persistência de dados; e de Recuperação, com atividades voltadas à fase em que estes dados passam a estar disponíveis para consulta e visualização. Foram descartadas as preocupações inerentes a fase de Descarte. Cada uma das fases é permeada por objetivos – delimitações temáticas de análise, sendo: Preservação, Disseminação, Direitos Autorais, Qualidade, Integração e Privacidade – explicitadas neste estudo para as fases de Coleta, Armazenamento e Recuperação.

Para o estudo de caso foi selecionada uma tese de doutorado apresentada ao PPGCI da UNESP na linha de pesquisa “Informação e Tecnologia” e o conjunto de documentos relacionados a ela. Esta escolha se deu em função de estar baseada em forte processo de coleta e geração de dados o que propiciou um conjunto relevante de documentos complementares e que fazem parte integrante da tese. Destaca-se que outro fator determinante para a escolha foi a não localização de outra tese que tenha tido como característica buscar a publicação de variados tipos de documentos.

A próxima seção apresenta as definições de Publicação Ampliada, seus requisitos e estruturas – identificadas no levantamento bibliográfico, bem como a análise de requisitos destes elementos estruturados pelo CVD e o estudo de caso. A quarta seção apresenta os resultados desta pesquisa que estão demonstrados na forma de análise de requisitos, com origem na área de Engenharia de Software, a partir do uso de grafos – com aporte da teoria matemática denominada Teoria dos Grafos (BIGGS; LLOYD; WILSON, 1986; WILSON, 1996) – para a elaboração das relações entre os documentos da Publicação Ampliada; e de requisitos nas fases de coleta, armazenamento e recuperação para a implementação das funcionalidades. A quinta seção apresenta as considerações finais e trabalhos futuros.

### **3 DISCUSSÃO**

Em cada um dos documentos identificados nesta pesquisa, foram encontrados, no mínimo, uma definição sobre a Publicação Ampliada (Quadro 2), tendo como núcleo comum a definição de que a formação destas publicações é composta pela publicação digital (*e-print*) e por documentos complementares contendo conjuntos de dados.

**Quadro 2** – Definições sobre Publicação Ampliada

<b>Autoria do Estudo</b>	<b>Definição</b>
Sales e Sayão (2012)	“Publicação ampliada pode ser definida como um modelo de informação que apoia a integração dos <i>e-prints</i> ao conjunto de dados a pesquisa, por exemplo, uma tese e os dados que foram gerados a partir dela”.
Sales, Sayão e Souza (2013)	“[...] objetos digitais compostos que combinam <i>e-prints</i> com um ou mais recursos de dados, um ou mais registros de metadados, ou qualquer combinação destes.” (Digital Repository Infrastructure Vision for European Research – II )
Sales, Souza e Sayão (2014)	“[...] objetos digitais compostos que combinam <i>e-prints</i> com um ou mais recursos de dados, um ou mais registros de metadados, ou qualquer combinação destes.” (Digital Repository Infrastructure Vision for European Research – II)
	“[...] é formada por objetos digitais (publicações digitais, dados digitais e outros documentos digitais) e relações conceituais se tornando dessa forma um objeto digital complexo e tem, entre suas funcionalidades, a possibilidade de registrar toda a memória digital de uma pesquisa”.
	“[...] é uma instância de um objeto digital complexo que integra, por meio de relações, recursos distribuídos por diversos repositórios digitais. Os componentes principais de uma publicação ampliada são então os objetos digitais ( <i>e-prints</i> , dados digitais de pesquisa, metadados e outros recursos) e relações – que quando possuem sua semântica explicitada, possibilita, entre outras coisas, que a publicação – integrada aos dados e documentos correlatos – seja recuperada e a disseminada maneira mais precisa e inteligente.
Sayão e Sales (2014)	“O documento abstrato que combina <i>e-prints</i> e dados de pesquisas – chamado de ‘publicação ampliada’ [...]”.
	“[...] é pensada como uma forma de objeto digital complexo que combina vários recursos heterogêneos que, porém, são relacionados.”
	“[...] pode ser considerada um veículo de comunicação científica de grande importância para a comunidade de pesquisadores.”
Araya (2014)	“[...] é a publicação melhorada com dados de pesquisa, materiais adicionais, dados posteriores à publicação, registros de bases de dados e que tem uma estrutura baseada no objeto com links explícitos entre os objetos. Um objeto pode ser (parte de) um artigo, um conjunto de dados, uma imagem, um filme, um comentário, um módulo ou um link para informação em uma base de dados”. (ELBAEK, PEDERSEN E SIEMAN 2009, tradução da autora)
	“[...] a publicação melhorada com os elementos informacionais provenientes do processo desenvolvido na pesquisa científica, sejam eles, por exemplo, artigos, conjuntos de dados, imagens, filmes, comentários, entre outros [...]”.
	“[...] um modelo de publicação que permita propiciar os reordenamentos necessários para que a comunicação científica possa conter os elementos informacionais gerados ao longo de qualquer pesquisa científica [...]”.

Autoria do Estudo	Definição
	<p>“[...] é um objeto composto digital caracterizado por um identificador, de preferência um identificador persistente, e por informação de metadados. Esse objeto composto digital deve conter obrigatoriamente uma parte narrativa equivalente à descrição da pesquisa (como na publicação científica tradicional), além dos diversos elementos informacionais, heterogêneos, mas interconectados, provenientes do processo de pesquisa. Tanto a narrativa quanto os elementos informacionais que constituem a publicação ampliada devem ser compreensíveis por si só. Isto é, a publicação ampliada é o conjunto formado pela narrativa dos resultados da pesquisa e pelos elementos informacionais inerentes ao próprio processo de pesquisa. Essas partes podem, ou não, ter um identificador e os metadados respectivos e se enlaçam por relações semânticas”. (BARDI e MANGHI, 2014)</p> <p>“[...] se compõe de uma parte narrativa estruturada em várias partes interconectadas (resumo, seções, figuras, tabelas, bibliografia, etc.) [...]”.</p>
Sales (2014)	<p>“[...] é um artefato tecnológico que visa reunir objetos digitais de diversas naturezas, porém de uma mesma temática [...]”.</p> <p>“[...] enquanto um objeto digital complexo, pode se configurar como a combinação de um <i>e-print</i> depositado em um repositório institucional, dados de pesquisa depositados em um repositório de dados e ainda outros documentos depositados em qualquer outra base de dados de informação técnico-científica [...]”.</p> <p>“[...] é uma instância de um objeto digital composto, cujo objetivo é ligar os resultados de pesquisa aos dados que o geraram, extrapolando o limite do documento físico [...]”.</p> <p>“[...] objetos digitais compostos que combinam <i>e-Prints</i> com um ou mais recursos de dados, um ou mais registros de metadados, ou qualquer combinação destes”. De acordo com o relatório de autoria de Peter Verhaar, <i>e-Prints</i> são entendidos como um recurso textual como trabalho acadêmico original, que se destina a ser lido por seres humanos, que apresenta algumas reivindicações acadêmicas e que geralmente contém uma interpretação ou uma análise de determinados dados primários.” (VEHAAR, 2008, p.11).</p> <p>“[...] é formada por objetos digitais, metadados e ainda por relações conceituais que conectam esses objetos formando um agregado de informações [...]”.</p>
Mucheroni, da Silva e Paletta (2015)	<p>“[...] uma forma de publicação eletrônica que amplia a divulgação e o compartilhamento dos resultados de pesquisa, de maneira que eles possam ser encontrados.” (VERNOOY-GERRITSEN, 2009)</p> <p>“[...] é uma nova forma de publicação onde a forma tradicional (um relatório, um artigo, um livro) é enriquecida com informações adicionais, ela depende das possibilidades de ligações-conexões da Web”.</p> <p>“[...] contempla os dados da pesquisa, de materiais extras, da postagem dos dados da publicação, ou, ainda, dos registros contidos em um banco de dados, e que tem algum objeto contendo uma estrutura baseada em ligações explícita entre ele (o objeto)” (HOGENAAR e HOOGERWERF, 2008).</p>

<b>Autoria do Estudo</b>	<b>Definição</b>
Sales e Sayão (2015)	“[...] agrega, de forma padronizada, publicações, dados, conjuntos de dados e metadados”. (tradução dos autores)
Machado (2015)	“[...] possibilidades de vinculação entre os dados de pesquisa e o trabalho final (publicação) relacionada a esses dados [...]”. (SALES e SAYÃO, 2012)
	“[...] contempla essas conexões entre dados e publicações [...]”. (SALES e SAYÃO, 2012)

Fonte: Elaborado pelos autores.

Neste estudo, adota-se a Publicação Ampliada como sendo uma composição de objetos digitais: documentos, dados, metadados e outros elementos descritivos, relacionados entre si, e que apresentem em seu todo, uma estrutura de corpo que lhe garanta uma identidade e características próprias que possam ser vinculadas ao conjunto.

Estas publicações ampliadas devem fazer uso de identificadores e, como forma de suporte, podem se valer de repositórios institucionais, de repositórios de dados ou outros sistemas de informação que permitam a sua recuperação em sua totalidade ou parte dos objetos que as compõem.

Os estudos de Sales, Sayão e Souza (2013), Mucheroni, da Silva e Paletta (2015), Sales, Souza e Sayão (2014), Machado (2015), Araya (2014) e Sales (2014) propõem delimitações sobre quais são os requisitos e as estruturas necessárias para a formação da composição de uma Publicação Ampliada.

O Quadro 3 apresenta a análise dos requisitos identificados nestes estudos, segmentando-os pelas fases e objetivos do CVD. Cada requisito está marcado com um número sobrescrito no final do período, indicando em qual estudo foi identificado-o.

**Quadro 3** – Organização no CVD dos requisitos e estruturas necessárias à Publicação Ampliada

<b>Objetivos</b>	<b>Fases</b>		
	<b>Coleta</b>	<b>Armazenamento</b>	<b>Recuperação</b>
Preservação	Identificação de elementos sobre os objetos que viabilizem a preservação <sup>1 2 3 4 5 6</sup>	Armazenamento dos metadados de preservação de cada objeto <sup>1 2 3 4 5 6</sup>	Disponibilização dos metadados de preservação dos objetos <sup>1 2 3 4 5 6</sup>

Objetivos	Fases		
	Coleta	Armazenamento	Recuperação
Disseminação		Geração de identificador para cada objeto: URI <sup>1 2 3 5 6</sup> , DOI <sup>2 5</sup> , PURLS <sup>2</sup> ou URN <sup>2</sup>	Permitir acesso a cada objeto por identificador: URI <sup>1 2 3 5 6</sup> , DOI <sup>2 5</sup> , PURL <sup>2</sup> ou URN <sup>2</sup>  Manter disponíveis as publicações em sítio web <sup>1 5</sup>  Manter acesso aos metadados dos objetos em versão legível por máquina <sup>1 2 3 6</sup>  Possibilitar a vinculação dos objetos em outros contextos ( <i>Linked Data</i> ) <sup>2</sup>
Direitos Autorais	Obter a descrição de elementos sobre a autoria dos objetos <sup>1 2 3 5 6</sup>	Persistir os metadados de autoria de cada objeto <sup>1 2 3 5 6</sup>	Permitir o acesso aos metadados de autoria dos objetos <sup>1 2 3 5 6</sup>  Garantir a autoria dos objetos <sup>5</sup>
Qualidade	Propiciar ambiente para inserção, alteração ou exclusão de objetos <sup>1 5</sup>	Manter registro sobre o controle de versões da publicação <sup>1 5</sup>  Manter registro sobre o controle de versões de cada objeto <sup>1 5</sup>	Permitir acesso a cada objeto univocamente <sup>1 2 3 5 6</sup>  Propiciar ambiente para a realização de comentários e avaliações <sup>1 2</sup>
Integração	Obter elementos que possibilitem o relacionamento entre objetos <sup>1 2 3 4 5 6</sup>	Persistir elementos que sustentem os relacionamentos entre os objetos <sup>1 2 3 4 5 6</sup>	Disponibilizar a identificação das relações entre os objetos <sup>1 2 3 4 5 6</sup>

<sup>1</sup> Sales, Sayão e Souza (2013)

<sup>2</sup> Mucheroni, da Silva e Paletta (2015)

<sup>3</sup> Sales, Souza e Sayão (2014)

<sup>4</sup> Sayão e Sales (2014)

<sup>5</sup> Araya (2014)

<sup>6</sup> Sales (2014)

Fonte: Elaborado pelos autores.

Não foram identificados nos textos estudados elementos vinculados a privacidade que pudessem ser vinculados às fases de coleta, armazenamento e recuperação do CVD.

Na fase de coleta é necessária a possibilidade de inserção e de identificação de elementos que viabilizem a preservação dos objetos, bem como informações de autoria que compõem a Publicação Ampliada. Também é importante que os objetos possam ser relacionados: por

exemplo, possibilitar a vinculação um conjunto de dados com um gráfico do texto. Nesta fase, também deve estar disponível um ambiente com as opções de inserção, alteração ou exclusão de objetos da Publicação Ampliada.

No armazenamento, as preocupações concentram na persistência de informações, principalmente no armazenamento dos metadados de preservação e de autoria, na persistência dos relacionamentos, e no controle de versões dos objetos e da Publicação Ampliada, além de garantir a unicidade de cada objeto (incluindo a diferenciação unívoca de cada versão) através do uso de identificadores, com preferência ao uso de *Uniform Resource Identifier (URI)*, ou outros sistemas de identificação tais como: o *Digital Object Identifier (DOI)*, o *Persistent Uniform Resource Locator (PURL)* ou o *Uniform Resource Name (URN)*.

Para a fase de recuperação, é necessária a publicação dos objetos – disponibilizando metadados de preservação e de autoria, suas relações e seus identificadores – em formato legível por humanos e por máquinas, através de sítio web que permita realizar comentários e avaliações dos conteúdos, para propiciar a disseminação da Publicação Ampliada e a vinculação dos objetos em outros contextos (como a vinculação de dados legíveis por máquina via *Linked Data*). Os objetos devem estar vinculados a identificadores, com preferência ao uso de *URI*, ou outros sistemas de identificação tais como: o *DOI*, o *PURL* ou o *URN*.

### **Características da Tese**

Os modelos de publicação impressa e digital para Teses e Dissertações impossibilitaram a inserção de todos os objetos que compõem a tese do estudo de caso. Por exemplo, uma das figuras da pesquisa possui tamanho de 5,2 m (cinco metros e dois décimos) de largura por 3,5 m (três metros e meio) de altura.

A tese é composta por um conjunto de 389 (trezentos e oitenta e nove) documentos, divididos em:

- (a) 1 (um) texto atual da tese, em formato *Open Document Text (ODT)*, com tamanho de arquivo de 69 (sessenta e nove) megabytes, 64122 (sessenta e quatro mil cento e vinte e dois) palavras e 428476 (quatrocentos e vinte e oito mil quatrocentos e setenta e seis) caracteres;

- (b) 18 (dezoito) versões anteriores do texto atual, em formato *ODT*, com tamanho total de 594,8 (quinhentos e noventa e quatro e oito décimos) megabytes;
- (c) 15 (quinze) conjuntos de planilhas eletrônicas, em formato *Open Document Spreadsheet (ODS)*, utilizadas na coleta de dados, com tamanho total de 0,9 (nove décimos) megabytes e 12128 (doze mil cento e vinte e oito) linhas;
- (d) 2 (duas) imagens em formato *Portable Network Graphics (PNG)*, contendo gráficos, com tamanho total de 0,1 (um décimo) megabytes;
- (e) 1 (um) conjunto de planilhas eletrônicas, em formato *ODS*, contendo dados sobre quadro de critérios, com tamanho total de 0,75 (setenta e cinco centésimos) megabytes e 45 (quarenta e cinco) linhas (posteriormente embarcado como tabela dentro do texto da tese);
- (f) 2 (dois) bancos de dados, armazenados em Sistema Gerenciador de Banco de Dados (SGBD) *MySQL*<sup>3</sup>, com tamanho total de 10 (dez) megabytes, contendo 25 (vinte e cinco) tabelas com 3977 (três mil novecentos e setenta e sete) linhas;
- (g) 4 (quatro) arquivos de backup do banco de dados, em formato *Structured Query Language (SQL) script*, com compressão em formato *Tarball (tar)*, com tamanho total de 0,25 (vinte e cinco centésimos) megabytes, contendo 25 (vinte e cinco) tabelas com 3977 (três mil novecentos e setenta e sete) linhas em cada arquivo;
- (h) 2 (duas) pastas de trabalho para elaboração de Modelagem Entidade-Relacionamento, em formato *MWB*, originário do aplicativo *MySQL Workbench*<sup>4</sup>, com tamanho total de 0,4 (quatro décimos) megabytes, contendo 3 (três) Diagramas Entidade-Relacionamento;
- (i) 2 (dois) Diagramas de Entidade-Relacionamento, em formato *PNG*, com tamanho total de 0,29 (vinte e nove décimos) megabytes (posteriormente embarcados como imagens dentro do texto da tese);
- (j) 1 (um) algoritmo para elaboração de Dicionário de Dados, em formato de *script* para execução, em linguagem de programação Lua, com 45 (quarenta e cinco) linhas de código e tamanho de 0,01 (um centésimo) megabyte;
- (k) 1 (um) algoritmo, em formato de *script* para execução, em linguagem de consulta a bancos de dados *SQL*, com 48 (quarenta e oito) linhas de código e tamanho de 0,03 (três centésimos) megabyte;

---

<sup>3</sup> *MySQL* é uma marca registrada de Oracle Corporation.

<sup>4</sup> *MySQL Workbench* é uma marca registrada de Oracle Corporation.

- (l) 2 (dois) Dicionários de Dados, em formato de tabela *HyperText Markup Language (HTML)*, contendo 28830 (vinte oito mil oitocentos e trinta) caracteres, com tamanho total de 0,29 (vinte e nove décimos) megabytes (posteriormente embarcados como imagens dentro do texto da tese);
- (m) 7 (sete) algoritmos, em formato de *script* para execução, em linguagem de marcação *Dot*, para elaboração automatizada de grafos pelo aplicativo *GraphViz*, contendo 693 (seiscentos e noventa e três) linhas, 47108 (quarenta e sete mil cento e oito) caracteres, com tamanho total de 0,47 (quarenta e sete décimos) megabytes;
- (n) 7 (sete) imagens em formato *PNG*, contendo os grafos elaborados pelo aplicativo *GraphViz*, com tamanho total de 4,4 (quatro e quatro décimos) megabytes;
- (o) 11 (onze) imagens em formato *PNG*, inseridas como recursos para a elaboração dos grafos pelo aplicativo *GraphViz*, com tamanho total de 2,9 (dois e nove décimos) megabytes;
- (p) 279 (duzentos e setenta e nove) capturas de tela, em formato *Portable Document Format (PDF)*, armazenados como parte da coleta de dados, com tamanho total de 104,7 (cento e quatro e sete décimos) megabytes;
- (q) 10 (dez) desenhos vetoriais, em formato *Open Document Graph (ODG)*, com tamanho total de 3,3 (três e três décimos) megabytes;
- (r) 10 (dez) imagens em formato *PNG*, geradas a partir dos desenhos vetoriais, com tamanho total de 24,1 (vinte e quatro e um décimo) megabytes (posteriormente embarcados como imagens dentro do texto da tese);
- (s) 1 (um) texto da tese, em formato *PDF*, enviado aos participantes da banca de qualificação, com tamanho de 17,3 (dezesete e três décimos) megabytes;
- (t) 1 (um) texto da tese, em formato *Electronic Publication (e-Pub)*, enviado aos participantes da banca de qualificação, com tamanho de 3,4 (três e quatro décimos) megabytes;
- (u) 7 (sete) capturas de tela, em formato *PNG*, armazenados como parte da coleta de dados, com tamanho total de 6,7 (seis e sete décimos) megabytes (posteriormente embarcados como imagens dentro do texto da tese);

- (v) 3 (três) esboços interligados (*wireframes*), em formato PNG, gerados no sítio *Moqups*, com tamanho total de 1,4 (um e quatro décimos) megabytes (posteriormente embarcados como imagens dentro do texto da tese).

A publicação possui um tamanho de 845,49 (oitocentos e quarenta e cinco e quarenta e nove décimos) megabytes, sendo identificados 1 (um) banco de dados (armazenado em um SGBD) e 12 (doze) tipos de arquivos: 5 (cinco) formatos vinculados com algoritmos para a execução por interpretadores de códigos e por SGBD (*lua*, *sql*, *sql+tarball*, *html* e *dot*); 3 (três) para a elaboração ou disponibilização de textos (*odt*, *pdf* e *epub*); 2 (dois) para elaboração de imagens (*png* e *odg*); 1 (um) para elaboração de planilhas eletrônicas (*ods*) e 1 (um) para interpretação por aplicativo de elaboração de Modelagem Entidade-Relacionamento (*mwb*).

Para cada formato de arquivo foram identificados um total de: 280 (duzentos e oitenta) arquivos em formato *pdf*; 42 (quarenta e duas) imagens em formato *png*; 19 (dezenove) textos em formato *pdf*; 18 (dezoito) planilhas eletrônicas em formato *ods*; 10 (dez) desenhos vetoriais em formato *odg*; 7 (sete) algoritmos em formato *dot*; 4 (quatro) *scripts* em formato *sql* com compressão *tarball*; 2 (dois) arquivos de Modelagem Entidade-Relacionamento em formato *mwb*; 2 (duas) tabelas em formato *html*; 1 (um) *script* em formato *sql*; 1 (um) *script* em formato *lua*, e; 1 (um) texto em formato *e-Pub*.

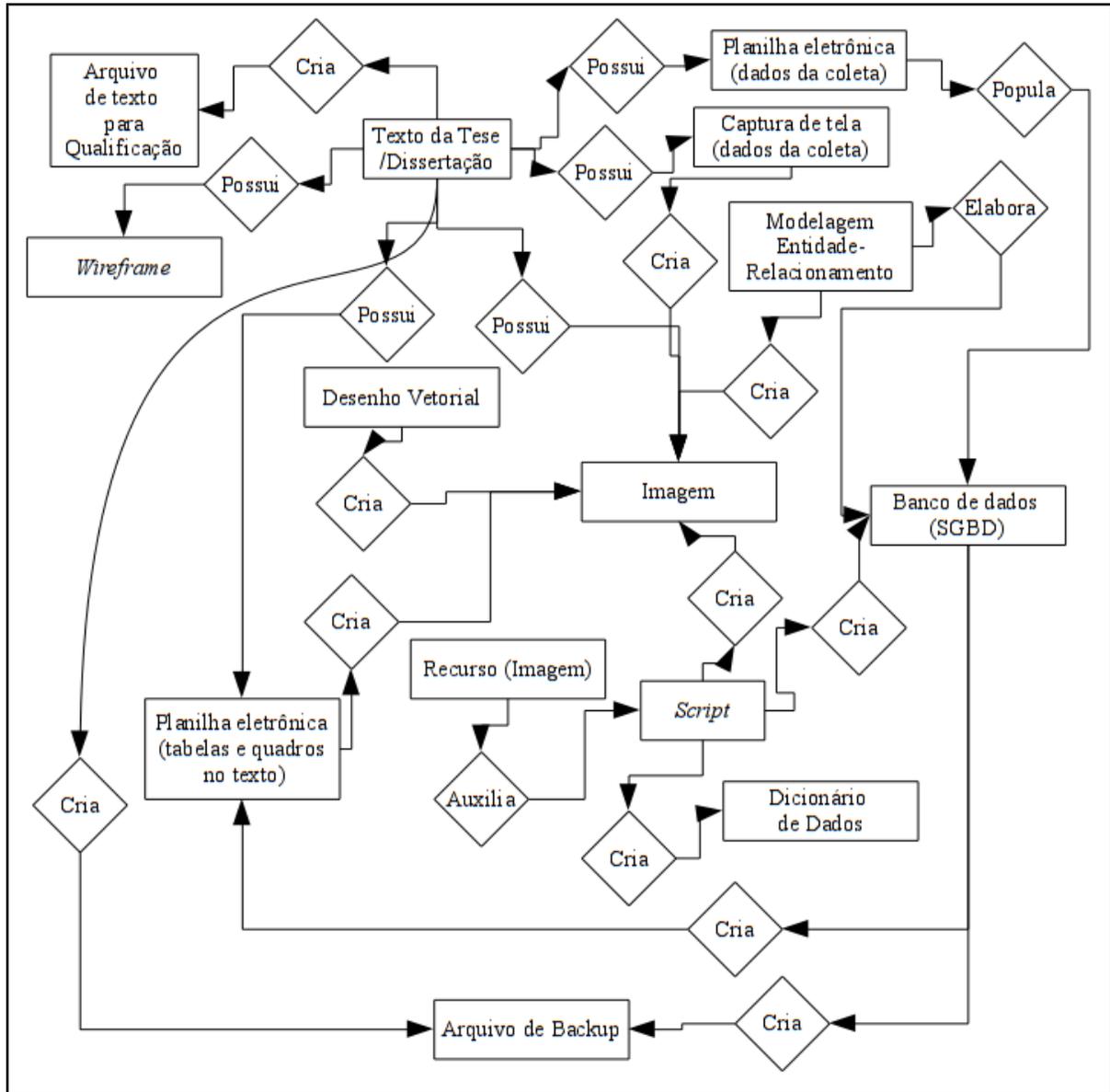
#### 4 RESULTADOS

A Ilustração 1 apresenta as relações entre os tipos de documentos identificados no estudo de caso. É composto por 19 (dezenove) relacionamentos entre 14 (catorze) tipos de documentos. Os tipos de documentos são:

- (a) “Texto da Tese/Dissertação”: texto da publicação (*e-print*), incluindo versões geradas conforme o desenvolvimento da pesquisa;
- (b) “Arquivo de texto para Qualificação”, versão do “Texto da Tese/Dissertação” exportada para o envio aos membros da banca e demais participantes, que podem ter formato de arquivo distinto do original (como o *PDF* ou o *e-Pub*);
- (c) “Arquivo de Backup”: arquivos com a finalidade de garantir ao pesquisador a possibilidade de recuperação dos dados do banco de dados e do texto da publicação;

- (d) “Captura de tela (dados da coleta)”: capturas de tela realizadas na coleta de dados da tese/dissertação;
- (e) “Planilha eletrônica (dados da coleta)”: planilhas eletrônicas utilizadas na coleta de dados da tese/dissertação;
- (f) “Planilha eletrônica (tabelas e quadros no texto)”: planilhas eletrônicas utilizadas para a estruturação de tabelas e quadros da tese/dissertação;
- (g) “Desenho Vetorial”: arquivos contendo desenhos vetoriais elaborados pelo pesquisador ou por terceiros, utilizados na tese/dissertação;
- (h) “Recurso (Imagem)”: recursos imagéticos utilizados para a elaboração de diagramas, gráficos e imagens. Por exemplo, o uso de ícone sob licença *Creative Commons* para a geração de um diagrama da tese/dissertação;
- (i) “Imagem”: arquivos de imagens utilizados no texto da tese/dissertação, gerados pelo pesquisador ou por terceiros;
- (j) “*Script*”: algoritmos elaborados para auxiliar a geração de imagens, gráficos e diagramas, complementares ao texto da tese/dissertação; para a geração automatizada do Dicionário de dados, e; para a exportação da estrutura e dos dados armazenados no SGBD;
- (k) “Modelagem Entidade-Relacionamento”: arquivos utilizados para elaborar a modelagem do banco de dados (tabelas, atributos, tipos de dados, relacionamentos), bem como diagramas e outros artefatos, elaborado para a análise dos dados da tese/dissertação;
- (l) “Dicionário de Dados”: documentos contendo as tabelas que explicitam os as colunas das tabelas do banco de dados, elaborado para a análise dos dados da tese/dissertação; “Banco de dados (SGBD)”: arquivos dos bancos de dados criados para a análise de dados da tese/dissertação, que integram o SGBD;
- (m) “*Wireframe*”: esboços interligados utilizados na tese/dissertação que foram gerados por aplicativos ou web sites.

**Ilustração 1** – Relações entre tipos de documentos no estudo de caso



Fonte: Elaborado pelos autores.

A partir da explicitação do conjunto de documentos que compõem a tese do estudo de caso, os tipos de documentos identificados (e seus relacionamentos), associados aos requisitos e as estruturas necessárias à Publicação Ampliada nos estudos analisados, o Quadro 4 apresenta os requisitos nas fases de coleta, armazenamento e recuperação para a implementação das funcionalidades de uma Publicação Ampliada.

**Quadro 4** – Análise de requisitos para cada tipo de documento, dividido nas fases de CVD

Tipo de Documento	Fases		
	Coleta	Armazenamento	Recuperação
Texto da Tese/ Dissertação	<ul style="list-style-type: none"> <li>– Identificação de elementos de preservação, como a data do envio e a data da última modificação;</li> <li>– Identificação de elementos sobre a autoria da tese/dissertação, como informações sobre o autor e o orientador;</li> <li>– Permitir ao pesquisador a inserção, a alteração e a exclusão de versões do texto, com a possibilidade de recuperação de versões anteriores;</li> <li>– Permitir o vínculo com outros tipos de documento.</li> </ul>	<ul style="list-style-type: none"> <li>– Persistência de elementos de preservação na forma de metadados;</li> <li>– Persistência de elementos de autoria na forma de metadados;</li> <li>– Geração, persistência e vinculação de identificador <i>URI</i> para cada versão do objeto;</li> <li>– Persistência dos relacionamentos de cada versão com seus objetos.</li> </ul>	<ul style="list-style-type: none"> <li>– Disponibilizar objeto em sítio web;</li> <li>– Permitir o acesso direto ao objeto pelo identificador (<i>URI</i>);</li> <li>– Disponibilizar os metadados de autoria e de preservação, legíveis para humanos e máquinas;</li> <li>– Possibilitar a vinculação a objetos em outros contextos;</li> <li>– Diferenciar por identificadores cada versão do objeto no momento do acesso;</li> <li>– Disponibilizar os relacionamentos do objeto com outros objetos;</li> <li>– Possibilitar a realização de comentários e avaliações do conteúdo do objeto.</li> </ul>
Arquivo de texto para Qualificação	<ul style="list-style-type: none"> <li>– Identificação de elementos de preservação, como a data do envio e a data da qualificação;</li> <li>– Identificação de elementos sobre a autoria da tese/dissertação, como informações sobre o autor, orientador e membros da banca;</li> <li>– Permitir a conversão para formatos de arquivo apropriados ao compartilhamento aos pares, como o <i>PDF/A</i> ou <i>e-Pub</i>;</li> <li>– Permitir o vínculo com outros tipos de documento.</li> </ul>	<ul style="list-style-type: none"> <li>– Persistência de elementos de preservação na forma de metadados;</li> <li>– Persistência de elementos de autoria na forma de metadados;</li> <li>– Geração, persistência e vinculação de identificador <i>URI</i> para o objeto;</li> <li>– Persistência dos relacionamentos do texto de qualificação com seus objetos.</li> </ul>	<ul style="list-style-type: none"> <li>– Disponibilizar objeto em sítio web;</li> <li>– Permitir o acesso direto ao objeto pelo identificador (<i>URI</i>);</li> <li>– Disponibilizar os metadados de autoria e de preservação, legíveis para humanos e máquinas;</li> <li>– Possibilitar a vinculação a objetos em outros contextos;</li> <li>– Disponibilizar os relacionamentos do objeto com outros objetos;</li> <li>– Possibilitar a realização de comentários e avaliações do conteúdo do objeto.</li> </ul>
Arquivo de Backup	<ul style="list-style-type: none"> <li>– Identificação de elementos de preservação, como a data da geração do arquivo;</li> <li>– Permitir o vínculo com outros tipos de documento.</li> </ul>	<ul style="list-style-type: none"> <li>– Persistência de elementos de preservação na forma de metadados;</li> <li>– Persistência dos relacionamentos de cada versão com seus objetos.</li> </ul>	<ul style="list-style-type: none"> <li>– Permitir o acesso direto ao objeto pelo identificador (<i>URI</i>);</li> <li>– Disponibilizar os metadados de autoria e de preservação, legíveis para humanos e máquinas;</li> <li>– Diferenciar por identificadores cada versão do objeto no momento do acesso;</li> <li>– Disponibilizar os relacionamentos do objeto com outros objetos.</li> </ul>
Captura de tela (dados da coleta)	<ul style="list-style-type: none"> <li>– Identificação de elementos de preservação, como a data da captura ou a data de coleta e informações da origem;</li> <li>– Permitir o vínculo com ou-</li> </ul>	<ul style="list-style-type: none"> <li>– Persistência de elementos de preservação na forma de metadados;</li> <li>– Persistência dos relacionamentos de cada versão com</li> </ul>	<ul style="list-style-type: none"> <li>– Disponibilizar objeto em sítio web;</li> <li>– Permitir o acesso direto ao objeto pelo identificador (<i>URI</i>);</li> <li>– Disponibilizar os metadados de autoria e de preservação, legíveis para</li> </ul>

Tipo de Documento	Fases		
	Coleta	Armazenamento	Recuperação
	tros tipos de documento.	seus objetos.	humanos e máquinas; – Possibilitar a vinculação a objetos em outros contextos; – Disponibilizar os relacionamentos do objeto com outros objetos; – Possibilitar a realização de comentários e avaliações do conteúdo do objeto.
Planilha eletrônica (dados da coleta)	– Identificação de elementos de preservação, como a data da coleta e informações sobre pesquisadores e participantes; – Permitir o vínculo com outros tipos de documento.	– Persistência de elementos de preservação na forma de metadados; – Persistência de elementos de autoria na forma de metadados; – Geração, persistência e vinculação de identificador <i>URI</i> para cada versão do objeto; – Persistência dos relacionamentos de cada versão com seus objetos.	– Disponibilizar objeto em sítio web; – Permitir o acesso direto ao objeto pelo identificador ( <i>URI</i> ); – Disponibilizar os metadados de autoria e de preservação, legíveis para humanos e máquinas; – Possibilitar a vinculação a objetos em outros contextos; – Disponibilizar os relacionamentos do objeto com outros objetos; – Possibilitar a realização de comentários e avaliações do conteúdo do objeto.
Planilha eletrônica (tabelas e quadros no texto)	– Identificação de elementos de preservação, como a data da geração do arquivo; – Permitir o vínculo com outros tipos de documento.	– Persistência de elementos de preservação na forma de metadados; – Persistência de elementos de autoria na forma de metadados; – Geração, persistência e vinculação de identificador <i>URI</i> para cada versão do objeto; – Persistência dos relacionamentos de cada versão com seus objetos.	– Disponibilizar objeto em sítio web; Permitir o acesso direto ao objeto pelo identificador ( <i>URI</i> ); – Disponibilizar os metadados de autoria e de preservação, legíveis para humanos e máquinas; – Possibilitar a vinculação a objetos em outros contextos; – Disponibilizar os relacionamentos do objeto com outros objetos; – Possibilitar a realização de comentários e avaliações do conteúdo do objeto.
Desenho Vetorial	– Identificação de elementos de preservação, como a data da geração do arquivo ou a data de coleta e informações da origem; – Permitir ao pesquisador a inserção, a alteração e a exclusão de versões do desenho vetorial, com a possibilidade de recuperação de versões anteriores; – Permitir o vínculo com outros tipos de documento.	– Persistência de elementos de preservação na forma de metadados; – Persistência de elementos de autoria na forma de metadados; – Geração, persistência e vinculação de identificador <i>URI</i> para cada versão do objeto; – Persistência dos relacionamentos de cada versão com seus objetos.	– Disponibilizar objeto em sítio web; Permitir o acesso direto ao objeto pelo identificador ( <i>URI</i> ); – Disponibilizar os metadados de autoria e de preservação, legíveis para humanos e máquinas; – Possibilitar a vinculação a objetos em outros contextos; – Disponibilizar os relacionamentos do objeto com outros objetos; Possibilitar a realização de comentários e avaliações do conteúdo do objeto.

Tipo de Documento	Fases		
	Coleta	Armazenamento	Recuperação
Recurso (Imagem)	<ul style="list-style-type: none"> <li>– Identificação de elementos de preservação, como a data de coleta, licenças e informações da origem;</li> <li>– Permitir o vínculo com outros tipos de documento.</li> </ul>	<ul style="list-style-type: none"> <li>– Persistência de elementos de preservação na forma de metadados;</li> <li>– Persistência de elementos de autoria na forma de metadados;</li> <li>– Geração, persistência e vinculação de identificador <i>URI</i> para cada versão do objeto;</li> <li>– Persistência dos relacionamentos de cada versão com seus objetos.</li> </ul>	<ul style="list-style-type: none"> <li>– Disponibilizar objeto em sítio web;</li> <li>– Permitir o acesso direto ao objeto pelo identificador (<i>URI</i>);</li> <li>– Disponibilizar os metadados de autoria e de preservação, legíveis para humanos e máquinas;</li> <li>– Possibilitar a vinculação a objetos em outros contextos;</li> <li>– Disponibilizar os relacionamentos do objeto com outros objetos;</li> <li>– Possibilitar a realização de comentários e avaliações do conteúdo do objeto.</li> </ul>
Imagem	<ul style="list-style-type: none"> <li>– Identificação de elementos de preservação, como a data da geração do arquivo ou a data de coleta e informações da origem;</li> <li>– Permitir ao pesquisador a inserção, a alteração e a exclusão de versões da imagem, com a possibilidade de recuperação de versões anteriores;</li> <li>– Permitir o vínculo com outros tipos de documento.</li> </ul>	<ul style="list-style-type: none"> <li>– Persistência de elementos de preservação na forma de metadados;</li> <li>– Persistência de elementos de autoria na forma de metadados;</li> <li>– Geração, persistência e vinculação de identificador <i>URI</i> para cada versão do objeto;</li> <li>– Persistência dos relacionamentos de cada versão com seus objetos.</li> </ul>	<ul style="list-style-type: none"> <li>– Disponibilizar objeto em sítio web;</li> <li>– Permitir o acesso direto ao objeto pelo identificador (<i>URI</i>);</li> <li>– Disponibilizar os metadados de autoria e de preservação, legíveis para humanos e máquinas;</li> <li>– Possibilitar a vinculação a objetos em outros contextos;</li> <li>– Disponibilizar os relacionamentos do objeto com outros objetos;</li> <li>– Possibilitar a realização de comentários e avaliações do conteúdo do objeto.</li> </ul>
<i>Script</i>	<ul style="list-style-type: none"> <li>– Identificação de elementos de preservação, como a data da geração do arquivo e a data da última modificação;</li> <li>– Permitir ao pesquisador a inserção, a alteração e a exclusão de versões do <i>script</i>, com a possibilidade de recuperação de versões anteriores;</li> <li>– Permitir o vínculo com outros tipos de documento.</li> </ul>	<ul style="list-style-type: none"> <li>– Persistência de elementos de preservação na forma de metadados;</li> <li>– Persistência de elementos de autoria na forma de metadados;</li> <li>– Geração, persistência e vinculação de identificador <i>URI</i> para cada versão do objeto;</li> <li>– Persistência dos relacionamentos de cada versão com seus objetos.</li> </ul>	<ul style="list-style-type: none"> <li>– Disponibilizar objeto em sítio web;</li> <li>– Permitir o acesso direto ao objeto pelo identificador (<i>URI</i>);</li> <li>– Disponibilizar os metadados de autoria e de preservação, legíveis para humanos e máquinas;</li> <li>– Possibilitar a vinculação a objetos em outros contextos;</li> <li>– Disponibilizar os relacionamentos do objeto com outros objetos;</li> <li>– Possibilitar a realização de comentários e avaliações do conteúdo do objeto.</li> </ul>
Modelagem Entidade-Relacionamento	<ul style="list-style-type: none"> <li>– Identificação de elementos de preservação, como a data da geração do arquivo e a data da última modificação;</li> <li>– Permitir ao pesquisador a inserção, a alteração e a exclusão de versões da Modelagem Entidade-Relacionamento, com a possibilidade de recupere-</li> </ul>	<ul style="list-style-type: none"> <li>– Persistência de elementos de preservação na forma de metadados;</li> <li>– Persistência de elementos de autoria na forma de metadados;</li> <li>– Geração, persistência e vinculação de identificador <i>URI</i> para cada versão do ob-</li> </ul>	<ul style="list-style-type: none"> <li>– Disponibilizar objeto em sítio web;</li> <li>– Permitir o acesso direto ao objeto pelo identificador (<i>URI</i>);</li> <li>– Disponibilizar os metadados de autoria e de preservação, legíveis para humanos e máquinas;</li> <li>– Possibilitar a vinculação a objetos em outros contextos;</li> <li>– Disponibilizar os relacionamentos</li> </ul>

Tipo de Documento	Fases		
	Coleta	Armazenamento	Recuperação
	<p>ração de versões anteriores;</p> <ul style="list-style-type: none"> <li>– Permitir o vínculo com outros tipos de documento.</li> </ul>	<p>do objeto;</p> <ul style="list-style-type: none"> <li>– Persistência dos relacionamentos de cada versão com seus objetos.</li> </ul>	<p>do objeto com outros objetos;</p> <ul style="list-style-type: none"> <li>– Possibilitar a realização de comentários e avaliações do conteúdo do objeto.</li> </ul>
Dicionário de Dados	<ul style="list-style-type: none"> <li>– Identificação de elementos de preservação, como a data da geração do arquivo e a data da última modificação;</li> <li>– Permitir ao pesquisador a inserção, a alteração e a exclusão de versões do Dicionário de Dados, com a possibilidade de recuperação de versões anteriores;</li> <li>– Permitir o vínculo com outros tipos de documento.</li> </ul>	<ul style="list-style-type: none"> <li>– Persistência de elementos de preservação na forma de metadados;</li> <li>– Persistência de elementos de autoria na forma de metadados;</li> <li>– Geração, persistência e vinculação de identificador <i>URI</i> para cada versão do objeto;</li> <li>– Persistência dos relacionamentos de cada versão com seus objetos.</li> </ul>	<ul style="list-style-type: none"> <li>– Disponibilizar objeto em sítio web;</li> <li>– Permitir o acesso direto ao objeto pelo identificador (<i>URI</i>);</li> <li>– Disponibilizar os metadados de autoria e de preservação, legíveis para humanos e máquinas;</li> <li>– Possibilitar a vinculação a objetos em outros contextos;</li> <li>– Disponibilizar os relacionamentos do objeto com outros objetos;</li> <li>– Possibilitar a realização de comentários e avaliações do conteúdo do objeto.</li> </ul>
Banco de dados (SGBD)	<ul style="list-style-type: none"> <li>– Identificação de elementos de preservação, como a data da geração do banco de dados e a data da última modificação;</li> <li>– Permitir ao pesquisador a recuperação de versões anteriores;</li> <li>– Permitir o vínculo com outros tipos de documento.</li> </ul>	<ul style="list-style-type: none"> <li>– Persistência de elementos de preservação na forma de metadados;</li> <li>– Persistência de elementos de autoria na forma de metadados;</li> <li>– Geração, persistência e vinculação de identificador <i>URI</i> para cada versão do objeto;</li> <li>– Persistência dos relacionamentos de cada versão com seus objetos.</li> </ul>	<ul style="list-style-type: none"> <li>– Disponibilizar objeto em sítio web;</li> <li>– Permitir o acesso direto ao objeto pelo identificador (<i>URI</i>);</li> <li>– Disponibilizar os metadados de autoria e de preservação, legíveis para humanos e máquinas;</li> <li>– Possibilitar a vinculação a objetos em outros contextos;</li> <li>– Disponibilizar os relacionamentos do objeto com outros objetos;</li> <li>– Possibilitar a realização de comentários e avaliações do conteúdo do objeto.</li> </ul>
<i>Wireframe</i>	<ul style="list-style-type: none"> <li>– Identificação de elementos de preservação, como a data da geração do arquivo e informações da origem;</li> <li>– Permitir ao pesquisador a inserção, a alteração e a exclusão de versões do <i>wireframe</i>, com a possibilidade de recuperação de versões anteriores;</li> <li>– Permitir o vínculo com outros tipos de documento.</li> </ul>	<ul style="list-style-type: none"> <li>– Persistência de elementos de preservação na forma de metadados;</li> <li>– Persistência de elementos de autoria na forma de metadados;</li> <li>– Geração, persistência e vinculação de identificador <i>URI</i> para cada versão do objeto;</li> <li>– Persistência dos relacionamentos de cada versão com seus objetos.</li> </ul>	<ul style="list-style-type: none"> <li>– Disponibilizar objeto em sítio web;</li> <li>– Permitir o acesso direto ao objeto pelo identificador (<i>URI</i>);</li> <li>– Disponibilizar os metadados de autoria e de preservação, legíveis para humanos e máquinas;</li> <li>– Possibilitar a vinculação a objetos em outros contextos;</li> <li>– Disponibilizar os relacionamentos do objeto com outros objetos;</li> <li>– Possibilitar a realização de comentários e avaliações do conteúdo do objeto.</li> </ul>

Fonte: Elaborado pelos autores.

Foram identificados 170 (cento e setenta) requisitos para os 14 (catorze) tipos de documentos identificados, sendo: 39 (trinta e nove) relacionados a fase de coleta; 52 (cinquenta e

dois) relacionados a fase de armazenamento, e; 79 (setenta e nove) relacionados a fase de recuperação. Os requisitos destas fases podem variar de acordo com os tipos de documentos que compõem a Publicação Ampliada, ou seja, para cada tipo de documento há um conjunto próprio de requisitos para cada fase.

## 5 CONSIDERAÇÕES FINAIS

Com a explicitação da análise de requisitos das fases de coleta, armazenamento e recuperação, começam a surgir mais preocupações na forma em que as publicações ampliadas serão apresentadas, no momento da implementação de funcionalidades, como a necessidade de explicitação de aspectos de privacidade.

Entende-se que o quadro conceitual de elementos necessários para a instanciação de Publicações Ampliadas (Quadro 3) identificados a partir dos estudos selecionados ainda carece de maior detalhamento e explicitação de ações, o que demonstra o estágio inicial que este contexto se encontra e que, portanto, merece um número maior de pesquisas e iniciativas concretas de implementação.

Como trabalhos futuros, propomos um maior aprofundamento das lacunas de quais são os requisitos, as funcionalidades e os objetos que farão a composição de um sistema de informação apropriado a um repositório de publicações ampliadas.

## ***Enhanced Publication aspects of Thesis and Dissertations***

### **ABSTRACT**

The aim of this research is to study aspects involved in enhanced publication context, especially for cases of thesis and dissertations and several documents that became part of the new range of elements of these results, providing subsidies for the development of a conceptual basis to support proposals to structures of gathering, storage and retrieval of this new set of documents that compose thesis and dissertations, from concepts already established in Information Science. The methodology used was the methodological triangulation, exploratory and descriptive, consisting of (i) identification of theoretical framework through literature about enhanced publication in Portuguese, through studies available in Google Scholar and SciELO databases, and Google Search; (ii) requirements analysis, with the use of phases and objectives of the Data Life Cycle for Information Science, and; (iii) case study, from the sets of documents from a thesis on process of Doctorate and Master Degree Program in Information Science from São Paulo State University. The results of this study are presented in the form of requirements analysis, with use of graphs for the development of relations between documents of the enhanced publication; and requirements in

phases of gathering, storage and retrieval for the implementation of features. It was concluded that with explanation of the requirements analysis of collection, storage and retrieval stages, begin to emerge more concerns in the way which enhanced publications will be presented at the time of implementation of features, such as requirements to clarify aspects of privacy, and a more detailed explanation and action on this issue, and elements necessary for instantiating enhanced publications.

**Keywords:** Enhanced Publication. Data. Dissertation. Thesis. Information Science.

---

## REFERÊNCIAS

ARAYA, E. R. M. **Comunicação científica: agregação, compartilhamento e reúso de elementos informacionais**. Tese—Marília: Universidade Estadual Paulista, 2014.

BIGGS, N.; LLOYD, E. K.; WILSON, R. J. **Graph theory, 1736-1936**. Oxford [Oxfordshire] ; New York: Clarendon Press, 1986.

BULLINGER, H.-J. et al. **Berlin declaration on open access to knowledge in the sciences and humanities** Open Access to Knowledge in the Sciences and Humanities, 2003. Disponível em: <[http://old.bsu.edu.ru/\\_files/Nauka/CoopOnLine/Berlin\(eng\).doc](http://old.bsu.edu.ru/_files/Nauka/CoopOnLine/Berlin(eng).doc)>. Acesso em: 8 set. 2016.

CASTELLS, M. **A sociedade em rede**. São Paulo: Paz e Terra, 2008.

DUARTE, T. A possibilidade da investigação a 3: reflexões sobre triangulação (metodológica). **CIES e-working paper**, v. 60, p. 24, 2009.

HEY, A. J. G. (ED.). **The fourth paradigm: data-intensive scientific discovery**. Redmond, Washington: Microsoft Research, 2009.

HEY, T. The Big Idea: The Next Scientific Revolution. **Harvard Business Review**, v. 88, n. 11, p. 56–63, nov. 2010.

HEY, T. The Fourth Paradigm – Data-Intensive Scientific Discovery. In: KURBANOĞLU, S. et al. (Eds.). **E-Science and Information Management**. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012. v. 317p. 1–1.

MACHADO, D. R. **Dados de pesquisa em repositório institucional: o caso do Edinburgh DataShare**. Dissertação—Porto Alegre: Universidade Federal do Rio Grande do Sul, 2015.

MUCHERONI, M. L.; SILVA, F. J. M. DA; PALETTA, C. F. **Entre a publicação ampliada e a multimodalidade**. Anais do XVI Encontro Nacional de Pesquisa em Ciência da Informação – XVI ENANCIB 2015. **Anais...**: 16. In: XVI ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO – XVI ENANCIB 2015. João Pessoa: ANCIB, 2015. Disponível em: <<http://www.ufpb.br/evento/lti/ocs/index.php/enancib2015/enancib2015>>. Acesso em: 12 nov. 2015.

RODRIGUES, F. DE A.; SANT'ANA, R. C. G. **Restrições tecnológicas e de acesso a dados disponíveis sobre destinos de repasses financeiros federais para a saúde pública em ambientes informacionais digitais**. Anais do XIII Encontro Nacional de Pesquisa em Ciência da Informação – XIII ENANCIB 2012. **Anais...**: 13. In: XIII ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO – XIII ENANCIB 2012. Rio de Janeiro: ANCIB, 28 out. 2012. Disponível em: <<http://repositorios.questoesemrede.uff.br/repositorios/handle/123456789/2054>>. Acesso em: 12 nov. 2015.

SALES, L. F. **Integração semântica de publicações científicas e dados de pesquisa: proposta de modelo de publicação ampliada para a área de Ciências Nucleares**. Tese—Rio de Janeiro: Universidade Federal do Rio de Janeiro, 2014.

SALES, L. F.; SAYÃO, L. F. **Modelo triádico de relações: um protótipo de modelagem conceitual para a área nuclear**. Anais do XIII Encontro Nacional de Pesquisa em Ciência da Informação – XIII ENANCIB 2012. **Anais...**: 13. In: XIII ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO – XIII ENANCIB 2012. Rio de Janeiro: ANCIB, 28 out. 2012. Disponível em: <<http://www.enancib2012.icict.fiocruz.br/>>. Acesso em: 8 out. 2016.

SALES, L. F.; SAYÃO, L. F. Enhanced publication: a new model of scientific publication for the nuclear area. **IEN – Progress Report 2013-2014 n.2**, v. 2, n. 2015, p. 1, 2015.

SALES, L. F.; SAYÃO, L. F.; SOUZA, R. F. **Modelo triádico de relações: um protótipo de modelagem conceitual para a área nuclear**. Anais do XIV Encontro Nacional de Pesquisa em Ciência da Informação – XIV ENANCIB 2013. **Anais...**: 14. In: XIV ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO – XIV ENANCIB 2013. Florianópolis: ANCIB, 2013. Disponível em: <<http://enancib2013.ufsc.br/index.php/enancib2013/XIVenancib>>. Acesso em: 12 nov. 2015.

SALES, L. F.; SOUZA, R. F.; SAYÃO, L. F. **Publicação ampliada: um novo modelo de publicação científica voltada para os desafios de uma ciência orientada por dados**. Anais do XV Encontro Nacional de Pesquisa em Ciência da Informação – XV ENANCIB 2014. **Anais...**: 15. In: XV ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO – XV ENANCIB 2014. Belo Horizonte: ANCIB, 2014. Disponível em: <<http://ridi.ibict.br/handle/123456789/824>>. Acesso em: 8 out. 2016.

SANT'ANA, R. C. G. **Ciclo de Vida dos Dados e o papel da Ciência da Informação**. Anais do XIV Encontro Nacional de Pesquisa em Ciência da Informação. **Anais...** In: XIV ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO. Florianópolis: ANCIB, 2013. Disponível em: <<http://enancib2013.ufsc.br/index.php/enancib2013/XIVenancib/paper/viewFile/284/319>>. Acesso em: 15 ago. 2015.

SAYÃO, L. F.; SALES, L. F. Dados abertos de pesquisa: ampliando o conceito de acesso livre. **Revista Eletrônica de Comunicação, Informação & Inovação em Saúde**, v. 8, n. 2, p. 76–92, 2014.

WILSON, R. J. **Introduction to graph theory**. 4. ed., [Nachdr.] ed. Harlow: Prentice Hall, 1996.

WIVES, Leandro K. **Indexação de documentos textuais**. 1997. 19 f. Trabalho Monográfico - Disciplina de Sistemas de Banco de Dados (Programa de Pós-Graduação em Ciência da Computação)- Instituto de Informática, Universidade Federal do Rio Grande do Sul, Porto Alegre, 1997.