

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL  
FACULDADE DE BIBLIOTECONOMIA E COMUNICAÇÃO  
PROGRAMA DE PÓS-GRADUAÇÃO EM COMUNICAÇÃO E INFORMAÇÃO

Daiane Barrili dos Santos

**PUBLICAÇÕES AMPLIADAS: ASPECTOS DA INTEGRAÇÃO DE DADOS DE  
PESQUISA**

Porto Alegre

2017

Daiane Barrili dos Santos

**PUBLICAÇÕES AMPLIADAS: ASPECTOS DA INTEGRAÇÃO DE DADOS DE  
PESQUISA**

Dissertação apresentada ao Programa de Pós-Graduação em Comunicação e Informação da Faculdade de Biblioteconomia e Comunicação da Universidade Federal do Rio Grande do Sul, como requisito parcial para a obtenção do grau de Mestre em Comunicação e Informação.

Orientador: Prof. Dr. Moisés Rockembach.

Porto Alegre

2017

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL

Reitor: Prof. Dr. Rui Vicente Oppermann

Vice-Reitora: Prof<sup>a</sup>.Dr<sup>a</sup>. Jane Fraga Tutikian

FACULDADE DE BIBLIOTECONOMIA E COMUNICAÇÃO

Direção: Prof<sup>a</sup>. Dr<sup>a</sup>. Karla Maria Müller

Vice-Direção: Prof<sup>a</sup>. Dr<sup>a</sup>. Ilza Maria Tourinho Girardi

PROGRAMA DE PÓS-GRADUAÇÃO EM COMUNICAÇÃO E INFORMAÇÃO

Coordenador: Prof. Dr. Rudimar Baldissera

Coordenadora substituta: Prof<sup>a</sup>. Dr<sup>a</sup>. Nísia Martins do Rosário

CIP - Catalogação na Publicação

Santos, Daiane Barrili dos

Publicações ampliadas: aspectos da integração de dados de pesquisa / Daiane Barrili dos Santos. -- 2017.

200 f.

Orientador: Moisés Rockembach.

Dissertação (Mestrado) -- Universidade Federal do Rio Grande do Sul, Faculdade de Biblioteconomia e Comunicação, Programa de Pós-Graduação em Comunicação e Informação, Porto Alegre, BR-RS, 2017.

1. Publicações ampliadas. 2. Dados de Pesquisa. 3. Repositórios de dados de pesquisa. 4. Aspectos da integração de dados de pesquisa. 5. Requisitos para publicações ampliadas. I. Rockembach, Moisés, orient. II. Título.

Elaborada pelo Sistema de Geração Automática de Ficha Catalográfica da UFRGS com os dados fornecidos pelo(a) autor(a).

PPGCOM/UFRGS

Rua Ramiro Barcelos 2705

CEP: 90.035-007 – Porto Alegre, RS

Telefone: (51) 3308-5116

E-mail: fabico@ufrgs.br

Daiane Barrili dos Santos

**PUBLICAÇÕES AMPLIADAS: ASPECTOS DA INTEGRAÇÃO DE DADOS DE  
PESQUISA**

Dissertação apresentada ao Programa de Pós-Graduação em Comunicação e Informação da Faculdade de Biblioteconomia e Comunicação da Universidade Federal do Rio Grande do Sul, como requisito parcial para a obtenção do grau de Mestre em Comunicação e Informação.

Orientador: Prof. Dr. Moisés Rockembach

Aprovada em: \_\_\_\_\_ de janeiro de 2018.

**BANCA EXAMINADORA:**

---

Profa. Dra. Ana Maria Mielniczuk de Moura  
Universidade Federal do Rio Grande do Sul

---

Profa. Dra. Sônia Elisa Caregnato  
Universidade Federal do Rio Grande do Sul

---

Prof. Dr. Rafael Port da Rocha  
Universidade Federal do Rio Grande do Sul

---

Profa. Dra. Samile Andréa de Souza Vanz (Suplente)  
Universidade Federal do Rio Grande do Sul

## **DEDICATÓRIA**

À Nossa Sra. do Trabalho que me proporcionou esta conquista

## AGRADECIMENTOS

À Universidade Federal do Rio Grande do Sul, pela oportunidade de ampliar meus conhecimentos na área da Ciência da Informação.

Ao meu orientador Prof. Dr. Moisés Rockembach, pelas sugestões, contribuições e parceria.

À profa. Dra. Ana Maria Mielniczuk de Moura, por ter me incentivado em dar continuidade aos estudos, pelo conhecimento transmitido e especialmente pelas palavras de motivação.

À profa. Dra. Sônia Elisa Caregnato, pelas oportunidades de aprendizado que me foram dadas durante as aulas, durante o estágio em docência e coorientações, e principalmente pela especial atenção nos momentos em que mais precisei de palavras de incentivo.

À Caterina Groposo Pavão, por todas as contribuições desde a época da graduação, por estar sempre presente nos momentos de aflição no desenvolvimento das pesquisas e especialmente pelas palavras de consolo, carinho e pela a amizade.

Aos professores Sônia Elisa Caregnato e Rafael Port da Rocha, pelas contribuições na qualificação do projeto de pesquisa.

Aos colegas da PPGCOM e ex-colegas do curso de biblioteconomia, pela disponibilidade para me auxiliar com as dúvidas durante essa etapa e pela atenção que me foi dada.

À minha família e aos meus amigos, pela paciência, por demonstrarem tanto orgulho pelas minhas conquistas e por me darem força para nunca desistir dos meus sonhos.

*A ciência nunca resolve um problema sem criar pelo  
menos outros dez.  
(George Bernard Shaw).*

## RESUMO

Esta pesquisa teve como objetivo principal propor um panorama de integração entre repositórios de dados de pesquisa e publicações científicas baseado no modelo de publicações ampliadas. Do ponto de vista de sua natureza, é básica. Segundo a abordagem é considerada qualitativa, e caracteriza-se por ser uma pesquisa exploratória e bibliográfica. O modelo de análise deste estudo foi composto por três fases. Na primeira fase de análise, foram coletadas literaturas específicas da área por meio da revisão sistemática e, a partir das literaturas, foram extraídos os requisitos indicados e sugeridos para um modelo de publicação ampliada. Na segunda fase de análise, foram selecionados os repositórios de dados de pesquisa através do diretório re3data.org e averiguadas as informações disponibilizadas por meio dos filtros que demonstrassem a utilização de ligações baseadas no modelo de publicação ampliada. Na terceira fase de análise, foi realizada a inspeção das interfaces e identificados os elementos que cada repositório contempla para posteriormente iniciar as análises, observações e comparações dos elementos utilizados nos repositórios com os requisitos estabelecidos por meio da revisão sistemática. Percebeu-se na literatura que há uma tentativa de adicionar uma estrutura utilizando-se requisitos relacionados a metadados, relacionamentos semânticos, entre outros. Verificou-se que parte dessa estrutura indicada na literatura já encontra-se nas interfaces de alguns repositórios identificados no diretório re3data.org. No entanto estes repositórios que contemplam parte da estrutura indicada foram caracterizados como uma minoria. Desta forma, verificou-se uma necessidade crescente, de uma infraestrutura compatível com as publicações ampliadas, tornando-se de extrema importância a utilização de requisitos para as interfaces dos repositórios de dados de pesquisa. O cruzamento de informações realizado neste estudo permitiu um maior entendimento do quanto é importante o uso de cada um dos requisitos estipulados, como também foi possível compreender as consequências caso estes requisitos não estejam contemplados nas interfaces de repositórios. Os repositórios de dados de pesquisa deveriam fornecer em suas interfaces as ligações e o acesso às respectivas publicações científicas dos dados registrados para que fosse possível compreender os estudos realizados a partir destes dados. Foi comprovado que esse processo não ocorre na grande parte dos repositórios, o que dificulta a compreensão sobre a origem e o contexto em que foram utilizados esses dados. Percebe-se que os repositórios que não estavam de acordo com os principais requisitos estabelecidos apresentavam características de publicações convencionais como, por exemplo, documentos em PDF, textos em HTML e que não estavam de acordo com padrões da web semântica. Esta proposta de um panorama de



integração de dados de pesquisa baseado em publicações ampliadas visa sugerir uma interface que contemple todos os requisitos especificados nesta pesquisa para que se tenha uma infraestrutura compatível com as publicações ampliadas. A partir da investigação e do aprofundamento sobre o tema, foi possível ampliar o conhecimento e detectar os elementos apropriados para propor um panorama de integração entre um repositório de dados de pesquisas e publicações científicas baseando-se em um modelo de publicação ampliada, como também destacar as vantagens da utilização deste modelo. Considera-se que um conjunto de dados disponíveis em um repositório digital, que contemple os requisitos, melhora a recuperação da informação e, conseqüentemente, aumenta a visibilidade da publicação e de autores.

**Palavras-chave:** Publicações ampliadas. Dados de pesquisa. Repositórios de dados de pesquisa. Aspectos da integração de dados de pesquisa. Requisitos para publicações ampliadas.

## ABSTRACT

This research had the main objective to propose an integration landscape between research data repositories and scientific publications based on the model for enhanced publications. From its nature's point of view, it is basic. According to the approach it is considered a qualitative, is characterized as an exploratory, bibliographic research and a case study as well. The analysis model of this study was constituted of three phases. In first phase of the analyses, specific literatures were collected through the systematic review and from literatures, in addition the suggested requirements and indicated for a model for enhanced publications were extracted, in addition the suggested requirements and indicated for an enhanced publication model were extract. In second phase of the analyses, research data repositories were select through the re3data.org directory and were verified the information provided through the filters which demonstrated the application of links based on the model for enhanced publications. In the third and last phase of the analyses, an inspection of interfaces were realized, and also elements that each repository comprehend were identified to subsequently initiate the analyses, observations and comparisions of the elements used on repositories with the established requirements through systematic review. It was perceived in the literature, that there are an attempt to add a structure by using related requirements to metadata, semantic relationship, among others. Also verified that part of this estructure indicated on literature is available on interfaces of some identified repositories on re3data.org. Although these repositories which comprehend part of the indicated structure were characterized as a minority. Therefore, there has been a growing need for an infrastructure compatible with the enhanced publications, aware of the importance for the application of requirements to the interfaces of research data repositories. The crossover of informations realized in this study allowed a greater understanding about the importance to use each one of the stipulated requirements, as well as was possible to understand the consequences in case of these requirements are not be included in the repositories interfaces. Research data repositories should provide on their interfaces the links and access to their respective scientific publications of the recorded data in order to be able to understand the studies carried out from this data. It has been proven that this process does not occur in most repositories, which difficults to understand the origin and context in which these data were used. It was noticed that repositories not in agreement with the main presented established requirements characteristics of conventional publications, as for example PDF documents, HTML texts and

which was not in accordance to patterns for semantic web. This proposal for a research data integration model based on enhanced publications aims to suggest an interface that fulfills all the specified requirements in this research to obtain an infrastructure compatible to enhanced publications. From the investigation and deepening the theme, it was possible to increase the knowledge and to detect the appropriate elements to propose an integration model between a researches data repository and scientific publications based on a model for enhanced publication as well as highlighting the advantages of applying this model. Thus, it is concluded that a set of available data into a digital repository that comprehend the requirements improves the information recovery and, therefore, increases the visibility of publication and authors.

**Keywords:** Enhanced publications. Search data. Research data repositories. Aspects of integrating search data. Requirements for enhanced publications.

## LISTA DE FIGURAS

Figura 1 –	Perda de dados de pesquisa.....	26
Figura 2 –	Objetivos da Ciência Aberta.....	28
Figura 3 –	Modelo de publicação ampliada.....	41
Figura 4 –	Descrição da publicação científica indexada no repositório OpenAIRE.....	42
Figura 5 –	Descrição dos dados de pesquisa indexados no repositório OpenAIRE.....	43
Figura 6 –	Utilização da publicação ampliada em periódicos científicos.....	43
Figura 7 –	Download dos dados de pesquisa.....	44
Figura 8 –	Dados de pesquisa – Informações suplementares do artigo publicado.....	44
Figura 9 –	Interface principal do diretório re3data.org.....	57
Figura 10 –	Opções de filtros do diretório re3data.org.....	58
Figura 11 –	Exemplo de repositório com recursos web identificados por URI.....	81
Figura 12 –	Exemplo de repositório com identificador persistente.....	82
Figura 13 –	Exemplo de repositório com publicações vinculadas.....	83
Figura 14 –	Exemplo de repositório com registro e controle de versões da publicação..	84
Figura 15 –	Exemplo de repositório com registro das propriedades básicas da publicação.....	85
Figura 16 –	Exemplo de repositório com registro das propriedades básicas dos recursos adicionados.....	85
Figura 17 –	Exemplo de repositório que possui vocabulário controlado e padronizado.	86
Figura 18 –	Exemplo de descrição dos metadados utilizados pelo repositório.....	87
Figura 19 –	Exemplo de repositório com descrições de conjunto de dados e objetos digitais.....	88
Figura 20 –	Exemplo de repositório com descrições de conjunto de dados e objetos digitais.....	88
Figura 21 –	Exemplo de repositório com registro de autoria.....	89
Figura 22 –	Informações sobre políticas de preservação de dados.....	90

Figura 23 – Registro/descrição das relações entre os recursos que fazem parte da publicação.....	91
Figura 24 – Exemplo de repositório em plataforma open access.....	92
Figura 25 – Exemplo de repositório baseado em padrões da web semântica – RDF.....	93
Figura 26 – Estrutura da publicação ampliada para integração de dados de pesquisa.....	96

## LISTA DE QUADROS

Quadro 1 –	Elementos para citação de dados de pesquisa.....	28
Quadro 2 –	Elementos essenciais de um Plano de Gestão de Dados.....	34
Quadro 3 –	Elementos opcionais de um Plano de Gestão de Dados.....	36
Quadro 4 –	Fontes dos dados de pesquisa.....	37
Quadro 5 –	Formas dos dados de pesquisa.....	37
Quadro 6 –	Categorias para análise.....	56
Quadro 7 –	Listagem de repositórios de dados de pesquisa nacionais ( <i>Brazil</i> ).....	59
Quadro 8 –	Quantidade de repositórios de dados de pesquisa quanto ao tipo de instituição.....	63
Quadro 9 –	Quantidade de repositórios de dados de pesquisa e os respectivos tipos de repositórios.....	67
Quadro 10 –	Quantidade de repositórios de dados de pesquisa que utilizam ou não as publicações ampliadas.....	67

## LISTA DE GRÁFICOS

Gráfico 1 –	Quantidade de itens recuperados na Scopus e seus respectivos anos de publicação.....	54
Gráfico 2 –	Quantidade de repositórios de dados de pesquisa por tipos de conteúdo.	58
Gráfico 3 –	Quantidade de repositórios de dados de pesquisa por países.....	59
Gráfico 4 –	Quantidade de repositórios de dados de pesquisa e os identificadores digitais para identificação de autores.....	60
Gráfico 5 –	Quantidade de repositórios de dados de pesquisa que utilizam padrões de programação API ( <i>Application Programming Interface</i> ).....	61
Gráfico 6 –	Quantidade de repositórios de dados de pesquisa quanto às restrições de acesso a dados ( <i>DATA ACCESS RESTRICTIONS</i> ).....	61
Gráfico 7 –	Quantidade de repositórios de dados de pesquisa e a forma de carregamento de dados ( <i>DATA UPLOAD</i> ).....	62
Gráfico 8 –	Quantidade de repositórios de dados de pesquisa e as restrições para o carregamento de dados ( <i>DATA UPLOAD RESTRICTIONS</i> ).....	62
Gráfico 9 –	Tipo de responsabilidade da instituição e seu respectivo número de repositórios relacionados ( <i>INSTITUTION RESPONSIBILITY TYPE</i> ).....	63
Gráfico 10–	Quantidade de repositórios de dados de pesquisa quanto aos padrões de metadados ( <i>METADATA STANDARDS</i> ).....	64
Gráfico 11–	Quantidade de repositórios de dados de pesquisa quanto ao uso de sistemas de identificação de objetos digitais ( <i>PID SYSTEMS</i> ).....	65
Gráfico 12–	Quantidade de repositórios de dados de pesquisa quanto aos idiomas....	65
Gráfico 13–	Quantidade de repositórios de dados de pesquisa e os respectivos softwares utilizados.....	66
Gráfico 14–	Total de repositórios com P.A recuperados no diretório.....	68
Gráfico 15–	Classificação das áreas mais abrangentes.....	70
Gráfico 16–	Conteúdos mais abrangentes entre os 136 repositórios.....	71
Gráfico 17–	Identificadores digitais de autoria mais utilizados.....	72
Gráfico 18–	Padrões de programação utilizados entre os repositórios selecionados...	73
Gráfico 19–	Restrições de acesso aos dados.....	74

Gráfico 20–	Tipos de Instituições vinculada aos repositórios.....	74
Gráfico 21–	Padrões de metadados utilizados entre os repositórios selecionados.....	75
Gráfico 22–	Sistema de identificação de objetos digitais.....	76
Gráfico 23–	Idiomas detectados entre os repositórios selecionados.....	77
Gráfico 24–	Softwares utilizados entre os 136 repositórios selecionados.....	78
Gráfico 25–	Tipos de repositórios identificados.....	78
Gráfico 26–	Países relacionados aos 136 repositórios selecionados.....	79



## SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO.....</b>	<b>19</b>
1.1	OBJETIVOS.....	21
<b>1.1.1</b>	<b>Objetivo geral.....</b>	<b>22</b>
<b>1.1.2</b>	<b>Objetivos específicos.....</b>	<b>22</b>
1.2	JUSTIFICATIVA.....	22
<b>2</b>	<b>REVISÃO DE LITERATURA.....</b>	<b>24</b>
2.1	DADOS DE PESQUISA.....	24
2.2	REPOSITÓRIOS DIGITAIS.....	29
2.3	REPOSITÓRIOS DE DADOS DE PESQUISA.....	30
2.4	GESTÃO DE DADOS DE PESQUISA.....	31
2.5	PUBLICAÇÃO AMPLIADA.....	38
<b>3</b>	<b>METODOLOGIA.....</b>	<b>47</b>
3.1	ABORDAGEM E TIPO DE PESQUISA.....	47
3.2	INSTRUMENTOS E COLETA DE DADOS.....	48
3.3	ANÁLISE DOS DADOS.....	51
3.4	LIMITAÇÕES DO ESTUDO.....	52
<b>4</b>	<b>ANÁLISE E APRESENTAÇÃO DOS RESULTADOS.....</b>	<b>53</b>
4.1	PRIMEIRA FASE DA ANÁLISE.....	53
4.2	SEGUNDA FASE DA ANÁLISE.....	57
4.3	TERCEIRA FASE DA ANÁLISE.....	68
<b>4.3.1</b>	<b>Recursos disponíveis nos repositórios selecionados.....</b>	<b>70</b>
4.3.1.1	Áreas mais abrangentes.....	70
4.3.1.2	Conteúdos mais abrangentes.....	71
4.3.1.3	Identificadores digitais de autoria.....	71
4.3.1.4	Padrões de programação – API.....	72
4.3.1.5	Restrições de acesso aos dados.....	73
4.3.1.6	Tipos de instituições vinculadas.....	74
4.3.1.7	Padrões de metadados.....	75
4.3.1.8	Sistema de identificação de objetos digitais.....	75
4.3.1.9	Idiomas detectados.....	76
4.3.1.10	Softwares mais utilizados.....	77

4.3.1.11	Tipos de repositórios.....	78
4.3.1.12	Países destacados.....	79
<b>4.3.2</b>	<b>Requisitos de publicação ampliada definidos na pesquisa.....</b>	<b>79</b>
4.3.2.1	Requisito 1 – Recursos web identificados por URI.....	80
4.3.2.2	Requisito 2 – Identificador persistente para cada objeto.....	81
4.3.2.3	Requisito 3 – Vincular objetos digitais à publicação.....	82
4.3.2.4	Requisito 4 – Registro e controle de versões da publicação e das partes constituintes.....	83
4.3.2.5	Requisito 5 – Registro das propriedades básicas da publicação e dos recursos que são adicionados.....	84
4.3.2.6	Requisito 6 – Vocabulário controlado e padronizado.....	86
4.3.2.7	Requisito 7 – Atributos – descrição.....	87
4.3.2.8	Requisito 8 – Registro da autoria da publicação e dos objetos digitais.....	89
4.3.2.9	Requisito 9 – Políticas de preservação.....	89
4.3.2.10	Requisito 10 – Registrar as relações entre os recursos web que fazem parte da publicação ampliada.....	90
4.3.2.11	Requisito 11 – Documentos baseados no modelo OAI-ORE .....	91
4.3.2.12	Requisito 12 – Disponibilizar em plataforma <i>open access</i> .....	92
4.3.2.13	Requisito 13 – Estar de acordo com os padrões da Web Semântica.....	93
4.3.2.14	Requisito 14 – Recomenda-se que os objetos digitais estejam conectados por relacionamentos semânticos.....	94
4.3.2.15	Requisito 15 – Recomenda-se que os objetos digitais devam ser fornecidos através de upload.....	94
4.3.2.16	Requisito 16 – Cada arquivo de dados tem seus próprios metadados, incluindo informações de objeto.....	94
4.4	PROPOSTA DE UM PANORAMA PARA PUBLICAÇÕES AMPLIADAS BASEADO EM REQUISITOS.....	95
<b>5</b>	<b>CONSIDERAÇÕES FINAIS.....</b>	<b>98</b>
	<b>REFERÊNCIAS.....</b>	<b>104</b>
	<b>APÊNDICE A – Protocolo de revisão sistemática.....</b>	<b>111</b>
	<b>APÊNDICE B – Formulário de condução da revisão.....</b>	<b>114</b>
	<b>APÊNDICE C – Formulário de seleção de estudos.....</b>	<b>118</b>
	<b>APÊNDICE D – Formulário de extração de dados.....</b>	<b>122</b>
	<b>APÊNDICE E – Checklist dos requisitos e de validação /Inspeção.....</b>	<b>162</b>

<b>APÊNDICE F – Listagem dos repositórios que possuem publicação ampliada.....</b>	<b>166</b>
<b>APÊNDICE G – Listagem de repositórios de dados de pesquisa e áreas abrangentes.....</b>	<b>179</b>
<b>APÊNDICE H – Formulário de identificação de recursos constatados nos repositórios com publicação ampliada.....</b>	<b>185</b>

## 1 INTRODUÇÃO

A Comunicação Científica vem sofrendo transformações com o advento da web e os pesquisadores têm se apropriado das tecnologias oferecidas para disseminar as produções científicas através de diversas alternativas de publicações. Da mesma forma que há o interesse em dar o acesso livre às publicações científicas por meio de repositórios digitais, atualmente também existe vantagem em dar acesso livre aos dados provenientes da pesquisa científica.

Na comunidade científica utilizam-se diferentes tipos de publicações para a disseminação de trabalhos científicos, os quais podem estar disponíveis em ambientes digitais. A divulgação de trabalhos científicos e o desenvolvimento da ciência estão diretamente ligados à melhoria contínua de tecnologias da informação. Deste modo, o aperfeiçoamento dessas tecnologias possibilita o processo de disponibilização de materiais digitais por meio de ferramentas que possuem a finalidade de organizar a informação e de disseminar o uso do conhecimento.

Essas inovações da comunicação científica também estão relacionadas à ciência aberta (*Open Science*). Conforme Sayão e Sales (2014), a ideia de ciência aberta tem muitos significados, porém o mais convincente é o que reconhece que o conhecimento científico é um patrimônio da humanidade e, portanto, precisa estar disponível livremente para que as pessoas possam usá-lo, reusá-lo e distribuí-lo. Por esta razão, a ciência aberta pode proporcionar maior dinamismo às atividades que envolvem a ciência, a tecnologia e a inovação. Albagli, Maciel e Abdo (2015) enfatizam que o movimento pela ciência aberta se insere nas novas formas de produção colaborativa, interativa e compartilhada da informação, do conhecimento e da cultura.

Nesse contexto, a internet torna-se uma aliada da pesquisa e da comunicação científica devido ao seu potencial de ampliar o alcance de resultados de pesquisas e publicações científicas. Com a popularização da internet, surgiram diversas melhorias nos canais de comunicação científica como, por exemplo, as publicações on-line que revolucionaram as possibilidades de descoberta e de acesso a recursos informacionais acadêmicos e científicos aumentando, desta forma, a amplitude e a facilidade de acesso a todo tipo de material, até mesmo conjuntos de dados científicos (SALES; SAYÃO; SOUZA, 2013).

Na Web está disponível uma imensurável quantidade destes dados. Na tentativa de estruturar e tornar acessível estes dados surge a Web Semântica que “[...] é o uso de uma linguagem formal para representar o conhecimento conceitual de um determinado domínio de conhecimento presente na Web de forma estruturada e compreensível.” (SOUZA, 2013, p. 1).

Com o aprimoramento de tecnologias que possibilitam a armazenagem e o compartilhamento de informações, iniciativas de acesso aberto vêm alterando as formas de disseminação e uso dessas informações. O movimento de Acesso Aberto (*Open Access*) ou acesso livre, que surgiu com a Declaração de Budapest<sup>1</sup> em 2002, só se deu a partir da contribuição da Iniciativa dos Arquivos Abertos – OAI (*Open Archives Initiative*). No entanto, a Declaração de Berlin (2003) ampliou o propósito do que se entendia por acesso aberto. Conforme a Declaração de Berlin (2003), as contribuições de acesso livre incluem resultados de pesquisas científicas originais, ou seja, os dados não processados e as fontes originais. Desta forma, considera-se que o conceito de acesso aberto não está somente relacionado a publicações acadêmicas tradicionais como também pode abranger coleções de dados de pesquisa.

As tecnologias digitais têm transformado a forma como esses conjuntos de dados eram caracterizados. Após o término de pesquisas e publicações, os dados eram muitas vezes descartados ou armazenados em servidores sem a devida gestão (SAYÃO; SALES, 2014). Nesse seguimento, outros movimentos e iniciativas surgem nos fazendo repensar os modelos atuais de disseminação de resultados e a compreender a otimização de pesquisas. Alguns modelos de publicações científicas como, por exemplo, a publicação ampliada, está sendo desenvolvida para que haja um maior compartilhamento e reuso de informações de forma que estas informações se relacionem.

Nesta pesquisa será apresentada uma nova forma de publicação – Publicação Ampliada, do inglês *Enhanced Publication* – que depende das possibilidades de ligações da web. Esse tipo de publicação proporciona a colaboração entre pesquisadores, a conectividade, o compartilhamento de dados e as convergências de linguagens. Nesse sentido, acredita-se que um conjunto de dados abertos de pesquisa, disponíveis em um repositório digital, possibilita o reuso, o aumento do índice de citações de autores, potencializa a troca de conhecimento e evita que futuras pesquisas sobre um determinado tema se repitam, evitando desperdícios financeiros e de tempo.

Iniciativas referentes à abertura de dados de pesquisa são apoiadas por uma ampla gama de políticas. Agências internacionais de financiamento à pesquisa científica como, por exemplo, a *Organization for Economic Cooperation and Development* (OECD) e revistas científicas como a *Science and Nature* recomendam que os dados de pesquisas científicas

---

<sup>1</sup> Movimento que definiu pela primeira vez a iniciativa de acesso livre. (BUDAPEST, [20-?]).

estejam em acesso aberto ao público. Tornar dados disponíveis publicamente já faz parte da cultura em algumas áreas como a física, a astronomia, a biologia entre outras.

Desta forma há o interesse das instituições em aderir às recomendações e desenvolver planos de gestão de dados de pesquisa que estabeleçam de forma clara todas as ações e requisitos para a abertura dos dados de pesquisa em seus repositórios. Percebe-se que essas inovações já estão presentes em outros países e que no Brasil ainda é bastante recente. Devido a este motivo, organizações que possuem a incumbência de promover o desenvolvimento tecnológico e apoiar pesquisas no Brasil como, por exemplo, a Rede Nacional de Ensino e Pesquisa (RNP), estão criando projetos relacionados ao desenvolvimento de repositórios de dados de pesquisa para a publicação e partilha desses dados. A RNP organiza a integração global e a colaboração apoiada em tecnologias de informação e comunicação para a geração do conhecimento e a excelência da educação e da pesquisa (RNP, [201-]). Desde 2002 é uma Organização Social (OS) vinculada ao Ministério da Ciência, Tecnologia, Inovações e Comunicações (MCTIC) e mantida por esse em conjunto com os ministérios da Educação (MEC), Cultura (MinC), Saúde (MS) e Defesa (MD), que participam do Programa Interministerial da RNP (PI-RNP). Grandes instituições de ensino estão se empenhando para promover o acesso aberto aos dados das pesquisas científicas e desenvolvendo diretrizes para a implementação de repositórios e planos de gestão de dados.

Entre as ações envolvidas no desenvolvimento de um repositório de dados de pesquisa, está a publicação ampliada com o propósito de conectar os dados das pesquisas com suas respectivas publicações. Observou-se que temas relacionados aos dados abertos ainda carecem de estudos que apresentem de forma clara como esses dados podem ser usados, reutilizados e compartilhados necessitando rever o modelo atual de disseminação de resultados de pesquisas. Desta maneira, apresenta-se como problema de pesquisa a seguinte questão:

Como desenvolver um panorama de integração entre repositórios de dados de pesquisa e publicações científicas baseado no modelo de publicações ampliadas?

## 1.1 OBJETIVOS

Os objetivos deste projeto estão divididos em objetivo geral e objetivos específicos da seguinte forma:

### 1.1.1 Objetivo geral

Propor um panorama da integração entre repositórios de dados de pesquisa e publicações científicas baseado no modelo de publicações ampliadas.

### 1.1.2 Objetivos específicos

São objetivos específicos desta investigação:

- a) levantar na literatura os principais requisitos que compõem a estrutura de uma publicação ampliada;
- b) identificar os repositórios de dados de pesquisa nacionais e internacionais e analisar, dentre os repositórios selecionados, os que interligam os dados de pesquisas às publicações científicas;
- c) investigar as características dos repositórios indexados no diretório re3data.org.
- d) verificar nas interfaces dos repositórios que interligam seus dados, se estes contemplam os principais requisitos de uma publicação ampliada;
- e) elaborar um panorama de integração entre repositórios de dados de pesquisa e publicações científicas baseado no modelo de publicações ampliadas.

## 1.2 JUSTIFICATIVA

Neste estudo pretende-se propor um panorama de integração entre repositórios de dados de pesquisa e publicações científicas baseado no modelo de publicações ampliadas e apontar as vantagens da utilização. A escolha do tema se deu após verificar que revistas como a *Science and Nature* e organizações que apoiam pesquisas e promovem a inovação como, por exemplo, a Comissão Europeia (*European Commission*, que abrange os países da União Europeia); a Fundação Nacional da Ciência (*National Science Foundation*) e o Instituto Nacional de Padrões e Tecnologia (*National Institute of Standards and Technology*), que caracterizam-se como agências governamentais dos Estados Unidos, estão recomendando que os dados provenientes das pesquisas científicas sejam submetidos a repositórios de dados e ligados às suas respectivas publicações para que haja maior transparência em relação ao estudo realizado. O Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT), ao

lançar o Manifesto de Acesso Aberto a Dados da Pesquisa Brasileira para Ciência Cidadã, recomenda diretrizes para que editores de revistas e periódicos científicos incluam em suas políticas a necessidade do registro de dados de pesquisa em repositórios abertos (IBICT, 2016).

Observa-se que alguns periódicos, que aderiram ao modelo de publicação ampliada, fornecem de forma clara o acesso a artigos e aos dados provenientes da pesquisa que foram registrados em um repositório de dados de pesquisa. No entanto, da mesma maneira que os periódicos, os repositórios de dados de pesquisa deveriam fornecer em suas interfaces as ligações e o acesso às respectivas publicações científicas dos dados registrados para que seja possível compreender os estudos realizados a partir destes dados. Supõe-se que esse processo não ocorra na grande parte dos repositórios, o que dificulta a compreensão sobre a origem e o contexto em que foram utilizados os dados registrados. Pretende-se realizar este estudo, pois utilizando ligações e integrando repositórios de dados de pesquisa às publicações científicas, acredita-se que o reuso, o compartilhamento e a visibilidade dos dados de pesquisa conseqüentemente poderão aumentar como também facilitarão o acesso à publicação científica relacionada para uma maior transparência das metodologias utilizadas e desenvolvidas durante a pesquisa. Por estes motivos, esta pesquisa terá como objetivo propor um panorama de integração baseado na publicação ampliada para repositórios de dados de pesquisa.

Em sua tese, Sales (2015) propôs um modelo de publicação científica para a área de Ciências Nucleares, em que os dados são ligados às publicações acadêmicas por meio de relações semânticas sistematizadas em taxonomia construída. O pressuposto da tese de Sales (2015, p. 27) “[...] é que a publicação ampliada será mais rica semanticamente, se a criação de elos entre a pesquisa e seus dados for feita por meio de relações conceituais explícitas”. Neste sentido, o estudo visou investigar a identificação de uma forma semântica para ampliar a publicação científica integrando-a aos dados que a fundamentam, ou seja, o estudo buscou responder de que forma é possível garantir semântica e interoperabilidade a esses novos modelos de publicação. Em contrapartida ao estudo de Sales, esta pesquisa tem como propósito a identificação dos principais requisitos para desenvolver uma infraestrutura compatível com os padrões associados às publicações ampliadas.

Portanto, a presente pesquisa se justifica pela necessidade de interligar os dados de pesquisas às suas respectivas publicações de acordo com os requisitos estabelecidos e identificados na literatura, associados aos padrões das publicações ampliadas.



## 2 REVISÃO DE LITERATURA

Esta seção tem por objetivo apresentar o contexto de estudo e definir os conceitos que darão sustentação à pesquisa. Serão abordados temas como: dados de pesquisa, repositórios digitais, repositórios de dados de pesquisa, gestão de dados de pesquisa, primórdios da publicação científica e publicação ampliada.

### 2.1 DADOS DE PESQUISA

Os dados de pesquisa são fundamentais para as pesquisas científicas e o acesso a eles de forma permanente é um desafio para a ciência e para a comunidade científica. São caracterizados como os registros científicos que embasam os resultados de uma pesquisa.

A OECD define dados de pesquisa como “[...] registros factuais utilizados como fontes primárias da pesquisa científica e que são comumente aceitos na comunidade científica para validar os resultados de pesquisa.”<sup>2</sup> (OECD, 2007, p. 13, tradução nossa).

Na literatura desenvolvida por alguns autores, como Araya (2014) e Seeber (2008), são encontrados outros termos equivalentes a dados de pesquisa, alguns citados neste estudo: informações suplementares (*supplementary information*); dados primários e elementos informacionais. Araya (2014) afirma que esses elementos informacionais justificam, apoiam, ilustram ou esclarecem as reivindicações científicas que são expostas em uma publicação.

Sayão e Sales (2015, p. 10) também afirmam que “[...] os dados são gerados para diferentes propósitos, por diferentes comunidades acadêmicas e científicas e por meio de diferentes processos [...]”. Há uma diversidade de tipos de dados como, por exemplo, números, imagens, textos, vídeos, áudio, software, algoritmos, equações, animações, modelos, simulações entre outros. Os dados podem assumir várias formas e são classificados, conforme Sayão e Sales (2015), de acordo com a natureza, origem e fases da pesquisa:

- a) classificação segundo a origem: dados observacionais são dados obtidos por meio de observação direta; dados computacionais são resultados da execução de modelos computacionais ou de simulações; dados experimentais são provenientes de situações controladas em laboratórios;

---

<sup>2</sup> “[...] factual records (numerical scores, textual records, images and sounds) used as primary sources for scientific research, and that are commonly accepted in the scientific community as necessary to validate research findings.”

- b) classificação segundo a natureza: números; imagens; vídeos; áudio; software; algoritmos; equações; animações; simulações;
- c) classificação segundo a fase da pesquisa: dados brutos, crus ou preliminares são os dados provenientes dos instrumentos científicos; dados derivados são resultados do processamento ou combinação de dados brutos e de outros dados; dados canônicos ou dados referenciais são coleções de dados consolidados e arquivados em centros de dados.

Alguns autores como, por exemplo, Seeber (2008) elucidam o quanto importante é a citação de informações suplementares – *supplementary information*, para os estudos referentes ao fator de impacto, índice- h e rankings de pesquisadores e revistas. Seeber (2008) afirma que o fato de não haver a praxe de submeter, disponibilizar e citar os dados de pesquisa pode afetar os cálculos de classificações em estudos sobre rankings. Portanto, entende-se que a citação de dados de pesquisa também contribui para maximizar a visibilidade de pesquisadores e de revistas científicas.

Nesse contexto, com a aceleração da ciência, um novo conceito surge: a *e-Science*, com o objetivo de compartilhamento de dados científicos primários. A *e-Science* pode ser entendida como uma infraestrutura que permite que cientistas e pesquisadores possam ter acesso a dados científicos primários. *e-Science* é uma forma abreviada para o conjunto de ferramentas e tecnologias necessárias para suportar ciência colaborativa em rede. Toda a infraestrutura de *e-Science* é destinada a capacitar cientistas para fazer suas pesquisas de maneiras mais rápidas, melhores e diferentes<sup>3</sup> (HEY; HEY, 2006, tradução nossa).

Além do uso de grande volume de dados, a *e-Science* abrange a pesquisa colaborativa e o uso de recursos compartilhados. Considera-se que os dados primários “[...] são todos os materiais registrados durante uma pesquisa, reconhecido pela comunidade científica e que servem para certificar os resultados alcançados.”<sup>4</sup> (TORRES-SALINAS; ROBINSON-GARCÍA; CABEZAS-CLAVIJO, 2012, p. 175, tradução nossa).

O compartilhamento de dados científicos primários permite o reuso que, segundo Medeiros e Caregnato (2012, p. 319), pode ser visto como “[...] a utilização de dados já utilizados, ou seja, são dados utilizados mais de uma vez, mas com intenções originais”. Com

---

<sup>3</sup> *e-Science is shorthand for the set of tools and technologies required to support collaborative, networked science. The entire e-Science infrastructure is intended to empower scientists to do their research in faster, better and different ways.*

<sup>4</sup> *todo aquel material que ha sido registrado durante la investigación, reconocido por la comunidad científica y que sirve para certificar los resultados de la investigación que se realiza.*

o reuso, além de agilizar processos científicos, é possível economizar recursos financeiros. Muitos repositórios estão sendo implementados para armazenar os dados de pesquisas, no entanto, é preciso encontrar maneiras para maximizar a reutilização e o compartilhamento desses dados e garantir a confiabilidade e qualidade para que seja possível a reutilização.

O reconhecimento do potencial informacional dos dados digitais, distribuídos em rede de computadores, para a ciência contemporânea transforma a visão que caracterizava dados de pesquisa, registrados em mídia impressa ou mesmo em formatos digitais, como simples subprodutos dos processos de pesquisa. Nesse contexto, os dados eram considerados somente na sua configuração final, sem considerar os seus ciclos de vida, versões e linhagens e, via de regra, eram descartados ou armazenados em mídias ou servidores sem a devida gestão quando os projetos eram concluídos. Quase sempre eram tragados silenciosamente pelo tempo: pela obsolescência tecnológica, pela efemeridade dos formatos e pela fragilidade das mídias digitais (SAYÃO; SALES, 2014, p. 78).

Na figura 1 é possível visualizar como os dados de pesquisa estão sujeitos a perdas.

Figura 1 – Perda de dados de pesquisa



Fonte: Sayão e Sales (2015).

Portanto, a inexistência de uma integração entre recursos tecnológicos pode influenciar no reuso de dados científicos e no extravio desses dados que poderiam ser reutilizados em novas pesquisas. Conforme Medeiros (2015, p. 25), “Poder realizar testes em

dados previamente coletados e anotados com metadados, isto é, reutilizar o que já foi mapeado, possibilita maiores e melhores progressos na ciência, além de oportunizar que recursos financeiros e humanos sejam poupados”. Isto é, ao invés de realizar novas coletas, pode-se reduzir esforços focando na análise e no tratamento dos dados já compartilhados, o que permitirá um progresso mais acelerado da ciência. O reuso e compartilhamento de dados, segundo Berman, Wilkinson e Wood (2014), permite descobrir conexões no que estava antes desconectado.

O avanço da web e as novas tecnologias possibilitaram a inovação, o compartilhamento de informações e a disponibilização de um grande volume de dados. Desta maneira, permitiu a interoperabilidade entre sistemas de informação, impulsionando a cooperação técnico-científica.

O compartilhamento de dados científicos permite que as possibilidades de colaboração e a interação entre os cientistas aumentem, o que contribui para o avanço da ciência. A colaboração científica é o trabalho conjunto de pesquisadores para atingir o objetivo comum de produzir novos conhecimentos científicos (KATZ; MARTIN, 1997). Em busca do reconhecimento, cientistas priorizam o trabalho com pesquisadores que possuem uma boa reputação na área. Maltrás Barba (2003) considera o reconhecimento como uma das principais razões para a colaboração, deste modo, o reconhecimento pode vir também ao se trabalhar juntamente com um pesquisador de renome.

A iniciativa da *e-Science*, além do objetivo de compartilhamento de dados científicos primários, como já foi dito, pretende impulsionar o progresso da ciência fazendo que resultados de pesquisas se tornem públicos.

Neste contexto, a possibilidade de compartilhamento de dados e a capacitação dos cientistas no uso de novas tecnologias têm influenciado as práticas colaborativas. A Ciência Aberta e um ambiente baseado em redes promove uma maior transparência das metodologias utilizadas e desenvolvidas durante uma pesquisa. Existem quatro objetivos fundamentais da ciência aberta: a transparência na metodologia experimental, observação e coleta de dados; a disponibilidade pública e reutilização de dados científicos; a acessibilidade pública e transparência da comunicação científica e o uso de ferramentas baseadas na web para facilitar a colaboração científica (GEZELTER, 2009). Os quatro objetivos são demonstrados na figura 2.

Figura 2 – Objetivos da Ciência Aberta



Fonte: Costa e Braga (2016, p. 82).

Neste âmbito, a divulgação de dados de pesquisa de forma aberta (Dados Abertos de Pesquisa) chamará a atenção para a necessidade de realizar estudos relacionados às redes de citação de dados de pesquisa, o que também envolve e possibilita os estudos de redes de colaboração e utilização de métodos de Análise de Redes Sociais. Em relação à citação de dados de pesquisa, o *DataCite* recomenda os seguintes elementos para uma citação de dados, conforme consta no quadro 1.

Quadro 1 – Elementos para citação de dados de pesquisa

Criadores: Título, Editor / Repositório, Identificador, Ano de Publicação.

**Exemplo:** Loew, A., Bennartz, R., Fell, F., Lattanzio, A., Doutriaux-Boucher, M. e Schulz, Sites de Validação, EUMETSAT, [http://dx.doi.org/10.15770/EUM\\_SEC\\_CLM\\_1001](http://dx.doi.org/10.15770/EUM_SEC_CLM_1001), 2015.

Fonte: Adaptado de *DataCite* (2017).

Os dados digitais distribuídos em redes de computadores em um ambiente *linked data* têm como premissa maximizar a visibilidade das pesquisas desenvolvidas, aumentar o índice de citações de autores, potencializar a troca de conhecimento e possibilitar que futuras pesquisas sejam realizadas com o reuso de dados evitando-se, desta forma, investigações repetidas e os custos a elas inerentes.

Segundo Oliveira, Dias e Padua (2014), o atual cenário da comunicação científica demanda interatividade e convergência de informações entre as publicações eletrônicas, o que

permite, além da geração e do uso da publicação, que ocorra a possibilidade de reuso, o compartilhamento de dados e resultados de pesquisas. Em relação à convergência, Jenkins (2009, p. 29) a define como:

[...] fluxo de conteúdos através de múltiplos suportes midiáticos, à cooperação entre múltiplos mercados midiáticos e ao comportamento migratório dos públicos dos meios de comunicação, que vão a quase qualquer parte em busca das experiências de entretenimento que desejam. Convergência é uma palavra que consegue definir transformações tecnológicas, mercadológicas, culturais e sociais, dependendo de quem está falando e do que imaginam estar falando.

O processo tradicional de publicação da comunicação científica pode ser aperfeiçoado por meio da utilização de tecnologias da internet. Conforme Jenkins, Ford e Green (2013), o ambiente da internet é favorável para a ampliação da interação social e como um espaço de criação colaborativa. A Web e a Internet oferecem deste modo, uma infraestrutura para uma publicação mais dinâmica com a inclusão de visualizações de textos, animações, dados de pesquisa entre outros, de forma integrada ou agregada. Também permite estabelecer relacionamentos entre publicações distintas (MUCHERONI; SILVA; PALETTA, 2015).

A partilha de dados científicos caracteriza-se por ser uma área ainda em consolidação e formação. No Brasil, percebe-se a falta de padronização, infraestruturas e tecnologias que estimulem a disponibilização e compartilhamento de dados.

## 2.2 REPOSITÓRIOS DIGITAIS

Com os avanços da internet, os repositórios digitais surgiram possibilitando que as instituições de pesquisa publicassem suas pesquisas, proporcionando maior visibilidade às produções dos pesquisadores. As iniciativas de acesso aberto incentivam que os resultados de investigações científicas sejam disponibilizados de forma aberta e para que não haja limitações quanto ao acesso. Nesse seguimento, Tomaél e Silva (2007) mencionam que a concepção de repositório está relacionada aos conceitos de acesso aberto (*open access*) ou acesso livre; de arquivos abertos (*open archives*); e de softwares livres (*open source*), conceitos que vêm sendo incrementados de modo significativo com o advento de tecnologias de última geração. Existem diversas definições para os repositórios digitais e uma delas é dada por Camargo e Vidotti (2008) que os descrevem como ambientes informacionais digitais para gerenciamento e controle da produção acadêmica e científica de instituições e/ou comunidades.

Os repositórios digitais oferecem vantagens como acesso irrestrito, interoperabilidade dos dados e preservação da informação em longo prazo podendo ser categorizados como temáticos e institucionais. “Um repositório temático se constitui em um conjunto de trabalhos de pesquisa de uma determinada área do conhecimento, disponibilizados na Internet. Esses repositórios utilizam tecnologias abertas e seguem a filosofia da Iniciativa dos Arquivos Abertos.” (CAFÉ et al., 2003, p.2). No entanto, um “[...] repositório institucional é a reunião de todos os repositórios temáticos hospedados em uma organização. No caso de uma universidade, cada departamento trata de uma área do conhecimento e, portanto, seu repositório temático será específico no assunto deste departamento.” (CAFÉ et al., 2003, p.4). Portanto, seu objetivo é preservar e disponibilizar a produção intelectual da instituição. Para tanto, se faz necessário o compromisso contínuo de garantir que o processo de preservação digital seja a longo prazo. Souza, Bezerra e Costa (2016, p. 122) destacam que “A disponibilização e a disseminação de qualquer recurso no ambiente digital envolvem não somente a inclusão da publicação na Web, mas também questões tecnológicas que garantam a organização, descrição, preservação e acessibilidade aos documentos publicados.” Possuir políticas de preservação digital e um conjunto de requisitos é essencial em relação à gestão do material informacional de um repositório digital porque garante a autenticidade de objetos digitais.

Desta forma, os repositórios preservam, disponibilizam e compartilham documentos em formato digital e são considerados ferramentas que permitem maximizar a visibilidade das pesquisas desenvolvidas, aumentam o índice de citações de autores e potencializam a troca de conhecimento. Para atender as inovações tecnológicas e diferentes comunidades, a tipologia dos repositórios pode alterar, surgindo, por exemplo, os repositórios de dados de pesquisa.

### 2.3 REPOSITÓRIOS DE DADOS DE PESQUISA

Da mesma forma que nos repositórios de publicações científicas, diversas plataformas podem ser utilizadas na implementação de um repositório de dados de pesquisa como, por exemplo, *DSpace* e o *EPrints*, que caracterizam-se por ser ferramentas bastante divulgadas e utilizadas com licença de código aberto.

No entanto, observa-se na literatura, que diversas plataformas disponíveis ainda possuem pouco desenvolvido o item relacionado à descrição dos conjuntos de dados, ou seja, os metadados para descrever dados científicos. Na curadoria digital existem ações para a descrição e representação de um conjunto de dados que garantem metadados relacionados à

produção, direitos de uso e características técnicas desses conjuntos de dados. Há também a descrição ao nível de item e pode incluir características técnicas como, por exemplo, enumerar grandezas, etc.

Referente aos repositórios de dados, pode concluir-se que há muito a fazer tanto no estabelecimento de práticas de descrição adequada como na normalização dos descritores a usar (RODRIGUES; SARAIVA, 2010). Verifica-se que existe um número baixo de repositórios e investigadores que possuem conhecimentos técnicos importantes para a disponibilização dos dados recolhidos de modo que sejam aceitos por outros utilizadores. Segundo Rodrigues e Saraiva (2010), os repositórios institucionais têm sido apontados como uma das infraestruturas passíveis de dar resposta às necessidades de curadoria de dados de pesquisa para as quais não existem outras soluções no imediato.

Nesse sentido, a curadoria de dados científicos poderia ser assumida pelos repositórios institucionais e pelos seus administradores. “No entanto, subsistem dúvidas que as instituições em moldes individuais possuam o interesse, os recursos e os profissionais com as competências necessárias para lidar com esse desafio.” (RODRIGUES; SARAIVA, 2010, p. 34).

Atividades como, por exemplo, o manuseio diário e o fluxo de trabalho de dados de investigação durante a fase ativa do projeto, bem como as práticas que sustentam a preservação em longo prazo devem incluir o planeamento, a documentação de dados, a formatação, o armazenamento e o controle de acesso a esses dados (COURSERA, 2016). Para a divulgação de dados abertos de pesquisa são necessárias algumas ações como o tratamento e o gerenciamento, de modo que o acesso à autenticidade, o reuso e a preservação desses dados sejam garantidos.

Para a elaboração de um modelo baseado nas publicações ampliadas é necessária uma abordagem técnica que envolva alguns elementos como, por exemplo, a gestão de dados de pesquisa, que é essencial para a composição desse novo conceito.

## 2.4 GESTÃO DE DADOS DE PESQUISA

Todas as atividades envolvidas na gestão de dados e na implantação de repositórios digitais que tratam das melhores práticas de digitalização, disponibilização e preservação fazem parte da curadoria digital. Para a divulgação e abertura dos dados de pesquisa em um repositório, deve-se obedecer a padrões mínimos de qualidade, incluindo-se padrões de organização, formatação e catalogação. Conforme Beagrie (2004, p. 7):



O termo ‘curadoria digital’ é cada vez mais usado para as ações necessárias para manter dados de pesquisa em meio digital e outros materiais ao longo de todo o seu ciclo de vida e do tempo para as gerações atuais e futuras. Implícitas nessa definição estão os processos de arquivamento digital e preservação digital, mas também inclui todos os processos necessários para o gerenciamento de dados e a capacidade de agregar valor aos dados para gerar novas fontes de informação e conhecimento.

O Centro de Curadoria Digital (Digital Curation Centre – DCC) é uma instituição reconhecida internacionalmente por fornecer informações e capacitar organizações que desejam armazenar, compartilhar, gerenciar dados de pesquisa. Em algumas das publicações editadas pela DCC, Higgins (2008) descreve a sequência de ações da curadoria digital:

- a) **conceituar:** conceber e planejar a criação do dado (métodos de captura e as opções de armazenamento);
- b) **criar e receber:** criar o dado incluindo o elenco de metadados necessários à sua gestão e compreensão;
- c) **avaliar e selecionar:** avaliar o dado e selecionar o que será objeto dos processos de curadoria e de preservação por longo prazo;
- d) **capturar:** transferir o dado para um arquivo, repositório, centro de dados ou outro custodiante apropriado;
- e) **ação de preservação:** promover ações para assegurar a preservação de longo prazo;
- f) **armazenar:** armazenar o dado de forma segura, mantendo a aderência aos padrões relevantes;
- g) **acessar, usar e reusar:** assegurar que o dado pode ser cotidianamente acessado tanto pela sua comunidade alvo quanto pelos demais usuários interessados no reuso do dado; isso pode ser realizado na forma de informação publicada disponível publicamente; controle de acesso robusto e procedimento de autenticação podem ser aplicados;
- h) **transformar:** criar novos dados a partir do original, por exemplo, pelo processo de migração para diferentes formatos ou pela criação de subconjuntos - realizada por meio de seleção ou formulação de consultas – derivando novos resultados que podem ser publicados;
- i) **eliminar:** eliminar o dado que não foi selecionado para curadoria e preservação de longo prazo de acordo com políticas documentadas, diretrizes e exigências legais;

- j) **reavaliar:** retornar ao dado cujos procedimentos de avaliação foram falhos para nova avaliação e possível seleção;
- k) **migrar:** migrar os dados para um formato diferente; isso pode ser feito no sentido de compatibilizá-lo com o ambiente de armazenamento ou para assegurar a imunidade do dado em relação à obsolescência de hardware e de software.

Aspectos relacionados à representação e ao tratamento da informação também devem estar descritas no plano de gestão. ( HIGGINS, 2008).

O plano de gestão de dados de pesquisa é um documento que possui a finalidade de orientar as etapas de implementação de um repositório. Tem a função de descrever que dados serão coletados, estipular os padrões que serão utilizados e como compartilhar e tornar abertos os dados científicos. Sayão e Sales (2015, p. 15) afirmam que um plano de gestão de dados:

[...] descreve o ciclo de vida de gestão para todos os dados que serão coletados, processados ou gerados por um projeto de pesquisa. De uma forma abreviada, ele se constitui em um documento formal que estabelece um compromisso de como esses dados serão tratados durante todo o desenvolvimento do projeto, e também após a sua conclusão.

Os processos de preservação e curadoria deverão estar presentes no plano e desta forma, entende-se que “Um plano de gerenciamento de dados é um documento formal que descreve o que você fará com seus dados durante e após um projeto de pesquisa.” (ICPSR, 2017).

Agências internacionais de financiamento à pesquisa, além da OECD, podemos citar a *National Science Foundation* (NSF), o *National Institute of Health* (Estados Unidos) e a *The Royal Society* (Reino Unido), também começaram a estabelecer políticas para disponibilizar esses dados de pesquisas. Estas agências estão recomendando que as instituições divulguem junto às publicações os dados provenientes de pesquisas científicas por meio de repositórios digitais. Além disso, recomendam a utilização de planos de gestão de dados e oferecem suporte aos pesquisadores para a construção e desenvolvimento de repositórios de dados de pesquisa. A NSF é uma agência federal independente criada pelo Congresso Americano que promove o progresso da ciência e apoia a pesquisa. Como parte da proposta de financiamentos, a agência exige um plano de gerenciamento de dados<sup>5</sup> (ICPSR, 2017). Da

---

<sup>5</sup> *The National Institutes of Health (NIH) now requires a data sharing plan for large projects, and in 2011 the National Science Foundation (NSF) began to require a data management plan as part of every grant application. (ICPSR, 2017).*

mesma forma que a NSF, acredita-se que a maioria das agências financiadoras irá exigir algum tipo de plano de gerenciamento de dados.

No Brasil, até o momento, não existem agências de fomento que tenham desenvolvido iniciativas para incentivar a disponibilização e compartilhamento de dados de pesquisa. No entanto, o Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT), estende a visão sobre o tema e reconhece a importância do acesso aberto a dados de pesquisa disponibilizando um repositório de dados de pesquisas usando a ferramenta *Dataverse*.

O repositório suporta o compartilhamento de dados de pesquisa com citação de dados persistentes, permitindo que eles sejam reproduzidos. O *Dataverse* é um grande repositório aberto de dados de todas as disciplinas, criado pelo *Institute for Quantitative Social Science* da Universidade de Harvard. O repositório *Dataverse* do IBICT fornece um meio disponível gratuitamente para depositar e descobrir conjuntos de dados específicos arquivados por colaboradores das instituições participantes da rede Cariniana. (IBICT, 2017).

Existem alguns critérios e elementos essenciais que devem constar em um plano de gestão de dados de pesquisa. O consórcio internacional - *Inter-University Consortium for Political and Social Research* (ICPSR) fornece treinamento em acesso a dados, curadoria e métodos de análise. Além disso lidera e participa de iniciativas e atividades que incentivam o uso de dados no ensino de graduação (ICPSR, 2017).

De acordo com alguns exemplos descritos pelo ICPSR (2012, p.10), um plano de gestão de dados deve conter os seguintes elementos, conforme o quadro 2.

Quadro 2 – Elementos essenciais de um Plano de Gestão de Dados

<b>Descrição dos dados:</b>	Fornecer uma breve descrição das informações a serem coletadas, incluindo a natureza, escopo e escala dos dados a serem produzidos. Isso ajudará os revisores a entender os dados, sua relação com os dados existentes e possíveis riscos de divulgação.
<b>Revisão de conjuntos de dados existentes:</b>	Uma análise completa dos dados existentes em revistas e arquivos de dados relacionados tornará claro o valor da pesquisa proposta e por que os conjuntos de dados atualmente disponíveis são inadequados para responder às suas perguntas de pesquisa.
<b>Formato:</b>	Descrever os formatos dos dados nas fases de submissão, distribuição e preservação. A escolha de formatos preferidos para arquivamento pode tornar o processamento e a liberação de dados mais rápidos e eficientes. O arquivamento e a disseminação de conjuntos de dados derivados também devem ser considerados.
<b>Metadados:</b>	Descrever os metadados a serem fornecidos juntamente com os dados gerados. Como os metadados são frequentemente a única forma de comunicação entre o analista secundário e o produtor de dados, bons metadados descritivos são essenciais para o uso efetivo dos dados.

<b>Armazenamento e backup:</b>	Indicar como e onde vai armazenar cópias de seus arquivos de pesquisa para garantir a sua segurança, bem como quantas cópias você vai manter e como você irá sincronizá-los.
<b>Segurança:</b>	Descrever as medidas para garantir que os dados estejam seguros. Os dados brutos podem incluir identificadores diretos dos participantes do estudo e devem ser bem protegidos durante a coleta e processamento. As práticas de segurança incluem restrições de acesso, como senhas, criptografia, backup de energia e proteção contra vírus e intrusos.
<b>Responsabilidade:</b>	O administrador responsável, ao longo do ciclo de vida dos dados deverá descrever circunstâncias atípicas como, por exemplo, a divisão de responsabilidades caso haja mais de um pesquisador.
<b>Direitos de propriedade intelectual e propriedade de dados:</b>	Indicar quem detém direitos de propriedade intelectual sobre os dados e outras informações criadas pelo projeto e se esses direitos serão transferidos para outra organização para distribuição e arquivamento de dados. Arquivos de dados precisam de uma declaração clara do produtor de dados sobre quem possui os dados antes que eles possam ser divulgados.
<b>Acesso e compartilhamento:</b>	Indicar como pretende arquivar e compartilhar seus dados, e por que você escolheu essa opção específica – (repositórios de domínio, autodivulgação, preservação com a divulgação tardia, repositórios institucionais, coleções de uso restrito, etc.).
<b>Público:</b>	A proposta de subvenção deve especificar os usuários prováveis (acadêmicos ou não acadêmicos) dos conjuntos de dados. A maioria dos usuários potenciais estará dentro da comunidade de pesquisa do ensino superior, mas cada vez mais os formuladores de políticas e os profissionais estão usando dados de pesquisa. Se o conjunto de dados tiver usos comerciais ou outros, isso também deve ser indicado no pedido de financiamento.
<b>Períodos de seleção e retenção:</b>	Descrever como os dados serão selecionados para arquivamento, quanto tempo eles serão mantidos e planos para eventual transição ou encerramento da coleta de dados no futuro.
<b>Arquivamento e preservação:</b>	Descrever como os dados serão preservados a longo prazo. Os dados digitais precisam ser ativamente gerenciados ao longo do tempo para garantir que eles estejam sempre disponíveis e utilizáveis. O conteúdo digital requer ação de preservação contínua para permanecer legível, compreensível e significativo.
<b>Ética e privacidade:</b>	Se for o caso, indicar como será tratado o consentimento informado com relação a informar os entrevistados que as informações pessoais que eles fornecem permanecerão confidenciais quando os dados forem compartilhados ou disponibilizados para análise secundária.

Fonte: Adaptado de ICPSR (2012).

Alguns elementos são considerados opcionais, como mostra o quadro 3.

Quadro 3 – Elementos opcionais de um Plano de Gestão de Dados

<b>Despesas:</b>	O investigador deve delinear os planos e custos de preparação dos dados e documentação para arquivamento. Idealmente, isso deve ser planejado em conjunto com um arquivo.
<b>Organização de dados:</b>	Descrever como os dados serão gerenciados durante o projeto, incluindo informações sobre controle de versão, convenções de nomenclatura, etc. Indicar como os dados podem ser diferentes da norma ajudará outros pesquisadores durante a análise secundária.
<b>Garantia da qualidade:</b>	Descrever procedimentos para garantir a qualidade dos dados durante o projeto;
<b>Requerimentos legais:</b>	Uma listagem de todos os requisitos federais ou de financiamento relevantes para o gerenciamento de dados e compartilhamento de dados.
<b>Importância da boa gestão de dados:</b>	Uma vez que o financiamento é recebido e o projeto de pesquisa começou, o pesquisador vai querer continuar a pensar e planejar a forma final da coleção, incluindo metadados, que será depositado em um arquivo. Planejar a gestão e o arquivamento de uma coleta de dados no início é fundamental para o sucesso do projeto. O custo de um projeto pode ser significativamente reduzido se o planejamento cuidadoso ocorrer no início do projeto.

Fonte: Adaptado de ICPSR (2012).

Acima foram citados alguns elementos considerados importantes para o desenvolvimento de um plano de gestão de dados. Existem outras questões iniciais a serem consideradas e algumas decisões importantes a ser tomadas referentes à documentação, como por exemplo: Que medidas serão tomadas para documentar as decisões que são tomadas à medida que o projeto se desenrola? Como a informação será registrada em procedimentos de campo? Como serão as decisões de codificação, construção variável e similares? Entre outras dúvidas que devem ser pensadas e inseridas como elementos a ser considerados. Em um plano de gestão de dados também deverá constar os tipos de dados e a classificação quanto ao formato, fonte, estabilidade e volume (ICPSR, 2017). A fonte dos dados e as formas podem ser categorizadas conforme os quadros 4 e 5:

Quadro 4 – Fontes dos dados de pesquisa

<b>Categorias</b>	<b>Descrição</b>
<b>Texto</b>	(Word, pdf.).
<b>Numéricos</b>	Tabelas, contagens, medições.
<b>Audiovisuais</b>	Imagens, gravações de som, vídeo.
<b>Modelos</b>	Código de computador.
<b>Disciplina específica</b>	FITS em Astronomia; CIF em Química.
<b>Instrumento específico</b>	Saídas de equipamentos.

Fonte: Adaptado de ICPSR (2012).

Quadro 5 – Formas dos dados de pesquisa

<b>Categorias</b>	<b>Descrição</b>
<b>Observacional</b>	Capturado em tempo real, normalmente fora do laboratório. Normalmente insubstituível e, portanto, o mais importante para salvar.
<b>Experimental</b>	Gerado tipicamente no laboratório ou em circunstâncias controladas. Muitas vezes reproduzível, mas pode ser caro ou demorado.
<b>Simulação</b>	Máquina gerada a partir de modelos de teste. Provavelmente será reproduzível se o modelo e os insumos forem preservados.
<b>Derivado / Compilado</b>	Gerado a partir de conjuntos de dados existentes; Reproduzível, mas pode ser muito caro e demorado.

Fonte: Adaptado de ICPSR (2012).

No plano de gestão também deverá constar informações referentes à preservação digital e às diretrizes, estipulando as convenções, nomenclaturas utilizadas e níveis de acesso. Para descobrir futuros problemas nesta fase inicial do desenvolvimento de um repositório de dados se faz necessário a realização de pré-testes e estudos pilotos. (ICPSR, 2012).

## 2.5 PUBLICAÇÃO AMPLIADA

As publicações científicas têm como propósito disseminar a pesquisa para a comunidade de modo que permita que outros pesquisadores possam avaliá-la, utilizá-la e reutilizá-la sob outro ponto de vista. Através de uma publicação científica, é possível ter o conhecimento dos resultados de uma pesquisa. A discussão proposta nesta seção é aprofundar o conhecimento dos modelos de publicações científicas em um momento em que o modelo tradicional tem sido crescentemente questionado, apontando para a necessidade de discussões sobre novos modelos. De acordo com Nassi-Calò (2016):

O artigo científico, seu formato e meios de publicação, disseminação e compartilhamento passou por significativas alterações desde o surgimento dos primeiros periódicos científicos no século XVII. A Internet, nos anos 1990, mudou drasticamente o paradigma da comunicação da ciência, evento comparável apenas à invenção da imprensa por Gutenberg em 1440, que possibilitou a disseminação dos artigos e periódicos para outras instâncias, além da academia. (NASSI-CALÒ, 2016).

Nota-se que a impressão em papel por meio de tipos móveis trouxe, além da possibilidade de reprodução das publicações em larga escala, uma maneira específica de ler o texto. Conforme Castedo (2007), sempre que há a criação de um novo suporte, cria-se também um novo hábito de leitura.

Em meados do século XIX, Paul Otlet, “pai da Documentação e da internet”, já criticava o modelo tradicional de publicação científica e influenciou pesquisadores que o sucederam no desenvolvimento de ferramentas para o acesso à informação. Em sua obra *Traité de Documentation: le livre sur le livre: théorie et pratique*, publicada em 1934, o autor trata de muitas ideias, hoje adotadas, sobre a organização de redes de cooperação para tratamento e troca de informações documentadas. Nesta obra, Paul Otlet apresenta os produtos do desenvolvimento tecnológico de sua época e suas aplicações à documentação, propondo uma rede universal de informação e documentação. (SANTOS, 2007).

Na década de 1980, o desenvolvimento das tecnologias da informação e comunicação alcançou a etapa em que podia começar a competir com a impressão em papel como meio universal para difundir informações científicas. Nos últimos anos, portanto, passou a ser importante examinar a possibilidade de se transferir informações científicas do meio impresso para o meio eletrônico. (MEADOWS, 1999).

Com o advento da Web, em 1990, diversas possibilidades de melhorias nos canais de comunicação científica surgiram, o que ocasionou a necessidade de novos modelos de publicações. Neste contexto é importante a reflexão sobre as mudanças que a comunicação de informações vêm sofrendo na última década. As tecnologias da informação possibilitaram o acesso às publicações científicas eletrônicas, facilitando a disseminação da pesquisa, permitindo de forma eficiente a recuperação da informação, oferecendo acesso ilimitado, diminuindo custos com impressão, facilitando e modificando o fluxo contínuo da informação científica (BOMFÁ; CASTRO, 2004).

De acordo com Gonçalves e Barbosa (2015, p. 703), “A literatura contrapõe o universo do impresso ao universo do digital e apresenta o primeiro como sendo linear e o segundo como não-linear.” “Os textos impressos teriam um caráter linear, sequencial, e seriam apresentados em objetos eles mesmos lineares, tais como livros. Dariam lugar a práticas de leitura lineares, de modo que a leitura não linear seria um produto do eletrônico e, sobretudo, do digital.” (GONÇALVES; BARBOSA, 2015, p. 703). Desta forma, o texto disponibilizado pelas redes de computadores, composto por conexões e que pode ser acessado por máquina e por usuários em qualquer lugar e simultaneamente, pode ser definido como hipertexto. Os hipertextos teriam uma estruturação em rede, se dariam a ler em objetos não-lineares e possibilitariam práticas de leitura não-lineares.

Passarelli (2008, p.7) destaca que:

O hipertexto institui a possibilidade da narrativa não-linear com os links de trechos de um documento para outros documentos (que podem ser infinitos), desconstruindo a narrativa linear instituída em 500 anos de comunicação por meio da palavra escrita impressa em papel. Desta forma o hipertexto aliado às potencialidades da conexão em rede na Web imprime à informação na sociedade contemporânea em rede características como: instantaneidade, transitoriedade, interoperabilidade e interatividade. (PASSARELLI, 2008, p.7).

Os modelos de publicação tradicionalmente aceitos pela comunidade científica têm se mostrado defasados para as novas formas de se comunicar ciência. Neste cenário, a Ciência Aberta vem possibilitando novos métodos para a divulgação, compartilhamento e reuso, proporcionando um aumento significativo da acessibilidade e da visibilidade. Nesse seguimento, novas práticas para publicação estão sendo desenvolvidas devido ao uso intensivo de dados como, por exemplo, as publicações ampliadas que possui como características a ampliação do escopo de uma publicação tradicional, ligando-a aos dados



provenientes da própria pesquisa, como também outros documentos e objetos digitais relacionados.

Conforme relatado anteriormente neste estudo, outro processo que está constituindo-se a partir da necessidade de conectar os dados das pesquisas com suas respectivas publicações é o que chamamos de publicação ampliada (*Enhanced Publication*). Conforme Araya (2014, p. 120), “A função de ligação/enlace dos elementos informacionais com a narrativa que trata sobre os resultados da pesquisa é a principal característica da publicação ampliada”. Esta prática possibilita o compartilhamento de dados, aumenta a visibilidade da pesquisa e contribui para o movimento mundial de acesso livre à produção de pesquisa na internet. Sayão e Sales (2016, p.69) afirmam que:

Uma publicação ampliada é uma instância de um objeto digital complexo que se caracteriza pela vinculação de artigos aos dados que subsidiaram a sua criação, bem como à outras informações que auxiliem na contextualização da pesquisa, por exemplo, informações sobre os autores, sobre os equipamentos utilizados, projetos, etc. Essas vinculações podem ter valores semânticos e estarem associados à diversas ontologias.

Podem ser consideradas 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 (MUCHERONI; SILVA; PALETTA, 2015). No modelo de publicação ampliada, uma publicação científica pode incluir: a) os próprios dados da pesquisa para evidenciar o estudo; b) materiais adicionais usados para ilustrar ou esclarecer a pesquisa; c) dados posteriores à publicação, tais como comentários, classificações etc. (ARAYA, 2014).

Segundo Verhaar <sup>6</sup> (2008, p. 11, tradução nossa) os dados de pesquisa:

[...] estão cada vez mais sendo armazenados em repositórios de dados confiáveis. O objetivo da curadoria de dados é garantir que os materiais acadêmicos e científicos possam ser preservados e reutilizados. Entretanto, na atual infraestrutura de comunicação científica, esses conjuntos de dados não estão conectados às publicações onde são discutidos.

As publicações ampliadas são criadas com o objetivo de ligar os conteúdos dos repositórios institucionais aos conteúdos dos repositórios de dados.

---

<sup>6</sup> [...] the data that are produced are increasingly stored in trusted data repositories. The aim of such data curation is to ensure that scholarly and scientific materials can be preserved and reused. However, a major shortcoming in the current infrastructure for academic communication is that these datasets are usually not connected to the scientific publications in which they are discussed.<sup>6</sup> Enhanced publications are created with the aim of bridging this imminent gap between the contents of institutional repositories and the contents of data repositories.

Desta maneira, o potencial dos repositórios de dados de pesquisa quando interligados a publicações científicas é elevado, pois esses dados expostos dentro de um contexto mais amplo permitem aos usuários ter uma visão que se perde quando os conjuntos de dados são vistos isoladamente (CLARIVATE ANALYTICS, 2017). Através de um estudo na área de Ciências Nucleares, realizado por Sales (2014), é possível observar (na figura 3) um modelo de publicação ampliada em que uma publicação científica é ligada a diferentes tipos de dados de pesquisa. Nota-se que os dados de pesquisa são armazenados em repositórios de dados.

Figura 3 – Modelo de publicação ampliada.



Fonte: Sales (2014).

Uma publicação tradicional oferece uma interpretação dos dados coletados; e o dado em si dá sustentação às argumentações e análises. Desta forma, conclui-se que os dados abertos de pesquisa ligados às publicações tradicionais de maneira integrada garantem a identificação de sua autoria como também agrega valor a uma nova pesquisa.

No campo da ciência, não basta somente gerar e disponibilizar uma quantidade maior desses elementos. Para poder atribuí-les valor científico, eles devem ser comunicados formalmente como, por exemplo, em periódicos científicos. Elementos informacionais desvinculados do processo de pesquisa que os originou perdem muito do seu valor informacional (ARAYA, 2014). Para que haja essa integração, o conjunto de dados de pesquisa deve carregar um identificador associado a ele pelo sistema de informação que

compôs a agregação/composição, ou seja, um identificador persistente como, por exemplo, o *Digital Object Identifier System* (DOI).

Como exemplo de repositório baseado no modelo de publicação ampliada podemos citar o OpenAIRE, que caracteriza-se por ser um projeto europeu que fornece meios para promover o *Open Access* e que possui a missão de facilitar o acesso a toda produção científica. Fornece ligações cruzadas entre publicações e os sistemas de dados e de financiamento. (OPENAIRE, 2016). A infraestrutura do repositório se constitui num sistema integrado que é construído de forma participativa com base num serviço de rede de repositórios. Na interface do repositório é possível delimitar as buscas por publicações, dados de pesquisa, projetos, entre outros. Na figura 4 é possível observar um exemplo de um resumo de um artigo indexado no repositório OpenAIRE com alguns recursos que interligam aos dados da pesquisa como também à publicação científica na íntegra. Para recuperar o artigo a busca foi delimitada utilizando a opção *Publications*. Na interface existem sugestões de como citar o artigo e opções de download que remete ao repositório de origem.

Figura 4 – Descrição da publicação científica indexada no repositório OpenAIRE

The screenshot displays the OpenAIRE interface for a specific publication. At the top, the OpenAIRE logo is visible on the left, and navigation links (PARTICIPATE, SEARCH, MONITOR, SUPPORT, OPEN ACCESS) are on the right. The main content area shows the title "cells employing methylated DNA immunoprecipitation coupled to next-generation sequencing" and the author "Grimm, C.; Adjaye, J. (2012)". Below the title, there are sections for "Languages: English", "Types: Part of book or chapter of book", "Subjects:", and "Identifiers: doi:10.1007/978-1-61779-794-1\_19". A detailed abstract follows. To the right of the abstract, there are sections for "SHARE - BOOKMARK" (with social media icons), "DOWNLOAD FROM" (MPG.PuRe), and "CITE THIS ARTICLE" (with a BibTeX dropdown and a citation template). Below these are "COLLECTED FROM" (MPG.PuRe) and "References (0)", "Related Research Data (0)", and "Similar Publications (7)". A sidebar on the right contains a "Search" section with a dropdown menu set to "Publications, data, projects, ..." and a list of filters: "all", "publications", "research data", "projects", "people", "organizations", "data providers", "Data Providers", and "General information". Two buttons, "LINK TO PROJECT" and "LINK TO RESEARCH DATA", are circled in red in the lower part of the main content area.

Fonte: *Screenshot* da interface do repositório OpenAIRE.

Refazendo a busca e delimitando por *Research Data* é possível recuperar uma listagem de dados de pesquisa com suas respectivas descrições. Na figura 5 é possível observar a descrição de dados de pesquisa indexados no repositório. Nota-se que o repositório disponibiliza para download os arquivos referentes aos dados de pesquisa como também

oferece links de acesso para a publicação científica. Observa-se que o repositório introduz na descrição da publicação os identificadores persistentes como, por exemplo, o DOI.

Figura 5 – Descrição dos dados de pesquisa indexados no repositório OpenAIRE

The screenshot displays the OpenAIRE repository interface. At the top, there are navigation links for 'PARTICIPATE', 'SEARCH', 'MONITOR', 'SUPPORT', and 'OPEN ACCESS'. The main content area shows a record for 'Palmetto Position Storing Lucene Index Of Dutch Wikipedia' by Janneke M. van der Zwaan, Marx, Maarten; Kamps, Jaap (2016). The publisher is Zenodo, and the type is 'dataset'. The subjects are 'topic modeling, topic coherence, Palmetto, Dutch, Wikipedia'. The identifiers include 'doi:10.5281/zenodo.46377'. A description follows, mentioning a Dutch language resource for calculating topic coherence with Palmetto. There are four references listed, including a paper by M. Roeder, A. Both, and A. Hinneburg. On the right side, there are options to 'SHARE - BOOKMARK', 'DOWNLOAD FROM Zenodo', and 'FUNDED BY PROJECTS' (showing 'No projects found'). A 'CITE THIS RESEARCH DATA' section provides a BibTeX citation. At the bottom, there are sections for 'Related Publications (0)' and 'Related Research Data (0)'. A 'LINK TO PUBLICATION' button is visible.

Fonte: Screenshot da interface do repositório OpenAIRE.

Observa-se que o modelo de publicação ampliada também está presente em alguns periódicos científicos. Como exemplo, podemos citar a *Atmospheric Chemistry and Physics*, que se caracteriza por ser uma revista científica internacional voltada à publicação de estudos de alta qualidade que investigam a atmosfera da Terra e os processos químicos e físicos. Na figura 6 é possível visualizar algumas características que demonstram a utilização da publicação ampliada.

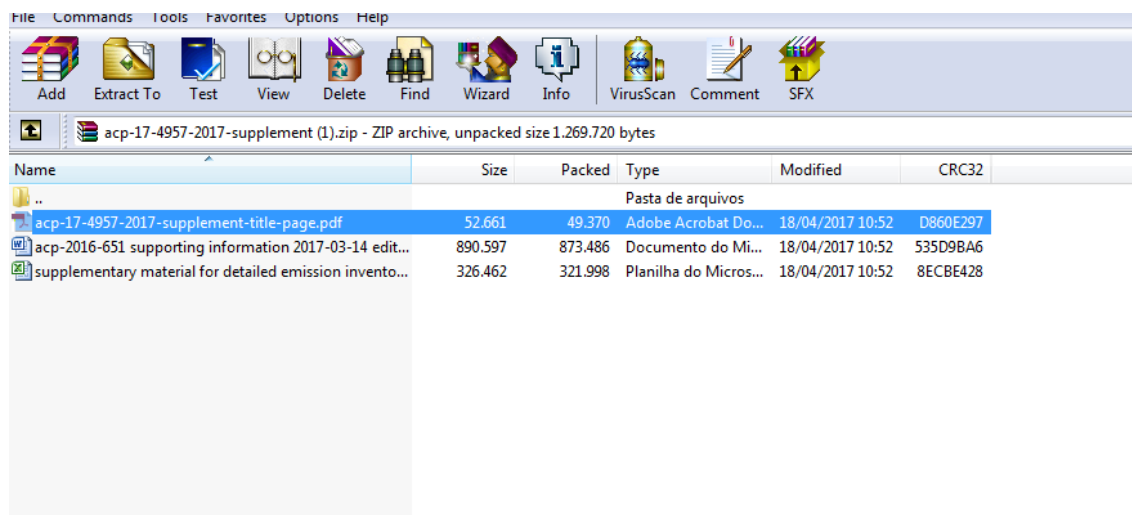
Figura 6 – Utilização da publicação ampliada em periódicos científicos

The screenshot shows the journal website for 'Atmospheric Chemistry and Physics'. The header includes the journal title and 'An interactive open-access journal of the European Geosciences Union'. Navigation links for 'EGU.eu', 'EGU Journals', 'EGU Highlight Articles', 'Contact', and 'Imprint' are present. The main content area features a research article titled 'Multi-pollutant emissions from the burning of major agricultural residues in China and the related health-economic effects' by Chunlin Li et al., dated 18 Apr 2017. Below the article title, there are 'Download' options for PDF and XML, with a red arrow pointing to the PDF icon. A 'Short summary' section provides a brief overview of the article's content. Other sections include 'Related authors' with a list of authors and their works, and 'Journal metrics' showing an IF of 5.114 and an IF 5-year of 5.626. The footer includes 'Copernicus Publications' and 'Share' options for various social media platforms.

Fonte: Screenshot da interface do periódico *Atmospheric Chemistry and Physics*.

Nota-se que na interface existem recursos disponíveis que possibilitam o acesso ao texto completo como também aos dados de pesquisa identificados. A figura 7 mostra os arquivos baixados.

Figura 7 – Download dos dados de pesquisa



Fonte: Screenshot da interface do periódico *Atmospheric Chemistry and Physics*.

Após a realização do download, é possível verificar na figura 8 a identificação dos dados de pesquisa referente ao artigo publicado.

Figura 8 – Dados de pesquisa – Informações suplementares do artigo publicado

AD	AE	AF	AG	AH	AI	AJ	AK	AL	AM	AN	AO	AP	AQ	AR	AS	AT	AU	AV	
<b>2012 autumn harvest emission from filed burning</b>																			
	<b>Phenols</b>	<b>THM</b>	<b>WSI</b>			<b>Gg</b>	<b>PM2.5</b>	<b>PM1.0</b>	<b>OC</b>	<b>EC</b>	<b>SO42-</b>	<b>NO3-</b>	<b>NH4+</b>	<b>K+</b>	<b>WSOA</b>	<b>WSA</b>	<b>PAHs</b>	<b>Phenols</b>	<b>THM</b>
1	0,000	0,000	0,009			Beijing	0,192	0,172	0,077	0,025	0,007	0,001	0,006	0,016	0,002	0,001	0,000	0,001	0,002
2	0,000	0,000	0,009			Tianjin	0,143	0,127	0,060	0,018	0,006	0,001	0,004	0,011	0,003	0,001	0,000	0,000	0,001
3	0,004	0,008	0,185			Hebei	1,836	1,631	0,753	0,230	0,071	0,010	0,056	0,151	0,024	0,010	0,001	0,005	0,015
4	0,002	0,003	0,078			Shanxi	1,917	1,714	0,771	0,248	0,070	0,010	0,062	0,163	0,017	0,011	0,001	0,005	0,017
5	0,000	0,000	0,005			Inner Mongolia	17,928	16,068	7,223	2,370	0,640	0,093	0,576	1,504	0,178	0,097	0,008	0,049	0,158
6	0,000	0,000	0,000			Liaoning	2,390	2,141	0,973	0,321	0,082	0,012	0,080	0,185	0,036	0,014	0,001	0,006	0,021
7	0,000	0,000	0,000			Jilin	3,079	2,758	1,248	0,411	0,107	0,015	0,102	0,245	0,041	0,018	0,001	0,008	0,027
8	0,000	0,000	0,001			Heilongjiang	7,890	7,078	3,270	1,094	0,256	0,036	0,262	0,563	0,162	0,047	0,004	0,019	0,065
9	0,000	0,000	0,004			Shanghai	0,099	0,089	0,043	0,014	0,003	0,000	0,004	0,005	0,004	0,001	0,000	0,000	0,001
10	0,002	0,004	0,100			Jiangsu	1,269	1,134	0,544	0,182	0,038	0,005	0,044	0,070	0,044	0,009	0,001	0,003	0,011
11	0,001	0,002	0,052			Zhejiang	1,948	1,745	0,838	0,286	0,056	0,007	0,069	0,101	0,072	0,014	0,001	0,004	0,017
12	0,004	0,008	0,184			Anhui	1,410	1,258	0,603	0,197	0,045	0,006	0,047	0,084	0,045	0,009	0,001	0,003	0,012
13	0,001	0,006	0,132			Fujian	2,021	1,813	0,866	0,299	0,057	0,007	0,073	0,105	0,074	0,014	0,001	0,004	0,018
14	0,003	0,013	0,302			Jiangxi	2,316	2,069	1,001	0,336	0,068	0,008	0,081	0,119	0,088	0,016	0,001	0,004	0,020
15	0,004	0,008	0,178			Shandong	1,326	1,178	0,545	0,166	0,051	0,007	0,040	0,108	0,018	0,008	0,001	0,004	0,011
16	0,005	0,010	0,234			Henan	1,132	1,011	0,464	0,149	0,040	0,006	0,036	0,088	0,018	0,007	0,001	0,003	0,010
17	0,003	0,007	0,168			Hubei	2,904	2,582	1,252	0,399	0,095	0,011	0,095	0,167	0,099	0,020	0,001	0,006	0,024
18	0,002	0,011	0,259			Hunan	3,390	3,028	1,455	0,486	0,102	0,013	0,118	0,185	0,120	0,023	0,002	0,007	0,025
19	0,003	0,015	0,341			Guangdong	2,149	1,926	0,915	0,314	0,062	0,008	0,077	0,119	0,073	0,015	0,001	0,004	0,018
20	0,003	0,016	0,370			Guangxi	3,020	2,706	1,269	0,430	0,092	0,012	0,106	0,187	0,085	0,020	0,001	0,007	0,027
21	0,001	0,003	0,063			Hainan	0,425	0,381	0,181	0,062	0,012	0,002	0,015	0,023	0,014	0,003	0,000	0,001	0,004
22	0,000	0,001	0,013			Chongqing	1,467	1,315	0,615	0,208	0,045	0,006	0,051	0,093	0,039	0,009	0,001	0,003	0,012

Fonte: Screenshot da interface do periódico *Atmospheric Chemistry and Physics*

No site do periódico são disponibilizadas algumas diretrizes para a submissão de artigos e algumas recomendações referentes à publicação de dados científicos. Aconselham que qualquer conjunto de dados utilizado no manuscrito seja submetido a um repositório de dados confiável e que seja vinculado através de um DOI. Durante a etapa de submissão do manuscrito, o periódico solicita o carregamento de alguns itens importantes como, por exemplo, o resumo; o texto completo e os dados da pesquisa.

Para propor um modelo de publicações ampliadas, é importante seguir alguns requisitos e recomendações para desenvolver uma infraestrutura compatível com os padrões complexos associados às publicações ampliadas. Verhaar (2009) descrevem alguns desses requisitos:

- a) deve ser possível gerar um identificador persistente para cada objeto;
- b) os objetos de uma publicação ampliada devem estar disponíveis com recursos da web para que possam ser consultados através de URIs (*Uniform Resource Identifier*);
- c) deve ser possível vincular objetos digitais à publicação;
- d) deve-se ter o registro e controle de versões da publicação e das partes constituintes;
- e) deve ser possível registrar propriedades básicas da publicação e dos recursos que são adicionados a ele (títulos, breve descrição, data da última modificação, data de criação, etc.);
- f) registrar a autoria da publicação e dos objetos digitais: as pesquisas científicas são cada vez mais colaborativas, por este motivo deve-se registrar as contribuições individuais;
- g) garantir a preservação a longo prazo das publicações ampliadas;
- h) deve ser possível registrar as relações entre os recursos web que fazem parte da publicação ampliada (os componentes precisam ser descritos e classificados utilizando-se um vocabulário padrão);
- i) as instituições que oferecem acesso às publicações ampliadas devem manter os metadados dos objetos em versão legível por máquinas para que sejam descobertos;
- j) as instituições que fornecem acesso às publicações ampliadas devem certificar-se de que estejam disponíveis como documentos baseados no modelo OAI-ORE (*Open Archives Initiative - Object Reuse and Exchange*).

Nesta seção foram descritos alguns dos requisitos que auxiliam no desenvolvimento de uma infraestrutura compatível com as publicações ampliadas. Contudo considera-se necessário um aprofundamento a partir de outros estudos sobre os requisitos necessários para a composição de um modelo. Nota-se a necessidade de esquemas mais ricos de metadados e uma maior complexidade das infraestruturas que estão permeando os dados, diferente do gerenciamento das publicações digitais convencionais. As publicações convencionais apresentam características como, por exemplo, documentos em PDF; textos em HTML; textos com links onde os dados não são curados; os textos não estão de acordo com normas e padrões relacionados a Web Semântica e os textos não são integrados com outras informações de outros repositórios. (SALES, 2014), características estas que diferem do novo modelo de publicações ampliadas.

### 3 METODOLOGIA

Nesta seção apresentou-se a metodologia que foi adotada para a realização do estudo com definições referentes ao tipo de pesquisa e a amostra utilizada. Foram apresentados os instrumentos para a coleta de dados, o tratamento dos dados para análise e as limitações de estudo.

#### 3.1 ABORDAGEM E TIPO DE PESQUISA

Esta pesquisa se caracteriza por ser de natureza básica, que objetiva gerar novos conhecimentos para o avanço da ciência. Segundo a abordagem, é considerada qualitativa, para compreensão e aprofundamento de uma determinada questão. Neste estudo, o objeto de pesquisa foi um panorama de integração entre um repositório de dados de pesquisa e publicações científicas baseado no modelo de publicações ampliadas. Segundo Martins (2008), uma pesquisa com abordagem qualitativa caracteriza-se pela descrição, compreensão e interpretação de fatos. Ou seja, a partir da coleta dos repositórios de dados de pesquisa existentes foi possível observar e comparar os repositórios quanto ao uso de ligações dos dados de pesquisa às publicações científicas. Também foram analisados e comparados os elementos disponíveis nas interfaces dos repositórios que indiquem as ligações e o uso de publicações ampliadas para que fosse possível propor um panorama posteriormente.

Segundo o objetivo, caracteriza-se por ser uma pesquisa exploratória que tem como propósito “[...] proporcionar maior familiaridade com o problema, com vistas a torná-lo mais explícito ou a construir hipóteses.” (GIL, 2010, p. 27). Portanto, foi realizada uma investigação preliminar em busca de modelos indicados na literatura e modelos utilizados em alguns repositórios de dados de pesquisa para que se tenha um maior conhecimento sobre publicações ampliadas.

Quanto aos procedimentos, esta pesquisa é classificada como uma pesquisa bibliográfica, ou seja, a pesquisa “[...] é elaborada com base em material já publicado.” (GIL, 2010, p. 29). Desta forma, é possível identificar o estágio atual do conhecimento referente ao tema, de evidenciar na literatura os principais requisitos e as vantagens da utilização de um modelo de publicação ampliada em repositórios de dados de pesquisa.

Portanto, a partir da investigação e do aprofundamento sobre o tema, foi possível ampliar o conhecimento e detectar os elementos apropriados para propor um panorama de



integração entre um repositório de dados de pesquisas e publicações científicas baseando-se em um modelo de publicação ampliada, como também destacar as vantagens da utilização deste modelo.

### 3.2 INSTRUMENTOS E COLETA DE DADOS

Para que fosse possível verificar como seria um adequado modelo de publicação ampliada, foi necessário identificar na literatura referente à área os principais elementos utilizados para a ligação dos dados às publicações científicas. Da mesma forma, foi necessário identificar nas interfaces dos repositórios esses elementos para que fosse possível realizar comparações entre as informações coletadas na literatura e os dados coletados nas interfaces dos repositórios. A verificação de requisitos presentes nas interfaces possibilitou a identificação de aspectos relacionados à publicação ampliada utilizado entre os repositórios.

Segundo Gil (2009), encontrar material relevante para a pesquisa é a primeira etapa do processo que deriva do problema de pesquisa. Considera-se uma etapa que deve possuir um prévio planejamento para que não haja desperdício de tempo e para que as etapas sejam cumpridas no prazo programado. Deve haver também um rigoroso controle na aplicação dos instrumentos.

A coleta de dados, segundo Marconi e Lakatos (2003, p. 165) é a “Etapa da pesquisa em que se inicia a aplicação dos instrumentos elaborados e das técnicas selecionadas, a fim de se efetuar a coleta dos dados previstos.”. São vários os procedimentos para a realização da coleta de dados que variam de acordo com o tipo de investigação (MARCONI; LAKATOS, 2003, p. 166). Nesta pesquisa, ainda na execução do projeto, foi elaborado um protocolo de revisão sistemática da literatura, que é uma metodologia de estudo que visa estabelecer as etapas devidamente delineadas e definidas em todos os momentos, ou seja, deve esclarecer o modo como serão selecionadas e apuradas as fontes, de maneira que as conclusões a extrair sobre os assuntos em estudo possam ser cientificamente consistentes. É importante esclarecer a diferença entre uma pesquisa bibliográfica e uma revisão sistemática. A pesquisa bibliográfica é feita a partir do levantamento de referências teóricas já analisadas, e publicadas por meios escritos e eletrônicos. O trabalho científico inicia-se com uma pesquisa bibliográfica, que permite ao pesquisador conhecer o que já se estudou sobre o assunto (FONSECA, 2002). Esse tipo de pesquisa, em algumas áreas do conhecimento, refere-se à seção dedicada à revisão bibliográfica (GIL, 2010). No entanto em áreas como, por exemplo, a Ciência da Informação a pesquisa bibliográfica também pode ser considerada um método

bastante detalhado com características equivalentes a uma revisão sistemática. O que difere na revisão sistemática são os parâmetros adotados, que devem ser divulgados de forma que outros pesquisadores possam refazer o método.

Uma revisão sistemática deve conter um protocolo de busca executado de forma a permitir que a revisão seja repetida por outros pesquisadores interessados (KITCHENHAM, 2004). Assim sendo, foi necessário definir métodos e critérios para que fosse possível elaborar o protocolo. A elaboração foi baseada no modelo desenvolvido por Munzlinger, Narcizo e Queiroz (2012). (APÊNDICE A).

De acordo com Kitchenham (2004, p. 3), o processo para a condução de revisões sistemáticas envolve três etapas: planejamento da revisão, condução da revisão e publicação (relatório) dos resultados.

a) os estágios associados ao planejamento da revisão são:

- identificação da necessidade de uma revisão;
- desenvolvimento de um protocolo de revisão.

b) os estágios associados à condução da revisão são:

- identificação da pesquisa;
- seleção de estudos primários;
- estudar avaliação de qualidade;
- extração e monitoramento de dados;
- síntese de dados.

No decorrer do planejamento da revisão, o protocolo de revisão foi definido. Durante a condução da revisão, as fontes para a revisão sistemática foram selecionadas e os estudos primários foram identificados, selecionados e avaliados de acordo com critérios de exclusão, inclusão e de qualidades. Os critérios de inclusão e exclusão servem para indicar por quais motivos um estudo foi incluído ou excluído da revisão ou porque são considerados relevantes ou não para o estudo. Com os critérios de qualidade é possível avaliar aspectos metodológicos dos estudos, buscando o aumento da confiabilidade nos resultados que serão obtidos. Após a seleção dos estudos, os dados foram extraídos e sintetizados para serem estabelecidas as categorias para análise.

Para auxiliar nos registros durante a etapa da condução da revisão, foram utilizados formulários elaborados no Excel: formulário de condução da revisão (APÊNDICE B);

formulário de seleção de estudos (APÊNDICE C) e formulário de extração de dados (APÊNDICE D). Estes formulários foram elaborados baseando-se em estudos já realizados por pesquisadores da área de Engenharia de Softwares como, por exemplo, Mafra e Travassos (2005).

Através do método de revisão sistemática, foram identificados estudos sobre o tema em questão para extrair os principais requisitos de uma publicação ampliada e averiguar a aplicabilidade. Para a realização da análise e comparações nos repositórios de acordo com os requisitos recuperados durante a revisão sistemática, foram estabelecidas as categorias para a análise. As categorias são classes que reúnem um grupo de elementos sob um título genérico, agrupamento esse efetuado em razão dos caracteres comuns destes elementos (BARDIN, [19--]).

A partir das categorias definidas, um checklist (APÊNDICE E) foi elaborado com base em critérios e elementos considerados essenciais no desenvolvimento de um modelo de publicação ampliada. O checklist auxiliou nas observações dos repositórios e identificação dos requisitos presentes em suas interfaces e a comparar com os requisitos estabelecidos neste estudo, através da literatura recuperada na revisão sistemática.

Para obter os registros e fazer a seleção dos repositórios de dados de pesquisa, foi utilizado o diretório re3data.org (*Registry of Research Data Repositories*) que caracteriza-se por ser uma ferramenta de busca de repositórios de dados. De 2012 a 2016, mais de 1.500 repositórios foram indexados. O re3data.org é um registro global de repositórios de dados de pesquisa que cobre repositórios de dados de pesquisa de diferentes disciplinas acadêmicas. Ele apresenta repositórios para o armazenamento permanente e acesso de conjuntos de dados para pesquisadores, organismos de financiamento, editores e instituições acadêmicas. Ele promove uma cultura de partilha, maior acesso e melhor visibilidade dos dados da investigação. O registro entrou em funcionamento no outono de 2012 e é financiado pela Fundação Alemã de Pesquisa Científica (DFG). (RE3DATA.ORG, 2017). A escolha do diretório se deu a partir da revisão de literatura e a consulta de suporte ao pesquisador dos sistemas de bibliotecas como, por exemplo, o Sistema Integrado de Bibliotecas da Universidade de São Paulo, onde há a indicação de utilização do diretório.

Foi elaborado um quadro com a listagem dos repositórios nacionais e internacionais de dados de pesquisa. Utilizando os filtros do diretório foi possível identificar os repositórios por assunto, conteúdo, tipos de instituições, licenças entre outros, mas, sobretudo, os repositórios que possuem publicações ampliadas.

Para o registro das observações durante a etapa de análise dos repositórios, foi utilizado um checklist de validação de requisitos (APÊNDICE E). Para realizar a inspeção, foi utilizado como roteiro de observação um checklist com os requisitos essenciais para a implementação de um modelo de publicações ampliadas, conforme a indicação da literatura da área. Um checklist é uma técnica que pode ser utilizada para verificar se uma determinada questão atende aos requisitos especificados. Neste estudo, essa técnica demonstrou se os repositórios selecionados foram desenvolvidos de acordo com padrões pré-definidos.

### 3.3 ANÁLISE DOS DADOS

Segundo Gil (2009, p. 91), “A análise e interpretação de dados nos estudos de caso é uma atividade complexa, pois não há um consenso acerca dos procedimentos a serem adotados.”. Afirma que, “[...] a análise e interpretação é um processo que de certa forma se dá simultaneamente à sua coleta”. O modelo de análise deste estudo é composto por três fases:

Na primeira fase de análise, foram coletadas literaturas específicas da área por meio da revisão sistemática. A partir das literaturas, foram extraídos os requisitos indicados e sugeridos para um modelo de publicação ampliada.

Na segunda fase de análise, foram selecionados os repositórios de dados de pesquisa através do diretório re3data.org e averiguadas as informações disponibilizadas por meio dos filtros que demonstrassem a utilização de ligações baseadas no modelo de publicação ampliada.

Na terceira fase de análise, foram armazenados os *screenshots* das interfaces dos repositórios coletados na segunda fase, que utilizam o modelo de publicações ampliadas. Durante a inspeção das interfaces, foram identificados os elementos que cada repositório contempla para posteriormente iniciar as análises, observações e comparações dos elementos utilizados nos repositórios com os requisitos estabelecidos por meio da revisão sistemática. As observações foram registradas no checklist de validação de requisitos (APÊNDICE E). A partir dos registros das observações no checklist de validação, foram elaborados quadros comparativos para verificação dos requisitos essenciais e comuns utilizados entre os repositórios.

Com o auxílio dos instrumentos utilizados para a coleta, foi possível dar início as análises dos dados. As informações coletadas, que foram registradas, foram analisadas e comparadas para que fosse possível evidenciar os elementos essenciais para o desenvolvimento de um modelo padrão baseado na publicação ampliada.

Cruzando as informações existentes na literatura com os dados coletados nos repositórios, foi proposto um panorama de integração para repositórios de dados de pesquisa e as publicações científicas relacionadas, baseando-se na publicação ampliada.

### 3.4 LIMITAÇÕES DO ESTUDO

Os requisitos destacados na literatura não são mencionados como obrigatórios, mas retratam que a utilização melhora a recuperação da informação e, conseqüentemente, aumenta a visibilidade da publicação e de autores. Desta forma, o panorama proposto neste estudo é baseado em requisitos, que conforme a literatura, não são considerados obrigatórios.

Devido a um número baixo de documentos que a revisão sistemática recuperou, foi necessário utilizar documentos citados no referencial teórico além das referências utilizadas nos documentos recuperados na revisão sistemática de modo que fosse possível a elaboração de categorias para análise proposta neste estudo.

Tem-se como limitação a recuperação de um número alto de repositórios (654 repositórios) caracterizados, de acordo com o diretório, como publicações ampliadas. No entanto, após as análises, foi necessário excluir um grande número destes, pois não constatou-se o uso das publicações ampliadas, restringindo desta forma a pesquisa. Além disso, a análise do requisito 9 – Políticas de preservação, retornou um número baixo de repositórios. O motivo se deu por não serem disponibilizadas as políticas de forma visível em suas interfaces, dificultando desta maneira a sua detecção.

Outro fator limitante foi a dificuldade de realizar as buscas nas ferramentas disponíveis. Para que fosse possível ter acesso ao conjunto de dados, foi necessário descobrir termos utilizados em determinadas áreas (termos técnicos cuja autora desta pesquisa não possuía conhecimento) para ter acesso às coleções disponíveis e, após, partir para a análise das descrições na interface.

## 4 ANÁLISE E APRESENTAÇÃO DOS RESULTADOS

Nesta seção são apresentadas as análises da investigação de acordo com os objetivos predefinidos para este estudo.

### 4.1 PRIMEIRA FASE DA ANÁLISE

Foram coletadas literaturas específicas da área por meio da revisão sistemática. A partir disto, foram extraídos os requisitos indicados e sugeridos para um modelo de publicação ampliada.

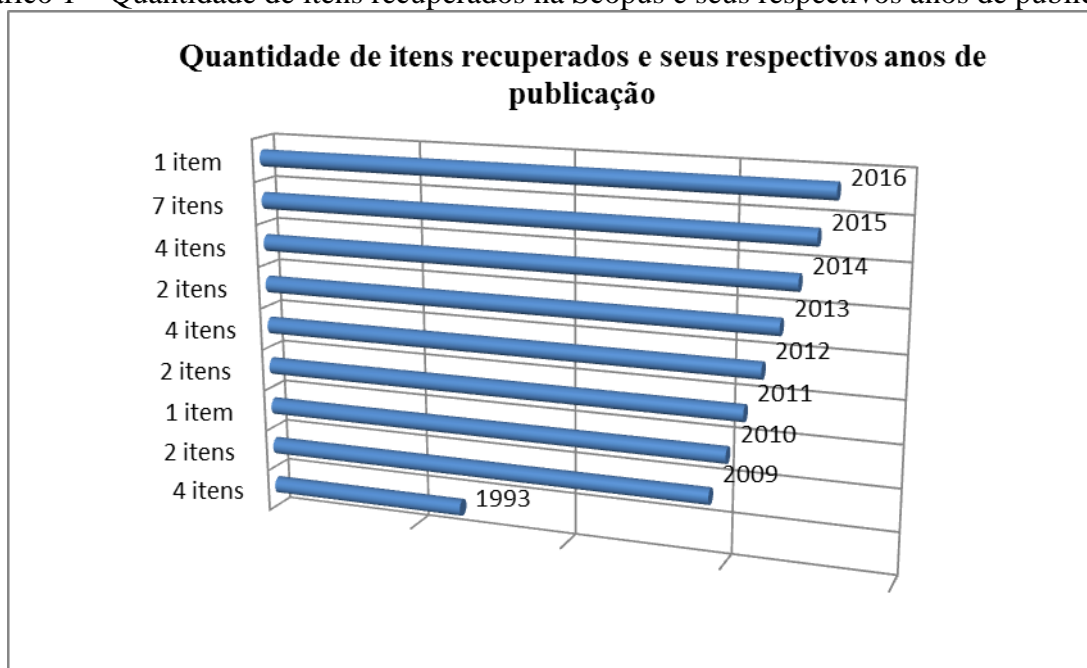
A revisão sistemática da literatura foi realizada em bases de dados eletrônicas utilizando a *string* de busca, que é a combinação das palavras-chave e termos relacionados utilizando os operadores lógicos de forma que a maior quantidade de estudos seja recuperada.

Foram escolhidas algumas bases: a SCOPUS por ser considerada a maior base de dados de citações e resumos revisada por pares; a WoS (*Web of Science*), que permite acesso a referências e resumos em todas as áreas do conhecimento; e a LISA (*Library Information Science Abstracts*), que é um serviço internacional de resumo e indexação direcionado à Ciência da Informação.

A busca na base de dados SCOPUS foi realizada no dia 12 de setembro e foram utilizados os limitadores: “*article title, abstract e keywords*” e foi selecionada a opção “*documents*”.

Na primeira rodada de teste, utilizaram-se os termos de busca “*enhanced publication*” e “*requirements*”. Utilizou-se também o operador booleano “*And*”. Foram recuperados nesta base de dados um total de 439 documentos sendo que grande parte não possuía os termos da expressão “*enhanced publication*” unidos. Desta forma, foi necessário realizar uma nova busca utilizando desta vez aspas (“”) para que fosse possível recuperar documentos referentes aos termos equivalentes e propostos neste estudo. Na segunda rodada do teste, com a utilização de operadores booleanos como, por exemplo, “*And*” apenas 1 documento foi recuperado. Foi realizada a 3ª rodada de teste utilizando apenas a expressão “*enhanced publication*” e aspas (“”). Nesta rodada foram recuperados 27 documentos que estão de acordo com a proposta desta pesquisa. Demonstramos no gráfico 1 os anos das publicações encontradas.

Gráfico 1 – Quantidade de itens recuperados na Scopus e seus respectivos anos de publicação



Fonte: Elaborado pela autora

Os documentos recuperados caracterizam-se como artigos, documentos de conferências, capítulo de livros no idioma inglês.

Dos 27 itens, apenas 6 itens estão em acesso aberto (*Open Access*). Observou-se que os seguintes títulos estavam repetidos: “*A framework supporting the shift from traditional digital publications to enhanced Publications*”; “*Enabling living systematic reviews and clinical guidelines through semantic technologies*”; “*Linking full-text grey literature to underlying research and post-publication data: An Enhanced Publications Project 2011-2012*”; “*Enhanced publication services to members*” (este último título apareceu 4 vezes pois foi publicado em fontes diferentes. Desta forma, foi considerado apenas um item nesta pesquisa). Assim sendo, foram retirados dos 27 itens recuperados, 6 itens, totalizando, desta maneira, 21 itens.

Para a realização da busca na base de dados *Web of Science* no dia 18 de setembro, foi utilizada a pesquisa básica; foi delimitada por “tópico” e selecionou-se “todos os anos”. Foi usada a expressão “*enhanced publication*” com aspas (“”), pois utilizando sem a pontuação - aspas (“”) a base recuperaria um número elevado de itens que não corresponderia ao termo proposto neste estudo. No entanto, limitando a busca com a utilização de aspas, a base recuperou um número bastante restrito de resultados: 5 itens, que tinham como data de publicação de 2007, 2010, 2012, 2015 e 2016.

Após o registro no formulário de condução da revisão (APÊNDICE B), foi possível observar que havia dois títulos idênticos à pesquisa anterior realizada na base SCOPUS: *Enhanced Publication Management Systems A systemic approach towards modern scientific communication; Linking full-text grey literature to underlying research and post-publication data: An Enhanced Publications Project 2011-2012*, desta maneira restaram apenas 3 itens.

Na mesma data, 18 de setembro de 2017, foi realizada a pesquisa na base LISA. Na primeira rodada utilizou-se a busca básica, o limitador “em qualquer lugar”, “todas as datas” e a expressão de busca *enhanced publication*. O sistema recuperou 6.368 resultados. No entanto, os termos da expressão não apareciam unidos, o que demonstrou ter outros significados que não estavam de acordo com o que foi estipulado nesta pesquisa. Na segunda rodada, optou-se em realizar novamente a busca, só que desta vez utilizando aspas (“”) para unir os termos que na pesquisa anterior apareceram separados. Desta vez, o sistema recuperou 14 resultados. Para verificar a existência de termos relacionados aos requisitos de uma publicação ampliada, na terceira rodada utilizou-se a expressão com aspas “*enhanced publication*”, o operador booleano “*And*” juntamente com o termo “*requirements*”. Desta vez, o resultado foi de 2 itens recuperados que possuíam a expressão *enhanced publication e requirements*. No entanto, optou-se em utilizar a forma de busca da segunda rodada que havia recuperado 14 resultados entre eles, os 2 itens que tratavam sobre os requisitos. Por haver títulos idênticos às pesquisas anteriores, foram eliminadas desta rodada 5 itens, totalizando 9 documentos.

Após a condução da revisão, foi realizada a etapa de seleção dos estudos de acordo com os critérios de inclusão e exclusão. Foram incluídas as publicações que descrevem as aplicações e sugerem os principais requisitos de um modelo de publicação ampliada; estudos experimentais que indiquem a utilização de requisitos e estudos que apresentem metodologias de avaliação dos principais requisitos. Foram excluídas as publicações cujo foco não correspondia à questão de pesquisa e que não forneciam informações sobre os principais requisitos utilizados em uma publicação ampliada. No formulário de seleção de estudos (APÊNDICE C) foram registrados os títulos dos 33 documentos recuperados durante a etapa de condução da revisão. Também foram registrados os autores e datas de publicações. Nesta etapa foram analisados os documentos quanto ao conteúdo disponível nas bases utilizadas para a coleta e se os documentos recuperados tratavam sobre o assunto pesquisado. Foi observado, através da leitura dos resumos e análise de conteúdo que, entre os 33 documentos recuperados, 5 itens não tratavam sobre o assunto proposto, desta forma, não foram registrados os resumos. Neste mesmo formulário, foram identificados 16 itens que não



possuem o texto completo disponível nas bases de dados (foram registrados somente os resumos), resultando 12 itens com resumos e textos completos. Sendo estes últimos armazenados em pastas para uma posterior análise. Desta forma, na próxima etapa de extração de dados esses 6 itens, que não correspondiam ao tema e os 16 itens que não possuíam textos completos, foram retirados da análise, totalizando 11 itens a serem analisados. Destes 11 itens recuperados, 5 documentos abordavam assuntos relacionados a infraestrutura de uma publicação ampliada e seus principais requisitos. Embora tenham sido feitos numerosos esforços para desenvolver filtros de pesquisa que recuperassem estudos relevantes com maior precisão, observou-se um número baixo de documentos referentes ao assunto *enhanced publication* e que tratassem sobre seus principais requisitos.

Desta forma, após a seleção dos estudos, os dados foram extraídos e sintetizados para que fosse possível estabelecer as categorias para análise. Além dos requisitos coletados durante a revisão sistemática, também foram incluídos nas categorias os requisitos citados anteriormente neste estudo, como também os estudos referenciados nas obras recuperadas durante a revisão sistemática. Durante a elaboração das categorias para análise foi identificado nos documentos recuperados que os estudos realizados foram fundamentados nos conceitos desenvolvidos por Verhaar (2009), na obra *Enhanced Publications: linking publications and research data in digital repositories*. Neste livro, é fornecida uma visão geral de última geração dos elementos estruturais de uma Publicação Ampliada, bem como modelos de publicação, inter-relações e problemas de repositório.

A partir das categorias definidas, um checklist (APÊNDICE E) foi elaborado baseado em critérios e requisitos considerados essenciais no desenvolvimento de um modelo de publicação ampliada. No quadro 6 estão descritas as categorias estabelecidas para a análise.

Quadro 6 – Categorias para análise

<b>Categorias para análise</b>
Recursos WEB – URI
Identificador persistente
Vínculo de objetos digitais as publicações
Registro das versões das publicações e dados
Registro das propriedades básicas (publicação e recursos adicionados)
Vocabulário controlado
Atributos – descrição
Registro de autoria da publicação e dados de pesquisa
Preservação
Registro das relações que fazem parte da publicação ampliada
Modelo OAI-ORE
<i>Open access</i>

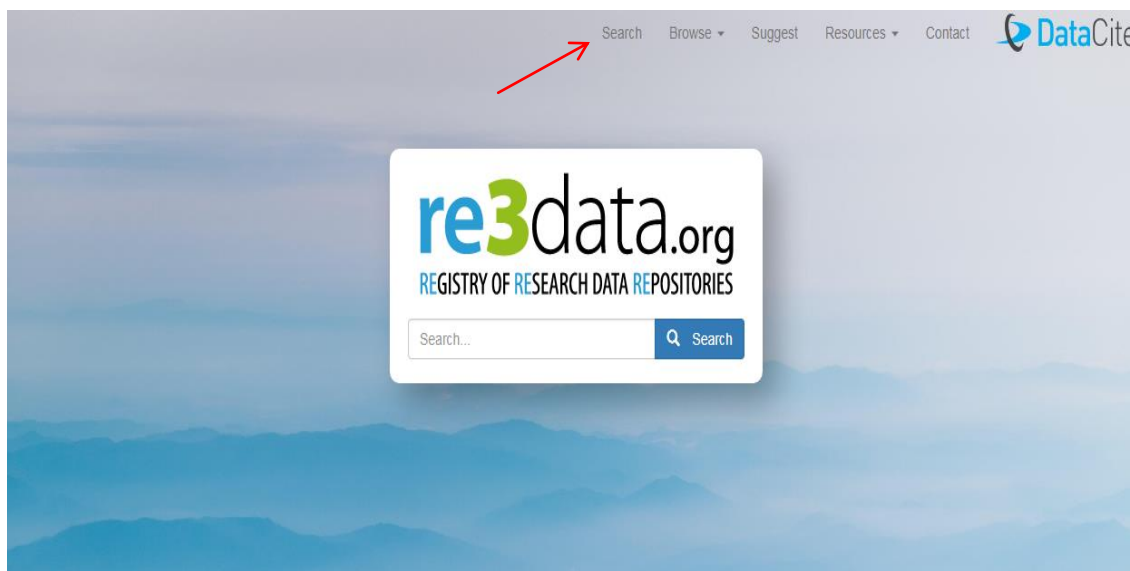
Padrões web semântica
Relacionamento semântico
Upload de objetos digitais
Metadados

Fonte: Elaborado pela autora

## 4.2 SEGUNDA FASE DA ANÁLISE

Foram selecionados os repositórios de dados de pesquisa através do diretório re3data.org e averiguadas as informações disponibilizadas por meio dos filtros que demonstrassem a utilização de ligações baseadas no modelo de publicação ampliada. Foi realizada a identificação de repositórios de dados de pesquisa indexados no diretório por assunto, conteúdo, tipos de instituições, licenças entre outros e descritas as quantidades de cada item. As informações referentes aos repositórios foram coletadas em 08 de setembro de 2017. Na interface principal do diretório, foi selecionada a opção “Search” para que fosse possível utilizar os filtros e realizar a seleção, conforme demonstra a figura 9.

Figura 9 – Interface principal do diretório re3data.org



Fonte: Screenshot da interface do diretório re3data.org

Após, identificou-se as opções de filtragem, conforme mostra a figura 10:

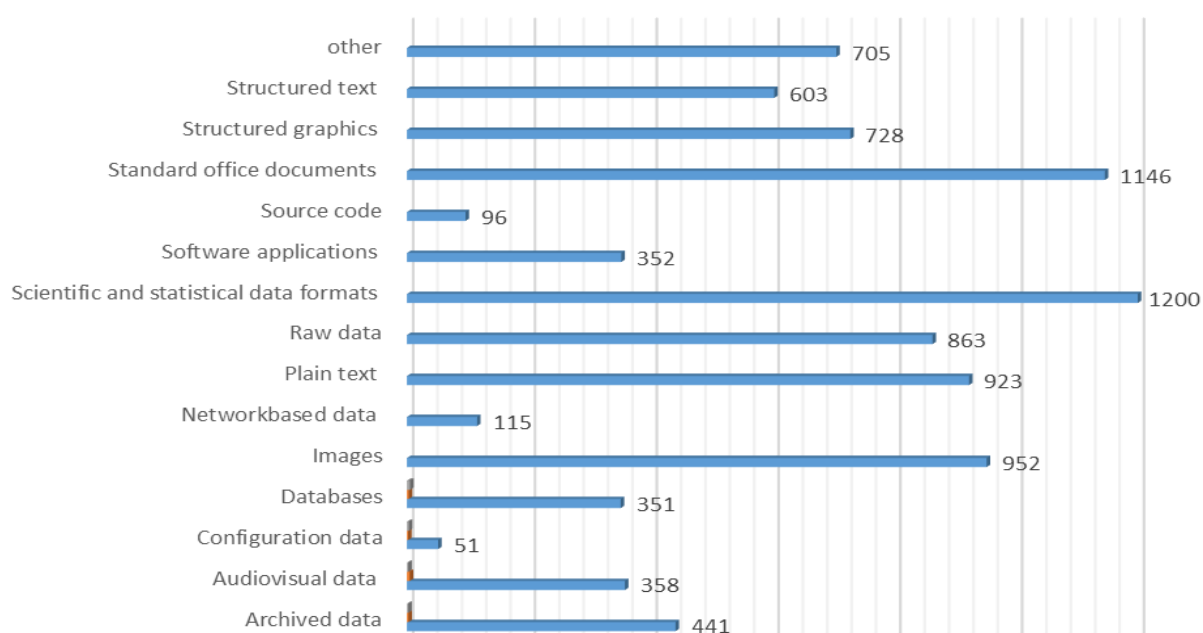
Figura 10 – Opções de filtros do diretório re3data.org

The screenshot shows the re3data.org website interface. On the left, there is a sidebar with a 'Filter' button and a list of filter categories: Subjects, Content Types, Countries, AID systems, API, Certificates, Data access, Data access restrictions, Database access, Database access restrictions, Database licenses, Data licenses, Data upload, Data upload restrictions, Enhanced publication, Institution responsibility type, Institution type, Keywords, Metadata standards, PID systems, Provider types, Quality management, Repository languages, Software, Syndications, Repository types, and Versioning. A red arrow points to the 'Filter' button. The main content area shows search results for 'UniProtKB/Swiss-Prot' and 'Khazar University Institutional Repository'. The UniProtKB/Swiss-Prot result shows subject categories like 'Basic Biological and Medical Research', 'General Genetics', 'Biology', and 'Life Sciences', and content types like 'Networkbased data', 'Structured graphics', 'Plain text', and 'other'. The Khazar University Institutional Repository result shows subject categories like 'Humanities and Social Sciences', 'Life Sciences', 'Natural Sciences', and 'Engineering Sciences', and content types like 'Standard office documents', 'Images', 'Audiovisual data', 'Plain text', and 'other'.

Fonte: Screenshot da interface do diretório re3data.org

Selecionando-se a opção “Subjects” foi possível identificar os repositórios de dados de pesquisa por “assunto” (APÊNDICE G). Selecionando-se a opção “Content types” foi possível identificar os repositórios de dados de pesquisa por “tipos de conteúdo”. O gráfico 2 apresenta a quantidade.

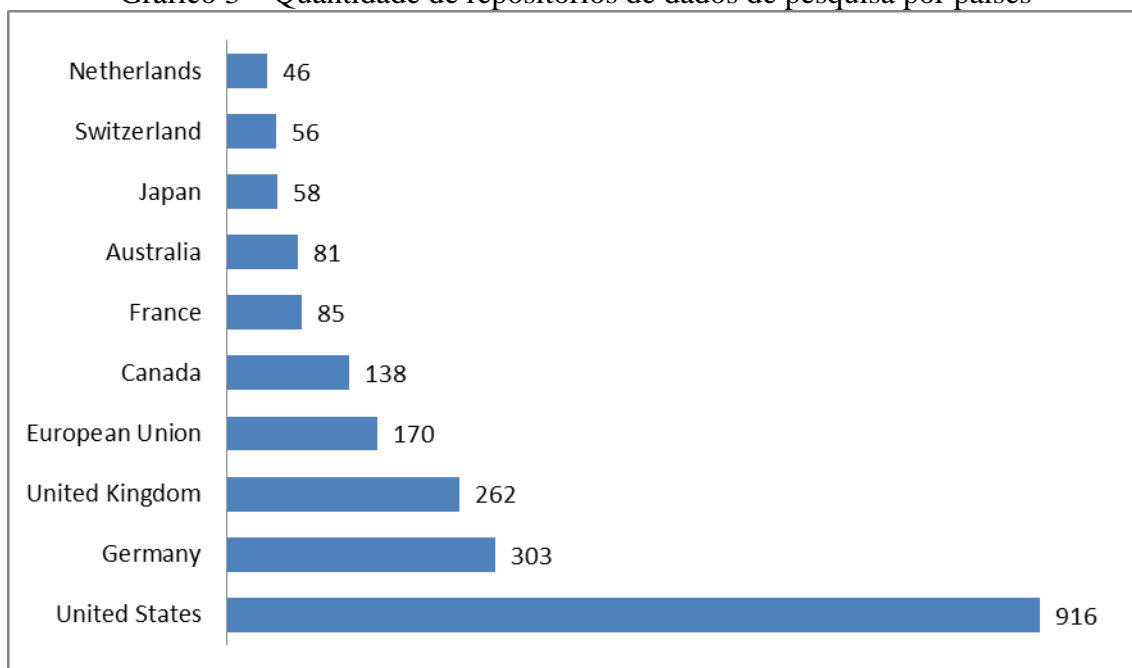
Gráfico 2 – Quantidade de repositórios de dados de pesquisa por tipos de conteúdo



Fonte: Informações coletadas no diretório re3data.org

Selecionando-se a opção “*Countries*” foram identificados os repositórios de dados de pesquisa por “países”. O gráfico 3 demonstra um recorte dos 10 (dez) países e o respectivo número de repositórios.

Gráfico 3 – Quantidade de repositórios de dados de pesquisa por países



Fonte: Informações coletadas no diretório re3data.org

No quadro 7 é possível visualizar os repositórios de dados de pesquisas relacionados ao Brasil, conforme informações do diretório.

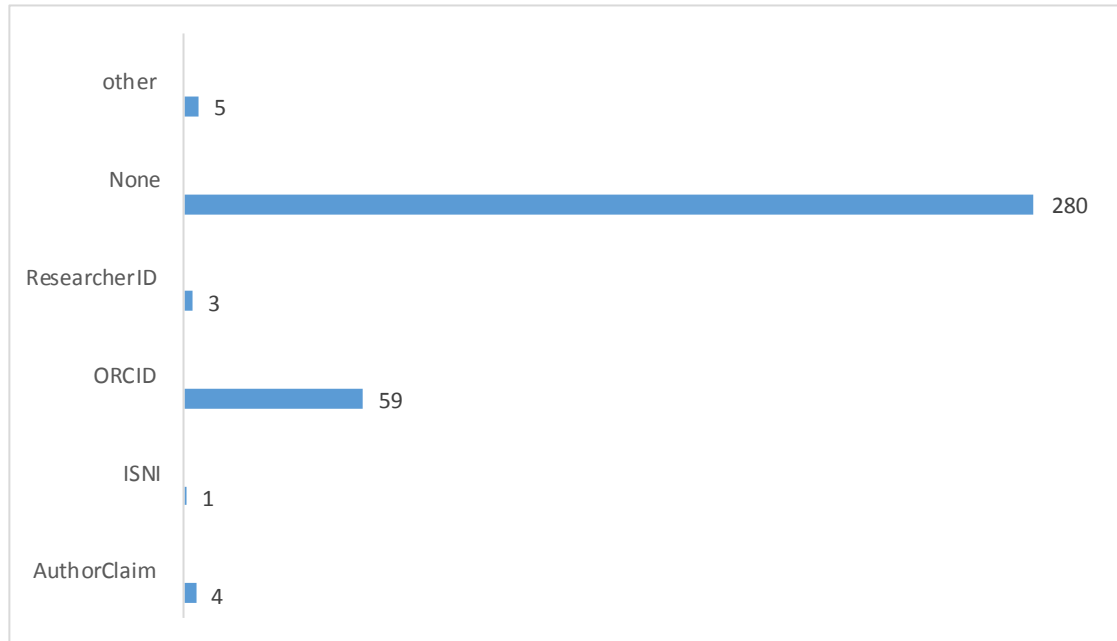
Quadro 7 – Listagem de repositórios de dados de pesquisa nacionais (*Brazil*)

1	<i>Brazil Exploration and Production Database</i> (Banco de Dados de Exploração e Produção)
2	<i>WorldClim - Global Climate Data</i>
3	<i>GLOBE</i>
4	<i>IBICT Dataverse Network</i> (Instituto Brasileiro de Informação em Ciência e Tecnologia Dataverse Network).
5	<i>International Ocean Discovery Program</i>
6	<i>PPBio Data Repository</i> (Repositório de Dados de Levantamentos Biológicos)

Fonte: Informações coletadas no diretório re3data.org

Selecionando-se a opção “*AID systems*” foi possível identificar os repositórios de dados de pesquisa e os identificadores digitais utilizados para identificação de autores, como mostra o gráfico 4.

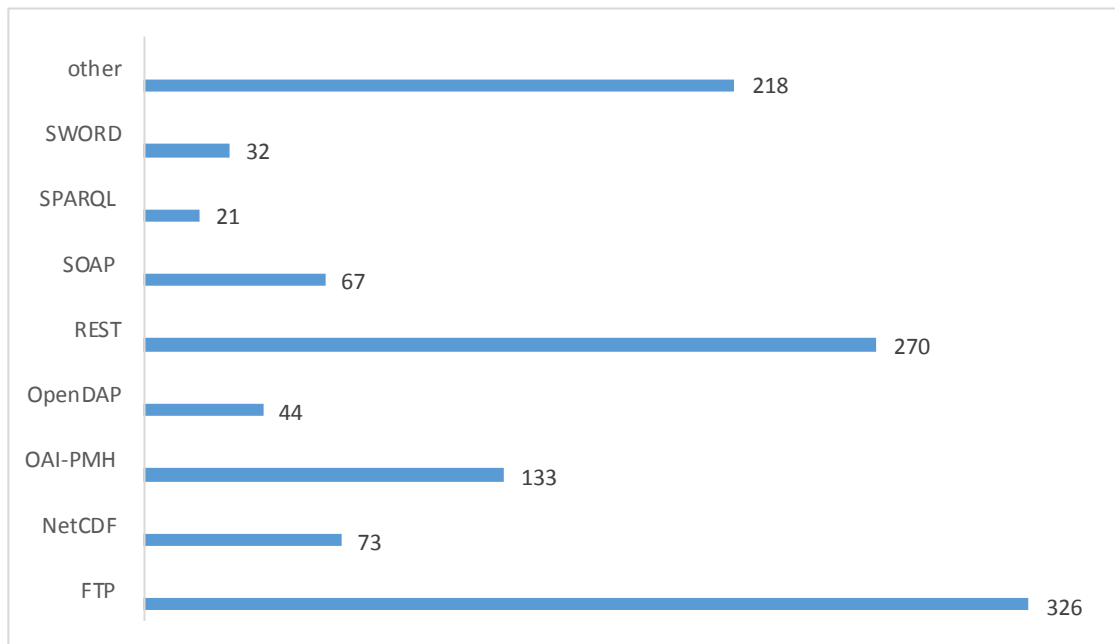
Gráfico 4– Quantidade de repositórios de dados de pesquisa e os identificadores digitais para identificação de autores



Fonte: Informações coletadas no diretório re3data.org

Selecionando-se a opção “*API*” identificou-se a quantidade de repositórios de dados de pesquisa que utilizam um determinado tipo de padrões de programação. *API (Application Programming Interface)* caracteriza-se por ser um conjunto de padrões estabelecido por um software para a utilização das suas funcionalidades por aplicativos que não pretendem envolver-se em detalhes da implementação, mas apenas usar seus serviços. É possível observar no gráfico 5 os tipos utilizados pelos repositórios indexados no diretório.

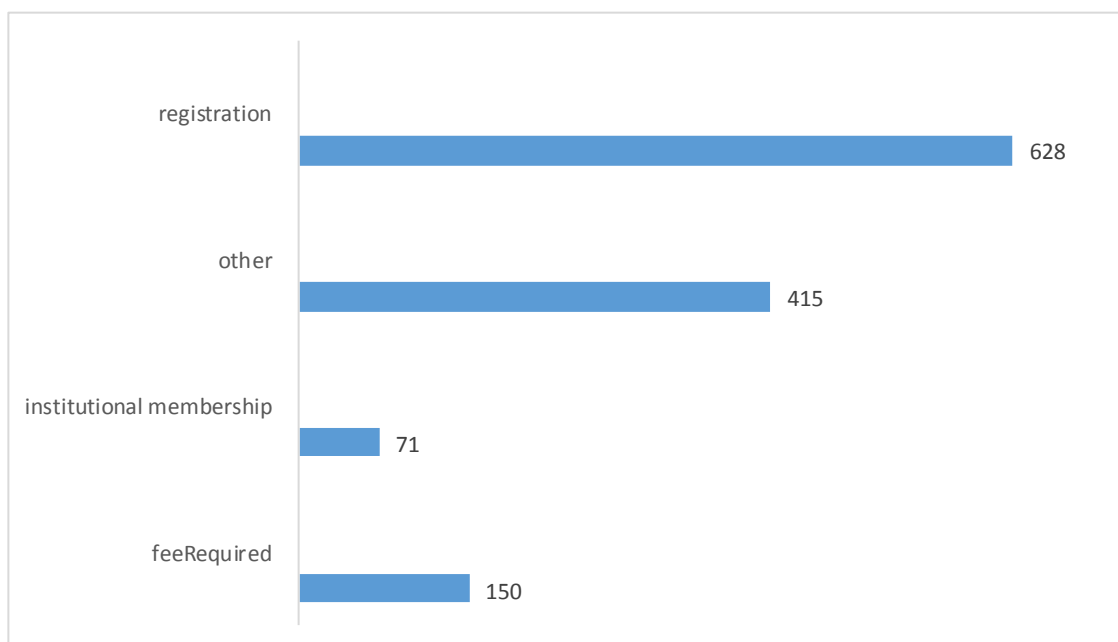
Gráfico 5 – Quantidade de repositórios de dados de pesquisa que utilizam padrões de programação *API* (*Application Programming Interface*)



Fonte: Informações coletadas no diretório re3data.org

Selecionando-se a opção “*Data access restrictions*” foi possível identificar a quantidade de repositórios de dados de pesquisa e suas respectivas restrições de acesso aos dados, conforme o gráfico 6.

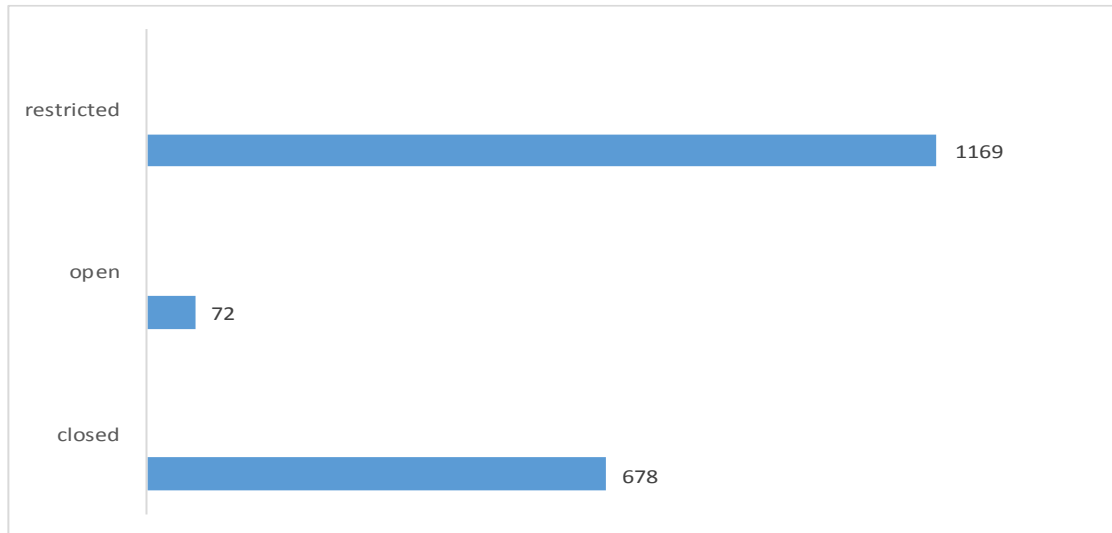
Gráfico 6 – Quantidade de repositórios de dados de pesquisa quanto às restrições de acesso a dados (*Data Access Restrictions*)



Fonte: Informações coletadas no diretório re3data.org

Selecionando-se a opção “*Data upload*” foi identificada a quantidade de repositórios de dados de pesquisa e a forma como é realizado o carregamento de dados. O gráfico 7 demonstra os detalhes.

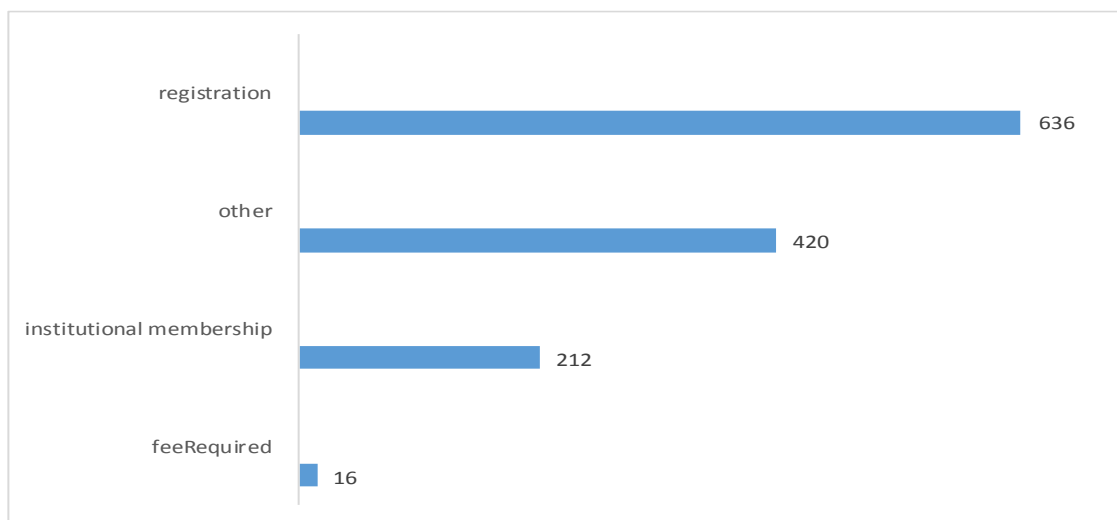
Gráfico 7 – Quantidade de repositórios de dados de pesquisa e a forma de carregamento de dados (*Data Upload*)



Fonte: Informações coletadas no diretório re3data.org

Selecionando-se a opção “*Data upload restrictions*” foi possível identificar a quantidade de repositórios de dados de pesquisa e as respectivas restrições para o carregamento de dados. O gráfico 8 mostra isso.

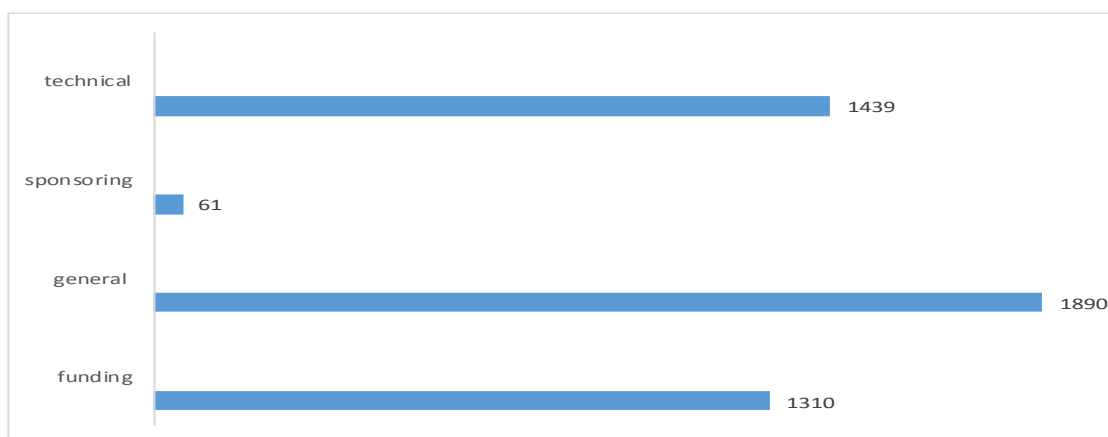
Gráfico 8– Quantidade de repositórios de dados de pesquisa e as restrições para o carregamento de dados (*Data Upload Restrictions*)



Fonte: Informações coletadas no diretório re3data.org

Conforme o gráfico 9, selecionando-se a opção “*Institution responsibility type*” foi possível identificar o tipo de responsabilidade da instituição e seu respectivo número de repositórios relacionados.

Gráfico 9 –Tipo de responsabilidade da instituição e seu respectivo número de repositórios relacionados (*Institution Responsibility Type*)



Fonte: Informações coletadas no diretório re3data.org

Selecionando-se a opção “*Institution type*” identificou-se os números de repositórios de dados de pesquisa relacionados ao tipo de instituição, de acordo com o quadro 8.

Quadro 8 – Quantidade de repositórios de dados de pesquisa quanto ao tipo de instituição

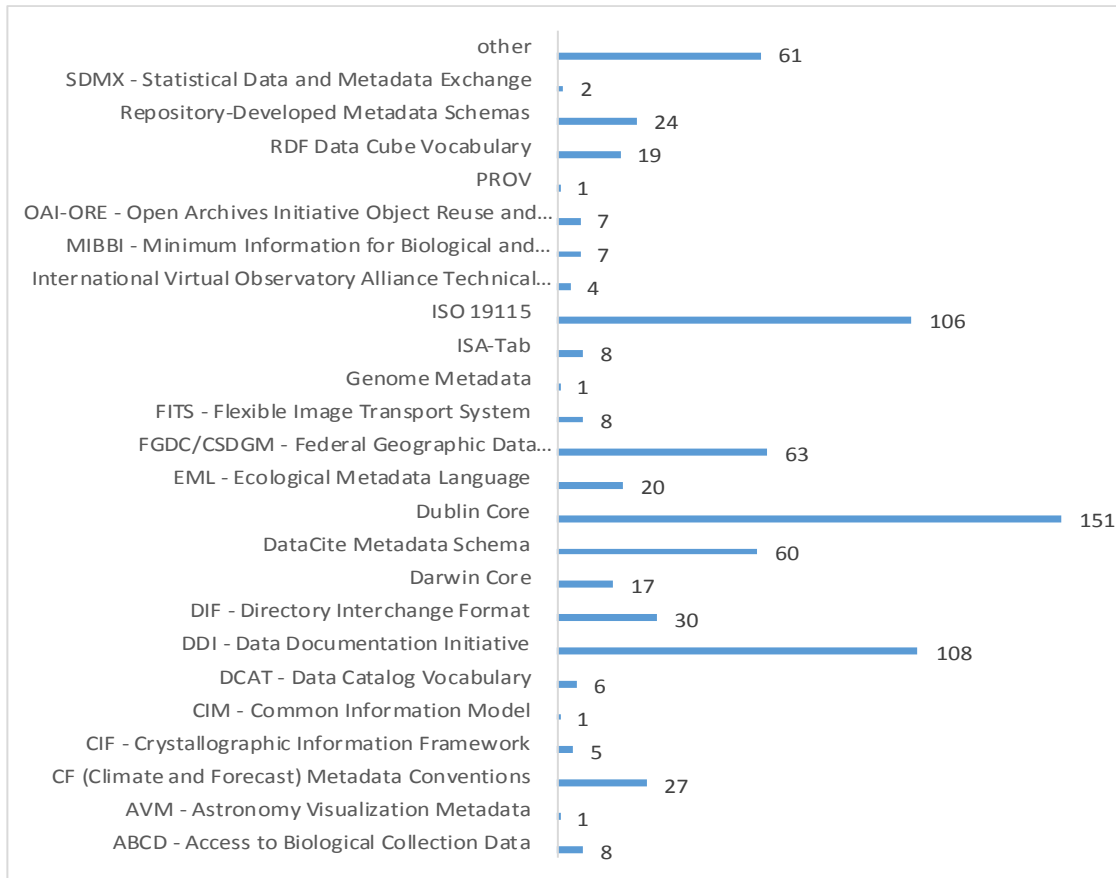
TIPO DE INSTITUIÇÃO	Qtd.
Comercial	103
Sem fins lucrativos	1895

Fonte: Informações coletadas no diretório re3data.org

Selecionando-se a opção “*Metadata standards*” foi possível identificar a quantidade de repositórios de dados de pesquisa e os padrões de metadados, que são fundamentais para o acesso e reutilização de dados de pesquisa, assim sendo, se faz necessário estabelecer padrões para os repositórios (DCC, 2017). No gráfico 10, é possível observar a listagem dos tipos de padrões existentes nos repositórios disponíveis no diretório.



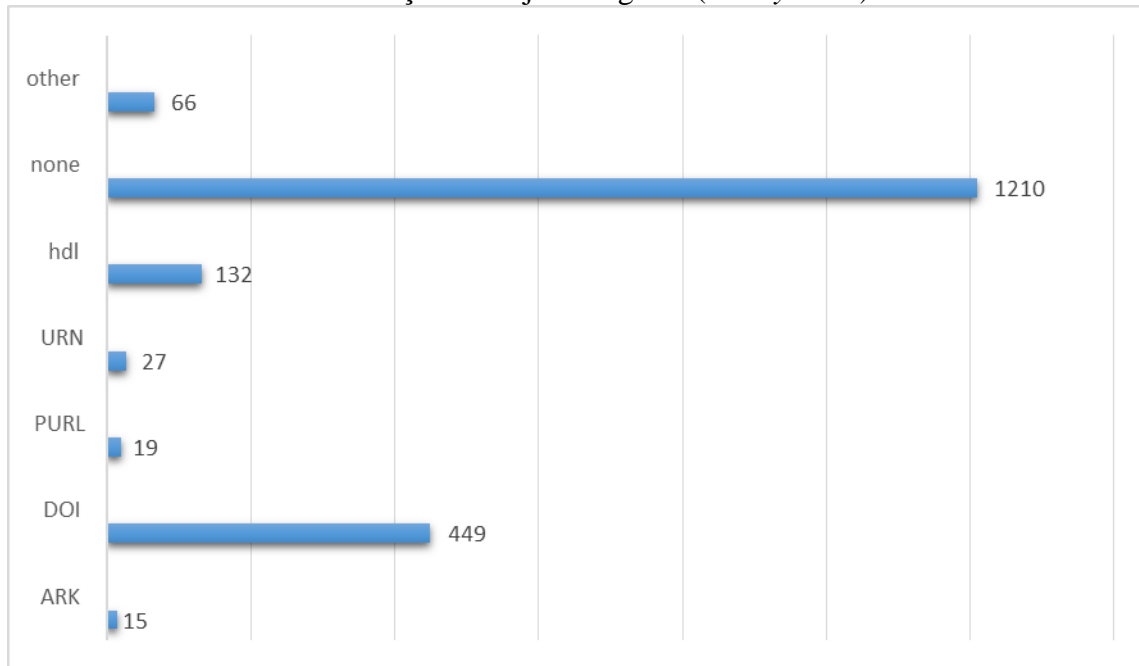
Gráfico 10 – Quantidade de repositórios de dados de pesquisa quanto aos padrões de metadados (*Metadata Standards*)



Fonte: Informações coletadas no diretório re3data.org

Selecionando-se a opção “*PID systems*” foram identificados os repositórios de dados de pesquisa e o uso de sistemas de identificação de objetos digitais. Esse recurso permite não apenas a identificação e a localização, mas também a recuperação e possibilita as referências cruzadas, ou seja, remete aos objetos digitais que estão relacionados a um texto específico por meio de um endereço Web. O gráfico 11 apresenta as opções de identificadores de objetos digitais encontrados no diretório.

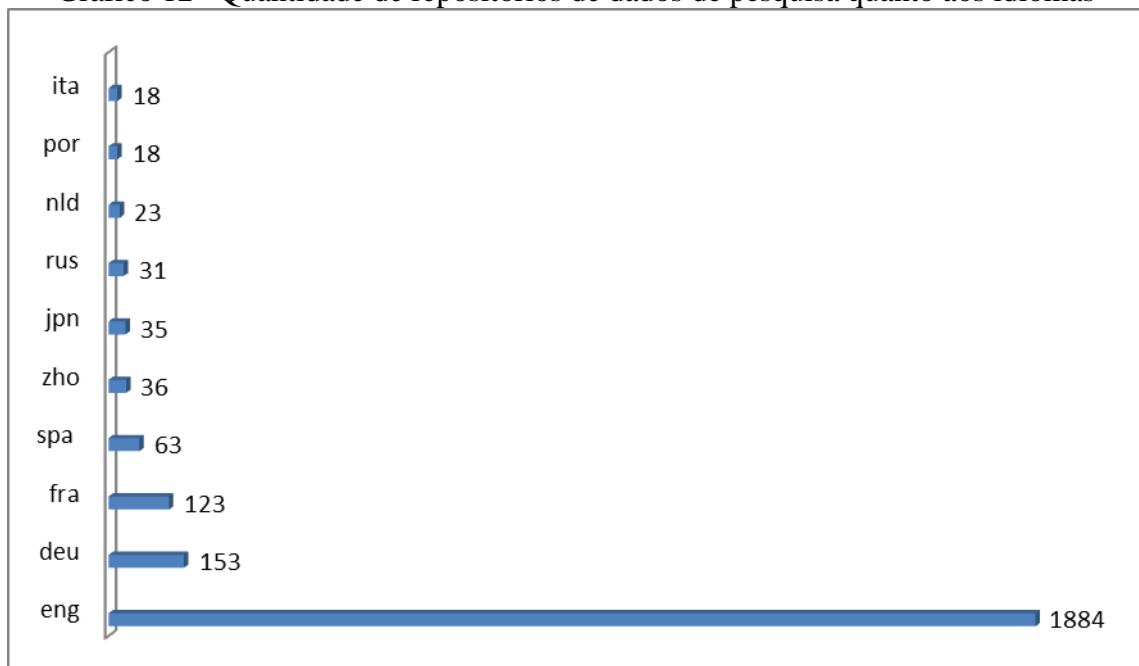
Gráfico 11– Quantidade de repositórios de dados de pesquisa quanto ao uso de sistemas de identificação de objetos digitais (*Pid Systems*)



Fonte: Informações coletadas no diretório re3data.org

Selecionando-se a opção “*Repository languages*” identificou-se o número de repositórios de dados de pesquisa e seus respectivos idiomas. O gráfico 12 demonstra um recorte de dez idiomas com um maior número de repositórios.

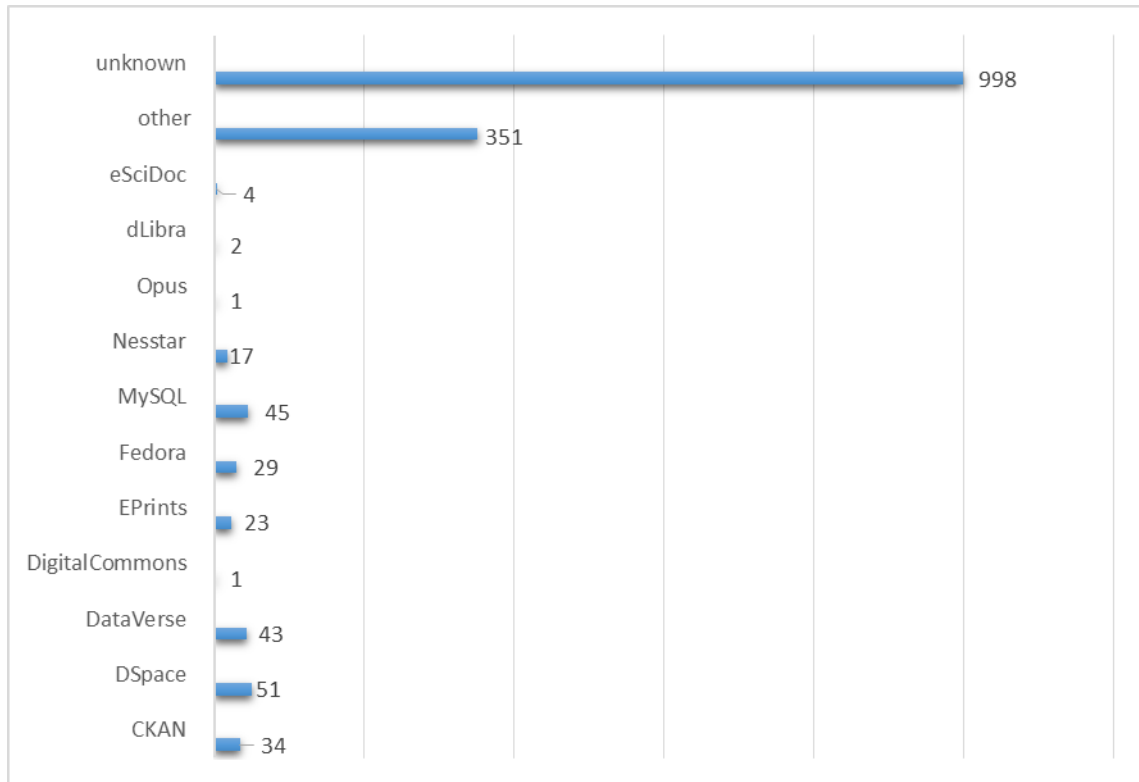
Gráfico 12– Quantidade de repositórios de dados de pesquisa quanto aos idiomas



Fonte: Informações coletadas no diretório re3data.org

Selecionando-se a opção “Software” foram identificados os números de repositórios de dados de pesquisa e os softwares utilizados, conforme mostra o gráfico 13.

Gráfico 13 – Quantidade de repositórios de dados de pesquisa e os respectivos softwares utilizados



Fonte: Informações coletadas no diretório re3data.org

Selecionando-se a opção “*Repository types*” identificou-se a quantidade repositórios de dados de pesquisa e seus respectivos tipos. O quadro 9 demonstra:

Quadro 9 – Quantidade de repositórios de dados de pesquisa e os respectivos tipos de repositórios

<b>TIPOS DE REPOSITÓRIOS</b>	<b>Qtd.</b>
Temático	1647
Institucional	476
Outros	192

Fonte: Informações coletadas no diretório re3data.org

Selecionando-se a opção “*Enhanced publication*” foi possível verificar a quantidade de repositórios de dados de pesquisa que utilizam ou não esse tipo de publicação, conforme o quadro 10.

Quadro 10– Quantidade de repositórios de dados de pesquisa que utilizam ou não as publicações ampliadas

<b>PUBLICAÇÕES AMPLIADAS</b>	<b>Qtd.</b>
Não	184
Desconhecido	1030
Sim	654

Fonte: Informações coletadas no diretório re3data.org

Observa-se que foram recuperados um total de 654 repositórios de dados de pesquisa que utilizam as publicações ampliadas. Foram registrados (APÊNDICE F) os nomes dos respectivos repositórios. Foi recuperado somente 1 repositório de dados de pesquisa indexado como “*Brazil*” que utiliza a publicação ampliada: *GLOBE - Global Collaboration Engine*.

A seleção dos repositórios que serão analisados será descrita na seção 4.3, na terceira fase da análise.

#### 4.3 TERCEIRA FASE DA ANÁLISE

Durante a inspeção das interfaces, foram identificados os elementos que cada repositório contempla a partir dos requisitos estabelecidos por meio da revisão sistemática, para posteriormente iniciar as análises, observações e comparações dos elementos utilizados nos repositórios.

Nesta seção foram apresentados diversos recursos (APÊNDICE H), que constituem os repositórios caracterizados com publicações ampliadas:

Gráfico 14 – Total de repositórios com Publicações Ampliadas recuperados no diretório



Fonte: Dados coletados pela autora

Durante a inspeção das interfaces, foram realizadas 3 (três) tentativas de buscas utilizando as ferramentas de consulta, termos específicos e listas disponíveis em suas coleções de dados de pesquisa. A partir dos resultados encontrados, foram realizadas as análises e descartados, conforme o gráfico 14, um total de 518 repositórios pelos seguintes motivos:

- publicações não relacionadas aos conjuntos de dados de pesquisa;
- páginas com erros e problemas de usabilidade (avisos de falhas);
- dificuldades de acesso ao conjunto de dados;

- d) solicitação de autenticação, registro e *login* para acesso a descrição e ao conjunto de dados;
- e) repositórios que apenas descrevem dados compilados, sem disponibilizar a publicação relacionada;
- f) disponibilização apenas de trabalhos acadêmicos sem dados de pesquisa (conjunto de dados não encontrados);
- g) disponibilização de bases de dados onde estão hospedados os dados (remete para fontes externas). Estas fontes externas solicitam *login* para acesso ao conjunto de dados;
- h) dificuldades de identificar os links que remetem ao conjunto de dados de pesquisa;
- i) repositórios que apenas possuem instruções sobre como citar, organizar e disponibilizar dados de pesquisa;
- j) páginas removidas;
- k) indexação pelo diretório de repositórios que não possuem publicações como também não possuem conjunto de dados (trata-se de um site de games);
- l) repositórios não recuperados no diretório (apesar de estarem disponibilizados na lista de filtros sobre publicação ampliada);
- m) repositórios suspensos;
- n) incompatibilidade com o navegador (foram utilizados os navegadores Mozilla Firefox, Google Chrome e Internet Explorer).

Conforme relatado nesta pesquisa, para que haja uma publicação ampliada, se faz necessária a conexão dos dados com suas respectivas publicações. Essa ligação ilustra e possibilita esclarecer a pesquisa, evidenciando o estudo e garante a identificação da autoria. Esses dados vinculados à publicação que os originou, proporciona o compartilhamento de dados e o aumento da visibilidade da pesquisa. Nesta etapa da análise, constatou-se um grande número de repositórios com publicações não relacionadas aos conjuntos de dados de pesquisa e por este motivo, optou-se em realizar a exclusão dos 518 repositórios que não contemplavam este requisito estabelecido, entre outros motivos já citados.

A partir desta seleção, foram inseridos ao estudo 136 repositórios, sendo identificadas as áreas, conteúdos mais abrangentes entre outros, como será demonstrado na seção 4.3.1.

### 4.3.1 Recursos disponíveis nos repositórios selecionados

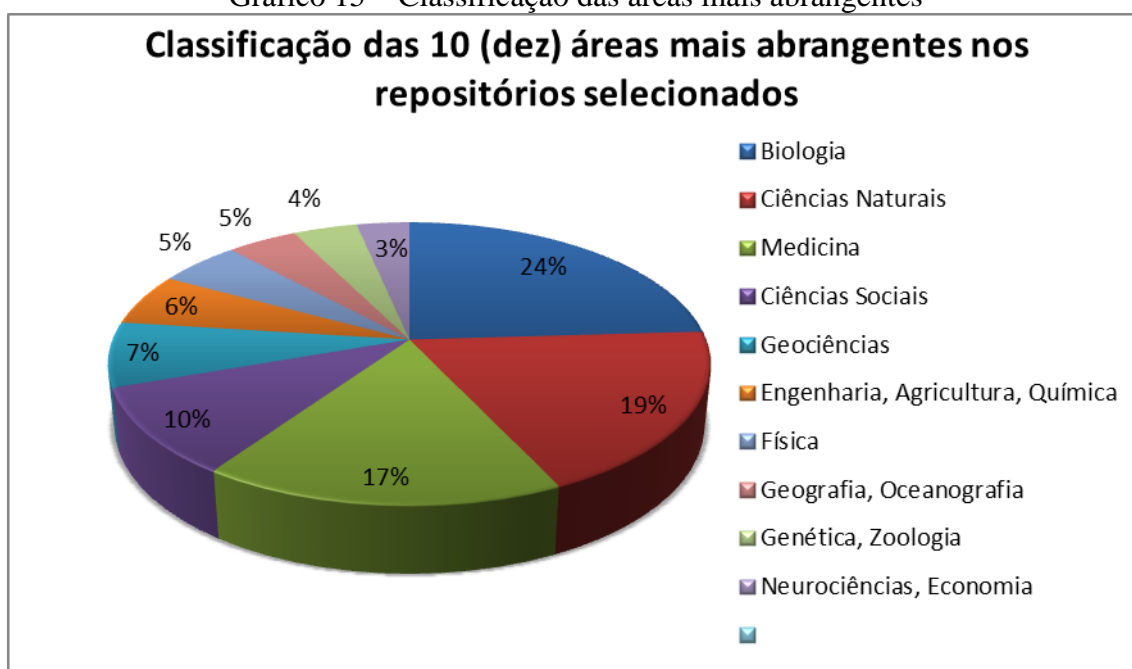
Nesta seção foram relatadas, a partir de informações disponibilizadas pelo diretório, características presentes em cada um dos 136 repositórios inseridos nesta pesquisa. Após a verificação e contagem de cada recurso encontrado, destacou-se: as 10 (dez) áreas e formas mais abrangentes; identificadores digitais de autoria; padrões de programação (API); restrições de acesso aos dados; instituições vinculadas aos repositórios; padrões de metadados; identificadores digitais de objetos digitais; idiomas mais presentes; softwares mais utilizados; tipos de repositórios e países mais destacados.

#### 4.3.1.1 Áreas mais abrangentes

A inspeção foi realizada em cada um dos 136 repositórios inseridos no estudo e foi possível verificar quais áreas do conhecimento se destacaram. De acordo com o gráfico 15, 24% dos repositórios disponibilizavam dados de pesquisas da área da Biologia; 19% na área de Ciências Naturais; 17% na Medicina; 10% em Ciências Sociais; 7% em Geociências. Com 6% estavam as áreas de Engenharia, Agricultura e Química; 5% a área da Física, Geografia e Oceanografia. A Genética e Zoologia tiveram uma taxa de 4% dos repositórios selecionados.

Percebe-se que o maior número de repositórios de dados de pesquisa com publicações ampliadas estão direcionados às áreas da Biologia, Ciências Naturais e Medicina.

Gráfico 15 – Classificação das áreas mais abrangentes



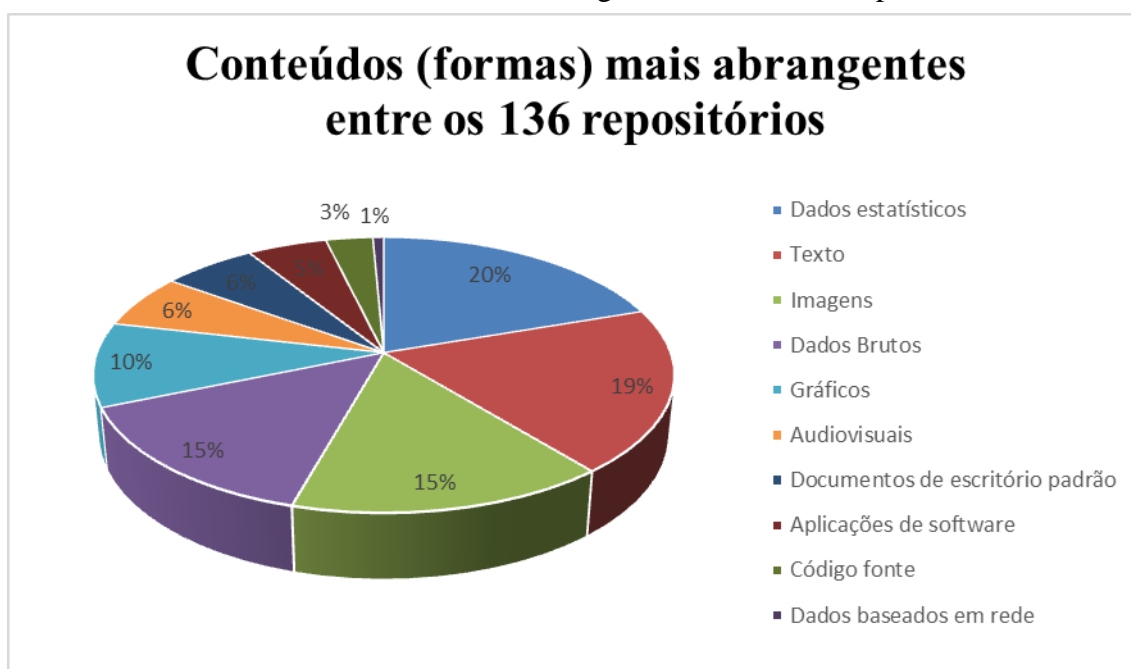
Fonte: Dados coletados pela autora

#### 4.3.1.2 Conteúdos mais abrangentes

Referente aos conteúdos (formas dos dados) mais disponibilizados entre os repositórios com publicações ampliadas, o gráfico 16 mostra os dados estatísticos com 20%; 19% de dados em forma textual; 15% em imagens; 15% de dados brutos (não tratados); 10% em gráficos; 6% em formato de audiovisuais e documentos de escritório; 5% em formato de aplicações de software; 3% em código fonte e apenas 1% de dados baseados em rede.

Observou-se que as formas de dados mais disponibilizadas entre os repositórios de dados de pesquisa com publicações ampliadas são de dados estatísticos, em texto e em imagens.

Gráfico 16 – Conteúdos mais abrangentes entre os 136 repositórios



Fonte: Dados coletados pela autora

#### 4.3.1.3 Identificadores digitais de autoria

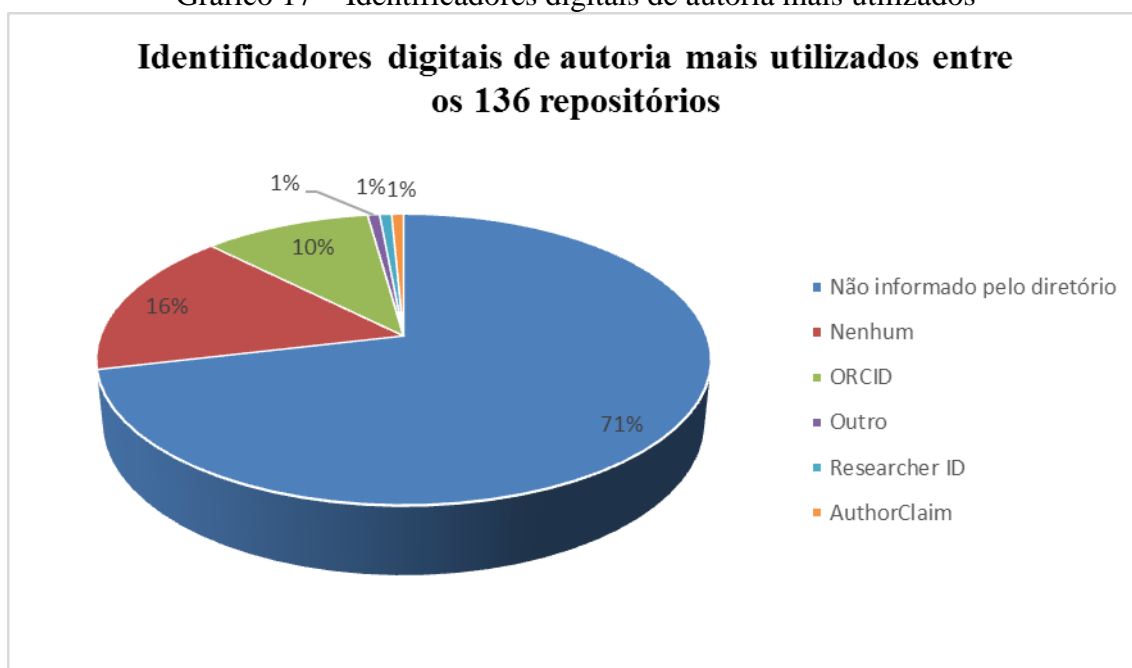
Um identificador de autor é um sistema que atribui um código numérico persistente a um pesquisador e permite diferenciá-lo dos restantes no meio digital. Estes sistemas possuem como objetivo atribuir uma identificação única a um autor e reunir todas as publicações e afiliação através de um código persistente. Grupos editoriais internacionais, agências de fomento, universidades, associações e institutos de pesquisa já utilizam esse sistema, que permite a obtenção de crédito pela pesquisa desenvolvida.



Durante a análise, conforme demonstra o gráfico 17, foram destacados os identificadores mais utilizados. Constatou-se que 16% dos repositórios não utilizam nenhum tipo de identificador; 10% utilizam o ORCID (*Open Researcher and Contributor ID*); 1% utilizam *AuthorClaim*, *Researcher ID* e outros. O diretório não informa a utilização de um identificador para 71% dos repositórios selecionados.

Foi concluído que um grande índice de repositórios de dados de pesquisa com publicações ampliadas ainda não utilizam um identificador digital persistente para autor.

Gráfico 17 – Identificadores digitais de autoria mais utilizados



Fonte: Dados coletados pela autora

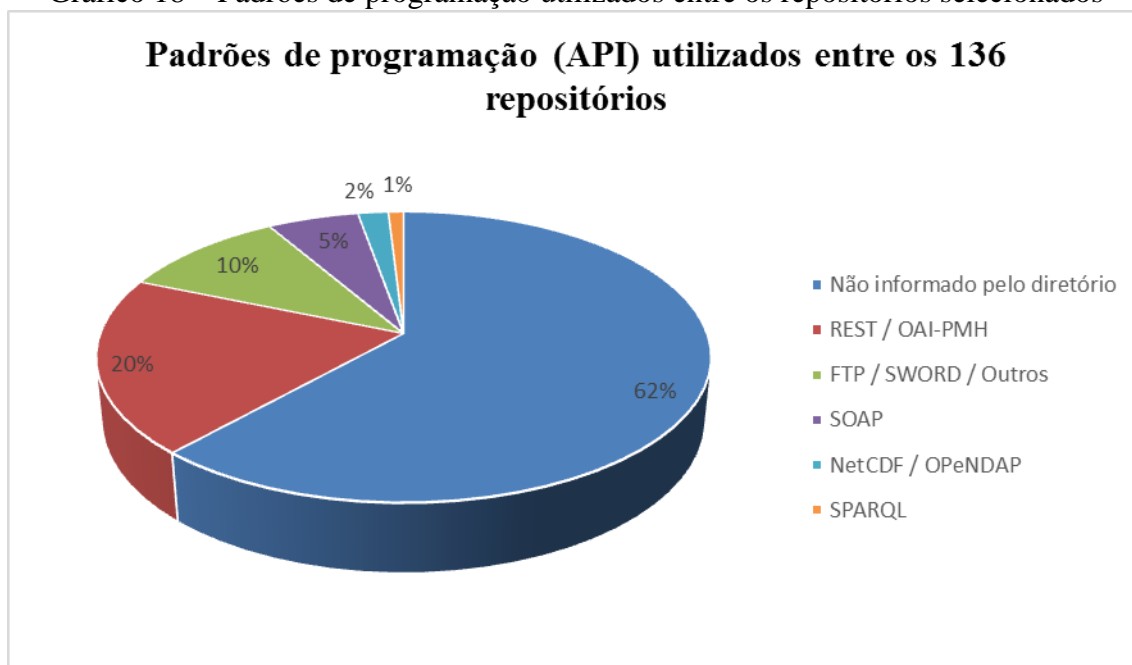
#### 4.3.1.4 Padrões de programação – API

API é considerado um conjunto de rotinas e padrões de programação para acesso a um aplicativo de software ou plataforma baseado na Web. A sigla API refere-se ao termo em inglês *Application Programming Interface* que significa em tradução para o português "Interface de Programação de Aplicativos". Através de APIs, os aplicativos podem se comunicar uns com os outros sem conhecimento ou intervenção dos usuários, funcionando através da comunicação de diversos códigos e definindo comportamentos específicos de determinado objeto em uma interface. (CANALTECH, 2017).

No gráfico 18 é possível observar que 20% dos repositórios utilizam os padrões REST e OAI-PMH; que 10% utilizam os padrões FTP, SWORD e outros não identificados. 5%

utilizam SOAP; 2% NetCDF e OPeNDAP e somente 1% SPARQL. O diretório não disponibiliza informações de 62% dos repositórios selecionados.

Gráfico 18 – Padrões de programação utilizados entre os repositórios selecionados



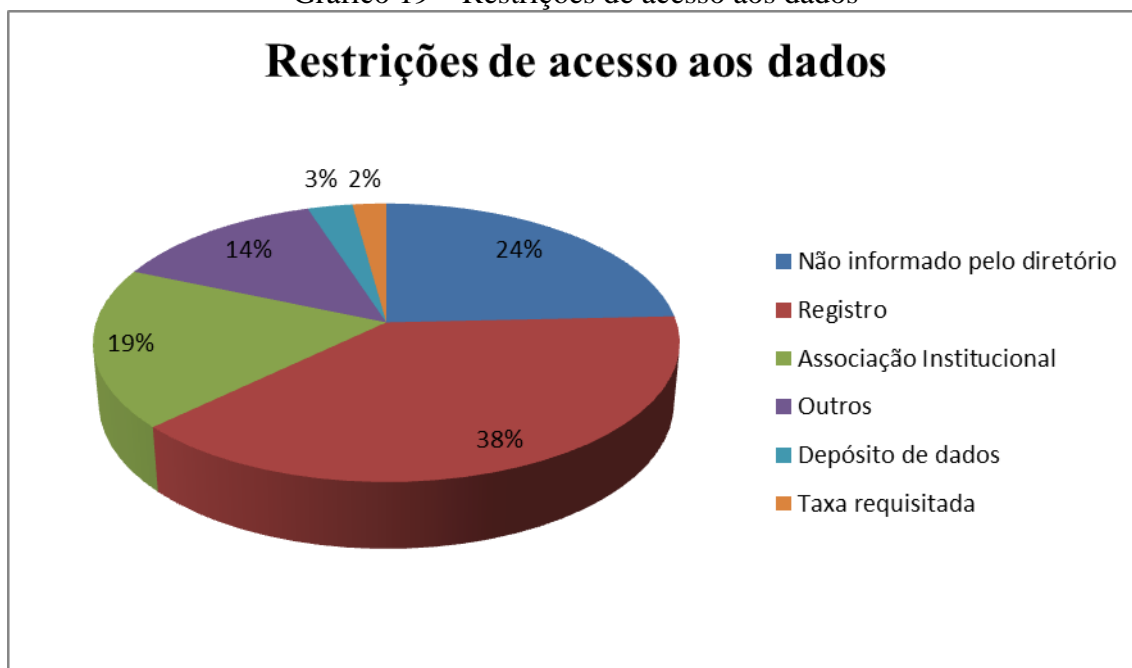
Fonte: Dados coletados pela autora

#### 4.3.1.5 Restrições de acesso aos dados

As restrições de acesso são um conjunto de regras que definem quais registros serão visualizados pelo usuário. Para acessar, pode ser solicitado um registro, alguma taxa requisitada, entre outros.

Durante a análise, e conforme o gráfico 19, constatou-se que 38% dos repositórios selecionados possibilita o acesso por meio de registro; 19% possui acesso através de alguma associação institucional; 3% através de depósito de dados; 2% através de taxa requisitada. O diretório informa 14% para outros tipos de restrições e 24% dos repositórios aparecem como não informado pelo diretório.

Gráfico 19 – Restrições de acesso aos dados

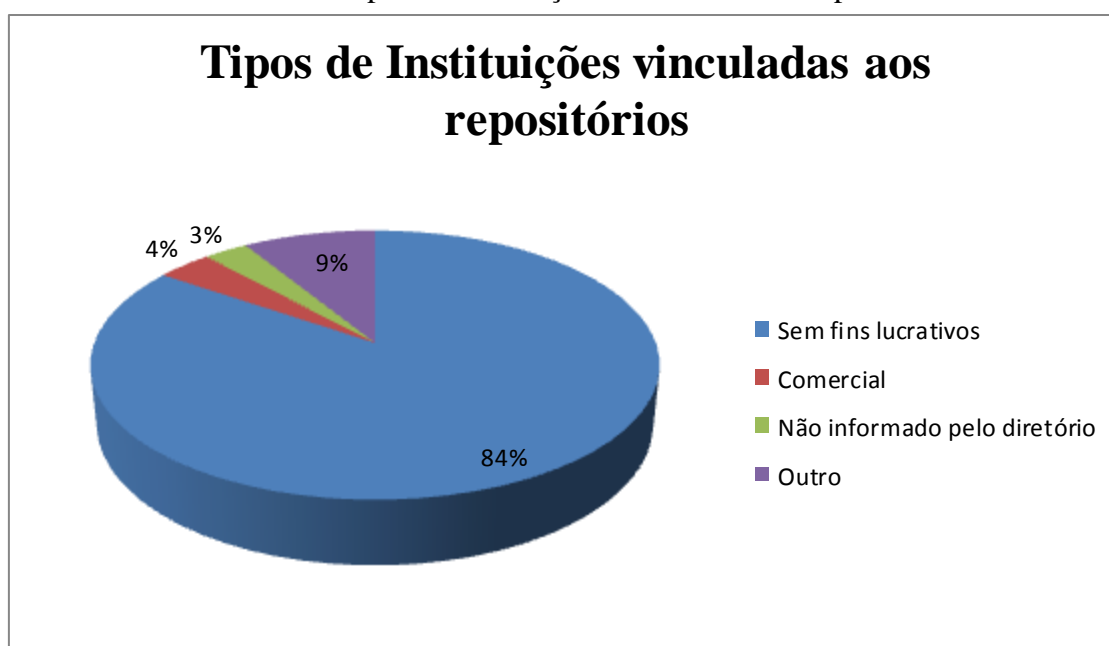


Fonte: Dados coletados pela autora

#### 4.3.1.6 Tipos de instituições vinculadas

Conforme o gráfico 20, 84% dos repositórios estão vinculados a instituições sem fins lucrativos; 4% a instituições comerciais e 9% a outros tipos. 3% não são informados pelo diretório.

Gráfico 20 – Tipos de Instituições vinculadas aos repositórios

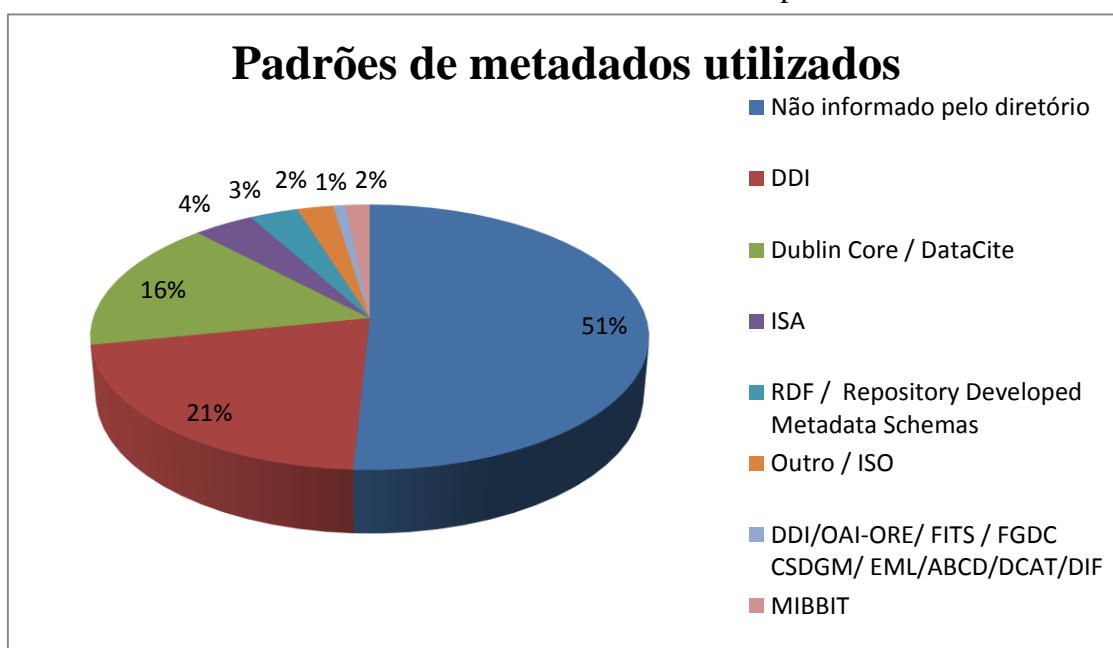


Fonte: Dados coletados pela autora

#### 4.3.1.7 Padrões de metadados

Metadados são considerados um conjunto de elementos planejados para facilitar a descrição de recursos eletrônicos, existindo uma variedade de tipos e padrões. Desta forma, se faz necessário conhecer as particularidades e diferenças entre suas estruturas e níveis de especificidade. Neste estudo, apenas foram destacados alguns tipos utilizados entre os 136 repositórios de dados de pesquisa que possuem publicações ampliadas, conforme é possível visualizar no gráfico 21. Foi identificado uma taxa de 21% de repositórios que utilizam DDI; 16% utilizam *Dublin Core* e *DataCite*; 4% utilizam *ISA*; apenas 3% utilizam *RDF* e *Repository Developed Metadata Schemas*; 2% utilizam *MIBBIT*, *ISO* e outros tipos não descritos; 1% utilizam *OAI-ORE*; *FITS*, *FGDC*, *CSDGM*, *EML*, *ABCD*, *DCAT* e *DIF*. O diretório não informa os tipos utilizados para 51% dos repositórios selecionados para este estudo.

Gráfico 21 – Padrões de metadados utilizados entre os repositórios selecionados



Fonte: Dados coletados pela autora

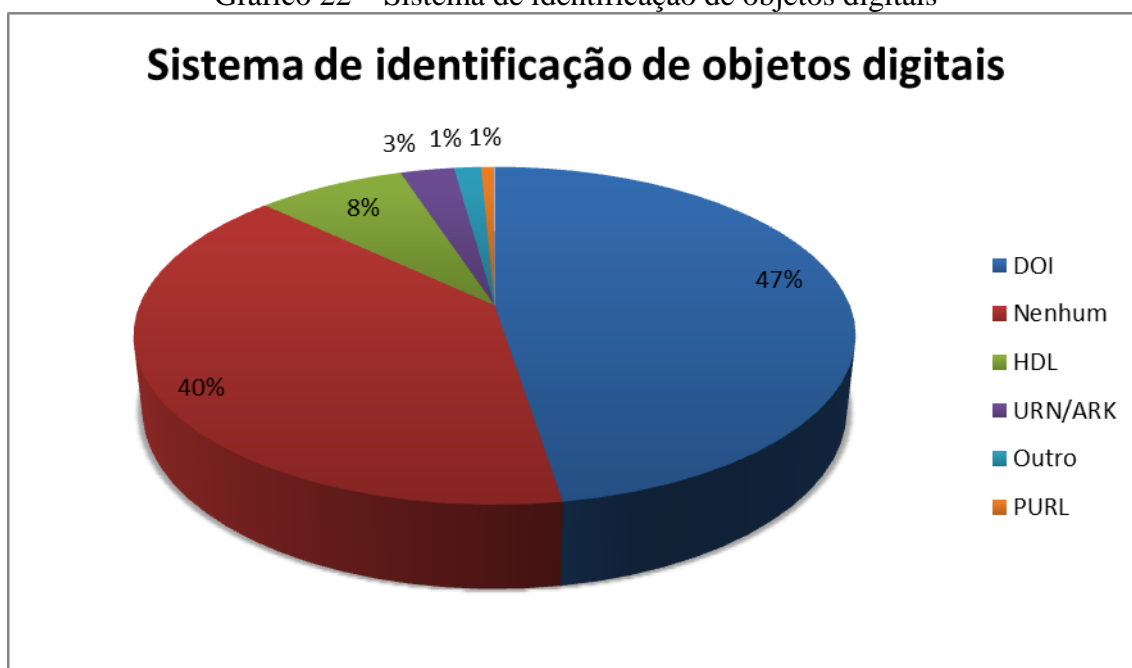
#### 4.3.1.8 Sistema de identificação de objetos digitais

Os identificadores persistentes de objetos digitais auxiliam na localização, recuperação e referências cruzadas de publicações científicas, caracterizando-se por ser um número atribuído a conteúdos em ambiente digital que permanece o mesmo, independente se o documento for movido, ou seja, os links continuarão acionáveis. Existem diversos sistemas de

identificação persistente baseados em padrões abertos, mas com objetivos diferentes. Nesta pesquisa, foram identificados diferentes identificadores utilizados pelos repositórios de dados de pesquisa selecionados, como é possível visualizar no gráfico 22.

Conforme informações do diretório, 47% dos repositórios utilizam o identificador DOI; 40% dos repositórios não utilizam nenhum tipo de identificador persistente; 8% utilizam o HDL; 3% utilizam URN e ARK e a taxa percentual para o PURL e outros tipos foi de 1% dos repositórios selecionados.

Gráfico 22 – Sistema de identificação de objetos digitais

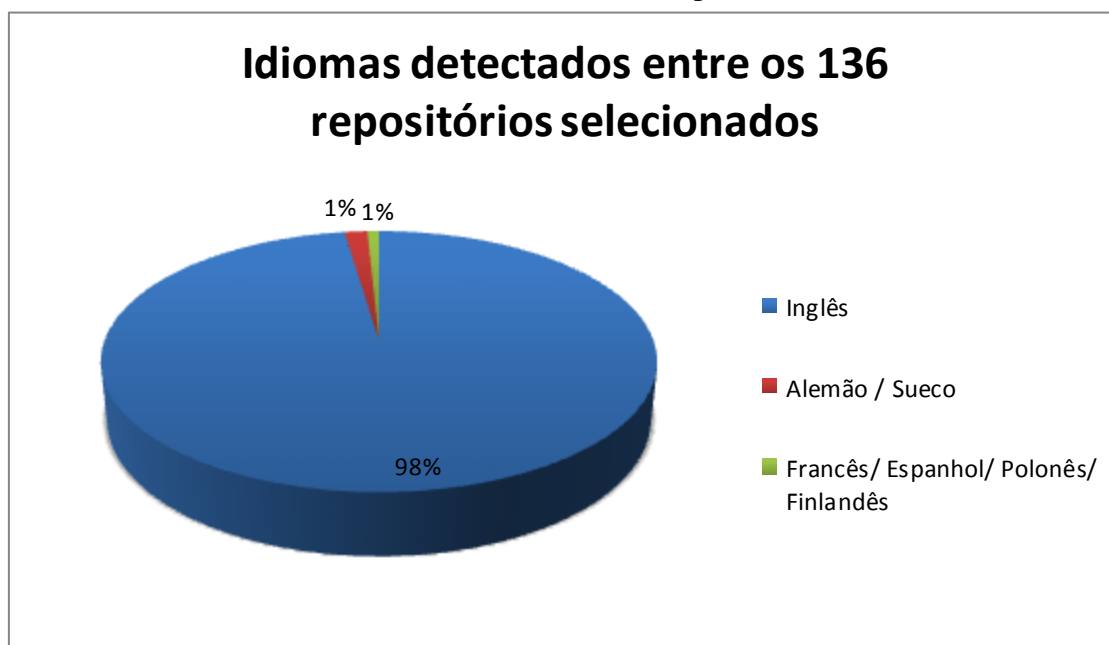


Fonte: Dados coletados pela autora

#### 4.3.1.9 Idiomas detectados

O principal idioma destacado entre os 136 repositórios é o inglês, com 98% e os restantes dos idiomas como, por exemplo, o alemão, Sueco, Francês, Espanhol, Polonês e Finlandês aparece com um percentual de 1% dos repositórios. No gráfico 23, é possível observar esses dados.

Gráfico 23 – Idiomas detectados entre os repositórios selecionados

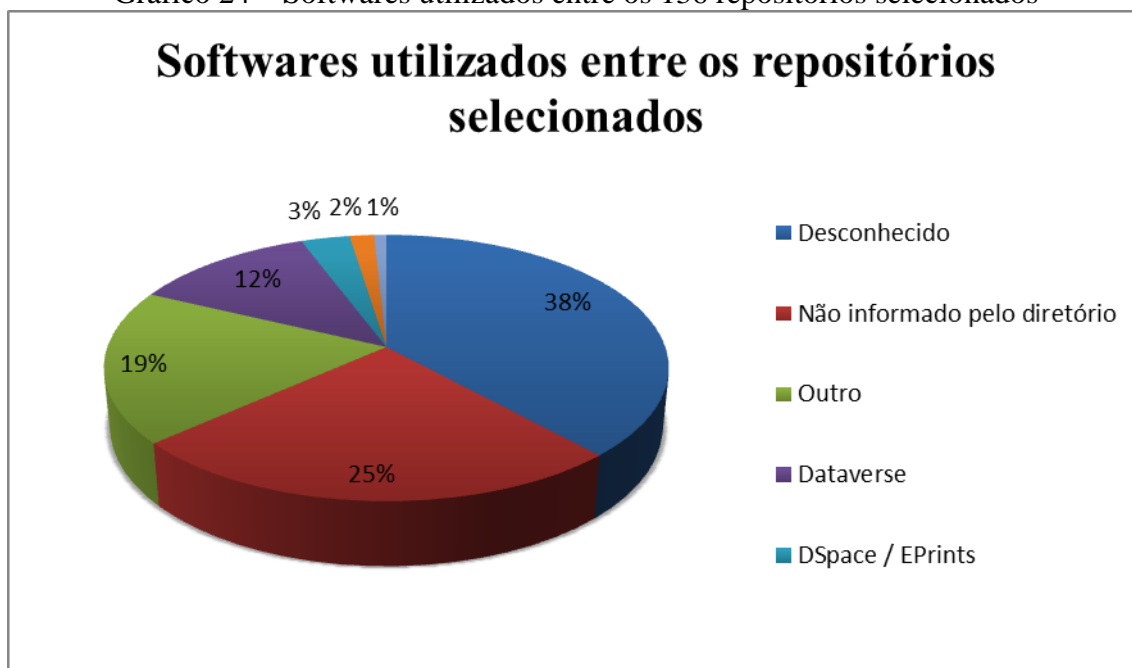


Fonte: Dados coletados pela autora

#### 4.3.1.10 Softwares mais utilizados

Existem diferentes softwares desenvolvidos para a criação de repositórios digitais, com funções de armazenamento, gerenciamento, preservação e disponibilização da produção intelectual. Na análise, conforme pode ser observado no gráfico 24, 12% dos repositórios utilizam o *Dataverse*; 3% utilizam o *DSpace* e *EPrints*. O diretório informa 38% de softwares desconhecidos, 25% não informado e 19% outros tipos não especificados.

Gráfico 24 – Softwares utilizados entre os 136 repositórios selecionados

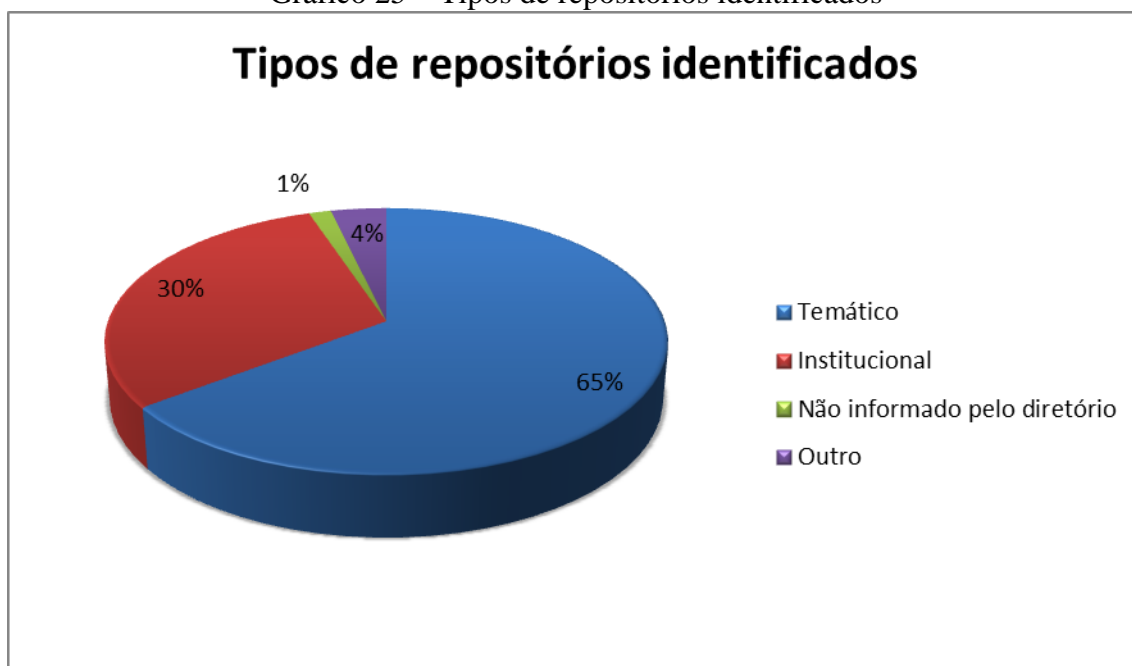


Fonte: Dados coletados pela autora

#### 4.3.1.11 Tipos de repositórios

Os tipos de repositórios identificados neste estudo foram os temáticos, com 65%; os institucionais com 30%; outros não especificados 4%; e não informado pelo diretório, 1%. Conforme pode ser visualizado no gráfico 25, os repositórios temáticos e institucionais são os que mais se destacam entre os repositórios com publicações ampliadas.

Gráfico 25 – Tipos de repositórios identificados

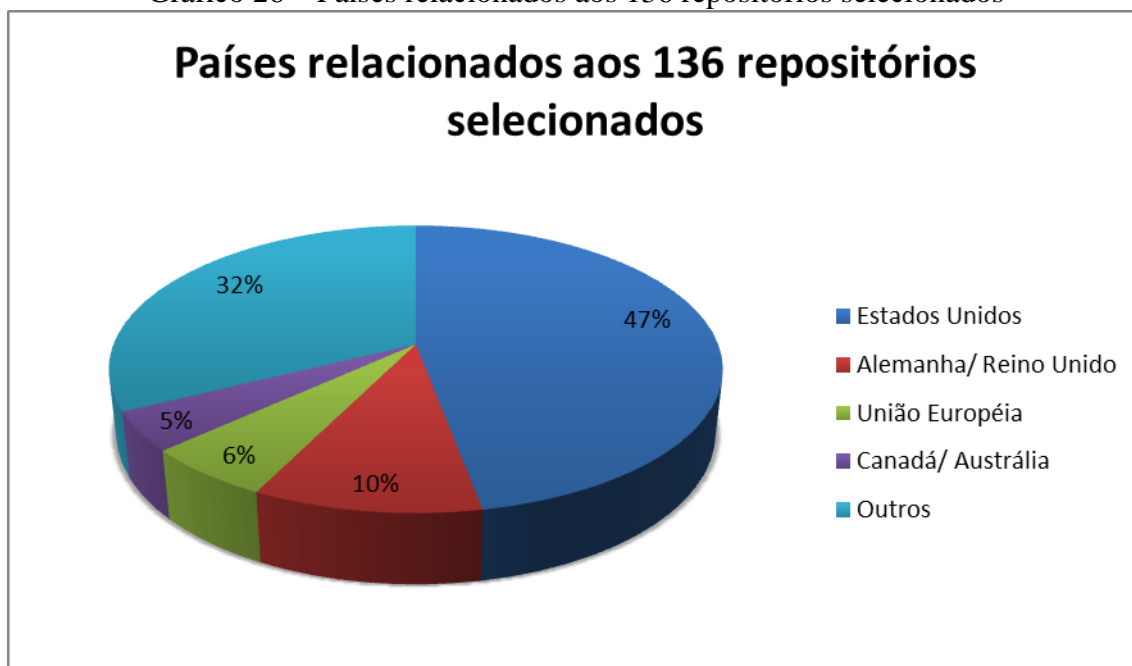


Fonte: Dados coletados pela autora

#### 4.3.1.12 Países destacados

Conforme o gráfico 26, 47% dos repositórios são provenientes dos Estados Unidos; 10% da Alemanha e Reino Unido; 6% da União Europeia; 5% do Canadá e Austrália. Entre outros países há um percentual de 32% dos repositórios inseridos no estudo.

Gráfico 26 – Países relacionados aos 136 repositórios selecionados



Fonte: Dados coletados pela autora

#### 4.3.2 Requisitos de publicação ampliada definidos na pesquisa

Após a revisão sistemática, foram definidos 16 (dezesseis) requisitos de publicação ampliada (APÊNDICE E) para a proposta deste estudo. As interfaces dos 136 repositórios foram inspecionadas para a verificação da ligação de seus dados e se contemplavam os 16 (dezesseis) requisitos.

As observações de cada um dos repositórios foram registradas no checklist de validação de requisitos/ inspeção (APÊNDICE E). Posteriormente, foi realizada a estimativa para verificar a quantidade de repositórios que estavam de acordo com a infraestrutura e padrões associados às publicações ampliadas propostos nesta pesquisa. Também foram armazenados os *screenshots* das interfaces dos repositórios que contemplam o maior número de requisitos de publicação ampliada, para que fosse possível realizar as comparações e demonstrar as funcionalidades.



O objetivo da demonstração das diferenças entre interfaces de repositórios, que contemplam ou não os requisitos de publicação ampliada destacados nesta pesquisa, foi provar que estes requisitos são essenciais e que deveriam ser incorporados nas interfaces dos repositórios de dados de pesquisa. O que se propôs foi investigar quais elementos são importantes constar nas interfaces de repositórios de dados de pesquisa para que se tenha, de fato, um modelo de publicação ampliada que possibilite a ligação de uma publicação tradicional aos dados de pesquisa.

Nas seções seguintes serão demonstrados os resultados e o detalhamento de cada requisito existente ou inexistente nos repositórios inseridos no estudo.

#### 4.3.2.1 Requisito 1 – Recursos web identificados por URI

É aconselhável certificar-se de que manuscritos e dados de pesquisa que são depositados em repositórios também se tornam parte da arquitetura web, o que implica que todos esses objetos estarão disponíveis como recursos da web que podem ser referenciados através de um *Uniform Resource Identifier* (URI<sup>7</sup>) e identificados como uma imagem, página, etc. Essas especificações foram analisadas em cada um dos repositórios.

*Uniform Resource Name* (URN) é o nome uniforme de recurso, ou seja, é o que define a identidade de um item. *Uniform Resource Locator* (URL) é o localizador uniforme de recurso, ou seja é o método utilizado para encontrar um item específico. Um exemplo de uma estrutura de URI, conforme Oliveira ([201-]), seria:

<http://woliveiras.com.br/desenvolvedor-front-end/>

Pode-se notar que a URI une o Protocolo (<http://>), a localização do recurso (URL - [woliveiras.com.br](http://woliveiras.com.br)), e o nome do recurso (URN - [/desenvolvedor-front-end/](http://woliveiras.com.br/desenvolvedor-front-end/)).

Após a contagem, nota-se que praticamente a totalidade dos repositórios possuem o requisito 1, ou seja, os recursos web identificados por URI.

---

<sup>7</sup> “[...] é uma cadeia de caracteres compacta usada para identificar ou denominar um recurso na internet. O principal propósito desta identificação é permitir a interação com representações do recurso por intermédio de uma rede, tipicamente a internet, usando protocolos específicos.” Disponível em: <<https://www.portaleducacao.com.br/conteudo/artigos/informatica/detalhando-as-diferencas-entre-uri-e-url/20377>>. Acesso em: 05 set. 2017.

Como exemplo, utilizou-se alguns *screenshots* de repositórios que contemplaram o maior número de requisitos estipulados nesta pesquisa. A partir da definição de URI, ou seja, uma cadeia de caracteres usada para identificar um recurso na internet permitindo interações através de uma rede, foi destacado na interface de um dos repositórios esse detalhamento. O repositório utilizado como exemplo possui o nome de “e-cienciaDatos”, vinculado às instituições na Espanha sem fins lucrativos. Observou-se que este repositório utiliza o software *Dataverse* e há a união do localizador do recurso (URL) com o nome do recurso (URN) que formam o URI, como pode ser visualizado na figura 11.

Figura 11 – Exemplo de repositório com recursos web identificados por URI

The screenshot shows a web browser window with the address bar displaying the URL: <https://edatos.consociomadrono.es/dataset.xhtml?persistentId=doi:10.21950/ZPRADY>. The page header includes the logo for 'Madroño CONSORCIO BETA' and 'ciencia Datos'. The main content area features the 'Dataverse Repositorio de Datos UC3M' logo and the title of the dataset: 'GAMS Transient Stability-Constrained Optimal Power Flow Model of the IEEE 39 Bus Test Case Including Non-Synchronous Generation'. Below the title, there is a download button showing '5 Descargas' and a citation button labeled 'Citar Dataset'. A detailed description of the dataset is provided in a table format.

<b>Descripción</b>	This file is a supplementary material of the paper "Optimal curtailment of non-synchronous renewable generation in the island of Tenerife considering steady state and transient stability constraints". It contains a transient stability-constrained optimal power flow model written in GAMS. The model represents the IEEE 39 Bus Test System with the addition of ten non-synchronous generators connected at different buses. To execute the file, open it in GAMS or type "gams tscopf_nonsyn.gms" from the command line.
<b>Materia</b>	Ingeniería
<b>Palabra Clave</b>	Power system control , Power system transient stability, Optimal Power Flow, Ingeniería Eléctrica

Fonte: *Screenshot* da interface do repositório

#### 4.3.2.2 Requisito 2 – Identificador persistente para cada objeto

Uma prática recomendada é associar identificadores globais únicos e persistentes com cada um dos recursos. Para uma instituição que utiliza as publicações ampliadas sugere-se associar à publicação e ao conjunto de dados um identificador persistente, pois irá garantir

que os dados publicados possam ser citados (VERHAAR, 2009). Entre os 136 repositórios inseridos no estudo, 81 destes apresentam identificadores persistentes para objetos digitais.

Foi possível visualizar o identificador persistente na maior parte dos repositórios selecionados. Como mostra a figura 12, o DOI é o identificador utilizado no repositório e verificou-se que está relacionado ao conjunto de dados e não a cada objeto digital disponível para download.

Figura 12 – Exemplo de repositório com identificador persistente

The screenshot shows a web browser displaying a dataset page from the 'Madroneo Consorcio' repository. The URL is <https://edatos.consociomadroneo.es/dataset.xhtml?persistentId=doi:10.21950/ZPRADY>. The page features a red header with the repository logo and navigation links. Below the header, there are tabs for 'Ficheros', 'Metadatos', 'Condiciones', and 'Versiones'. A red button labeled 'Exportar Metadatos' is visible on the right. The main content area is titled 'Metadatos de Cita' and contains the following information:

ID Persistente del Dataset	doi:10.21950/ZPRADY
Fecha de Publicación	2017-11-14
Título	GAMS Transient Stability-Constrained Optimal Power Flow Model of the IEEE 39 Bus Test Case Including Non-Synchronous Generation
Autor	Ledesma Larrea, Pablo (Universidad Carlos III de Madrid) - ORCID: 0000-0002-6637-7498 Arrendondo Rodríguez, Francisco (Universidad Carlos III de Madrid) - ORCID: 0000-0001-7447-6325 Castronuovo, Edgardo Daniel (Universidad Carlos III de Madrid) - ORCID: 0000-0003-2292-4928
Contacto	Utilice el botón de e-mail de arriba para contactar. Ledesma Larrea, Pablo (Universidad Carlos III de Madrid)
Descripción	This file is a supplementary material of the paper "Optimal curtailment of non-synchronous renewable generation in the island of Tenerife considering steady state and transient stability constraints". It contains a transient stability-constrained optimal power flow model written in GAMS. The model represents the IEEE 39 Bus Test System with the addition of ten non-synchronous generators connected at different buses. To execute the file, open it in GAMS or type "gams tscopf_nonsyn.gms" from the command line. (2017)

Fonte: *Screenshot* da interface do repositório

#### 4.3.2.3 Requisito 3 – Vincular objetos digitais à publicação

Os dados de pesquisa são caracterizados como elementos digitais, ou seja, podem ser representados com um único bit<sup>8</sup>, no entanto, esses dados podem ser agrupados em um objeto composto maior. Esses conjuntos de dados compostos consistem em vários arquivos de dados e de múltiplos registros de metadados e podem ser adicionados à publicação. As publicações ampliadas são, portanto, objetos altamente complexos, que agrupam os textos de artigos de revistas às imagens, arquivos de vídeos etc. (VERHAAR, 2009). Verificou-se que 120 repositórios estão de acordo com o requisito número 2, ou seja, vinculam objetos digitais às publicações.

<sup>8</sup> É a menor unidade de medida de transmissão de dados usada na computação e informática. Disponível em: <<https://www.tecmundo.com.br/programacao/227-o-que-e-bit-.htm>>. Acesso em: 05 set. 2017.

Nos repositórios de dados de pesquisa com publicações ampliadas, as publicações relacionadas estão disponíveis de forma visível nas interfaces dos repositórios que contemplaram o requisito número 3. No caso do repositório utilizado como exemplo, este contemplava os 16 requisitos propostos pela pesquisa. Na figura 13 é possível observar o vínculo do conjunto de dados e a publicação relacionada através de links.

Figura 13 – Exemplo de repositório com publicações vinculadas

The screenshot shows a web browser window with the URL <https://edatos.consortiomadrono.es/dataset.xhtml?persistentId=doi:10.21950/MJFM2>. The page header features the Madroño Ciencia Datos logo and navigation links: 'Acerca de', 'Guías', 'Soporte', and 'Iniciar Sesión'. The main content area displays a dataset description in Spanish, detailing an experiment on liquid bridges. Below the description, there are three sections: 'Materia' (Ingeniería; Física), 'Palabra Clave' (Spacelab, Microgravity, Experiment, Liquid bridge, Capillarity, Fluid mechanics), and 'Publicación Relacionada' (I. Martínez, J.M. Perales, Experimental analysis of the breakage of a liquid bridge under microgravity conditions, Advances in Space Research, Volume 48, Issue 3, 3 August 2011, Pages 465-471, ISSN 0273-1177, <http://dx.doi.org/10.1016/j.asr.2011.03.012>). A black arrow points from the 'Publicación Relacionada' label to the publication title. At the bottom, there are tabs for 'Ficheros', 'Metadatos', 'Condiciones', and 'Versiones'.

Fonte: *Screenshot* da interface do repositório

#### 4.3.2.4 Requisito 4 – Registro e controle de versões da publicação e das partes constituintes

Recomenda-se o registro e controle de versões da publicação e das suas partes constituintes. As publicações ampliadas são recursos dinâmicos e podem ser adicionados, atualizados e removidos e, por este motivo, gerentes de repositórios de dados devem referir as versões específicas do objeto composto. Será útil registrar a data da última modificação, no caso de recém-criado, esta data coincidirá com a data da criação. (VERHAAR, 2009). Após a contagem observou-se que dentre os 136 repositórios, 77 apresentam o requisito.

Nota-se, na figura 14, um exemplo de repositório que possui o controle das versões da publicação dos dados de pesquisa.

Figura 14 – Exemplo de repositório com registro e controle de versões da publicação

Seguro | <https://edatos.consociomadrono.es/dataset.xhtml?persistentId=doi:10.21950/MJFJM2>

**Madroño** BETA  
CONSORCIO  
ciencia  
datos

Palabra Clave: Spacelab, Microgravity, Experiment, Liquid bridge, Capillarity, Fluid mechanics

Publicación Relacionada: I. Martínez, J.M. Perales, Experimental analysis of the breakage of a liquid bridge under microgravity conditions, *Advances in Space Research*, Volume 48, Issue 3, 3 August 2011, Pages 465-471, ISSN 0273-1177, <http://dx.doi.org/10.1016/j.asr.2011.03.012>, doi: [10.1016/j.asr.2011.03.012](https://doi.org/10.1016/j.asr.2011.03.012)

Ficheros Metadatos Condiciones **Versiones** ←

Ver Diferencias

<input type="checkbox"/>	<a href="#">3.1</a>	<b>Metadatos de Citas Adicionales:</b> (2 Cambiado/a); <a href="#">Ver detalles</a>	Dataverse Admin	noviembre 6, 2017
<input type="checkbox"/>	<a href="#">3.0</a>	<b>Metadatos de Citas:</b> Description (1 Cambiado/a); <b>Ficheros (Añadidos: 1); Condiciones de Uso/Acceso Cambiadas</b> <a href="#">Ver detalles</a>	Jose Ignacio Gonzalez, Dataverse Admin	octubre 9, 2017
<input type="checkbox"/>	<a href="#">2.1</a>	<b>Metadatos de Citas:</b> Author (4 Cambiado/a); <b>Metadatos de Citas Adicionales:</b> (2 Añadido/a, 1 Cambiado/a); <a href="#">Ver detalles</a>	Jose Ignacio Gonzalez	noviembre 30, 2016
<input type="checkbox"/>	<a href="#">2.0</a>	<b>Ficheros (Añadidos: 9Eliminados: 7);</b> <a href="#">Ver detalles</a>	Jose Ignacio Gonzalez	noviembre 28, 2016
<input type="checkbox"/>	<a href="#">1.0</a>	Esta es la primera versión publicada.	Dataverse Admin, Jose Ignacio Gonzalez	noviembre 25, 2016

Fonte: *Screenshot* da interface do repositório

#### 4.3.2.5 Requisito 5 – Registro das propriedades básicas da publicação e dos recursos que são adicionados

Recomenda-se registrar propriedades básicas da publicação e dos recursos que são adicionados a ele (títulos, breve descrição, data da última modificação, data de criação, etc.). Para conjuntos de dados atômicos ou compostos, uma breve descrição pode ser dada. É aconselhável fornecer um título, o que torna explícito que estes recursos particulares são objetos de dados. (VERHAAR, 2009). Verificou-se que 108 repositórios estavam de acordo com o requisito 5.

Observa-se que no repositório utilizado como exemplo, as propriedades básicas da publicação e dos recursos que são adicionados a ele (títulos, breve descrição, data da última modificação, data de criação, etc.) estão visíveis na interface, conforme mostram as figura 15 e 16.

Figura 15 – Exemplo de repositório com registro das propriedades básicas da publicação

Seguro | <https://edatos.consorciomadrono.es/dataset.xhtml?persistentId=doi:10.21950/MJFJM2>

**Madroño** CONSORCIO **BETA**  
**ciencia** Datos

82 Descargas

**Experiment STABILITY of liquid COLUMNS (STACO) on the Advanced Fluid Physics Module (FPM) aboard Spacelab-D2**

Martínez Herranz, Isidoro; Perales Perales, José Manuel; Meseguer Ruiz, José; Ulrich Walter, 2016, "Experiment STABILITY of liquid COLUMNS (STACO) on the Advanced Fluid Physics Module (FPM) aboard Spacelab-D2", doi:10.21950/MJFJM2, Dataverse e-cienciaDatos, V3 Citar Dataset

Revise los Estándares de Citas de Datos.

**Descripción**

During the second German Mission of Spacelab (1993, Spacelab D-2) an experiment on the mechanical stability of liquid columns was performed. A liquid column is a liquid mass held between to solid (circular) supports. A 30 mm in diameter and 90 mm long liquid drop was established in the Advanced Fluid Physics Module (AFPM) and was subjected to different mechanical stimuli (rotation of the supports, oscillation of one of the supports, change in volume and in length) to check whether the resulting drop was stable or not. Results were obtained in terms of images and recorded in a video sequence. Telemetry of the different stimuli and parameter values was also recorded. Objective: The aim of this experiment was to measure the outer shape deformation of long liquid bridges near their stability limit under microgravity, caused by g-jitters and by some controlled mechanical disturbances (change of geometry, change of volume, rotation and vibration). The liquid used was a silicone-oil 10 times more viscous than water (5 times for the last run). The working length of the liquid column was 85 mm. The two solid supports are made of aluminium, of 30 mm in diameter, with a sharp cutback (30 deg. edge) to prevent liquid spreading over the edges. This choice of geometry allows a direct comparison with other TEXUS experiments where two discs of 30 mm in diameter and 86 mm apart were used to hold a cylindrical liquid column (35 mm discs were used in SL-D1 and 40 mm discs on SL-1). A top priority was to verify the results of a previous experiment performed on SL-D1, and because an equivalent Bond number  $Bo=0.007$  was deduced from this experiment and there was no reason to expect a different behaviour, it was planned, in a second run to make use of unequal discs of 30 mm and 28 mm in a second run to precisely counterbalance the expected deformation and better quantify this effect. Procedures: The actual procedures executed were: -- Run 1: background sensing, disc oscillation at 0.40 Hz, 0.41 Hz, 0.42 Hz and again 0.42 Hz. -- Run 2: background sensing and disc separation until breakage of the column. -- Run 3: background sensing. Results: An example of free oscillations of a liquid column (diameter 30 mm, length 85 mm, volume 60 cc), due to g-jitter, is shown in Video1, recorded during the background sensing phase of Run 2. Video2 shows the breakage

Fonte: Screenshot da interface do repositório

Figura 16 – Exemplo de repositório com registro das propriedades básicas dos recursos adicionados

Seguro | <https://edatos.consorciomadrono.es/dataset.xhtml?persistentId=doi:10.21950/MJFJM2>

**Madroño** CONSORCIO **BETA**  
**ciencia** Datos

43 Ficheros Descargar

	<b>00_INFORMATION_OF_FILES_MP4_ODT_M_DAT.txt</b> Plain Text - 1.3 KB - 28-nov-2016 - 10 Descargas MD5: 929fe5c639e8478e790fed67ce50fc5 INFORMATION ABOUT FILES: NAME, FORMAT, ETC <span>Data</span>	<span>Descargar</span>
	<b>112_1416 ods</b> Unknown - 36.4 KB - 25-nov-2016 - 4 Descargas MD5: 70d9e1f93a0d8e0e2619714e123e7e06 Telemetria <span>Data</span>	<span>Descargar</span>
	<b>118_09_50 ods</b> Unknown - 221.7 KB - 25-nov-2016 - 4 Descargas MD5: 62df590726513c801613d11d7f8f767 Telemetria <span>Data</span>	<span>Descargar</span>
	<b>118_12_55 ods</b> Unknown - 188.4 KB - 25-nov-2016 - 2 Descargas MD5: 04f52385cfb74ef6047c35f7d9d74886 Telemetria <span>Data</span>	<span>Descargar</span>
	<b>118_1419 ods</b> Unknown - 295.7 KB - 25-nov-2016 - 3 Descargas MD5: 62e74497ee5db00d11260b530d8ea827 Telemetria <span>Data</span>	<span>Descargar</span>

Fonte: Screenshot da interface do repositório

#### 4.3.2.6 Requisito 6 – Vocabulário controlado e padronizado

As publicações ampliadas podem ser interoperáveis utilizando-se vocabulário controlado e padronizado. (VERHAAR, 2009). Iniciativas como o *Dublin Core Metadata Initiative*, entre outros modelos já citados nesta pesquisa, oferecem um vocabulário que pode ser usado para registrar a relação entre os objetos digitais. Este requisito foi identificado em 60 repositórios.

A figura 17 demonstra a interface de um repositório que está de acordo com o requisito 6, ou seja, possui vocabulário controlado e padronizado. As análises desse requisito foram concluídas a partir da observação do uso de metadados como, por exemplo, o *Dublin Core*, DDI, *Data Cite* e pela lista de termos disponibilizada na interface. Na figura 18 há o exemplo de descrição dos metadados utilizados pelo repositório

Figura 17 – Exemplo de repositório que possui vocabulário controlado e padronizado.

The screenshot shows the Madroño repository interface. At the top, there is a red navigation bar with the logo and name 'Madroño BETA' and 'Ciencia Datos'. Below the navigation bar, the URL '10.1016/j.asr.2011.03.012' is displayed. A horizontal menu contains tabs for 'Ficheros', 'Metadatos', 'Condiciones', and 'Versiones'. The 'Metadatos' tab is active, showing a table of metadata for a dataset. A dropdown menu is open, showing options for exporting metadata: 'Dublin Core', 'DDI', 'DataCite Metadata Schema 4.0', and 'JSON'. An arrow points from the 'Metadatos de Cita' header to the dropdown menu.

Metadatos de Cita	
ID Persistente del Dataset	doi:10.21950/MJFJM2
Fecha de Publicación	2016-11-25
Título	Experiment STAbility of liquid COlumnS (STACO) on the Advanced Fluid Physics Module (FPM) aboard Spacelab-D2
Subtítulo	liquid bridge under microgravity
Título Alternativo	SL-D2-WL-AFPM-STACO
Otro ID	European Space Agency (ESA): 0000 0001 2163 0931 0000 0004 0623 6946
Autor	Martínez Herranz, Isidoro (UPM) - ORCID: 0000-0003-0994-5898 Perales Perales, José Manuel (UPM) - ORCID: 0000-0001-9873-8115 Meseguer Ruiz, José (UPM) Ulrich Walter (Technische Universität München)
Contacto	Utiliza el botón de e-mail de arriba para contactar

Fonte: *Screenshot* da interface do repositório





Figura 19 – Exemplo de repositório com descrições de conjunto de dados e objetos digitais

The screenshot shows the Madroño Consorcio BETA repository interface. The header includes the logo, 'BETA' label, and navigation links: 'Acerca de', 'Guías', 'Soporte', and 'Iniciar Sesión'. The main content area displays a list of digital objects, each with a checkbox, a file icon, a title, a description, a 'Data' label, and a 'Descargar' button.

Checkbox	File Icon	Title	Description	Label	Download Button
<input type="checkbox"/>		118_12_55 ods	Unknown - 168.4 KB - 25-nov-2016 - 2 Descargas MD5: 04f52385cfb74ef6047c35f7d9d74886	Data	Descargar
<input type="checkbox"/>		118_1419 ods	Unknown - 295.7 KB - 25-nov-2016 - 3 Descargas MD5: 62e74497ee5db0d11260b530d8ea827	Data	Descargar
<input type="checkbox"/>		1995_Martinez_Perales_Meseguer.pdf	Adobe PDF - 3.2 MB - 09-oct-2017 - 0 Descargas MD5: 83714c54e6484815972cc81d58380821	Documentation	Descargar
<input type="checkbox"/>		ColumbiaD2_xvid.mp4	video/mp4 - 70.8 MB - 28-nov-2016 - 5 Descargas MD5: 1cb67fa776b47163ceefe7b3a17aa840	Documentation	Descargar
<input type="checkbox"/>		LI_ramp1_xvid.mp4	video/mp4 - 89.5 MB - 25-nov-2016 - 3 Descargas MD5: df3661aa76d5eff040863d496c244cf5	Data	Descargar

Fonte: Screenshot da interface do repositório

Figura 20 – Exemplo de repositório com descrições de conjunto de dados e objetos digitais

The screenshot shows a detailed view of dataset(s) in the repository. On the left, there is a 'Quick Download' button and a 'Table of Contents' section with links to 'Access Notes', 'Dataset(s)', 'Study Description', 'Citation', 'Funding', 'Scope of Study', 'Methodology', 'Version(s)', 'Related Publications', 'Variables', 'Instructional Resources', 'Utilities', 'Metadata Exports', and 'Download Statistics'. The main content area is titled 'Dataset(s)' and lists five datasets (DS0 to DS5) with their respective documentation and download options.

**Dataset(s)**

**DS0: Study-Level Files**  
Documentation: [User guide.pdf](#)

**DS1: Wave I: In-Home Questionnaire, Public Use Sample - Download All Files (259,5 MB) large file**  
Documentation: [Codebook/Questionnaire.pdf](#) [Frequencies.pdf](#) [User guide.pdf \(MULTI\)](#)  
Download: [SAS](#) [SPSS](#) [Stata](#) [R](#) [ASCII](#) [Excel/TSV](#)  
[ASCII + SAS Setup](#) [SPSS Setup](#) [Stata Setup](#)

**DS2: Wave I: Public Use Contextual Database - Download All Files (9,9 MB)**  
Documentation: [Codebook/Questionnaire.pdf](#) [Frequencies.pdf](#) [User guide.pdf \(MULTI\)](#)  
Download: [SAS](#) [SPSS](#) [Stata](#) [R](#) [ASCII](#) [Excel/TSV](#)  
[ASCII + SAS Setup](#) [SPSS Setup](#) [Stata Setup](#)

**DS3: Wave I: Network Variables - Download All Files (107,4 MB)**  
Documentation: [Codebook/Questionnaire.pdf](#) [Frequencies.pdf](#) [User guide.pdf \(MULTI\)](#)  
Download: [SAS](#) [SPSS](#) [Stata](#) [R](#) [ASCII](#) [Excel/TSV](#)  
[ASCII + SAS Setup](#) [SPSS Setup](#) [Stata Setup](#)

**DS4: Wave I: Public Use Grand Sample Weights - Download All Files (4,5 MB)**  
Documentation: [Codebook/Questionnaire.pdf](#) [Frequencies.pdf](#) [User guide.pdf \(MULTI\)](#)  
Download: [SAS](#) [SPSS](#) [Stata](#) [R](#) [ASCII](#) [Excel/TSV](#)  
[ASCII + SAS Setup](#) [SPSS Setup](#) [Stata Setup](#)

**DS5: Wave II: In-Home Questionnaire, Public Use Sample - Download All Files (203,1 MB) large file**  
Documentation: [Codebook/Questionnaire.pdf](#) [Frequencies.pdf](#)

Fonte: Screenshot da interface do repositório

#### 4.3.2.8 Requisito 8 – Registro da autoria da publicação e dos objetos digitais

As pesquisas científicas são cada vez mais colaborativas, por este motivo recomenda-se registrar as contribuições individuais e fazer uma distinção clara entre o autor da publicação ampliada e os autores de suas partes componentes. Autores de *ePrints* e recursos de dados são os agentes responsáveis por seus conteúdos intelectuais e o autor da publicação ampliada como um todo é o agente que decidiu combinar os vários recursos em um único objeto composto. (VERHAAR, 2009). Conforme as análises, 105 repositórios possuem o registro da autoria da publicação e dos objetos digitais.

Na figura 21 é possível observar o registro das autorias do conjunto de dados de pesquisa do repositório e-cienciaDatos.

Figura 21 – Exemplo de repositório com registro de autoria

The screenshot shows the interface of the Madroño Consorcio e-cienciaDatos repository. The header includes the logo and navigation links. The main content area displays the following information:

<b>Título</b>	Experiment STABILITY of liquid COLUMNS (STACO) on the Advanced Fluid Physics Module (FPM) aboard Spacelab-D2
<b>Subtítulo</b>	liquid bridge under microgravity
<b>Título Alternativo</b>	SL-D2-WL-AFPM-STACO
<b>Otro ID</b>	European Space Agency (ESA): 0000 0001 2163 0931 0000 0004 0623 6946
<b>Autor</b>	Martínez Herranz, Isidoro (UPM) - ORCID: 0000-0003-0994-5898 Perales Perales, José Manuel (UPM) - ORCID: 0000-0001-9873-8115 Meseguer Ruiz, José (UPM) Ulrich Walter (Technische Universität München)
<b>Contacto</b>	Utilice el botón de e-mail de arriba para contactar. Archivo Digital UPM (UPM)
<b>Descripción</b>	During the second German Mission of Spacelab (1993, Spacelab D-2) an experiment on the mechanical stability of liquid columns was performed. A liquid column is a liquid mass held between to solid (circular) supports. A 30 mm in diameter and 90 mm long liquid drop was established in the Advanced Fluid Physics Module (AFPM) and was subjected to different mechanical stimuli (rotation of the supports, oscillation of one of the supports, change in volume and in length) to check whether the resulting drop was stable or not. Results were obtained in terms of images and recorded in a video sequence. Telemetry of the different stimuli and parameter values was also recorded. Objective: The aim of this experiment was to measure the outer shape deformation of long liquid bridges near their stability limit under microgravity, caused by g-jitters and by some controlled mechanical disturbances (change of geometry, change of volume, rotation and vibration). The liquid used was a silicone-oil 10 times more viscous than water (5 times for the last run). The working length of the liquid column was 85 mm. The two solid supports are made of aluminium, of 30 mm in diameter, with a sharp cutback (30 deg. edge) to prevent liquid spreading over the edges. This choice of geometry allows a direct comparison with other TEXUS experiments where two discs of 30 mm in diameter and 86 mm apart were used to hold a cylindrical

Fonte: Screenshot da interface do repositório

#### 4.3.2.9 Requisito 9 – Políticas de preservação

É imprescindível garantir a preservação a longo prazo das publicações ampliadas e devido a esta necessidade, um número crescente de instituições está desenvolvendo repositórios que visam preservar conteúdos digitais para futuras gerações e estão empregando uma combinação de técnicas para garantir a longevidade, incluindo migração, que envolve a transferência de bits de um formato ou meio para outro. (VERHAAR, 2009).

Um total de 126 repositórios não contemplava o requisito número 9. O motivo se deu pela falta de identificação em suas interfaces de informações relacionadas à preservação. Acredita-se que as instituições possuam documentos internos que relatem, no entanto nesta pesquisa não foram identificados.

O restante dos repositórios que apresentaram algum tipo de documentação ou instruções referentes à preservação possuíam o software *Dataverse* ou estavam relacionados à Associação ICPSR, que fornece recursos e treinamentos voltados aos dados de pesquisa. A figura 22 demonstra a interface inicial do projeto *Dataverse* e links para as políticas de gestão e preservação de dados.

Figura 22 – Informações sobre políticas de preservação de dados



Fonte: *Screenshot* da interface do repositório

#### 4.3.2.10 Requisito 10 – Registrar as relações entre os recursos web que fazem parte da publicação ampliada

Recomenda-se registrar as relações entre os recursos web que fazem parte da publicação ampliada (os componentes precisam ser descritos e classificados utilizando-se um vocabulário padrão). É necessário que, nas publicações ampliadas, as relações entre os objetos compostos estejam visíveis. Descrições entre os recursos relacionados ajudam a esclarecer os motivos pelos quais os objetos foram adicionados à publicação. Desta forma, se faz necessário, um vocabulário controlado e um padrão de classificação. (VERHAAR, 2009).

Durante as análises dos 136 repositórios foi possível observar que 103 deles possuíam, visivelmente, as relações entre os recursos web que faziam parte da publicação ampliada.

Conforme o requisito número 3, as publicações podem estar vinculadas aos conjuntos de dados de pesquisa de forma visível e por meio de links que interligam esses dados. Além de links, os componentes podem estar descritos de forma padronizada para a compreensão da forma como foram coletados, como mostra a figura 23.

Figura 23 – Registro/descrição das relações entre os recursos que fazem parte da publicação

The screenshot shows a web interface for a dataset. At the top, there is a red navigation bar with the logo of 'Madroño Consórcio e-ciencia' and a 'BETA' label. The main content area has a white background with a red border. The title of the dataset is 'Experimente a STABILIDADE de COnumnas líquidas (STACO) no Módulo Avançado de Física de Fluidos (FPM) a bordo do Spacelab-D2'. Below the title, there is a citation information box containing the text: 'Martínez Herranz, Isidoro; Perales Perales, José Manuel; Meseguer Ruiz, José; Ulrich Walter, 2016, "Experiência de STABILIDADE de COnumnas líquidas (STACO) no Módulo Avançado de Física de Fluidos (FPM) a bordo do Spacelab-D2", doi: 10.21950 / MJFJM2 , Dataverse e-cienciaDatos, V3'. To the right of this box is a red button labeled 'Dataset do Citar'. Below the citation box is a 'Descrição' section with a detailed text description of the experiment. The description mentions the Spacelab D-2 mission, the experiment setup with liquid columns, and the results obtained.

Fonte: *Screenshot* da interface do repositório

#### 4.3.2.11 Requisito 11 – Documentos baseados no modelo OAI-ORE

As instituições que fornecem acesso a publicações ampliadas devem certificar-se de que estejam disponíveis como documentos baseados no modelo OAI-ORE (*Open Archives Initiative - Object Reuse and Exchange*). O padrão OAI-ORE é um padrão estabelecido para capturar e trocar informações sobre objetos compostos e seu objetivo é desenvolver padrões interoperáveis e mecanismos legíveis por máquinas para expressar as informações dos objetos compostos na web. As instituições devem tomar medidas que garantam que as publicações ampliadas estejam disponíveis com documentos baseados no modelo OAI-ORE. (VERHAAR, 2009). Conforme demonstrado nos requisitos anteriores, grande parte dos repositórios utilizam metadados, ou seja, possuem documentos baseados no modelo OAI-

ORE, ou seja, troca e reusa objetos digitais em várias plataformas: ATOM, YADS, RDF, XML. (OPEN ARCHIVES INITIATIVE, 2008).

Dentre os repositórios selecionados, 78 apresentam documentos baseados no modelo OAI-ORE.

#### 4.3.2.12 Requisito 12 – Disponibilizar em plataforma *open access*

Diferentes componentes que fazem parte de uma publicação ampliada podem ter diferentes tipos de acesso, ou seja, podem ser originários de diferentes conjuntos de dados com diferentes direitos de acesso, o que poderá tornar a publicação ampliada inacessível. Por este motivo é importante que os dados estejam disponíveis em plataformas *open access*. Verificou-se que 133 repositórios está de acordo com o requisito 12.

Na interface inicial dos repositórios que estão de acordo com o requisito 12, existem informações sobre a forma de disponibilizar as informações em suas plataformas. No repositório utilizado como exemplo na figura 24, notam-se as informações existentes.

Figura 24 – Exemplo de repositório em plataforma open access

The screenshot shows a website interface with a navigation bar at the top containing the following items: Inicio, Acerca de, Recursos electrónicos, Servicios, and Investigam. The main content area is divided into two sections. The upper section, titled 'Plan de Gestión de Datos', contains text about data management and a list of links: 'Ciencia abierta', 'Declaracion Open Access', 'COAR', and 'SCOAP3'. Below this is a red button labeled 'Ver más ...'. The lower section, titled 'e-ciencia', contains text about a project in the Madrid region and a search bar with the placeholder 'Buscar en e-ciencia ...' and a red 'Buscar' button. Below the search bar is another red button labeled 'Ver más ...'. A black arrow points from the right side of the page towards the 'e-ciencia' section.

Fonte: *Screenshot* da interface do repositório

#### 4.3.2.13 Requisito 13 – Estar de acordo com os padrões da Web Semântica

A Web semântica é definida como a web de dados. O RDF é a base para a publicação e linkagem de dados. Outras tecnologias permitem inserir dados em documentos (RDFa, GRDDL) ou expor em um bancos de dados SQL, ou ainda, torná-lo disponível no formato RDF. (W3C BRASIL, 2011). Conforme já definido nesta pesquisa, a web semântica tem como objetivo organizar a informação de maneira legível por máquinas através de padrões de formatação de dados. Além de formatos RDF, a linguagem XML é recomendada pela W3C para a criação de dados organizados. Desta forma é uma das características dos padrões da web semântica. Entre os 55 repositórios que apresentavam o requisito 13, observou-se essas características.

Na figura 25 observa-se outro repositório chamado *Ag-Data-Commons*, que contempla o requisito número 13. Neste repositório é possível perceber que o conjunto de dados está disponível no formato de RDF.

Figura 25 – Exemplo de repositório baseado em padrões da web semântica - RDF

The screenshot shows the 'Ag Data Commons Beta' website. The header includes the USDA National Agricultural Library logo and the text 'United States Department of Agriculture'. The main navigation bar contains links for 'Datasets', 'About', 'News', 'Contact Us', a search bar, and 'Log in / Register'. The breadcrumb trail reads 'Home / Datasets / Legume (Fabaceae) Fruits and Seeds Version 2'. The main content area is titled 'Legume (Fabaceae) Fruits and Seeds Version 2'. On the left, there is a 'License' section showing 'U.S. Public Domain' with an 'OPEN DATA' button, and an 'Other Access' section stating that metadata is available in JSON and RDF formats. A black arrow points to the 'RDF' button. Below this, there is a 'ZIP' icon and the text 'Legume (Fabaceae) Fruits and Seeds Version 2' followed by 'This electronic database contains the following: 685 accepted...'. An 'Explore Data' button is located at the bottom right of the dataset information.

Fonte: *Screenshot* da interface do repositório

#### 4.3.2.14 Requisito 14 – Recomenda-se que os objetos digitais estejam conectados por relacionamentos semânticos

Enquanto os repositórios atualmente se concentram principalmente no armazenamento de objetos, deve-se reconhecer que vários recursos que deveriam estar relacionados podem ser encontrados em diferentes depósitos de dados, ou seja, os dados das pesquisas podem estar armazenados em outro local. É essencial, nas publicações ampliadas, que essas relações estejam declaradas explicitamente. (VERHAAR, 2009). A utilização de links entre os recursos permitirá o acesso aos dados provenientes. Este requisito assemelha-se aos requisitos número 3 e 10. Para que seja considerada uma publicação ampliada recomenda-se que os objetos digitais estejam relacionados à publicação e vice-versa. Estas relações devem estar visíveis através de links e de descrições. Percebeu-se que 121 repositórios estão de acordo com o requisito 14.

#### 4.3.2.15 Requisito 15 – Recomenda-se que os objetos digitais devam ser fornecidos através de upload

Upload caracteriza-se pela ação de enviar dados, neste caso, aos repositórios de dados de pesquisa (VERHAAR, 2009). Alguns repositórios possuem em suas interfaces de forma visível um local (link) para envio após o cadastramento do usuário no sistema. No entanto, verificou-se entre os 136 repositórios com publicações ampliadas, que 93 deles não possuíam em suas interfaces este item. Apenas forneciam um contato para solicitação de envio.

#### 4.3.2.16 Requisito 16 – Cada arquivo de dados tem seus próprios metadados, incluindo informações do objeto

Recomenda-se que as instituições que oferecem acesso às publicações ampliadas mantenham os metadados dos objetos em versão legível por máquinas para que sejam descobertos. Os conteúdos das publicações ampliadas devem ser acessíveis aos serviços como os rastreadores web, ferramentas de mineração de dados e mecanismos para a coleta de registros de metadados. (VERHAAR, 2009).

Entre os 136 repositórios, 72 apresentaram metadados e informações do objeto. Foram considerados os repositórios que possuíam a descrição dos metadados de forma visível. Este requisito assemelha-se ao requisito número 6, que sugere vocabulário controlado, e ao requisito número 11, que sugere que os documentos estejam baseados no modelo OAI-ORE.

#### 4.4 PROPOSTA DE UM PANORAMA PARA PUBLICAÇÕES AMPLIADAS BASEADO EM REQUISITOS

As diretrizes propostas por este estudo recomenda padrões de requisitos para a integração de dados de pesquisa baseados nas publicações ampliadas, de acordo com a literatura da área, para a utilização entre os repositórios no contexto atual. Após a realização da revisão sistemática foram destacadas categorias para análise e definidos alguns requisitos para publicação ampliada. Verificou-se que grande parte da literatura existente que trata sobre requisitos de publicações ampliadas foi fundamentada nos conceitos desenvolvidos por Verhaar (2009), na obra *Enhanced Publications: linking publications and research data in digital repositories*, que descreve algumas propostas de modelos para publicações ampliadas. Percebe-se que há uma tentativa de adicionar uma estrutura utilizando-se requisitos relacionados a metadados, relacionamentos semânticos, entre outros.

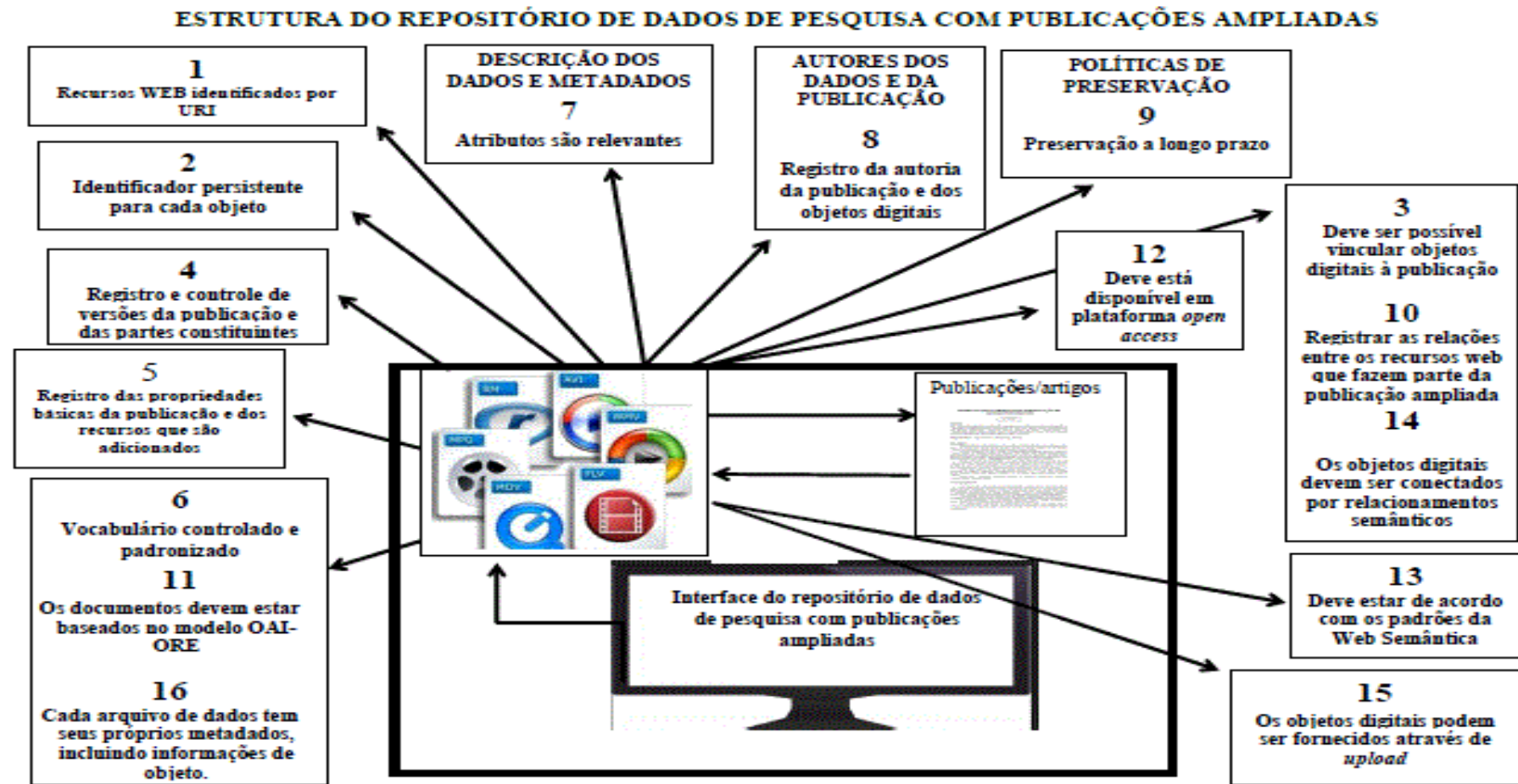
Desta forma, verificou-se uma necessidade crescente, de uma infraestrutura compatível com as publicações ampliadas, tornando-se de extrema importância a utilização de requisitos para as interfaces dos repositórios de dados de pesquisa. O cruzamento de informações realizado neste estudo permite um maior entendimento do quanto é importante o uso de cada um dos requisitos estipulados como também é possível compreender as consequências caso estes requisitos não estejam contemplados nas interfaces de repositórios.

Os repositórios de dados de pesquisa deveriam fornecer em suas interfaces as ligações e o acesso às respectivas publicações científicas dos dados registrados para que fosse possível compreender os estudos realizados a partir destes dados. Desta maneira entende-se que há a necessidade de interligar os dados de pesquisas às suas respectivas publicações de acordo com os requisitos estabelecidos e identificados na literatura, associados aos padrões das publicações ampliadas. Nesse sentido, acredita-se que um conjunto de dados disponíveis em um repositório digital que contemple os requisitos, melhora a recuperação da informação e, conseqüentemente, aumenta a visibilidade da publicação e de autores.

Esta proposta de um panorama de integração de dados de pesquisa baseado em publicações ampliadas visa sugerir uma interface que contemple todos os requisitos especificados nesta pesquisa para que se tenha uma infraestrutura eficiente. Após o detalhamento de cada requisito e demonstrações, decidiu-se elaborar um esquema, conforme a figura 28, para que se tenha uma visão geral do panorama proposto neste estudo.



Figura 26 – Estrutura da publicação ampliada para integração de dados de pesquisa



Fonte: Elaborado pela autora

No esquema apresentado, conforme mostra a figura 26, existe uma imagem expressando uma tela de computador com a interface de um repositório de dados de pesquisa que possui publicações ampliadas. Percebe-se que o conjunto de dados é composto por imagens, vídeos, textos, audiovisuais, entre outros e estão ligados à uma publicação científica. Este conjunto de dados possui os 16 requisitos estabelecidos nesta pesquisa (APÊNDICE E) e alguns desses requisitos demonstrados no esquema foram agrupados de acordo com as semelhanças já descritas neste estudo. Os requisitos propostos nesta pesquisa e demonstrados neste panorama consequentemente permitiu a recuperação da informação de forma eficiente e uma melhor compreensão dos dados provenientes da pesquisa científica publicada como também proporcionou uma maior visibilidade da publicação e dos autores.

## 5 CONSIDERAÇÕES FINAIS

Presume-se que um conjunto de dados de pesquisa disponíveis em um repositório digital permite o aumento da visibilidade, o reuso e potencializa a troca de conhecimento. Neste contexto, a publicação ampliada possui o propósito de conectar esses dados disponíveis nos repositórios com suas respectivas publicações científicas, sendo considerada uma nova forma de publicação.

O objetivo deste estudo foi não só propor um modelo de integração entre repositórios de dados de pesquisa e publicações científicas baseado no modelo de publicações ampliadas, mas também contribuir para a discussão e desenvolvimento de infraestruturas voltadas aos repositórios de dados de pesquisa que desejam implementar esse tipo de publicação.

O panorama proposto mostrou uma das formas possíveis de apresentação de uma interface de repositórios de dados de pesquisa caracterizada com publicação ampliada a partir de requisitos estabelecidos por meio da literatura recuperada na revisão sistemática. Foi decidido manter o princípio de que uma publicação tradicional deve estar relacionada aos objetos digitais para ser considerada uma publicação ampliada. Percebe-se na literatura que há uma tendência para seguir esta definição mais ampla. No entanto existem diversas características a serem consideradas para que, de fato, possa estar de acordo com os critérios associados a este tipo de publicação.

A revisão sistemática da literatura, realizada na primeira fase da análise, em bases de dados eletrônicas utilizando um *string* de busca, recuperou um total de 11 itens a serem analisados, sendo que somente 5 documentos abordavam assuntos relacionados à infraestrutura de uma publicação ampliada e seus principais requisitos. Mesmo com um grande empenho para desenvolver filtros de pesquisa que recuperassem estudos relevantes com maior precisão, observou-se um número baixo de documentos referentes ao assunto *enhanced publication* e que tratassem sobre seus principais requisitos. Desta maneira, após a seleção dos estudos, os dados foram extraídos e sintetizados para que fosse possível estabelecer as categorias para análise. Além dos requisitos coletados durante a revisão sistemática, também foram incluídos nas categorias os requisitos citados por Verhaar, na obra *Enhanced Publications: linking publications and research data in digital repositories*, como também os estudos referenciados nas obras recuperadas durante a revisão sistemática. Sendo assim, as categorias para análise foram estabelecidas.

Na segunda fase da análise foram selecionados os repositórios de dados de pesquisa através do diretório re3data.org e averiguadas as informações disponibilizadas por meio dos

filtros que demonstrassem a utilização de ligações baseadas no modelo de publicação ampliada. Foram descritos os tipos de conteúdos, quantidades de repositórios por países, identificadores digitais, tipos de padrões de programação, restrições de acesso aos dados, formas de upload, tipos de instituições vinculadas aos repositórios, padrões de metadados, idiomas, softwares utilizados e a quantidade destes repositórios que utilizam publicações ampliadas, totalizando 654 repositórios.

Na terceira fase da análise foi realizada a inspeção das interfaces e identificados os elementos que cada repositório contempla a partir dos requisitos estabelecidos por meio da revisão sistemática. Posteriormente iniciaram-se as análises, observações e comparações dos elementos utilizados nos repositórios, destacando-se os principais recursos. Entre os 654 repositórios definidos pelo diretório, com publicações ampliadas, somente 136 foram inseridos no estudo e 518 repositórios foram excluídos por motivos diversos como, por exemplo, publicações não relacionadas aos conjuntos de dados de pesquisa.

Entre os 136 repositórios selecionados foram evidenciados os recursos disponíveis como, as 10 (dez) áreas e formas mais abrangentes, destacando-se a Biologia, Ciências Naturais e Medicina. Observou-se que as formas de dados mais disponibilizadas entre os repositórios de dados de pesquisa com publicações ampliadas são de dados estatísticos, em texto e em imagens. Em relação aos identificadores digitais de autoria, foi concluído que um grande índice de repositórios de dados de pesquisa com publicações ampliadas ainda não utilizam um identificador digital persistente para autor. Contudo, entre os padrões de programação destacam-se os padrões REST e OAI-PMH. Referente às restrições de acesso aos dados; constatou-se que um percentual maior dos repositórios selecionados possibilita o acesso por meio de registro ou através de alguma associação institucional e um pequeno percentual através de depósito de dados e de taxa requisitada. Sobre as instituições vinculadas aos repositórios, observou-se que grande parte estão vinculados a instituições sem fins lucrativos. Para os padrões de metadados, foi identificado uma taxa de 21% de repositórios que utilizam DDI; 16% utilizam *Dublin Core* e *DataCite*; 4% utilizam *ISA*; apenas 3% utilizam *RDF* e *Repository Developed Metadata Schemas*.

Conforme informações do diretório, a maior parte dos repositórios utilizam o identificador digital de objetos digitais DOI ou não utilizam nenhum tipo de identificador.

O principal idioma destacado entre os 136 repositórios é o inglês, com 98% e o restante dos idiomas como, por exemplo, o alemão, Sueco, Francês, Espanhol, Polonês e Finlandês aparece com um percentual de 1% dos repositórios. Em relação aos softwares mais utilizados; 12% dos repositórios utilizam o *Dataverse*; 3% utilizam o *DSpace* e *EPrints*.

Os tipos de repositórios identificados neste estudo foram os temáticos, com 65% e os institucionais com 30%. Entre os países mais destacados, 47% dos repositórios são provenientes dos Estados Unidos; 10% da Alemanha e Reino Unido; 6% da União Europeia; 5% do Canadá e Austrália. Entre outros países há um percentual de 32% dos repositórios inseridos no estudo.

Após serem evidenciados os recursos disponíveis entre os 136 repositórios, foram destacados os 16 requisitos de publicações ampliadas definidos nesta pesquisa. As interfaces dos repositórios foram inspecionadas para a verificação da ligação de seus dados e se contemplavam os 16 (dezesesseis) requisitos. A proposta fundamentou-se em investigar quais elementos são importantes nas interfaces de repositórios de dados de pesquisa para que se tenha, de fato, um modelo de publicação ampliada. Constatou-se nas análises que praticamente a totalidade dos repositórios possuem o requisito 1, ou seja, tinham os recursos web identificados por URI. Conforme consta na literatura é importante constatar se os manuscritos e dados de pesquisa que são depositados em repositórios fazem parte da arquitetura web, o que implica que todos esses objetos deveriam estar disponíveis como recursos da web que podem ser referenciados através de URI. Para que fosse possível exemplificar cada um dos requisitos encontrado na análise, utilizou-se alguns *screenshots* para as demonstrações.

Outra prática indicada na literatura consiste em associar identificadores globais únicos e persistentes com cada um dos recursos, conforme destacado no requisito número 2. Na inspeção verificou-se que entre os 136 repositórios inseridos no estudo, 81 destes apresentam identificadores persistentes para objetos digitais e o mais destacado entre eles foi o DOI.

O requisito 3 descreve a possibilidade de vincular objetos digitais à publicação, ou seja, as publicações ampliadas são consideradas objetos altamente complexos, que agrupam os textos de artigos de revistas às imagens, arquivos de vídeos etc. Durante a inspeção foi possível identificar que 120 repositórios estão de acordo com o requisito número 2 e vinculam objetos digitais às publicações. No restante dos repositórios observou-se que apenas existem referências descritas mas sem nenhum link de ligação, desta maneira entende-se que não havia vínculos de forma clara.

O requisito número 4 trata sobre o registro e controle de versões da publicação e das partes constituintes. Dentre os 136 repositórios selecionados para esta pesquisa, 77 apresentaram o requisito 4.

Recomenda-se registrar as propriedades básicas da publicação e dos recursos que são adicionados a ele como, por exemplo, títulos, breve descrição, data da última modificação e

data de criação. Estas características estão relacionadas ao requisito número 5. A análise resultou em 108 repositórios que contemplam este requisito.

Foi constatado 60 repositórios que utilizam um vocabulário para registrar a relação entre os objetos digitais, conforme o requisito 6.

Conforme citado neste estudo, atributos são relevantes, isto é, recomenda-se que cada componente seja digitado semanticamente para deixar claro que tipo de recurso está sendo referido. Esta característica está contemplada no requisito número 7 e observou-se que 90 repositórios possuíam uma breve descrição de cada componente inserido ao conjunto de dados, o que possibilitou a identificação sobre que tipo de arquivo estava disponível antes de fazer o download. Entre os 46 repositórios que não contemplavam o requisito número 7, foi verificado que grande parte não utilizava metadados como também não havia descrições de cada objeto digital, fornecendo somente o link de acesso para download.

De acordo com as análises, 105 repositórios possuíam o registro da autoria da publicação e dos objetos digitais e estavam de acordo com o requisito número 8. Constatou-se que em alguns repositórios havia os registros das contribuições individuais e a distinção clara entre o autor da publicação científica e os autores de suas partes componentes.

Um total de 126 repositórios não contemplava o requisito número 9, que trata sobre políticas de preservação. O motivo se deu pela falta de identificação em suas interfaces de informações relacionadas à preservação. Acredita-se que as instituições possuam documentos internos que relatem suas políticas de preservação, no entanto nesta pesquisa não foram reconhecidos. Os 10 repositórios restantes apresentaram algum tipo de documentação ou instruções referentes à preservação.

O requisito 10 refere-se ao registro das relações entre os recursos web que fazem parte da publicação ampliada, isto é, os componentes precisam ser descritos e classificados utilizando-se um vocabulário padrão. Ao longo das análises dos 136 repositórios foi possível observar que 103 deles possuíam, visivelmente, as relações entre os recursos web que faziam parte da publicação ampliada.

O requisito 11 recomenda que os documentos estejam baseados no modelo OAI-ORE. A partir desta característica, verificou-se que entre os repositórios selecionados, 78 estavam de acordo com este requisito.

Conforme já mencionado nesta pesquisa, diferentes componentes que fazem parte de uma publicação ampliada podem ter diferentes tipos de acesso e se tornar inacessíveis, por este motivo, é importante que os dados estejam disponíveis em plataformas *open access*,

conforme recomendado no requisito número 12. Percebeu-se que praticamente, a totalidade dos repositórios está de acordo com o requisito.

Estar de acordo com os padrões da Web Semântica é o que recomenda o requisito número 13. Conforme já mencionado neste estudo, a web semântica tem como objetivo organizar a informação de maneira legível por máquinas através de padrões de formatação de dados. A partir dessa recomendação, observou-se na inspeção que 55 repositórios apresentavam as características descritas neste requisito.

O requisito 14 recomenda que os objetos digitais estejam conectados por relacionamentos semânticos. Foi percebido nesta pesquisa que este requisito assemelha-se aos requisitos número 3 e 10 e estabelece que para ser considerada uma publicação ampliada os objetos digitais devem estar relacionados à publicação e vice-versa. Estas relações devem estar visíveis através de links e de descrições. Portanto, verificou-se que 121 repositórios estavam de acordo com o requisito 14.

Os objetos digitais podem ser fornecidos através de upload, é o que estipula o requisito número 15, ou seja, seria ação de enviar dados aos repositórios de dados de pesquisa. Alguns repositórios possuem em suas interfaces visivelmente um link para envio após o cadastramento do usuário no sistema e neste estudo destacou-se que 93 repositórios não possuíam em suas interfaces este item. Apenas forneciam um contato para solicitação de envio.

Referente ao requisito 16, cada arquivo de dados tem seus próprios metadados, incluindo informações de objeto. Entre os 136 repositórios, 72 apresentaram metadados e informações do objeto e foram considerados os repositórios que possuíam a descrição dos metadados de forma visível. Notou-se que este requisito assemelha-se ao requisito número 6 e ao requisito número 11.

Pode-se concluir que a integração de dados de pesquisas às publicações científicas utilizando-se os requisitos associados às publicações ampliadas conseqüentemente melhora a recuperação da informação de modo que as informações relacionadas se tornem de fácil acesso, além disto, permite uma melhor interpretação dos dados vinculados podendo ser reutilizados em outros estudos. Observou-se a partir da revisão sistemática que os estudos referentes à publicação ampliada ainda carecem de maior detalhamento e explanação de ações, o que confirma o estágio inicial que se encontra o tema.

A partir de definições abordadas nas literaturas e após as análises dos repositórios concluiu-se que a utilização de requisitos referentes à web semântica permite realizar a busca da informação de forma mais ampla devido às várias relações, isto é, as informações não

estavam disponíveis de forma isolada. Requisitos que abordaram os elementos de relacionamentos proporcionaram a integração; e os requisitos equivalentes aos registros de autoria, descrições e controle de versões ofereceram qualidade destacando os direitos autorais, ou seja, aumentaram a visibilidade tanto de publicações quanto de autorias.

Foi possível perceber que os repositórios que contemplavam grande parte dos requisitos estabelecidos nesta pesquisa tinham usabilidade tornando as interfaces agradáveis, eficazes e eficientes durante a realização das buscas, ou seja, as buscas executadas foram de forma completa e com um menor tempo gasto para a recuperação, diferente das interfaces dos repositórios que não possuíam diversos requisitos.

Assim sendo, a partir do desenvolvimento das três etapas deste estudo pode-se ter uma visão geral para que fosse possível desenvolver um panorama de integração entre repositórios de dados de pesquisa e publicações científicas baseado no modelo de publicações ampliadas e espera-se que a presente pesquisa possa contribuir para a ampliação do conhecimento referente às publicações ampliadas e servir de referência para aperfeiçoar o desenvolvimento de repositórios de dados de pesquisas que possuem o intuito de implementar este tipo de publicação. Novas investigações poderão contribuir para o desenvolvimento de modelos e padrões associados às publicações ampliadas, desta maneira, outras propostas de infraestruturas e modelos podem ser sugeridas, assim como avaliações de interfaces de repositórios que já contemplam os requisitos associados às publicações ampliadas.



## REFERÊNCIAS

- ALBAGLI, Sarita. Ciência aberta em questão. In: ALBAGLI, Sarita; MACIEL, Maria Lucia; ABDO, Alexandre Hannud (Orgs.). **Ciência aberta: questões abertas**. Brasília: IBICT, 2015.
- ARAYA, Elizabeth Roxana Mass. **Comunicação científica: agregação, compartilhamento e reuso de elementos informacionais**. 2014. Tese (Doutorado) - Programa de Pós-Graduação em Ciência da Informação. Faculdade de Filosofia e Ciências. Universidade Estadual Paulista. Marília, 2014. Disponível em: <<https://repositorio.unesp.br/bitstream/handle/11449/121981/000812425.pdf?sequence=1&isAllowed=y>>. Acesso em: 02 maio. 2017.
- BARABÁSI, Albert-László. **Linked: a nova ciência dos networks**. 2009.
- BARDIN, L. **Análise de conteúdo**. Lisboa: Edições 70, [19--].
- BATISTA, Mateus Gondim Romão; LÓSCIO, Bernadette Farias. OpenSBBD: Usando Linked Data para Publicação de Dados Abertos sobre o SBBD. In: SIMPÓSIO BRASILEIRO DE BANCO DE DADOS – SBBD, 2013, Recife. **Anais...** Recife: UFPE, 2013. Disponível em: <[http://sbbd2013.cin.ufpe.br/Proceedings/artigos/pdfs/sbbd\\_shp\\_10.pdf](http://sbbd2013.cin.ufpe.br/Proceedings/artigos/pdfs/sbbd_shp_10.pdf)>. Acesso em: 03 ago. 2016.
- BEAGRIE, Neil. The digital curation centre. **Learned publishing**. v.17, n.1, jan. 2004. Disponível em: <<http://onlinelibrary.wiley.com/doi/10.1087/095315104322710197/full>>. Acesso em: 04 set. 2017.
- BERLIN **Declaration on open access to knowledge in the sciences and humanities**. Berlin, 2003. Disponível em: <<https://openaccess.mpg.de/Berlin-Declaration>>. Acesso em: 11 mar. 2017.
- BERMAN, Francine; WILKINSON, Ross; WOOD, John. Buiding Global Infrastructure for data sharing and exchange through the Research Data Alliance. **D-Lib Magazine**, Jan./Feb. 2014. Disponível em: <[http://www.dlib.org/dlib/january14/01guest\\_editorial.html](http://www.dlib.org/dlib/january14/01guest_editorial.html)>. Acesso em: 15 ago. 2016.
- BOMFÁ, Cláudia Regina Ziliotto.; CASTRO, João Ernesto E. Desenvolvimento de revistas científicas em mídia digital: o caso da Revista Produção on-line. **Ciência da Informação**, v. 33, n. 2, dez. 2004. Disponível em: <<http://www.scielo.br/pdf/ci/v33n2/a04v33n2>>. Acesso em: 26 out. 2017.
- CAFÉ, Lúgia et. al. **Repositórios institucionais: nova estratégia para publicação científica na Rede**. XXVI Congresso Brasileiro de Ciências da Comunicação – BH/MG, 2003. Disponível em: <[http://www.intercom.org.br/papers/nacionais/2003/www/pdf/2003\\_ENDOCOM\\_TRABALH\\_O\\_cafe.pdf](http://www.intercom.org.br/papers/nacionais/2003/www/pdf/2003_ENDOCOM_TRABALH_O_cafe.pdf)>. Acesso em: 01 set. 2017.
- CAMARGO, Liriane Soares de Araújo.; VIDOTTI, Silvana Aparecida Borsetti Gregorio. Uma estratégia de avaliação em repositórios digitais. In: XV SEMINÁRIO NACIONAL DE BIBLIOTECAS UNIVERSITÁRIAS, 2008, São Paulo. **Anais...** São Paulo, p.1-16, 2008.

CANALTECH. Disponível em: <<https://canaltech.com.br/software/o-que-e-api/>>. Acesso em: 07 dez. 2017.

CASTEDO, Raquel da Silva. Periódicos científicos on-line: novas interfaces, novos usos, novas práticas. In: XXX CONGRESSO BRASILEIRO DE CIÊNCIAS DA COMUNICAÇÃO – Santos, 2007. **Anais...** Santos, 2007.. Disponível em: <<http://www.intercom.org.br/papers/nacionais/2007/resumos/R1485-1.pdf>>. Acesso em: 27 out. 2017.

CEDAP – Centro de Documentação e Acervo Digital da Pesquisa. Disponível em: <<http://www.ufrgs.br/fabico/nucleos-e-laboratorios/cedap>>. Acesso em: 17 jan. 2017.

CLARIVATE ANALYTICS. **The data citation index**: connecting the data to the research it informs. 2017. Disponível em: <[http://wokinfo.com/products\\_tools/multidisciplinary/dci](http://wokinfo.com/products_tools/multidisciplinary/dci)>. Acesso em: 09 mar. 2017.

COSTA, Michelli; BRAGA, Tiago. Repositórios de dados de pesquisa no mundo. **Cadernos BAD**. Coimbra, n. 2, jul./dez. 2016. Disponível em: <<file:///C:/Users/daiane.santos/Downloads/1585-4033-1-PB.pdf>>. Acesso em: 04 maio. 2017.

COURSERA. **Research Data Management and Sharing**. 2016. Disponível em: <<https://www.coursera.org/>>. Acesso em 25 out. 2016.

DATAcite. Disponível em: <<https://www.datacite.org/>>. Acesso em: 03 maio. 2017.

DCC - Digital Curation Centre. **Disciplinary Metadata**. Disponível em: <<http://www.dcc.ac.uk/drupal/resources/metadata-standards>>. Acesso em: 21 out. 2017.

FONSECA, J. J. S. Metodologia da pesquisa científica. Fortaleza: UEC, 2002. Disponível em: <[http://leg.ufpi.br/subsiteFiles/lapnex/arquivos/files/Apostila\\_\\_METODOLOGIA\\_DA\\_PESQUISA\(1\).pdf](http://leg.ufpi.br/subsiteFiles/lapnex/arquivos/files/Apostila__METODOLOGIA_DA_PESQUISA(1).pdf)>. Acesso em: 07 out. 2017.

FRAGOSO, Suely; RECUERO, Raquel; AMARAL, Adriana. **Métodos de pesquisa para internet**. Porto Alegre: Sulina, 2012.

GEZELTER, Dan. **The open science project**. 2009. Disponível em: <<http://www.openscience.org/blog/?p=269>>. Acesso em: 10 mar. 2017.

GIL, Antonio Carlos. **Como elaborar projetos de pesquisa**. 5. ed. São Paulo: Atlas, 2010.

\_\_\_\_\_. **Estudo de caso**. São Paulo: Atlas, 2009.

GONÇALVES, Márcio Souza; BARBOSA, Rafael de Oliveira. Comunicação, linearidade e não linearidade: costurando conceitos e práticas. **Intexto**, n. 34, p. 694-712, set./dez. 2015. Disponível em: <<http://seer.ufrgs.br/index.php/intexto/article/viewFile/58500/35515>>. Acesso em: 26 out. 2017.

HEY, T.; HEY, J. **E-science and its implications for the library community**. Library Hi Tech, v. 24, n. 4, p. 515-528, 2006. Disponível em: <[http://conference.ub.uni-bielefeld.de/2006/proceedings/heyhey\\_final\\_web.pdf](http://conference.ub.uni-bielefeld.de/2006/proceedings/heyhey_final_web.pdf)>. Acesso em: 18 ago. 2017.

HIGGINS, Sarah. The DCC Curation Lifecycle Model. **The International Journal of Digital Curation**. v.3, n.1, 2008. Disponível em: <<http://www.ijdc.net/index.php/ijdc/article/viewFile/69/48>>. Acesso em: 15 ago. 2017.

IBICT. **Manifesto de Acesso Aberto a Dados da Pesquisa Brasileira para Ciência Cidadã**, Brasília, DF. 2016. Disponível em: <<http://www.ibict.br/Sala-de-Imprensa/noticias/2016/ibict-lanca-manifesto-de-acesso-aberto-a-dados-da-pesquisa-brasileira-para-ciencia-cidada>>. Acesso em: 17 abr. 2017.

\_\_\_\_\_. **IBICT Dataverse**. 2017. Disponível em: <<https://repositoriopesquisas.ibict.br/dataverse/ibict>>. Acesso em: 31 ago. 2017.

ICPSR (INTER-UNIVERSITY CONSORTIUM FOR POLITICAL AND SOCIAL RESEARCH)-, Ann Arbor: ICPSR, 2017. Disponível em: <<http://www.icpsr.umich.edu/icpsrweb/deposit/index.jsp>>. Acesso em: 22 fev. 2017.

\_\_\_\_\_. Guide to social science data preparation and archiving. 2012. Disponível em: <<http://www.icpsr.umich.edu/files/deposit/dataprep.pdf>>. Acesso em: 07 out. 2017.

JENKINS, Henry; FORD, Sam; GREEN, Joshua. **Spreadable media: creating value and meaning in a networked culture**. New York: NYU Press, 2013. p. 1-46. Disponível em: <[http://nyupress.org/webchapters/jenkins\\_intro.pdf](http://nyupress.org/webchapters/jenkins_intro.pdf)>. Acesso em: 15 jul. 2016.

JENKINS, Henry. **Cultura da Convergência**. 2. ed. São Paulo: Aleph, 2009.

KATZ, J. S.; MARTIN, B. R. What is research collaboration? **Research Policy**, v. 26, p. 1-18, 1997. Disponível em: <[http://users.sussex.ac.uk/~sylvank/pubs/Res\\_col9.pdf](http://users.sussex.ac.uk/~sylvank/pubs/Res_col9.pdf)>. Acesso em: 03 jun. 2016.

KITCHENHAM, Barbara. **“Procedures for Performing Systematic Reviews”**, Joint Technical Report Software Engineering Group, Department of Computer Science Keele University, United King and Empirical Software Engineering, 2004. Disponível em: <<http://www.ifs.tuwien.ac.at/~weippl/systemicReviewsSoftwareEngineering.pdf>>. Acesso em: 08 set. 2017.

LINKED DATA. 2013. Disponível em: <[www.linkeddata.org](http://www.linkeddata.org)>. Acesso em: 01 ago. 2016.

MAFRA, Sômulo Nogueira; TRAVASSOS, Guilherme Horta. **Estudos primários e secundários apoiando a busca por evidência em Engenharia de Software**. Rio de Janeiro: PESC, 2006. Disponível em: <<http://www.cos.ufrj.br/uploadfile/1149103120.pdf>>. Acesso em: 22 maio. 2017.

\_\_\_\_\_. **Técnicas de leitura de software: uma revisão sistemática**. Rio de Janeiro: COPPE/UFRJ, 2005. Disponível em: <<http://www.lbd.dcc.ufmg.br/colecoes/sbes/2005/005.pdf>>. Acesso em: 23 maio. 2017.

MALTRÁS BARBA, Bruno. **Los indicadores bibliométricos: fundamentos y aplicación al análisis de la ciência.** Gijón: Trea, 2003. Disponível em: <[https://www.researchgate.net/publication/31766528\\_Los\\_indicadores\\_bibliometricos\\_fundamentos\\_y\\_aplicacion\\_al\\_analisis\\_de\\_la\\_ciencia\\_B\\_Maltras\\_Barba](https://www.researchgate.net/publication/31766528_Los_indicadores_bibliometricos_fundamentos_y_aplicacion_al_analisis_de_la_ciencia_B_Maltras_Barba)>. Acesso em: 04 abr. 2017.

MARCONDES, Carlos Henrique. “Linked data”: dados interligados - e interoperabilidade entre arquivos, bibliotecas e museus na web. **Encontros Bibli: revista eletrônica de biblioteconomia e Ciência da Informação.** Florianópolis, v. 17, n. 34, p.171-192, maio./ago. 2012. Disponível em: <<https://periodicos.ufsc.br/index.php/eb/article/viewfile/1518-2924.2012v17n34p171/22782>>. Acesso em: 18 jul. 2016.

MARCONI, Marina de Andrade; LAKATOS, Eva Maria (Org.). **Fundamentos de metodologia científica.** 5. ed. São Paulo: Atlas, 2003.

MEADOWS, A. J. **A comunicação científica.** Brasília: Briquet de Lemos, 1999.

MEDEIROS, Jackson da Silva. **Uma investigação sobre a autoria de dados científicos: teias de uma rede em construção.** 2015. Tese (Doutorado) – Programa de Pós-Graduação em Ciência da Informação. Universidade Federal do Rio Grande do Sul. Porto Alegre, 2015. Disponível em: <<https://www.lume.ufrgs.br/bitstream/handle/10183/116504/000966998.pdf?sequence=1>>. Acesso em: 13 out. 2016.

MUCHERONI, Marcos Luiz; SILVA, Fernando José Modesto da.; PALETTA, Carlos Francisco. Entre a publicação ampliada e a multimodalidade. In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO - ENANCIB, 16., 2015, João Pessoa. **Anais...** João Pessoa, 2015.. Disponível em: <<http://www.brapci.ufpr.br/brapci/index.php/article/view/0000017566/524e08c83118ed11f73ea166974e6d52>>. Acesso em: 02 ago. 2016.

MUNZLINGER, Elizabete; NARCIZO, Fabricio Batista; QUEIROZ, José Eustáquio Rangel de. **Protocolo de Revisão Sistemática.** Campina Grande: UFCG, 2012. Disponível em: <[http://www.elizabete.com.br/rs/Tutorial\\_IHC\\_2012\\_files/ProtocoloRevisaoSistematica.pdf](http://www.elizabete.com.br/rs/Tutorial_IHC_2012_files/ProtocoloRevisaoSistematica.pdf)>. Acesso em: 23 maio. 2017.

NASSI-CALÒ, Lilian. Teses e dissertações: prós e contras dos formatos tradicional e alternativo. **SciELO em perspectiva,** 2016. Disponível em: <[http://blog.scielo.org/blog/2016/08/24/teses-e-dissertacoes-pros-e-contras-dos-formatos-tradicional-e-alternativo/#.WfOGq\\_lSxdg](http://blog.scielo.org/blog/2016/08/24/teses-e-dissertacoes-pros-e-contras-dos-formatos-tradicional-e-alternativo/#.WfOGq_lSxdg)>. Acesso em: 27 out. 2017.

OECD. **Principles and Guidelines for Access to Research Data from Public Funding.** 2007. Disponível em: <<https://www.oecd.org/sti/sci-tech/38500813.pdf>>. Acesso em: 10 mar. 2017.

OLIVEIRA, Adriana Carla Silva de.; DIAS, Guilherme Ataíde; PADUA, Mariana Cantisani. Desenvolvendo uma publicação científica ampliada na plataforma e-quilt. In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO - ENANCIB, 15., 2014, Belo Horizonte. **Anais...** Belo Horizonte, 2014. Disponível em: <<http://enancib2014.eci.ufmg.br/documentos/anais/anais-gt8>>. Acesso em: 03 ago. 2016.

OLIVEIRA, William. **URL OU URI, qual a diferença?** Disponível em: <<https://woliveiras.com.br/posts/url-uri-qual-diferenca/>>. Acesso em: 08 dez. 2017.

OPENAIRE. **H2020 Programme Guidelines on Open Access to Scientific Publications and Research Data in Horizon 2020**, 2016. Disponível em: <[file:///C:/Users/daiane.santos/Downloads/h2020-hi-oa-pilot-guide\\_en.pdf](file:///C:/Users/daiane.santos/Downloads/h2020-hi-oa-pilot-guide_en.pdf)>.

OPEN ARCHIVES INITIATIVE. **Open archives initiative object reuse and exchange: the OAI-ORE effort: progress, challenges, synergies**. JCDL: Vancouver, 2007. Disponível em: <<http://www.openarchives.org/ore/documents/ore-jcdl2007.pdf>>. Acesso em: 28 jan. 2018.

OPEN KNOWLEDGE INTERNATIONAL. **The Open Definition**. Disponível em: <<http://opendefinition.org/>>. Acesso em: 01 ago. 2016.

PALETTA, Francisco Carlos; MUCHERONI, Marcos Luiz. O desenvolvimento da WEB 3.0: Linked Data e DBPEDIA. **Revista PRISMA.COM**, 2014. Disponível em: <[http://revistas.ua.pt/index.php/prisma.com/article/view/3047/pdf\\_30](http://revistas.ua.pt/index.php/prisma.com/article/view/3047/pdf_30)>. Acesso em: 15 jul. 2016.

PASSARELLI, Brasilina. Do Mundaneum à WEB Semântica: discussão sobre a revolução nos conceitos de autor e autoridade das fontes de informação. **DataGramZero**, v.9 n.5 out. 2008. Disponível em: <<http://www.brapci.ufpr.br/brapci/index.php/article/download/7638>>. Acesso em: 24 out. 2017.

RECUERO, Raquel. **Redes sociais na internet**. Porto Alegre: Sulina, 2009. Disponível em: <<http://www.ichca.ufal.br/graduacao/biblioteconomia/v1/wpcontent/uploads/redessociaisnainternetrecuero.pdf>>. Acesso em: 26 jul. 2016.

RODRIGUES, Eloy; SARAIVA, Ricardo. Projeto RCAAP – **Os repositórios de dados científicos**: estado da arte. Universidade do Porto, 2010. Disponível em: <[https://repositorium.sdum.uminho.pt/bitstream/1822/10830/1/Relat%C3%B3rio%20Estado%20da%20Arte%20dos%20Reposit%C3%B3rios%20de%20Dados%20Cient%C3%ADficos%20Vers%C3%A3o%201.0\\_Final.pdf](https://repositorium.sdum.uminho.pt/bitstream/1822/10830/1/Relat%C3%B3rio%20Estado%20da%20Arte%20dos%20Reposit%C3%B3rios%20de%20Dados%20Cient%C3%ADficos%20Vers%C3%A3o%201.0_Final.pdf)>. Acesso em: 21 fev. 2017.

SALES, Luana Farias. **Integração semântica de publicações científicas e dados de pesquisa**: proposta de modelo de publicação ampliada para área de Ciências Nucleares. 2014. Tese (Doutorado) – Programa de Pós-Graduação em Ciência da Informação. Universidade Federal do Rio de Janeiro. Rio de Janeiro, 2014. Disponível em: <<http://repositorio.ibict.br/bitstream/123456789/874/1/LUANA%20SALES%20D.pdf>>. Acesso em: 07 abr. 2017.

\_\_\_\_\_. Publicação Ampliada: o que é e o que não é? In: **Publicação ampliada**: integração semântica da memória digital da pesquisa técnico-científica. 2014. Disponível em: <<http://documentoampliado.blogspot.com.br/2014/07/publicacao-ampliada-o-que-e-e-o-que-nao.html>>. Acesso em: 17 out. 2017.

SALES, Luana Farias; SAYÃO, Luís Fernando; SOUZA, Rosali Fernandez. Publicações ampliadas: um novo modelo de publicação acadêmica para o ambiente de e-science. In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO – ENANCIB, 2013, Florianópolis. **Anais ...** Florianópolis, 2013. Disponível

em:<<http://basessibi.c3sl.ufpr.br/brapci/index.php/article/view/0000013705/54d555cb665c858479ea1dfdc3f417ff>>. Acesso em: 23 jul. 2016.

\_\_\_\_\_. O impacto da curadoria digital dos dados de pesquisa na comunicação científica.

**Encontros Bibli**: revista eletrônica de biblioteconomia e Ciência da Informação.

Florianópolis, v. 17, n. esp. 2 – III SBCC, p.118-135, 2012. Disponível em:

<<https://periodicos.ufsc.br/index.php/eb/article/viewFile/15182924.2012v17nesp2p118/23573>>. Acesso em: 13 jul. 2016.

SANTOS, Paola. Paul Otlet: um pioneiro da organização das redes mundiais de tratamento e difusão da informação registrada. **Ciência da Informação**, v. 36, n. 2, p. 54-63, maio/ago. 2007. Disponível em: < <http://revista.ibict.br/ciinf/article/view/1176/1339>>. Acesso em: 24 out. 2017.

SAYÃO, Luís Fernando. Modelos teóricos em Ciência da Informação: abstração e método científico. **Ciência da Informação**, v. 30, n. 1, p. 82-91, 2001. Disponível em: < <http://www.scielo.br/pdf/ci/v30n1/a10v30n1.pdf>>. Acesso em: 07 abr. 2017.

SAYÃO, Luís Fernando; SALES, Luana Farias. Dados abertos de pesquisa: ampliando o conceito de acesso livre. **RECIIS** | 2014, jun., 8(2) – p.76-92, 2014. Disponível em: <[www.reciis.icict.fiocruz.br/index.php/reciis/article/download/611/1252](http://www.reciis.icict.fiocruz.br/index.php/reciis/article/download/611/1252)>. Acesso em: 02 ago. 2016.

\_\_\_\_\_. **Guia de gestão de dados de pesquisa para bibliotecários e pesquisadores**. CNEN: Rio de Janeiro, 2015.

\_\_\_\_\_. Curadoria digital e dados de pesquisa. **AtoZ**: novas práticas em informação e conhecimento. v.5, n.2, p. 67-71, 2016.

SEEBER, Frank. Citations in supplementary information are invisible. **Nature**, v. 451, n. 21, 2008. Disponível em: <<https://www.nature.com/nature/journal/v451/n7181/pdf/451887d.pdf>>. Acesso em: 26 abr. 2017.

SOUZA, Ricardo Augusto Teixeira de. **Predição de tags usando linked data**: um estudo de caso no banco de dados Arquigrafia. 2013. Dissertação (Mestrado) - Programa de Pós-Graduação em Ciência da Informação. Faculdade de Filosofia e Ciências. Universidade Estadual Paulista. Marília, 2013. Disponível em: <<http://www.teses.usp.br/teses/disponiveis/45/45134/tde-04062014-230955/pt-br.php>>. Acesso em: 11 ago. 2016.

SOUZA, Elisabete Gonçalves de; BEZERRA, Darlene Alves COSTA, Wellington Freire Cunha **Em Questão**, Porto Alegre, v. 22, n. 1, p. 113-136, jan/abr. 2016. Disponível em: < <http://seer.ufrgs.br/index.php/EmQuestao/article/view/58835/36968>.> Acesso em: 01 set. 2017.

TOMAÉL, Maria Inês; SILVA, Terezinha Elisabeth da. Repositórios Institucionais: diretrizes para políticas de informação. In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO – ENANCIB, 2007, 8., Salvador. **Anais...** Salvador, 2007. Disponível em: < <http://www.enancib.ppgci.ufba.br/artigos/GT5--142.pdf>.> Acesso em: 04 jun. 2016.

TORRES-SALINAS, D.; ROBINSON-GARCÍA, N.; CABEZAS-CLAVIJO, A. **Compartir los datos de investigación em ciência**: introducción al data sharing. El profesional de la información, v. 21, n. 2, p. 173-184, mar./abr. 2012. Disponível em: <<http://eprints.rclis.org/16786/1/data%20sharing.pdf>>. Acesso em: 20 ago. 2017.

VERHAAR, Peter. Report on object models and functionalities. In: PLACE, Thomas et al. **DRIVER II.**, 2008. Disponível em: <[https://openaccess.leidenuniv.nl/bitstream/handle/1887/16018/Report\\_on\\_Object\\_Models\\_and\\_Functionalities.pdf?sequence=1](https://openaccess.leidenuniv.nl/bitstream/handle/1887/16018/Report_on_Object_Models_and_Functionalities.pdf?sequence=1)>. Acesso em: 21 ago. 2017.

\_\_\_\_\_. Object models and functionalities VERNOOY-GERRITSEN, Marjan et al. In: **Enhanced Publications**: linking publications and research data in digital repositories. SURF Foundation: Amsterdam University Press, Amsterdam, 2009.

W3C BRASIL. Disponível em:< <http://www.w3c.br/Home/WebHome>>. Aceso em: 09 dez. 2017.

YIN, Robert K. **Estudo de caso**: planejamento e métodos. 3. ed. Porto Alegre: Bookman, 2005.

## APÊNDICE A – Protocolo de revisão sistemática

<b>Título da pesquisa</b>	<b>Publicações ampliadas: proposta de integração de dados de pesquisa</b>
<b>Objetivo</b>	Recomendar um modelo de integração entre um repositório de dados de pesquisa e publicações científicas baseado no modelo de publicações ampliadas.
<b>Questões de Pesquisa</b>	Como implementar um modelo de integração entre um repositório de dados de pesquisa e publicações científicas baseado no modelo de publicações ampliadas?
<b>Escopo da Pesquisa</b>	Foram utilizadas bases de dados internacionais para a coleta dos principais requisitos de um modelo de publicação ampliada. A partir da identificação e caracterização dos principais requisitos de um modelo de publicação ampliada foi possível partir para a etapa de inspeção das interfaces dos repositórios para que fosse exequível realizar as comparações e propor um adequado modelo de publicação ampliada.
<b>Crítérios Adotados para Seleção de Fontes</b>	Para a coleta dos dados foram utilizadas bases de dados eletrônicas. (WoS, LISA e SCOPUS).
<b>Restrições</b>	Após as 3 rodadas de testes em cada uma das bases de dados observou-se a existência de um número baixo de publicações referentes ao tema. Desta forma optou-se em não restringir as buscas utilizando outros termos como, por exemplo, “ <i>requirements</i> ”, como também optou-se em não utilizar os operadores booleanos pois constatou-se que utilizando as buscas avançadas e os operadores booleanos restringiam as buscas e recuperava-se um número inferior de resultados.
<b>Idiomas</b>	Os idiomas das publicações aceitos para esta pesquisa serão em inglês.
<b>Métodos de Busca de Publicações</b>	O termo utilizado na busca das publicações será “ <i>enhanced publication</i> ” e a utilização de aspas.
<b>Procedimentos de Seleção e Critérios</b>	Crítérios de inclusão: - Publicações que descrevem as aplicações e sugerem os



	<p>principais requisitos de um modelo de publicação ampliada;</p> <ul style="list-style-type: none"> <li>- Estudos experimentais que indiquem a utilização de requisitos;</li> <li>- Estudos que apresentem metodologias de avaliação dos principais requisitos.</li> </ul> <p>Critérios de exclusão:</p> <ul style="list-style-type: none"> <li>- Publicações cujo foco não corresponda à questão de pesquisa e que não forneçam informações sobre os principais requisitos utilizados em uma publicação ampliada.</li> </ul>
<b>Procedimento para Extração de Dados</b>	Leitura completa de cada estudo primário identificado para refinar os dados.
<b>Definição do registro Preliminar dos Dados Coletados</b>	Os dados foram registrados em formulários elaborados na planilha excel.
<b>Extração dos Dados</b>	São descritos todos os campos de dados que foram extraídos das publicações.
<b>Sumarização dos Resultados</b>	Por meio dos critérios definidos para seleção dos estudos, os artigos encontrados durante a revisão, sejam eles incluídos ou excluídos, foram todos agrupados em uma tabela.
<b>Procedimentos para Análise</b>	Durante a análise dos dados coletados, optou-se em realizar a leitura na íntegra de todos os documentos recuperados para que fosse possível identificar a existência de requisitos descritos nos mesmos.
<b>Análise Quantitativa</b>	Foram listadas as informações quantitativas que foram extraídas do montante de publicações.
<b>Análise Qualitativa</b>	Foram listadas as informações qualitativas que foram extraídas do montante de publicações.
<b>Planejamento e Execução da Revisão Sistemática</b>	Na etapa de planejamento e execução o protocolo foi testado e avaliado; Após foi executada a pesquisa e analisados os resultados.
<b>Definição do Escopo e Estudos Preliminares</b>	O processo de seleção final consistiu na leitura completa dos trabalhos selecionados na etapa de seleção preliminar. Após, foi realizada uma síntese geral e algumas considerações sobre os resultados observados nos trabalhos selecionados.
<b>Definição das Palavras-Chave e Calibração da Expressão de Busca</b>	A partir dos estudos primários/preliminares foi definida uma expressão de busca inicial que foi calibrada através das rodadas de teste. Como expressões de buscas utilizou-se “ <i>enhanced publication</i> ”, “ <i>requirements</i> ”. Utilizou-se também o operador booleano “ <i>And</i> ”. Durante a recuperação de documentos observou-se que os termos da expressão de busca “ <i>enhanced publication</i> ” apareciam separados, desta forma optou-se em utilizar aspas (“ “) para que fosse possível recuperar os termos

	unidos. Foi descartado também a utilização de operadores booleanos “And” e o termo “requirements” pois as buscas se tornaram bastante limitadas.
<b>Primeira Rodada</b>	Foi registrado a data, e o máximo possível de informações sobre cada palavra-chave adicionada à <i>string</i> , assim como o nº de publicações retornadas. Foi realizada uma leitura dinâmica destas publicações em busca de palavras-chaves, termos, sinônimos pertinentes à busca para serem adicionados à <i>string</i> .
<b>Segunda Rodada</b>	A segunda rodada serviu para novos ajustes na <i>string</i> de busca. Ocorreu pela leitura dinâmica das publicações selecionadas na rodada anterior em busca de sinônimos, termos e novas palavras-chave que possam representar resultados na busca. Foram registrados novamente todos os dados possíveis.
<b>Terceira Rodada</b>	A terceira rodada serviu para ajustes finais na <i>string</i> de busca seguindo os mesmos procedimentos dos passos anteriores.
<b>Definição das Máquinas de Busca</b>	Após as rodadas foi definido como expressão de busca o termo “ <i>enhanced publication</i> ”.
<b>Identificação do Período de Busca</b>	O período de busca foi de 12 de setembro de 2017 a 18 de setembro de 2017.
<b>Execução do Protocolo</b>	Foi descrito o número geral de retornos por tema e por base de dados.
<b>Considerações sobre o Resultado do Estudo</b>	Após a finalização de todo o processo, foi descrito os resultados, se satisfatórios ou não.
<b>Apoio Ferramental para a Condução do Estudo</b>	Foram utilizados formulários para a observação e registros.
<b>Dados Coletados</b>	Nesta seção foram apresentados os dados coletados no processo de extração de dados das publicações selecionadas na etapa do processo de seleção.
<b>Dados das Publicações Dentro do Escopo do Estudo</b>	Foram apresentados os dados extraídos da leitura das publicações.

**APÊNDICE B – Formulário de condução da revisão**

<b>REVISÃO SISTEMÁTICA</b>	<b>Fonte (base de dados):</b>	<b>Data da realização da busca</b>	<b>Palavras-chave</b>	<b>Lista de documentos encontrados</b>
<b>ETAPA 2 – Formulário de condução da revisão</b>				
<b>1ª rodada</b>	SCOPUS	12.09.2017	<i>enhanced publication and requirements</i>	439 documentos
<b>2ª rodada</b>	SCOPUS	12.09.2017	<i>"enhanced publication" and requirements</i>	1 documento
<b>3ª rodada</b>	SCOPUS	12.09.2017	<i>"enhanced publication"</i>	27 documentos
				<b>Restaram 21 itens (RETIRANDO OS REPETIDOS)</b>
<b>Títulos</b>				
			1	<i>Global Health and Graduate Medical Education: A Systematic Review of the Literature</i>
			2	<i>Enhanced publication management systems a systemic approach towards modern scientific communication</i>
			3	<i>Enhanced publications in v4 countries</i>
			4	<i>A framework supporting the shift from traditional digital publications to enhanced Publications</i>
			5	<i>A framework supporting the shift from traditional digital publications to enhanced publications</i>
			6	<i>Enabling living systematic reviews and clinical guidelines through semantic technologies</i>
			7	<i>Providing access to research data, publications and current research information at data archiving and networked services - DANS</i>
			8	<i>Enabling living systematic reviews and clinical guidelines through semantic technologies</i>
			9	<i>Survey of enhanced publications in the Czech Republic</i>
			10	<i>Enhanced publications: Data models and information systems</i>
			11	<i>Transforming a research paper into a rich internet publication</i>

			12	<i>EPUB3 for integrated and customizable representation of a scientific publication and its associated resources</i>
			13	<i>Full-text and topic based authorrank and enhanced publication ranking</i>
			14	<i>OAIzer: Configurable OAI Exports over Relational Databases</i>
			15	<i>Linking full-text Grey Literature to underlying research and post-publication data: An Enhanced Publications Project 2011-2012</i>
			16	<i>Information bulletin on variable stars - rich content and novel services for an enhanced publication</i>
			17	<i>Linking full-text grey literature to underlying research and post-publication data: An Enhanced Publications Project 2011-2012</i>
			18	<i>GL-Conference Series: Conference Proceedings</i>
			19	<i>Editor's note enhanced publications project (EPP)</i>
			20	<i>Institutional repositories, long term preservation and the changing nature of scholarly publications</i>
			21	<i>Rich internet publications: "Show what you tell"</i>
			22	<i>Supporting semantically enhanced web service discovery for enterprise application integration</i>
			23	<i>Commercializing public sector information: Privacy and security concerns</i>
			24	<i>Enhanced publication services to members</i>
			25	<i>Enhanced publication services to members</i>
			26	<i>Editorial: Enhanced publication services to members</i>
			27	<i>Enhanced publication services to member</i>
<b>1ª rodada</b>	WoS	18.09.2017	<i>enhanced publication</i>	7349 documentos
<b>2ª rodada</b>	WoS	18.09.2017	<i>"enhanced publication"</i>	5 documentos
<b>3ª rodada</b>	WoS	18.09.2017	<i>"enhanced publication"</i>	5 documentos
				<b>Restaram 3 itens (RETIRANDO OS REPETIDOS)</b>
<b>Títulos</b>				
			1	<i>A Vision for Open Cyber-Scholarly Infrastructures</i>
			2	<i>Enhanced Publication Management Systems A systemic approach towards modern scientific communication</i>

			3	<i>Linking full-text grey literature to underlying research and post-publication data: An Enhanced Publications Project 2011-2012</i>
			4	<i>The VeteranTapes: Research corpus, fragment processing tool, and enhanced publications for the e-Humanities</i>
			5	<i>Web service discovery in a semantically extended uddi registry: The case of fusion</i>
<b>1ª rodada</b>	LISA	18.09.2017	<i>enhanced publication</i>	6368 resultados
<b>2ª rodada</b>	LISA	18.09.2017	<i>"enhanced publication"</i>	14 resultados
<b>3ª rodada</b>	LISA	18.09.2017	<i>"enhanced publication" + And+ requirements</i>	2 resultados
				<b>Restaram 9 itens (RETIRANDO OS REPETIDOS)</b>
<b>Títulos</b>				
			1	<i>Information Bulletin on Variable Stars -- Rich Content and Novel Services for an Enhanced Publication</i>
			2	<i>Transforming a research paper into a rich internet publication</i>
			3	<i>Institutional Repositories, Long Term Preservation and the changing nature of Scholarly Publications</i>
			4	<i>Enhanced Publications: Data Models and Information Systems</i>
			5	<i>Linking Full-Text Grey Literature to Underlying Research and Post-Publication Data: An Enhanced Publications Project 2011-2012</i>
			6	<i>Production and use of information. Characterization of informetric distributions using effort function and density function Exponential informetric process</i>
			7	<i>XL: An XML programming language for Web service specification and composition</i>
			8	<i>The Enrichment of Publications</i>
			9	<i>A practical and robust inter-domain marking scheme for IP traceback</i>
			10	<i>Towards a more accessible e-government in Jordan: an evaluation study of visually impaired users and Web developers</i>
			11	<i>DRIVER: building a sustainable infrastructure for global repositories</i>

			12	<i>Scratching beneath the surface of communities of (mal)practice</i>
			13	<i>Non-tenured leadership appointments in Nigerian university libraries: Problems and prospects</i>
			14	Curadoria digital: um novo patamar para preservação de dados digitais de pesquisa
<b>Total= 33</b> resultados				

## APÊNDICE C – Formulário de seleção de estudos

REVISÃO SISTEMÁTICA	Título do documento	Autores	Data publicação	Periódico
<b>ETAPA 2 – Formulário de seleção de estudos</b>				
1	<i>Global Health and Graduate Medical Education: A Systematic Review of the Literature</i>	Bills, Corey B. ; Ahn, James	2016	
2	<i>Enhanced publication management systems a systemic approach towards modern scientific communication</i>	Bardi, A., Manghi, P.	2015	
3	<i>Enhanced publications in v4 countries</i>	Görögh, E., Kędzierska, E., Kavalchuk, N., Pejšová, P., Vyčítalová, H.	2015	
4	<i>A framework supporting the shift from traditional digital publications to enhanced publications</i>	Bardi, A., Manghi, P. / pesquisadores da Italia / Mangui é gerente da infra- do OpenAire	2015	<i>D-Lib Magazine</i>
5	<i>Providing access to research data, publications and current research information at data archiving and networked services - DANS</i>	Dijk, E., Doom, P.	2015	21(1), pp. 1-1
6	<i>Enabling living systematic reviews and clinical guidelines through semantic technologies</i>	Slaughter, L., Berntsen, C.F., Brandt, L., Mavergames, C.	2015	<i>D-Lib Magazine</i>
7	<i>Survey of enhanced publications in the Czech Republic</i>	Pejšová, P., Vyčítalová, H	2014	21(1), pp. 1-1
8	<i>Enhanced publications: Data models and information systems</i>	Bardi, A., Manghi, P.	2014	<i>LIBER Quarterly</i>
9	<i>Transforming a research paper into a rich internet publication</i>	Breure, L.	2014	<i>Information Services and Use</i>
10	<i>EPUB3 for integrated and customizable representation of a scientific publication and its associated resources</i>	Sigarchian, H.G., De Meester, B., De Nies, T., Mannens, E., Van De Walle, R.	2014	34(3-4), pp. 335- 344
11	<i>Full-text and topic based authorrank and enhanced publication ranking</i>	Zhang, J., Liu, X.	2013	

12	<i>OAIzer: Configurable OAI Exports over Relational Databases</i>	La Bruzzo, S., Manghi, P., Bardi, A.	2013	
13	<i>Information bulletin on variable stars - rich content and novel services for an enhanced publication</i>	Holl, A.	2012	D-Lib Magazine
14	<i>Linking full-text grey literature to underlying research and post-publication data: An Enhanced Publications Project 2011-2012</i>	Farace, D.J., Frantzen, J., Stock, C., Sesink, L., Rabina, D.	2012	18(5-6)
15	<i>GL-Conference Series: Conference Proceedings</i>		2012	
16	<i>Editor's note enhanced publications project (EPP)</i>	Farace, D.	2011	
17	<i>Institutional repositories, long term preservation and the changing nature of scholarly publications</i>	Doorenbosch, P., Sierman, B.	2011	
18	<i>Rich internet publications: "Show what you tell"</i>	Breure, L., Voorbij, H., Hoogerwerf, M.	2010	
19	<i>Supporting semantically enhanced web service discovery for enterprise application integration</i>	Kourtesis, D., Paraskakis, I	2009	
20	<i>Commercializing public sector information: Privacy and security concerns</i>	Burdon, M	2009	<i>IEEE Technology and Society Magazine</i>
21	<i>Enhanced publication services to members</i>	Williams, J.C.	1993	<i>MANUFACTURING ENGINEER</i>
22	<i>A Vision for Open Cyber-Scholarly Infrastructures</i>	Thanos, C	2016	<i>PUBLICATIONS</i> Volume: 4 Edição: 2 Número do artigo: 13
23	<i>The VeteranTapes: Research corpus, fragment processing tool, and enhanced publications for the e-Humanities</i>	van den Heuvel, Henk; van Horik, Rene; Sanders, Eric; et al.	2010	
24	<i>Web service discovery in a semantically extended uddi registry: The case of fusion</i>	Kourtesis, Dimitrios; Paraskakis, Iraklis; Friesen, Andreas; et al.	2007	



25	<i>Production and use of information. Characterization of informetric distributions using effort function and density function Exponential informetric process</i>	Lafouge, Thierry; Prime-Claverie, Camille.	2005	
26	<i>XL: An XML programming language for Web service specification and composition</i>	Florescu, Daniela; Grunhagen, Andreas; Kossmann, Donald.	2003	
27	<i>The Enrichment of Publications t</i>	Hogenaar, Arjan	2012	
28	<i>A practical and robust inter-domain marking scheme for IP traceback</i>	Gao, Zhiqiang; Ansari, Nirwan.	2007	
29	<i>Towards a more accessible e-government in Jordan: an evaluation study of visually impaired users and Web developers</i>	Abu-Doush, Iyad; Bany-Mohammed, Ashraf; Ali, Emad; Al-Betar, Mohammed Azmi.	2013	
30	<i>DRIVER: building a sustainable infrastructure for global repositories</i>	Peters, Dale; Lossau, Norbert.	2011	<i>The Electronic Library</i> 29. 2 (0, 2011): 249-260.
31	<i>Scratching beneath the surface of communities of (mal)practice</i>	Pemberton, Jon; Mavin, Sharon; Stalker, Brenda.	2007	<i>Learning Organizati on</i> 14.1 (0, 2006): 62-73.
32	<i>Non-tenured leadership appointments in Nigerian university libraries: Problems and prospects</i>	Agboola, A T	2001	
33	Curadoria digital: um novo patamar para preservação de dados digitais de pesquisa	Sayão, Luis Fernando; Luana Farias Sales	2012	<i>Informaçã o &amp; Sociedade ; João Pessoa</i> 22. 3 (2012).
<b>Total de documentos= 33</b>	33 resultados			
<b>TEXTO COMPLETO (NÃO ENCONTRADO )</b> 16 ITENS				

<b>TEXTO NÃO POSSUI RELAÇÃO COM O TEMA PESQUISADO</b> 6 ITENS				
<b>ITENS REstantes:</b> 11 ITENS				

**APÊNDICE D – Formulário de extração de dados**

REVISÃO SISTEMÁTICA	Resumo do documento	Extração dos requisitos para publicação ampliada
<b>ETAPA 2 – Formulário de extração de dados</b>		
<b>Títulos:</b> 11 documentos após a análise de critérios de inclusão e exclusão.		
<p align="center"><b><i>Global Health and Graduate Medical Education: A Systematic Review of the Literature</i></b></p>	<p><i>BACKGROUND: Global health (GH) interest is increasing in graduate medical education (GME). The popularity of the GH topic has created growth in the GME literature.</i></p> <p><i>OBJECTIVE: The authors aim to provide a systematic review of published approaches to GH in GME.</i></p> <p><i>METHODS : We searched PubMed using variable keywords to identify articles with abstracts published between January 1975 and January 2015 focusing on GME approaches to GH. Articles meeting inclusion criteria were evaluated for content by authors to ensure relevance. Methodological quality was assessed using the Medical Education Research Study Quality Instrument (MERSQI), which has demonstrated reliability and validity evidence.</i></p> <p><i>RESULTS: Overall, 69 articles met initial inclusion criteria. Articles represented research and curricula from a number of specialties and a range of institutions. Many studies reported data from a single institution, lacked randomization and/or evidence of clinical benefit, and had poor reliability and validity evidence. The mean MERSQI score</i></p>	

	<p>among 42 quantitative articles was 8.87 (2.79).  <b>CONCLUSIONS :</b> There is significant heterogeneity in GH curricula in GME, with no single strategy for teaching GH to graduate medical learners. The quality of literature is marginal, and the body of work overall does not facilitate assessment of educational or clinical benefit of GH experiences. Improved methods of curriculum evaluation and enhanced publication guidelines would have a positive impact on the quality of research in this area.</p>	
<p><b><i>Enhanced publication management systems a systemic approach towards modern scientific communication</i></b></p>	<p><i>Enhanced Publication Information Systems (EPISs) are information systems devised for the management of enhanced publications (EP), i.e. digital publications enriched with (links to) other research outcomes such as data, processing workflows, software. Today, EPISs are typically realised with a "from scratch" approach that entails non-negligible implementation and maintenance costs. This work argues for a more systemic approach to narrow those costs and presents the notion of Enhanced Publication Management Systems, software frameworks that support the realisation of EPISs by providing developers with</i></p>	

	<i>EP-oriented tools and functionalities.</i>	
<b><i>Enhanced publications in v4 countries</i></b>	<p><i>The article describes the project Enhancing scholarly communication: National initiatives to manage research data in V4 countries. The main goal of the project is a survey about state of research data management, repository contents, services and archiving policies in colleges, universities and research institutions on national levels. The results of the survey will be presented by representatives from Hungary, Czech Republic, Slovak Republic and Poland. This text is a collaborative work by authors participating in the “Enhancing scholarly communication: national initiatives to manage research data in the V4 countries” project. This paper contains five parts:</i></p> <ul style="list-style-type: none"> <li><i>• An introduction to Open Access in scholarly communication, research data and projects.</i></li> <li><i>• The report from the survey of Hungarian universities.</i></li> <li><i>• The report from the Survey of Polish Scientific and Research-Development Units.</i></li> <li><i>• The report from the survey of Slovak universities and scholarly institutions.</i></li> <li><i>• The report from the survey of Czech research institutions and universities.</i></li> </ul>	

<p style="text-align: center;"><b>1</b></p> <p style="text-align: center;"><b><i>A framework supporting the shift from traditional digital publications to enhanced publications</i></b></p>	<p><i>Enhanced publications (EPs) can be generally conceived as digital publications "enriched with" or "linking to" related research results, such as research data, workflows, software, and possibly connections among them. Enhanced Publication Information Systems (EPISs) are information systems devised for the management of EPs in specific application domains. Currently, no framework supporting the realization of EPISs is known, and EPISs are typically realized "from scratch" by integrating general-purpose technologies (e.g. relational databases, file stores, triple stores) and Digital Library oriented software (e.g. repositories, cataloguing systems). Such an approach is doomed to entail non-negligible realization and maintenance costs that could be decreased by adopting a more systemic approach. The framework proposed in this work addresses this task by providing EPIS developers with EP management tools that facilitate their efforts by hiding the complexity of the underlying technologies.</i></p>	<p><i>Referenced research data is connected to its generating workflow via the relationship class <b>LinksTo</b>. In this case the semantics of the relationship has been plugged in as simple free text.</i></p> <p><b>Requirement 1.</b> <i>Supporting different back-ends for data storage. The framework should provide a storage management module for the configuration of the storage back-ends to use. Depending on the functional requirements of the target EPIS, a type of back-end may be preferable to another. For example, a triple store could be more suitable to support the export of EPs via Linked Data, while a full-text index could be dedicated to support search on the text and on the metadata of publications.</i></p> <p><b>Requirement 2.</b> <i>Offering data definition, manipulation, and access languages. The framework should provide a language for the definition of EP data models (EP-DMDL, EP Data Model Definition Language). In addition, developers of EPISs should be able to operate on EP instances (compliant to the defined EP data model) with a dedicated domain-specific language that allows manipulation of resources</i></p>
--	--	--

		<p>whose types are defined in the EP data model (EP-DSML, EP Domain Specific Manipulation Language).</p> <p><b>Requirement 3.</b> Enabling data sharing. The framework should support the export of content via different standard APIs and protocols to serve third-party applications. Examples are OAI-PMH, OAI-ORE, Linked Data, OpenSearch, RDF.</p> <p><b>Requirement 4.</b> Supporting data portability.</p> <p>To further promote data sharing and re-use, the framework should support open standards for the representation of data. A data transformation module may take care of transforming data from one format to another, ensuring a high level of interoperability and portability.</p> <p><i>EP-specific requirements</i></p> <p><b>Requirement 5.</b> Support the integration of heterogeneous data sources.</p> <p>EPs are digital objects that aggregate research outputs and related material. Such content is already available from different types of data sources, such as literature and data repositories, archives, metadata aggregators, and scientific databases. Data sources</p>
--	--	--

		<p><i>export different typologies of content according to different formats and via different protocols. EPMSs should support developers in the integration of such diverse content, providing (i) built-in functionalities for the import of content via standard protocols and (ii) supporting the implementation of ad-hoc integration functions, for those data sources exporting content via idiosyncratic protocols and formats.</i></p> <p><b>Requirement 6.</b> <i>Support the management of dynamic data sources. Data sources are typically dynamic in terms of (i) supported protocols and formats and (ii) availability of services and content. Furthermore, the list of integrated data sources may change during the lifetime of an EPIS. EPMSs should therefore provide data source management functionality to ease the administrative operations needed to take care of the dynamic nature of the data sources.</i></p> <p><b>Requirement 7.</b> <i>Support the integration of content. Due to the heterogeneity of the content, a transformation and harmonization module is necessary in order to massage the incoming material and transform it in a homogeneous format, so that further</i></p>
--	--	---



		<p><i>operations can be performed on content without tackling again the peculiarities of each data source.</i></p> <p><b>Requirement 8.</b> <i>Enable the customization of the EP data model. Components of EPs in one application domain may be very different from the components of EPs in another domain, both in terms of structure and semantics. Therefore, an EPMS should not provide a pre-defined model, as it would be either too generic or too specific. A generic EP data model would be usable in any domains, but it would not cope with specific requirements of the target community. On the other hand, if the model captures some community-specific requirements, it would be not usable in different contexts. EPMSs should instead offer tools for the definition of EP data models, so that EPIS developers can define the structure and semantics of the EPs they want to manage.</i></p> <p><b>Requirement 9.</b> <i>Support the enrichment and curation of content. When we create EPs, we are not just creating a new digital object by composing other digital objects. We are also creating new relationships between objects that can be useful to</i></p>
--	--	---

		<p><i>better the quality of the EPs and enrich the original content. For example: let's suppose we have imported metadata records about the same scientific article from two different data sources. If we can understand that they are indeed about the same article, then it is possible to eliminate duplicates by merging the two records into one, so as to better the quality of the integrated content.</i></p>
--	--	--

<p><b><i>Providing access to research data, publications and current research information at data archiving and networked services - DANS</i></b></p>	<p><i>Data Archiving and Networked Services (DANS) promotes sustained access to digital research data in the Netherlands. Researchers can deposit their data through the online archiving system EASY. Via the portal NARCIS the research data are shown in context, namely in relation to e-publications, and other research information. Both EASY and NARCIS contain grey literature like archaeological reports, data documentation, doctoral thesis, conference papers, patents, and technical documentation of universities and other research institutes.</i></p>	
<p>2</p> <p><b><i>Enabling living systematic reviews and clinical guidelines through semantic technologies</i></b></p>	<p><i>In clinical medicine, secondary research that produces systematic reviews and clinical practice guidelines is key to sound decision-making and quality care. Having machine-readable primary study publications, namely the methods and results of published human clinical trials can greatly improve the process of summarizing and synthesizing knowledge in medicine. In this short introduction to the problem, we provide a brief review of the related literature on various efforts to produce semantic technologies for sharing and reusing content from clinical investigations (RCTs and other clinical primary studies). Using an illustrative case, we outline some of the necessary metadata that needs to be captured in order to achieve some initial automation in authorship of systematic reviews and clinical guidelines. In addition, we list desiderata that we believe are needed to reduce the time and costs of maintaining these documents. These include linking provenance</i></p>	<p>NÃO HÁ REQUISITOS SOBRE ENHANCED PUBLICATION</p>

	<p><i>information to a much longer scientific investigation lifecycle, one that incorporates a single study's role all the way through its use in clinical guideline recommendations for patient treatments..</i></p>	
<p><b><i>Survey of enhanced publications in the Czech Republic</i></b></p>	<p><i>The poster presents survey about the state of enhanced publications in the Czech Republic. The goal of this survey was to find out situation of enhanced publications at research institutions in the Czech Republic. What research data arise? How are they stored and archived? To whom they made available? Are there the real enhanced publications, i.e. research publications linked directly to research data? 113 Czech research institutions were contacted with online questionnaire during 11th September to 7th October 2013. The poster will show the results from 65 Czech research institutions which filled in the questionnaire.</i></p>	<p>NÃO HÁ REQUISITOS SOBRE ENHANCED PUBLICATION</p>

3

***Enhanced publications: Data models and information systems***

*Enhanced publications" are commonly intended as digital publications that consist of a mandatory narrative part (the description of the research conducted) plus related "parts", such as datasets, other publications, images, tables, workflows, devices. The state-of-the-art on information systems for enhanced publications has today reached the point where some kind of common understanding is required, in order to provide the methodology and language for scientists to compare, analyse, or simply discuss the multitude of solutions in the field. In this paper, we thoroughly examined the literature with a two-fold aim: firstly, introducing the terminology required to describe and compare structural and semantic features of existing enhanced publication data models; secondly, proposing a classification of enhanced publication information systems based on their main functional goals.*

*Enhanced publications are digital objects characterized by an identifier (possibly a persistent identifier) and by descriptive metadata information. The constituent components of an enhanced publication include one mandatory textual narration part (the description of the research) and a set of interconnected sub-parts. Parts may have or not have an identifier and relative metadata descriptions and are connected by semantic relationships. In general, enhanced publication data models vary in the way they define the structure of their parts, metadata, and relationships, which reflect and support the functional goals of a given enhanced publication information system. For example in some approaches an enhanced publication is a "package" embedding all its subparts, i.e. sub-parts cannot be shared by different enhanced publications; in other approaches, parts can instead be referenced, shared or*

		<p><i>passed as inputs to workflow engines. In some solutions the narrative part of an enhanced publication is intended in a traditional (for the digital world) sense as a readable file (PDF, DOCX, etc.); while in others, the text is structured into interconnected sub-parts, e.g. sections, figures, tables. Figure 1 shows two enhanced publications consisting of the narrative part, representing the scientific article, and supplementary material: the slides presented at a conference, the video of the presentation, and one spreadsheet of related research data. The two publications have an analogous structure for the supplementary material, but differ in the nature of the mandatory narrative: a single PDF file versus a “structured text”, i.e. the text is made of several interconnected sub-parts such as abstract, sections, figures, tables, etc. The first model is generally preferable in digital library settings where the traditional management of PDF articles is to be</i></p>
--	--	--

		<p><i>enriched with supplementary material. Depending on the data model, parts may be provided by uploading files from the file system or providing references to files (e.g. URLs, DOIs).</i></p>
--	--	--

4

***Transforming a research paper into a rich internet publication***

*Scientific communication is gradually changing. In the web era the genre of the printed paper has survived in digital format as static PDF, but annotations, multimedia and access to data are appearing as new interactive features of innovative scholarly journals. Together they justify the term Rich Internet Publication (RIP) as name for a new emerging genre. Because of the authoritative role of publications in highly ranked journals, a RIP is mostly used as a 'showcase' to publish supplementary material that could not be made available otherwise, and to generate a broader interest in a project. The traditional authoring of linear text does not adequately fit the new situation, in which the author has to deal with much more information in various formats and has access to new options for presentation. This paper describes a strategy to transform a regular paper into a RIP. The strategy is divided into a conceptual stage, in which focal points are created and a balanced mixture of different media are achieved divided over different levels, and the implementation.*

*Hyperlinking* pode ser usado como mecanismo para conectar um artigo com anexos,



<p><b><i>EPUB3 for integrated and customizable representation of a scientific publication and its associated resources</i></b></p>	<p><i>Scientific publications point to many associated resources, including videos, prototypes, slides, and datasets. However, discovering and accessing these resources is not always straightforward: links could be broken, readers may be offline, or the number of associated resources might make it difficult to keep track of the viewing order. In this paper, we explore potential integration of such resources into the digital version of a scientific publication. Specifically, we evaluate the most common scientific publication formats in terms of their capability to implement the desirable attributes of an enhanced publication and to meet the functional goals of an enhanced publication information system: PDF, HTML, EPUB2, and EPUB3. In addition, we present an EPUB3 version of an exemplary publication in the field of computer science, integrating and interlinking an explanatory video and an interactive prototype. Finally, we introduce a demonstrator that is capable of outputting customized scientific publications in EPUB3. By making use of EPUB3 to create an integrated and customizable representation of a scientific publication and its associated resources, we believe that we are able to augment the reading experience of scholarly publications, and thus the effectiveness of scientific communication.</i></p>	
--	---	--

<p><b><i>Full-text and topic based authorrank and enhanced publication ranking</i></b></p>	<p><i>The idea behind AuthorRank is that a content created by more popular authors should rank higher than the content created by less popular authors. This paper brings this idea into scientific publications analysis to test whether the optimized topical AuthorRank can replace or enhance topical PageRank for publication ranking. First, the PageRank with Priors (PRP) algorithm was employed to rank topic-based publications and authors. Second, the first author's reputation was used for generating an AuthorRank score. Additionally, linear combination method of topical AuthorRank and PageRank were compared with several baselines. Finally, as shown in our evaluation results, the performance of topical AuthorRank combined with topic-based PageRank is better than other baselines for publication ranking.</i></p>	
<p><b><i>OAIzer: Configurable OAI Exports over Relational Databases</i></b></p>	<p><i>Modern Digital Library Systems (DLSs) typically support information spaces of interconnected objects, whose graph-like document models surpass the traditional DL payload-metadata document models. Examples are repositories for enhanced publications, CRIS systems, cultural heritage archives. To enable interoperability, DLSs expose their objects and interlinks with other objects as "export packages", via standard exchange formats (e.g. XML, RDF encodings) and OAI-ORE or OAI-PMH protocols. This paper presents OAIzer, a tool for the easy configuration and automatic deploy of OAI interfaces over an RDBMSbased DLS. Starting from the given relational representation of a</i></p>	

	<p><i>document model, OAIzer provides DLS developers with user interfaces for drafting the intended structure of export packages and the automated deploy of OAI endpoints capable of exporting such packages.</i></p>	
<p style="text-align: center;"><b>5</b></p> <p><b><i>Information bulletin on variable stars - rich content and novel services for an enhanced publication</i></b></p>	<p><i>We describe the features of a small enhanced journal, the Information Bulletin on Variable Stars (IBVS). It was founded 50 years ago as a bulletin, changed to a refereed express journal, appeared on the web early on, and is now electronic only, with rich content and services. Technically skilled authors, freely available bibliographic services and discipline-wide standardization - all characteristic of astronomy - form the foundation on which this uniquely enhanced journal is built, enabling it to provide quality services to its research community.</i></p>	<p><i>Besides literature linking, we have links for astronomical databases. The authors (or the editors) might insert special LaTeX macros to produce such links. We provide astronomical object links to the SIMBAD database at CDS, the on-line version of the General Catalog of Variable Stars (GCVS, Moscow), the NASA Extragalactic Database (NED) and dataset (photographic plate) links to the Wide-Field Plate Database (WFPDB) in Sofia. Persistent data set identifiers are extremely desirable. The use of DOIs by DataCite is an excellent solution. The only reason we do not use them is that we have no revenue, and could not afford the cost, however small and reasonable they are. But we do have data set identifiers, which are resolvable. One can reach the meta-data and the data</i></p>

		<p><i>set itself by pasting the identifier to the end of an URL. IBVS makes the article meta-data visible to the readers and sends it via email, at the time of publication on the web, to the relevant data centers for indexing, the ADS and CDS. We plan to make the meta-data harvestable by OAI-PMH protocol too We maintain meta-data deeper than the article level. Each figure, and each data file, has its own meta-data, including object information. The readers might search for figures or data files using object names and special keywords</i></p>
<p><b><i>Linking full-text grey literature to underlying research and post-publication data: An Enhanced Publications Project 2011-2012</i></b></p>	<p><i>This project seeks to circumvent the data vs. documents camp in the grey literature community by way of a middle ground provided through enhanced publications. Enhanced publications allow for a fuller understanding of the process in which data and information are used and applied in the generation of knowledge. The enhanced publication of grey literature precludes the idea of a random selection of data and information, and instead focuses on the human intervention in data-rich environments. The definition of an enhanced publication is borrowed</i></p>	

	<p><i>from the DRIVER-II project, "a publication that is enhanced with three categories of information: research data, extra materials, and post-publication data". Enhanced publications combine textual resources i.e. documents intended to be read by human beings, which contain an interpretation or analysis of primary data. Enhanced publications inherently contribute to the review process of grey literature as well as the replication of research and improved visibility of research results in the scholarly communication chain.</i></p>	
<p><b><i>GL-Conference Series: Conference Proceedings</i></b></p>	<p><i>The proceedings contain 19 papers. The topics discussed include: GeoStoryteller: taking grey literature to the streets of New York; acquisition and distribution of technical reports and conference proceedings on science and technology in Korea; open is not enough: a case study on grey literature in an OAI environment; grey literature matters: the role of grey literature as a public communication tool in risk management practices of nuclear power plants; management of obsolete grey literature in engineering research institutions; audit DRAMBORA for trustworthy repositories: a study dealing with the digital repository of grey literature; federal information system on GL in Russia: a new stage of development in digital and network environment; enhancing diffusion of scientific contents: open data in repositories; research product repositories: strategies for data and metadata quality control; and linking full-text grey literature to underlying research and post-</i></p>	

	<i>publication data: an enhanced publications project 2011-2012.</i>	
<b><i>Editor's note enhanced publications project (EPP)</i></b>	<b>Sem resumo</b>	
<b><i>Institutional repositories, long term preservation and the changing nature of scholarly publications</i></b>	<i>The web offers new opportunities for scholars to publish the outcome of their research. One of these new forms is called Enhanced Publications. In an Enhanced Publication different objects and files that has a meaningful and close relation to each other are aggregated on the level of a resource map in witch not only the separate files are described, but also the relation between those files are. An example of an Enhanced Publication is a digital text publication and a dataset on which the publication is based. Preserving these compound entities in the existing infrastructures raises new issues. This article discusses these issues against the background of the Dutch long term preservation infrastructure and organisation.</i>	

***Rich internet publications: "Show what you tell"***

*The journal article is still the basis of scholarly communication. This genre, however, largely adheres to the rules of the printed publication and does not meet the requirements of this age of digital Web publishing. Today we do not need to restrict ourselves any longer to communicating the results of the research process only. We can also allow readers to inspect the underlying data online, to publish their own comments and, using a variety of multimedia content, to be witness to intermediary stages of the scientific discovery process. This development has stimulated the transformation of the conventional article: when published in a digital format, it is more and more enhanced with data sets, photos, videos, interactive maps and animations; these enhancements affect its structure and layout. A variety of new publication formats is appearing, some of which can be no longer adequately described as simply "enhanced" publications. They are rather to be conceived as a new genre, for which we propose the term Rich Internet Publication (RIP), analogue to the well-known concept of Rich Internet Application. Both share features of information integration, visualization and exploration (i.e. non-linear reading), typical for hypermedia products. This paper relates apparently quite different variants of scholarly publishing by highlighting common elements of data access and delineates the gradation of publication enhancement. RIPs do not constitute a sharply delimited category, but are part of a broad spectrum, which starts with*

*regular enhanced publications closely resembling their printed counterparts, and ends with high-quality multimedia presentations having more in common with Web applications than with the conventional journal article. We distinguish two subcategories: RIP type I is primarily based on a linear text, but fully integrated with multimedia content and tools to access and analyze data, while RIP type II is more image-driven, has a user interface with more graphic elements and encourages explorative, non-linear reading. The production of enhanced publications and RIPs is not yet a straightforward process. It requires extra effort from the author, which is currently insufficiently rewarded. This may change when funding agencies get more interested in research products that go beyond the level of textual publications. Dedicated tools for construction of RIPs are equally important, which requires consensus on architecture and infrastructure. Development of these tools could fit in with the recently started research line of adding semantic metadata to object-based enhanced publications. Moreover, the creation of a RIP will rely on the author's basic competencies of e-scholarship. When authors start creating RIPs on a larger scale, the process of exchanging and preserving them has to be supported. Usually, a RIP is not a single static file, which can be downloaded and attached to an email, but a set of related components. Preserving the content's integrity will be a major concern.*



<p><b><i>Supporting semantically enhanced web service discovery for enterprise application integration</i></b></p>	<p><i>The availability of sophisticated Web service discovery mechanisms is an essential prerequisite for increasing the levels of efficiency and automation in EAI. In this chapter, we present an approach for developing service registries building on the UDDI standard and offering semantically-enhanced publication and discovery capabilities in order to overcome some of the known limitations of conventional service registries. The approach aspires to promote efficiency in EAI in a number of ways, but primarily by automating the task of evaluating service integrability on the basis of the input and output messages that are defined in the Web service's interface. The presented solution combines the use of three technology standards to meet its objectives: OWL-DL, for modelling service characteristics and performing fine-grained service matchmaking via DL reasoning, SAWSDL, for creating semantically annotated descriptions of service interfaces, and UDDI, for storing and retrieving syntactic and semantic information about services and service providers.</i></p>	
<p><b>6</b></p> <p><b><i>Commercializing public sector information: Privacy and security concerns</i></b></p>	<p><i>Advances in information and communication technologies in government has produced new avenues for agencies to collect, share and re-use data. As such, through the commercial re-use of public sector information (PSI), the prospect of increasing revenue becomes enticing for governments and their agencies. They may be used to simplify the dissemination of residential property transaction details, land title information, ordinance survey data and street address registers.</i></p>	<p>NÃO HÁ REQUISITOS SOBRE ENHANCED PUBLICATION</p>

	<p><i>However, as with any other advances, the commercial re-use of PSI raises concerns about its possible effects. One is the fact that there is a tendency to for e-government organizations to move from being service providers to market-oriented, income generators. Another is that the enhanced publication of PSI could raise security responsibilities and may lead to conflicts within and between government agencies pertaining to the dissemination of information.</i></p>	
<p><b>7</b></p> <p><i>Enhanced publication services to members</i></p>	<p><b>Sem resumo</b></p>	<p>NÃO HÁ REQUISITOS SOBRE ENHANCED PUBLICATION</p>

<p style="text-align: center;"><b>8</b></p> <p style="text-align: center;"><b><i>A Vision for Open Cyber-Scholarly Infrastructures</i></b></p>	<p><i>The characteristics of modern science, i.e., data-intensive, multidisciplinary, open, and heavily dependent on Internet technologies, entail the creation of a linked scholarly record that is online and open. Instrumental in making this vision happen is the development of the next generation of Open Cyber-Scholarly Infrastructures (OCIs), i.e., enablers of an open, evolvable, and extensible scholarly ecosystem. The paper delineates the evolving scenario of the modern scholarly record and describes the functionality of future OCIs as well as the radical changes in scholarly practices including new reading, learning, and information-seeking practices enabled by OCIs</i></p>	<p><i>A dataset, once accepted for deposit and archived, is assigned by a Registration Agency a Digital Object Identifier (DOI) for registration. A Digital Object Identifier (DOI) is a unique name (not a location) within a name space of a networked data environment and provides a system for persistent and actionable identification of datasets. A dataset must be accompanied by metadata, which describes the information contained in the dataset, details of data formatting and coding, how the dataset was collected and obtained, associated publications, and other research information. The conceptual model of an enhanced publication includes a mandatory text body and a set of interconnected sub-parts. Several instantiations of this model have been proposed in the literature. These instantiations, essentially, regard the way the mandatory text body is organized, the type of the sub-parts, and the way they are connected to the text. A first instantiation regards the case where the sub-parts are essentially</i></p>
--	---	---

		<p><i>supplementary material along with the mandatory text. Examples include presentation slides, appendixes to the text, tables, etc. In this case, generally, the sub-parts do not have an identifier and are not described by metadata.</i></p> <p><i>A second instantiation regards the case where the mandatory text body is not a single block of text but is structured in a number of interconnected modules, such as abstract, sections, bibliography, etc.</i></p> <p><i>A third instantiation regards the case where the sub-parts are scientific datasets external to the publication, i.e., stored in discipline specific data centers/repositories with their own identity (DOIs). In this case, the scientific datasets are cited from within the text using a DOI system.</i></p> <p><i>A fourth instantiation regards the case where some sections or modules of the text body or some sub-parts are live, meaning that they can be activated in order to produce visual content, video streaming, etc.</i></p> <p><i>Finally, a fifth instantiation regards the case where some sections or modules of the text body or</i></p>
--	--	---

		<p><i>sub-parts can be dynamically executed at run time.</i></p> <p><i>A generalization of the concept of “Enhanced Publication” is the concept of Research Object (RO). Informally, a Research Object is intended as a semantically rich aggregation of resources that poses some scientific intent or supports some research objective. It should allow a principled publication of the results of research activity in a self-contained manner that facilitates the sharing and reuse of these objects. An RO bundles together all the essential information relating to a scientific investigation, i.e., article, data produced/used, methods used to produce and analyze that data, as well as the people involved in the investigation. In addition, an RO includes additional semantic information that allows one to link its components in a meaningful way.</i></p> <p><i>Scientific articles are increasingly being assigned DOIs that provide live links from online citing articles to the cited articles in their reference lists. In addition, they should be enriched with appropriate metadata.</i></p>
--	--	---

		<p><i>DOIs could logically be assigned to every single article module; having the possibility to make references to article modules would be highly desirable. Linked Scholarly Record</i></p> <p><i>The scholarly record is poorly interconnected. This is in opposition to modern science, which requires the establishment of discipline-specific linked scientific records in order to effectively support scholarly inquiry. In fact, scientists and scholars need to be able to move from hypotheses to evidence, from article to article, from dataset to dataset, and from article to dataset and conversely. They need to discover potentially significant patterns and ways to make meaningful connections between parts of the scholarly record.</i></p> <p><i>From a conceptual point of view, a linked scholarly record means that its single parties, i.e., a dataset, an article module, etc. constitute single nodes of a networked scholarly record that can be accessed by any scholar, anytime, anywhere.</i></p> <p><i>The two pillars of the modern scholarly communication are</i></p>
--	--	--

		<p><i>discipline-specific Data Centers and Research Digital Libraries, whose technologies and organizations allow researchers to store, curate, discover, and reuse the data and publications they produce. Made to implement complementary phases of the scientific research and publication process, they are poorly integrated with one another and do not adopt the strengths of the other. Such a dichotomy hampers the realization of a linked scholarly record. However, I am confident that the recent technological advances in many fields of information technology will make it happen. A linked disciplinary-specific data space should enjoy the following properties:</i></p> <p><i>it contains datasets specific to a scientific discipline;</i></p> <p><i>any scientific community belonging to this discipline can publish on the scientific data space;</i></p> <p><i>dataset creators are not constrained by the choice of vocabularies with which to represent them;</i></p>
--	--	---

		<p><i>datasets are connected by links creating a global data graph that spans datasets and enables the discovery of new datasets;</i></p> <p><i>datasets are self-describing;</i></p> <p><i>datasets are strictly separated from formatting and presentational aspects;</i></p> <p><i>the scientific data space is open, meaning that applications do not have to be implemented against a fixed set of datasets, but can discover new datasets at run time by following the data links. A managed linked data space will enable researchers to start browsing in one dataset and then navigate to related datasets; or it can support data search engines that crawl the data space by following links between datasets. However, in order to be able to implement a linked discipline-specific data space, the ability to meaningfully and formally describe the datasets that participate in the linked data space, as well as the links among them, is of paramount importance.</i></p> <p><i>Metadata is the descriptive information about datasets that</i></p>
--	--	---



		<p><i>explains the measured attributes, their names, units, precision, accuracy, data layout, and ideally a great deal more. Most importantly, metadata should include the dataset lineage, i.e., how the dataset was measured, acquired, or computed. Equally important is the concept of the dataset identifier, i.e., DOI (or URI) as mechanisms for referring to datasets, on which there exists some agreement among multiple data providers.</i></p> <p><i>Modeling the many kinds of relationships existing between datasets is equally important. We need to define metadata models for describing links. We must be able to model, for example, dataset B as a temporal/spatial abstraction of dataset A; or show that datasets A and B are generated independently but both reflect the same observational or experimental activity; or that datasets A and B were generated at the same time and by the same organization, etc.</i></p> <p><i>In order to be able to exploit the full potential of the linked data space, it is importance to make sense of heterogeneous datasets that</i></p>
--	--	--

		<p><i>constitute a linked data space. This can be achieved by adopting formalisms for representing discipline-specific ontologies. An initiative that implements the concept of linked data space by using the semantic Web technologies is Linked Data [4]. However, in order to be able to implement a linked, discipline-specific literature space, it is important to meaningfully and formally describe the article modules that participate in the linked literature space as well as the links among them. Publishers are beginning to embrace the opportunity to integrate data with scientific articles but barriers to the sustainability of this practice include the sheer volume of data and the huge variety of data formats. Several levels of integration can be achieved ranging from tight to weak integration. A tight integration is achieved when datasets are contained within peer-reviewed articles. In this publishing model, the publisher takes full responsibility for the publication of the article and the aggregated data embedded in it and</i></p>
--	--	--

		<p><i>the way it is presented. The embedding of the dataset into the publication makes it citable and retrievable. However, the reusability of the dataset is limited as it is difficult to find it separate from the publication. This publishing model is not appropriate when the embedded dataset is too large to fit into the traditional publication format.</i></p> <p><i>In addition, the preservation of these enhanced articles is more demanding than for traditional articles.</i></p> <p><i>A less tight integration is achieved when the datasets reside in supplementary files added to the scientific article. The publisher offers authors the option of adding supplementary files to their article containing any relevant material that will not fit the traditional article format or its narrative, such as datasets, multimedia files, large tables, animations, etc. There are some issues related to this publishing model: they mainly concern the preservation of the supplementary files as well as the ability to find them independently from the main publication.. A weak integration is achieved when the datasets reside in</i></p>
--	--	---

		<p><i>Institutional Data Repositories or in discipline-specific Data Centers with bi-directional linking to and from articles. In this publishing model the article should include a citation and links to the dataset. The data preservation is the responsibility of the administrators of the Institutional Repository or Data Center. In this model the datasets become better discoverable and can be reused separately from the publication and in combination with other datasets. However, this publishing model depends very much on the existence of proper and persistent linking mechanisms enabling bi-directional citation. In the Big Data era it is obvious that only the weak integration scheme is viable. Unfortunately, due to technological and policy reasons discipline-specific Data Centers and Research Libraries currently do not interoperate.</i></p>
--	--	--

<p><b><i>The VeteranTapes: Research corpus, fragment processing tool, and enhanced publications for the e-Humanities</i></b></p>	<p><i>Enhanced Publications are a new way to publish scientific and other results in an electronic article. The advantage of EPs is that the relation between the article and the underlying data facilitate the peer review process and other quality assessment activities. Due to the link between the publication and the research data the publication can be much richer than a paper edition permits. We present an example of EPs in which links are made to interview fragments that include transcripts, audio segments, annotations and metadata. EPs call for a new paradigm of research methodology in which digital persistent access to research data are a central issue. In this contribution we highlight 1. The research data as it is archived and curated, 2. the concept "enhanced publication" and its scientific value, 3. the "fragment fitter tool", a language processing tool to facilitate the creation of EPs, 4. IPR issues related to the re-use of the interview data.</i></p>	
<p><b><i>Web service discovery in a semantically extended uddi registry: The case of fusion</i></b></p>	<p><i>Service-oriented computing is being adopted at an unprecedented rate, making the effectiveness of automated service discovery an increasingly important challenge. UDDI has emerged as a de facto industry standard and fundamental building block within SOA infrastructures. Nevertheless, conventional UDDI registries lack means to provide unambiguous, semantically rich representations of Web service capabilities, and the logic inference power required for facilitating automated service discovery. To overcome this important limitation, a number of</i></p>	

	<p><i>approaches have been proposed towards augmenting Web service discovery with semantics. This paper discusses the benefits of semantically extending Web set-vice descriptions and UDDI registries, and presents all overview of the approach put forward in project FUSION, towards semantically-enhanced publication and discovery of services based on SAWSDL.</i></p>	
<p><b><i>The Enrichment of Publications</i></b></p>	<p><i>An enhanced publication is a new type of publication consisting of a traditional publication (a book, article or report) enriched with additional information, including underlying research data, audiovisual materials or commentary. An enriched publication can change over time. This enhancement makes research more transparent. The big challenge is: how can we ensure that a researcher without technical expertise can make a composite publication? This article gives an overview of the steps involved the creation of this new infrastructure and makes recommendations for future activities. Adapted from the source document.</i></p>	

9

***DRIVER: building a sustainable infrastructure for global repositories***

*Purpose - DRIVER embodies a bold vision - that of worldwide networks of scientific data repositories. This paper seeks to examine the aims of the European Union funded project, to explore the development of a distributed infrastructure that enables enhanced interoperability of data, resulting in a global knowledge infrastructure supporting the scholarly communication of the future. Design/methodology/approach - The primary objective of DRIVER was to establish a flexible, robust, and scalable infrastructure for all European and world-wide digital repositories, managing scientific information in an open access model increasingly demanded by researchers, funding organisations and other stakeholders. Adopting a result-driven approach, activities focused on the expansion of the content base with high quality research outputs, including textual research papers, data sets and other scholarly publications. Findings - The release of the D-NET v1.0 open source software proved a successful basis for a distributed service-oriented architecture, enabling enhanced interoperability of data and service-providers, and offering wide-ranging functionality including search; recommendation; collection building, and personal profiling as innovative tools for repository managers. In addition, it was found that in building a robust network of voluntary content providers, known as the DRIVER Confederation, the infrastructure came to support a durable organisational structure, now formally constituted as the Confederation of Open Access Repositories*

NÃO HÁ REQUISITOS SOBRE ENHANCED PUBLICATION

*(COAR). The international repository organisation enables further collaboration between research communities in a co-ordinated network comprising a growing number of institutional repositories, national federations and research institutions and data aggregators. Practical implications - The development of COAR is the extension of the EU-based infrastructure to global research communities in China, India, Africa and Latin America, deploying a vigorous awareness and advocacy programme. Evolving from the DRIVER Confederation, COAR aims to provide an ongoing support service for repository managers, in a dynamic set of guidelines aimed at data interoperability, and to provide the strategic support required to implement new forms of scholarly communication. These issues are addressed in terms of technical infrastructure developments but will focus on strategic issues of policy development, improved services and additional functionality offered to the scholarly community. Originality/value - This paper outlines DRIVER's unique response to the changing global information environment. Concepts of strategic international collaboration are pursued in COAR, based on the scientific and technical collaboration achieved in DRIVER. The paper addresses significant repository development goals that currently challenge repository managers, librarians, scholars and funders and that indicate the future of Open Access publication - in the ultimate goal of a global and interactive representation of human knowledge.*



10

*Scratching beneath the surface of communities of (mal)practice*

*This paper seeks to surface less positive aspects of communities of practice (CoPs), regardless of emergent or organisationally managed, grounded in political-power interactions. Examples are provided from the authors' experiences of a research-based CoP within UK higher education. The paper is primarily theoretical with empirical examples drawn from a descriptive CoP case study. The paper discusses the following themes: the impact of timing on CoP development; the impact of CoP leaders and managers in "managed" CoPs; the power-political interrelationship between emergent CoPs and formal organisation; the impact of dominant actors with position power; emotional containment and emotion work within CoPs; power implications of novices and masters and the implications when CoP practices diverge from organisational practices. It finds that to ignore such issues of power within CoPs is to limit the knowledge creation process. Further empirical research is necessary to investigate micro and macro power-political issues of CoPs. In particular, emotional containment and emotion work of CoP members and the impact of this on knowledge creation is worthy of future research. The paper has significant implications for CoPs in practice as the quest for pragmatic mechanisms to develop individual and organisational learning and knowledge creation for competitive advantage.*

NÃO HÁ REQUISITOS SOBRE ENHANCED PUBLICATION

<p style="text-align: center;">11</p> <p><b>Curadoria digital: um novo patamar para preservação de dados digitais de pesquisa</b></p>	<p>Uma parte considerável dos resultados das atividades de pesquisa está sendo criada em formatos digitais. Embora valiosos, esses dados correm o risco de serem perdidos pela obsolescência tecnológica e pela fragilidade inerente da mídia digital. Assim, a gestão de dados de pesquisa em um ambiente digital em rede e distribuído torna-se um desafio crescente para o mundo da pesquisa e para a ciência da informação. Em resposta a este desafio surge o conceito de curadoria digital, que envolve o gerenciamento de dados de pesquisa e seu planejamento, garantindo sua preservação, descoberta, interpretação e reutilização a longo prazo. Nesse sentido, este estudo examina brevemente a importância dos dados de pesquisa e da ideia de curadoria digital e seu impacto na formulação de novos documentos e comunicação científica.</p>	<p>NÃO HÁ REQUISITOS SOBRE ENHANCED PUBLICATION</p>
<p><i>Production and use of information. Characterization of informetric distributions using effort function and density function Exponential informetric process</i></p>	<p>Não tem relação com <i>enhanced publication</i></p>	
<p><i>XL: An XML programming language for Web service specification and</i></p>	<p>Não tem relação com <i>enhanced publication</i></p>	
<p><i>A practical and robust inter-domain marking scheme for IP</i></p>	<p>Não tem relação com <i>enhanced publication</i></p>	
<p><i>Towards a more accessible e-government in Jordan: an evaluation study of visually impaired users and Web</i></p>	<p>Não tem relação com <i>enhanced publication</i></p>	
<p><i>Non-tenured leadership appointments in Nigerian university libraries: Problems and prospects</i></p>	<p>Não diz respeito a <i>enhanced publication</i></p>	

**APÊNDICE E – Checklist dos requisitos e de validação /Inspeção**

<b>Item</b>	<b>Requisitos</b>	<b>Definição</b>	<b>Item para verificação</b>	<b>Sim</b>	<b>Não</b>
<b>1</b>	<b>Recursos WEB identificados por URI</b>	A publicação ampliada e seus componentes devem ser disponibilizados como recurso <i>Web</i> identificados por URI. O mesmo vale para seus componentes.			
<b>2</b>	<b>Identificador persistente para cada objeto</b>	Para que haja essa integração, o conjunto de dados de pesquisa deve carregar um identificador associado a ele pelo sistema de informação que compôs a agregação/composição.			
<b>3</b>	<b>Deve ser possível vincular objetos digitais à publicação</b>	Devem agrupar os textos de artigos de revistas às imagens, arquivos de vídeos etc.;			
<b>4</b>	<b>Registro e controle de versões da publicação e das partes constituintes</b>	Devem referir as versões específicas do objeto composto. Deve-se registrar a data da última modificação.			

5	<b>Registro das propriedades básicas da publicação e dos recursos que são adicionados</b>	Deve ser possível registrar propriedades básicas da publicação, e dos recursos que são adicionados a ele (títulos, breve descrição, data da última modificação, data de criação, etc.)			
6	<b>Vocabulário controlado e padronizado</b>	Os objetos digitais devem ser interoperáveis utilizando-se vocabulário controlado e padronizado			
7	<b>Atributos são relevantes</b>	Cada componente deve ser digitado semanticamente e uma breve descrição pode ser dada.			
8	<b>Registro da autoria da publicação e dos objetos digitais</b>	Deve ser feita uma distinção clara entre o autor da publicação ampliada e os autores de suas partes componentes).			
9	<b>Preservação a longo prazo</b>	Deve-se garantir a preservação a longo prazo (curadoria).			

10	<b>Registrar as relações entre os recursos web que fazem parte da publicação ampliada</b>	<p>Deve ser possível registrar as relações entre os recursos web que fazem parte da publicação ampliada (os componentes precisam ser descritos e classificados utilizando-se um vocabulário padrão).</p> <p>É necessário que, nas publicações ampliadas, as relações entre os objetos compostos estejam visíveis.</p>			
11	<b>Os documentos devem estar baseados no modelo OAI-ORE</b>	<p>OAI-ORE (<i>Open Archives Initiative Object Reuse and Exchange</i>)– define padrões para a descrição e troca de agregados de recursos Web.</p> <p>As instituições devem tomar medidas que garantam que as publicações ampliadas estejam disponíveis com documentos baseados no modelo OAI-ORE</p>			
12	<b>Deve está disponível em plataforma <i>open access</i></b>	<p>Disponibilização online e sem limitações dos resultados de investigação científica. Pode ser designado como: <i>grátis</i>, quando se refere ao acesso online gratuito, e <i>livre</i>, quando se refere ao acesso online gratuito com alguns direitos adicionais de utilização.</p>			
		<p>O RDF é um modelo padrão para o intercâmbio de dados na Web. A linkagem também é possível com</p>			

13	<b>Deve estar de acordo com os padrões da Web Semântica</b>	a utilização de SPARQL – que é a linguagem de consulta para a Web Semântica. OWL - para construir vocabulários, ou "ontologias"), SKOS - para projetar sistemas de gestão do conhecimento.			
14	<b>Os objetos digitais devem ser conectados por relacionamentos semânticos</b>	Além de apontar para o documento, o link especifica também que tipo de relação este documento tem com o outro.			
15	<b>Os objetos digitais podem ser fornecidos através de upload</b>	Arquivos do sistema ou fornecendo referências a arquivos (por exemplo, URLs, DOIs).			
16	<b>Cada arquivo de dados tem seus próprios metadados, incluindo informações de objeto.</b>	Metadados é a descrição dos conjuntos de dados. São os detalhes de formatação e codificação de dados, como o conjunto de dados foi coletado e obtido, publicações associadas e outras informações de pesquisa. Cada módulo possui seu conjunto de metadados, o que permite que cada parte da unidade de informação possa ser citada separadamente, reutilizada para diferentes propósitos ou ainda recombinação, formando novos documentos complexos.			

**APÊNDICE F – Listagem dos repositórios que possuem publicação ampliada**

<i>1.2 Meter CO Survey Dataverse</i>	<i>Ag-Data-Commons</i>	<i>Ancient Sundials</i>
<i>1000 Functional Connectomes Project</i>	<i>AHEAD</i>	<i>Androgen Receptor Mutations Database</i>
<i>45 and Up Study</i>	<i>Alaska Climate Research Center</i>	<i>Animal Genome Tracks on GBrowse</i>
<i>4TU.Centre for Research Data</i>	<i>Alberta Geological Survey</i>	<i>Antarctic Glaciological Data Center at NSIDC</i>
<i>Inter-Zwei-Dok</i>	<i>AlgaeBase</i>	<i>ANU Data Commons</i>
<i>Aberdeen Birth Cohorts</i>	<i>Allele Frequency Net Database</i>	<i>APID</i>
<i>ACEpepDB: Peptide Database</i>	<i>Allen Brain Atlas</i>	<i>ARACHNE IDAI.objects</i>
<i>ADS</i>	<i>Alzheimer Disease &amp; Frontotemporal Dementia Mutation Database</i>	<i>Archives of programs and data for physics</i>
<i>Adult Blood Lead Epidemiology and Surveillance Interactive Database</i>	<i>AMBDAS Bibliographic Database</i>	<i>arthistoricum.net</i>
<i>Advanced Cooperative Arctic Data and Information Service</i>	<i>American Mineralogist Crystal Structure Database</i>	<i>ASAP</i>
<i>AEKOS Data Portal</i>	<i>American National Election Studies</i>	<i>AspGD</i>
<i>Africa Centre for Population Health</i>	<i>American Type Culture Collection</i>	<i>ASTER JPL</i>
<i>AfricaRice Dataverse</i>	<i>AMODS</i>	<i>Astrophysics Source Code Library</i>
<i>ATNF Pulsar Catalogue</i>	<i>Amsterdam Cohort Studies</i>	<i>Atmospheric Chemistry Experiment</i>
<i>Atomic &amp; Molecular Database</i>	<i>BioDare</i>	<i>Brunel figshare</i>
<i>Australian Drosophila Ecology and Evolution Resource</i>	<i>Biodiversity Exploratories Information System</i>	<i>BioDare</i>
<i>Autism Chromosome Rearrangement Database</i>	<i>BioGRID</i>	<i>BSRN</i>
<i>Avon Longitudinal Study of Parents and Children</i>	<i>BioImages - Virtual Field-Guide</i>	<i>BsubCyc</i>
<i>AWI EXPEDITION</i>	<i>Biologic Specimen and</i>	<i>CaltechDATA</i>

	<i>Data Repository Information Coordinating Center</i>	
<i>Bacterial Carbohydrate Structure DataBase</i>	<i>BioModels Database</i>	<i>Canadian Biodiversity Information Facility</i>
<i>Barrow Observatory, Alaska</i>	<i>BioVel</i>	<i>Canadian Climate Data and Scenarios</i>
<i>BeeBase</i>	<i>Bolin Centre Database</i>	<i>Canadian Epigenetics, Environment and Health Research Consortium Network</i>
<i>Behavioral Risk Factor Surveillance System</i>	<i>Boola Boola Forest bird study</i>	<i>Canadian Opinion Research Archive</i>
<i>Benchmark Energy &amp; Geometry Database</i>	<i>Born in Bradford</i>	<i>Canadian Research Data Centre Network</i>
<i>Berman Jewish Data Bank</i>	<i>Brain Transcriptome Database</i>	<i>Canadian Wildland Fire Information System</i>
<i>BILS DOI Repository</i>	<i>BrainMaps.org</i>	<i>caNanoLab</i>
<i>BindingDB</i>	<i>British Antarctic Survey</i>	<i>Cancer Genome Anatomy Project</i>
<i>CANSIM</i>	<i>CfA Dataverses</i>	<i>ClinicalTrials.gov</i>
<i>CaPSURE</i>	<i>Chapman University Digital Commons Datasets</i>	<i>Code Ocean</i>
<i>Carcinogenic Potency Database</i>	<i>ChArMEx database</i>	<i>Cognitive Function and Ageing Study</i>
<i>CARIBIC</i>	<i>Chickpea Transcriptome Database</i>	<i>Coherent X-ray Imaging Data Bank</i>
<i>CARMEN</i>	<i>Child Language Data Exchange System</i>	<i>Collaborative Psychiatric Epidemiology Surveys 2001 - 2003</i>
<i>CASRdb</i>	<i>Chorotree</i>	<i>Collaborative Research Centre 806 Database</i>
<i>Catalytic Site Atlas</i>	<i>CIAT Dataverse</i>	<i>Collaborative Research Centre Transregio 32 Database</i>
<i>Catena</i>	<i>CISER Data Archive</i>	<i>Collaborative Research in Computational Neuroscience</i>
<i>CDC WONDER</i>	<i>Citrination</i>	<i>Columbia University Academic Commons</i>



<i>CEACS Data Library</i>	<i>CLARIN INL Portal</i>	<i>Combined QTL Map of Dairy Cattle Traits</i>
<i>Cellular Phenotype database</i>	<i>CLARIN-PL</i>	<i>Common Cold Project Data Sets</i>
<i>Center for Demography of Health and Aging</i>	<i>Climate Hazards Group Data Archive</i>	<i>Comparative Agendas Project</i>
<i>Center for Engineering Strong Motion Data</i>	<i>Clinical Genomic Database</i>	<i>Comparative Study of Electoral Systems</i>
<i>Center for International Earth Science Information Network</i>	<i>Clinical Proteomic Tumor Analysis Consortium Data Portal</i>	<i>ClinicalTrials.gov</i>
<i>Center of Estonian Language Resources</i>	<i>Clinical Trials Registry - India</i>	<i>Code Ocean</i>
<i>Centre National de Ressources Textuelles et Lexicales</i>	<i>CfA Dataverses</i>	<i>Cognitive Function and Ageing Study</i>
	<i>Danish National Birth Cohort</i>	<i>DEG</i>
<i>Comparative Toxicogenomics Database</i>	<i>Data and Service Center for the Humanities</i>	<i>DEPOD</i>
<i>Comparative Welfare Entitlements Dataset</i>	<i>Data Portal German Marine Research</i>	<i>DepositOnce</i>
<i>COMPLETE Dataverse</i>	<i>Data Repository for the University of Minnesota</i>	<i>Deutsches Register Klinischer Studien</i>
<i>Complete Genomics</i>	<i>Data Sharing for Demographic Research</i>	<i>dharmae</i>
<i>Computational and Information Systems Laboratory Research Data Archive</i>	<i>data.bris Research Data Repository</i>	<i>DHS Data Access</i>
<i>Conserved Domain database</i>	<i>Database of Genomic Variants Archive</i>	<i>Diabetes Study of Northern California</i>
<i>CorrDB</i>	<i>Databrary</i>	<i>Digibug:Repositorio Institucional de la Universidad de Granada</i>
<i>CosmoSim</i>	<i>DataFirst</i>	<i>DigiMorph</i>
<i>Coupled Model Intercomparison Project</i>	<i>DataSTORRE</i>	<i>DEG</i>

<i>cranach.net</i>	<i>datorium</i>	<i>DEPOD</i>
<i>cropPAL</i>	<i>Deakin Research Online</i>	<i>DepositOnce</i>
<i>Cystic Fibrosis Mutation Database</i>	<i>Danish National Birth Cohort</i>	<i>Deutsches Register Klinischer Studien</i>
<i>Dallas Heart Study</i>	<i>Data and Service Center for the Humanities</i>	<i>Electron-Impact Ionization of Multicharged Ions at ORNL</i>
<i>Digital Averroes Research Environment</i>	<i>e-cienciaDatos</i>	<i>Electronic Microscopy Databank</i>
<i>Digital Rocks Portal</i>	<i>eagle-i</i>	<i>Emage</i>
<i>DIGITAL.CSIC</i>	<i>Earth-prints Repository</i>	<i>EMEP</i>
<i>DipnET</i>	<i>EarthByte</i>	<i>Ensembl Bacteria</i>
<i>DisProt</i>	<i>EarthChem</i>	<i>Ensembl Fungi</i>
<i>DNASU plasmid repository</i>	<i>EarthChem Library</i>	<i>Ensembl Genomes</i>
<i>Dog Genome SNP Database</i>	<i>Earthref Magnetis Information Consortium</i>	<i>Electron-Impact Ionization of Multicharged Ions at ORNL</i>
<i>doRiNA</i>	<i>EBI Metagenomics</i>	<i>Electronic Microscopy Databank</i>
<i>DR-NTU (Data)</i>	<i>eBird</i>	<i>Emage</i>
<i>DrugBank</i>	<i>EchoBase</i>	<i>GAVO Data Center</i>
<i>DRYAD</i>	<i>EcoCyc Database</i>	
<i>DSPR</i>	<i>Ecological Archives</i>	
<i>Dutch Dataverse Network</i>	<i>Edmond</i>	
<i>e!DAL - electronic Data Archive Library</i>	<i>eFish</i>	<i>HomoMINT</i>
<i>e-Atlas</i>	<i>eLaborate</i>	<i>FlyReactome</i>
<i>Ensembl Metazoa</i>	<i>European Soil Data Centre</i>	<i>FoodB</i>
<i>Ensembl Plants</i>	<i>European Union Open Data Portal</i>	<i>ForestPlots.net</i>
<i>Ensembl Protists</i>	<i>Europeana</i>	<i>FORSbase</i>
<i>Environmental Change Network</i>	<i>Europäische Geschichte Online</i>	<i>Forschungsdaten- und Servicezentrum der Bundesbank</i>

<i>EOL</i>	<i>ExPASy Bioinformatics Resource Portal</i>	<i>Forschungsdatenzentrum PIAAC bei GESIS</i>
<i>ePapers Repository Birmingham</i>	<i>Experimental Nuclear Reaction Data</i>	<i>fossilworks</i>
<i>EPIC study</i>	<i>Experimental Tropical Watersheds</i>	<i>FlyReactome</i>
<i>ESTHER database</i>	<i>Exposome-Explorer</i>	<i>FoodB</i>
<i>Estonian Biocentre Public Data</i>	<i>FactSage</i>	<i>ForestPlots.net</i>
<i>ETH Data Archive</i>	<i>figshare</i>	<i>Fragile Families and Child Wellbeing Study</i>
<i>ETH Travel Data Archive</i>	<i>Finnish Social Science Data Archive</i>	<i>Freshwater Biodiversity Data Portal</i>
<i>euHCVdb</i>	<i>Fishbase</i>	<i>FungiDB</i>
<i>EuPathDB</i>	<i>Flora von Bayern</i>	<i>Gateway to Global Aging Data</i>
<i>European Research Community on Flow, Turbulence and Combustion Database - Classic Collection</i>	<i>GFZ Data Services</i>	<i>Griffith University Research Data Repository</i>
<i>Gazel</i>	<i>GGOS Atmosphere</i>	<i>GWAS Central</i>
<i>Gene Expression Nervous System Atlas</i>	<i>GigaDB</i>	<i>HAGR</i>
<i>Gene Expression Omnibus</i>	<i>GIN</i>	<i>Handbook of Basic Atomic Spectroscopic Data</i>
<i>Genome-Scale Metabolic Network DataBase</i>	<i>GISAID</i>	<i>Hardin.MD</i>
<i>GenomeRNAi</i>	<i>GISTEMP</i>	<i>Hardwood Genomics Project</i>
<i>Genomic Data Commons</i>	<i>Global carbon atlas</i>	<i>Harvard Medical School, Library of Integrated Network-Based Cellular Signatures Database</i>
<i>GeoReM</i>	<i>Global Change Research Data Publishing and Repository</i>	<i>Hazardous Substance Data Bank</i>
<i>GEOROC</i>	<i>Global Forest Change 2000–2013</i>	<i>Health and Medical Care Archive</i>
<i>Geospatial Data</i>	<i>Global Pesticides Release</i>	<i>Health and Retirement Study</i>

	<i>Data</i>	
<i>Geothermal Data Repository</i>	<i>Global Precipitation Climatology Centre</i>	<i>heiDATA</i>
<i>GERDA</i>	<i>Global Trade, Assistance, and Production</i>	<i>HEPData</i>
<i>German Federation for Biological Data</i>	<i>GLOBE</i>	<i>HIstome</i>
<i>GerManC project</i>	<i>GlyTouCan</i>	<i>HJ Andrews Experimental Forest</i>
<i>Host - Pathogen Interaction Database</i>	<i>ILL Data Portal</i>	<i>Integrated Resource for Reproducibility in Macromolecular Crystallography</i>
<i>Household, Income and Labour Dynamics in Australia Survey</i>	<i>Illinois Data Bank</i>	<i>Interface to Los Alamos Atomic Physics Codes</i>
<i>HSRC Research Data Service</i>	<i>Illinois State Water Survey</i>	<i>International Argo Project</i>
<i>HUGO Gene Nomenclature Committee</i>	<i>Image Data Resource</i>	<i>International Food Policy Research Institute Dataverse</i>
<i>Human Metabolome Database</i>	<i>IMGT/HLA Database</i>	<i>International Geodynamics and Earth Tide Service</i>
<i>Human Protein Reference Database</i>	<i>Imperial College High Performance Computing Service Data Repository</i>	<i>International HapMap Project</i>
<i>Human Proteinpedia</i>	<i>IMS Universität Stuttgart Repository</i>	<i>International Mouse Phenotyping Consortium</i>
<i>HumanCyc</i>	<i>InCHIANTI</i>	<i>International Network of Nuclear Reaction Data Centres</i>
<i>Huygens ING</i>	<i>IndExs</i>	<i>International Neuroimaging Data-sharing Initiative</i>
<i>ICRISAT Dataverse Network</i>	<i>Indian Genetic Disease Database</i>	<i>International Toxicity Estimates for Risk</i>
<i>IDEADB</i>	<i>Infefers</i>	<i>Interologous Interaction Database</i>
<i>IDEALS</i>	<i>Influenza Research Database</i>	<i>InterPro</i>
<i>IDEAS</i>	<i>Inspire-HEP</i>	<i>IonomicHub</i>

<i>IDS Repository</i>	<i>INTEGRALL</i>	<i>IPK Gatersleben</i>
<i>IRI/LDEO Climate Data Library</i>	<i>Ligand-Gated Ion Channel Database</i>	<i>Libra Data</i>
<i>ISPS Data Archive</i>	<i>LINCS Data Portal</i>	<i>Media Repository Humboldt-Universität zu Berlin</i>
<i>ITIS</i>	<i>Linguistic Data Consortium</i>	<i>megx</i>
<i>JGI MycoCosm</i>	<i>London Datastore</i>	<i>melbourne.figshare.com</i>
<i>Johns Hopkins Data Archive Dataverse Network</i>	<i>Loughborough Data Repository</i>	<i>MEMENTO</i>
<i>Johns Hopkins Turbulence Databases</i>	<i>LSHTM Data Compass</i>	<i>Mendeley Data</i>
<i>Journal of Cell Biology</i>	<i>MADATA - Mannheim research data repository</i>	<i>mentha</i>
<i>jPOSTrepo</i>	<i>MaizeGDB</i>	<i>META-SHARE</i>
<i>Kiel Ocean Science Information System</i>	<i>Malaria Atlas Project</i>	<i>MetaboLights</i>
<i>Kinetic models of biological systems</i>	<i>Manitoba Centre for Health Policy Population Health Research Data Repository</i>	<i>MetaCyc</i>
<i>Kinsources.net</i>	<i>MARGINS Data Portal</i>	<i>Meteoritical bulletin database</i>
<i>KNMI Data Centre</i>	<i>Marine Microbial Database of India</i>	<i>MG-RAST</i>
<i>KU ScholarWorks</i>	<i>Materials Data Facility</i>	<i>MiCroKitS</i>
<i>Lancaster University Research Directory</i>	<i>MatrixDB</i>	<i>MIDAS</i>
<i>Spectroradiomete</i>		
<i>Monarch</i>	<i>National Archive of Computerized Data on Aging</i>	<i>NCBI dbMHC</i>
<i>Monash University Data Repository</i>	<i>National Archive of Criminal Justice Data</i>	<i>NCBI dbVar</i>
<i>Monash University Research Repository</i>	<i>National Archive of Criminal Justice Data, Terrorism and Preparedness Data</i>	<i>NCBI Epigenomics</i>

	<i>Resource Center</i>	
<i>Morph·D·Base</i>	<i>National Biodiversity Data Centre - Biodiversity Maps</i>	<i>NCBI Gene</i>
<i>Mouse Genome Informatics</i>	<i>National Cancer Data Base</i>	<i>NCBI Genome</i>
<i>Movebank</i>	<i>National Data Archive on Child Abuse and Neglect</i>	<i>NCBI GSS</i>
<i>Multimodal Learning Corpus Exchange</i>	<i>National Geothermal Data System</i>	<i>NCBI HomoloGene</i>
<i>My Geo Hub</i>	<i>National Snow and Ice Data Center</i>	<i>NCBI Influenza Virus Resource</i>
<i>Mycobank</i>	<i>Natural History Museum, Data Portal</i>	<i>NCBI Probe</i>
<i>myExperiment</i>	<i>NAVDAT</i>	<i>NCBI Protein Clusters</i>
<i>NAGRP Blast Center</i>	<i>NCAA Student-Athlete Experiences Data Archive</i>	<i>NCBI Third Party Annotation</i>
<i>nanoHUB</i>	<i>NCAR Research Data archive</i>	<i>NCBI GSS</i>
<i>NASA Earth Exchange</i>	<i>NCBI BioSystems Database</i>	<i>NCBI HomoloGene</i>
<i>NASA Prognostics Data Repository</i>	<i>NIST Atomic Spectra Database</i>	<i>NCBI Influenza Virus Resource</i>
<i>NASA Socioeconomic Data and Applications Center</i>	<i>NIST Physical reference data</i>	<i>Open Data by Socrata</i>
<i>National Addiction &amp; HIV Data Archive Program</i>	<i>NIST Standard Reference Data</i>	<i>Open Exoplanet Catalogue</i>
<i>NCBI dbGaP</i>	<i>NITRC</i>	<i>Open Research Data Online</i>
<i>National Archive of Computerized Data on Aging</i>	<i>NoMaD Repository</i>	<i>Open Tree of Life</i>
<i>NCBI dbGaP</i>	<i>Nord-Trondelag Health Study</i>	<i>OpenKIM</i>
<i>NCBI Trace Archive</i>	<i>Nordicana D</i>	<i>OpenML</i>
<i>NCBI Virus Variation</i>	<i>NorStore Research Data Archive</i>	<i>ORA-Data</i>
<i>NCCOR</i>	<i>North American Breeding Bird Survey</i>	<i>Open Data by Socrata</i>

<i>Neotoma Paleoecology Database</i>	<i>Norwegian Polar Data Centre</i>	<i>Oral Cancer Gene Database</i>
<i>Network for the Detection of Atmospheric Composition Change</i>	<i>Nuclear Receptor Signaling Atlas</i>	<i>ORDA - The University of Sheffield Research Data Catalogue and Repository</i>
<i>Network Repository</i>	<i>Oak Ridge National Laboratory Distributed Active Archive Center for Biogeochemical Dynamics</i>	<i>Ortolang</i>
<i>NeuroElectro</i>	<i>Ocean Biogeographic Information System</i>	<i>OsteoArthritis Initiative</i>
<i>Neuroscience Information Framework</i>	<i>Ocean Data and Information System</i>	<i>Oxford University Research Archive</i>
<i>NGEE Arctic</i>	<i>Online Mendelian Inheritance in Animals</i>	<i>OsteoArthritis Initiative</i>
<i>Ningaloo Atlas</i>	<i>NIST Atomic Spectra Database</i>	<i>Population Services International Dataverse</i>
<i>Ozone Mapping and Profiler Suite</i>	<i>Phytozome</i>	<i>Portal to Los Alamos Opacity Codes</i>
<i>Pacific Islands Families Study</i>	<i>Plant and Fungal Carbohydrate Structure Database</i>	<i>PROFILES Registry</i>
<i>PAIN repository</i>	<i>Plant Genomics and Phenomics Research Data Repository</i>	<i>Project Achilles</i>
<i>Panel on household finances</i>	<i>Plant Organelles Database Version 3</i>	<i>Project Tycho®</i>
<i>Parkinson Disease Mutation Database</i>	<i>PLANTS Database</i>	<i>pSILAC</i>
<i>PATRIC</i>	<i>Plants of TAIWAN</i>	<i>PSLC DataShop</i>
<i>PDBj</i>	<i>PlasmoDB</i>	<i>PsychData</i>
<i>PeanutBase</i>	<i>PLEXdb</i>	<i>Population Services International Dataverse</i>
<i>PetDB</i>	<i>Polar Rock Repository</i>	<i>Portal to Los Alamos Opacity Codes</i>
<i>PhenoGen Informatics</i>	<i>Polinsky Language Sciences Lab Dataverse</i>	<i>PubChem</i>

<i>Phenylalanine Hydroxylase Locus Knowledgebase</i>	<i>Pombase</i>	<i>Purdue University Research Repository</i>
<i>PHI-base</i>	<i>Phytozome</i>	<i>QTL Archive</i>
<i>PhysioBank</i>	<i>Plant and Fungal Carbohydrate Structure Database</i>	<i>PubChem</i>
<i>PhysioNet</i>	<i>Plant Genomics and Phenomics Research Data Repository</i>	<i>Romani Morpho-Syntax Database</i>
<i>PhytoPath</i>	<i>Plant Organelles Database Version 3</i>	<i>RRUFF Project</i>
<i>Qualiservice</i>	<i>Research Data Exchange Platform - Johanna Mestorf Academy</i>	<i>RunMyCode</i>
<i>QUT Research Data Finder</i>	<i>Research Data Online</i>	<i>Rutgers/New Jersey Geological and Water Survey Core Repository</i>
<i>RADAR</i>	<i>ResearchGate</i>	<i>SABIO-RK</i>
<i>RADAR Luke</i>	<i>RESID Database of Protein Modifications</i>	<i>Scholar's Bank</i>
<i>RAM Legacy Stock Assessment Database</i>	<i>Rhea</i>	<i>ScholarsArchive@OSU</i>
<i>Reality Commons</i>	<i>Ridge 2000 Data Portal</i>	<i>Science3D</i>
<i>REasons for Geographic and Racial Differences in Stroke Study</i>	<i>Research Data Exchange Platform - Johanna Mestorf Academy</i>	<i>ScienceBase USGS</i>
<i>Reciprocal Net</i>	<i>Research Data Online</i>	<i>Scientific Data Repository</i>
<i>Recode</i>	<i>ResearchGate</i>	<i>Screening Unit Berlin-Buch</i>
<i>REFOLDdb</i>	<i>RESID Database of Protein Modifications</i>	<i>SEANOE</i>
<i>RefractiveIndex.INFO</i>	<i>Rhea</i>	<i>SedDB</i>
<i>RELMIN</i>	<i>Ridge 2000 Data Portal</i>	<i>SeedMe</i>
<i>ROAR Isolate Database</i>	<i>Selected constants energy levels and atomic spectra of actinides</i>	<i>Systematic Review Data Repository</i>
<i>SFB 882 Forschungsdatenzentrum</i>	<i>St. Edward's University</i>	<i>Taiwan Biodiversity Information</i>



	<i>institutional repository</i>	<i>Facility</i>
<i>Sheffield Hallam University Research Data Archive</i>	<i>Stanford Network Analysis Project</i>	<i>TalkBank</i>
<i>SilkDB</i>	<i>STARK-B</i>	<i>TAMBORA</i>
<i>SILVA</i>	<i>Stemformatics</i>	<i>TB Database</i>
<i>SIOR</i>	<i>Stockholm University repository for data</i>	<i>The Atomic Line List</i>
<i>SMOKA Science Archive</i>	<i>Store.Synchrotron Data Store</i>	<i>Systematic Review Data Repository</i>
<i>Socio Cognitive Processes Lab data</i>	<i>STOREDB</i>	<i>Taiwan Biodiversity Information Facility</i>
<i>SOFIA</i>	<i>Structural Biology Data Grid</i>	<i>TalkBank</i>
<i>SOL Genomics Network</i>	<i>Studie zur Gesundheit von Kindern und Jugendlichen in Deutschland</i>	<i>TAMBORA</i>
<i>Solar and Heliospheric Observatory</i>	<i>studyforrest.oth</i>	<i>TB Database</i>
<i>Solar Dynamics Observatory</i>	<i>Substance Abuse and Mental Health Data Archive</i>	<i>The Atomic Line List</i>
<i>SoyBase</i>	<i>Surrey Research Insight</i>	<i>Systematic Review Data Repository</i>
<i>SPECTR-W3</i>	<i>Swedish National Data Service</i>	<i>Synapse</i>
<i>Speech and Language Data Repository</i>	<i>Swiss HIV Cohort Study &amp; Swiss Mother and Child HIV Cohort Study</i>	<i>TRAILS</i>
<i>SPT Galaxy Cluster Followup Dataverse</i>	<i>The Lafora Progressive Myoclonus Epilepsy Mutation and Polymorphism Database</i>	<i>TreeBASE</i>
<i>The Cambridge Structural Database</i>	<i>The Population Research in Sexual Minority Health Data Archive</i>	<i>TreeGenes</i>
<i>The Cancer Imaging Archive</i>	<i>The SAO/NASA Astrophysics Data System Dataverse</i>	<i>TriTrypDB</i>

<i>The Cell Image Library</i>	<i>The World Atlas of Language Structures</i>	<i>TROLLing</i>
<i>The Chromosome 7 Annotation Project</i>	<i>THEREDA</i>	<i>TropFlux</i>
<i>The Comprehensive Resource of Mammalian protein complexes</i>	<i>ThermoML</i>	<i>Tropical Cyclone Information System</i>
<i>The Electron Microscopy Data Bank at PDBe</i>	<i>Tierstimmenarchiv - Museum für Naturkunde Berlin</i>	<i>Tropicos®</i>
<i>The European Genome-phenome Archive</i>	<i>Top Down Proteomics Repository</i>	<i>TwinsUK</i>
<i>The Global Agricultural Trial Repository and Database</i>	<i>Toxicology Data Network</i>	<i>TRAILS</i>
<i>The Health Improvement Network</i>	<i>TOXMAP</i>	<i>U.S. Geological Survey</i>
<i>The Human Protein Atlas</i>	<i>ToxoDB</i>	<i>U.S. National Archives and Records Administration Dataverse</i>
<i>The Infrared Space Observatory data archive</i>	<i>The Lafora Progressive Myoclonus Epilepsy Mutation and Polymorphism Database</i>	<i>Virtual Liver Network</i>
<i>The International Research Institute for Climate and Society</i>	<i>The Population Research in Sexual Minority Health Data Archive</i>	<i>Virus Pathogen Resource</i>
<i>UCD Digital Library</i>	<i>University of Reading Research Data Archive</i>	<i>VRP-REP</i>
<i>UCI Machine Learning Repository</i>	<i>University of Salford Data Repository</i>	<i>WASCAL Data Discovery Portal</i>
<i>UCSD Signaling gateway</i>	<i>University of Southampton Institutional Research Repository</i>	<i>Washington State University Data Center Dataverse</i>
<i>UMIN Individual Case Data Repository</i>	<i>University of Southern Queensland research data collection</i>	<i>West of Scotland Twenty-07 Study</i>
<i>UNESCO Institute for Statistics, Data Centre</i>	<i>UPF Digital Repository - Recursos i dades primàries</i>	<i>Work Histories Italian Panel</i>
<i>UniParc</i>	<i>UQ eSpace, Research Data Collections</i>	<i>Virtual Liver Network</i>

<i>UniProtKB</i>	<i>UWE Research Data Repository</i>	<i>Virus Pathogen Resource</i>
<i>UniRef</i>	<i>VecNet</i>	<i>VRP-REP</i>
<i>UniSA Research Data Access Portal</i>	<i>VectorBase</i>	<i>WASCAL Data Discovery Portal</i>
<i>Universal PBM Resource for Oligonucleotide Binding Evaluation</i>	<i>VertNet</i>	<i>World Agroforestry Centre - ICRAF Dataverse</i>
<i>University of Auckland Data Repository</i>	<i>Vienna Atomic Line Database</i>	<i>World Data Service for Oceanography</i>
<i>University of Liverpool Research Data Catalogue</i>	<i>VIPERdb</i>	<i>World Register of Marine Species</i>
<i>University of Minnesota Biocatalysis/Biodegradation Database</i>	<i>Viral Bioinformatics Resource Center</i>	<i>Zentrale Biomaterialbank der Charité</i>
<i>Worldwide Protein Data Bank</i>	<i>Yeast Resource Center</i>	<i>Zika Open-Research Portal</i>
<i>Xenbase</i>	<i>ZACAT</i>	<i>Zentrale Biomaterialbank der Charité</i>
<i>XSEDE</i>	<i>ZBW Journal Data Archive</i>	

**APÊNDICE G – Listagem de repositórios de dados de pesquisa e áreas abrangentes**

<b>SUBJECTS</b>	<b>Qtd.</b>	<b>SUBJECTS</b>	<b>Qtd.</b>
<i>Humanities and Social Sciences</i>	557	<i>European and American Literature</i>	2
<i>Humanities</i>	181	<i>General and Comparative Literature and Cultural Studies</i>	2
<i>Ancient Cultures</i>	34	<i>Non-European Languages and Cultures, Social and Cultural Anthropology, Jewish Studies and Religious Studies</i>	25
<i>Prehistory</i>	4	<i>Social and Cultural Anthropology and Ethnology/Folklore</i>	6
<i>Ancient History</i>	1	<i>Asian Studies</i>	2
<i>Classical Archaeology</i>	8	<i>African, American and Oceania Studies</i>	3
<i>Egyptology and Ancient Near Eastern Studies</i>	2	<i>Islamic Studies, Arabian Studies, Semitic Studies</i>	2
<i>History</i>	54	<i>Religious Studies and Jewish Studies</i>	6
<i>Medieval History</i>	1	<i>Theology</i>	9
<i>Early Modern History</i>	2	<i>Protestant Theology</i>	2
<i>Modern and Current History</i>	6	<i>Roman Catholic Theology</i>	1
<i>History of Science</i>	7	<i>Philosophy</i>	4
<i>Fine Arts, Music, Theatre and Media Studies</i>	40	<i>History of Philosophy</i>	1
<i>Art History</i>	11	<i>Social and Behavioural Sciences</i>	307
<i>Musicology</i>	5	<i>Education Sciences</i>	36
<i>Theatre and Media Studies</i>	3	<i>General Education and History of Education</i>	3
<i>Linguistics</i>	75	<i>Research on Teaching, Learning and Training</i>	6
<i>General and Applied Linguistics</i>	2	<i>Research on Socialization and Educational Institutions and Professions</i>	3
<i>Individual Linguistics</i>	2	<i>European and American Literature</i>	2
<i>Typology, Non-European Languages, Historical Linguistics</i>	6	<i>General and Comparative Literature and Cultural Studies</i>	2
<i>Literary Studies</i>	11	<i>Non-European Languages and Cultures, Social and Cultural Anthropology, Jewish Studies and Religious Studies</i>	25
<i>Social and Cultural</i>	6	<i>Plant Systematics and</i>	7

<i>Anthropology and Ethnology/Folklore</i>		<i>Evolution</i>	
<i>Psychology</i>	26	<i>Plant Ecology and Ecosystem Analysis</i>	40
<i>Developmental and Educational Psychology</i>	1	<i>Inter-organismic Interactions of Plants</i>	1
<i>Differential Psychology, Clinical Psychology, Medical Psychology, Methodology</i>	2	<i>Plant Physiology</i>	2
<i>Social Sciences</i>	198	<i>Plant Biochemistry and Biophysics</i>	8
<i>Sociological Theory</i>	2	<i>Plant Cell and Developmental Biology</i>	5
<i>Empirical Social Research</i>	78	<i>Plant Genetics</i>	56
<i>Communication Science</i>	13	<i>Zoology</i>	193
<i>Political Science</i>	36	<i>Systematics and Morphology</i>	9
<i>Economics</i>	143	<i>Evolution, Anthropology</i>	22
<i>Economic Theory</i>	1	<i>Animal Ecology, Biodiversity and Ecosystem Research</i>	55
<i>Economic and Social Policy</i>	40	<i>Sensory and Behavioural Biology</i>	5
<i>Public Finance</i>	10	<i>Biochemistry and Animal Physiology</i>	8
<i>Business Administration</i>	4	<i>Animal Genetics, Cell and Developmental Biology</i>	78
<i>Statistics and Econometrics</i>	52	<i>Medicine</i>	473
<i>Economic and Social History</i>	6	<i>Microbiology, Virology and Immunology</i>	171
<i>Jurisprudence</i>	32	<i>Metabolism, Biochemistry and Genetics of Microorganisms</i>	33
<i>Legal and Political Philosophy, Legal History, Legal Theory</i>	4	<i>Microbial Ecology and Applied Microbiology</i>	8
<i>Private Law</i>	1	<i>Medical Microbiology, Molecular Infection Biology</i>	8
<i>Public Law</i>	3	<i>Virology</i>	10
<i>Criminal Law and Law of Criminal Procedure</i>	1	<i>Immunology</i>	9
<i>Criminology</i>	6	<i>Medicine</i>	309
<i>Life Sciences</i>	985	<i>Epidemiology, Medical Biometry, Medical Informatics</i>	40
<i>Biology</i>	675	<i>Public Health, Health Services Research, Social Medicine</i>	101
<i>Basic Biological and Medical Research</i>	398	<i>Human Genetics</i>	101
<i>Biochemistry</i>	49	<i>Physiology</i>	12
<i>Biophysics</i>	9	<i>Nutritional Sciences</i>	2

<i>Cell Biology</i>	62	<i>Pathology and Forensic Medicine</i>	6
<i>Structural Biology</i>	20	<i>Plant Systematics and Evolution</i>	7
<i>General Genetics</i>	172	<i>Plant Ecology and Ecosystem Analysis</i>	40
<i>Developmental Biology</i>	3	<i>Inter-organismic Interactions of Plants</i>	1
<i>Bioinformatics and Theoretical Biology</i>	126	<i>Plant Physiology</i>	2
<i>Anatomy</i>	11	<i>Plant Biochemistry and Biophysics</i>	8
<i>Plant Sciences</i>	156	<i>Plant Cell and Developmental Biology</i>	5
<i>Clinical Chemistry and Pathobiochemistry</i>	3	<i>Cellular Neuroscience</i>	1
<i>Pharmacy</i>	10	<i>Developmental Neurobiology</i>	1
<i>Pharmacology</i>	16	<i>Systemic Neuroscience, Computational Neuroscience, Behaviour</i>	6
<i>Toxicology and Occupational Medicine</i>	10	<i>Cognitive Neuroscience and Neuroimaging</i>	12
<i>Cardiology, Angiology</i>	1	<i>Biological Psychiatry</i>	1
<i>Pneumology, Clinical Infectiology Intensive Care Medicine</i>	3	<i>Clinical Neurosciences III - Ophthalmology</i>	2
<i>Hematology, Oncology, Transfusion Medicine</i>	12	<i>Agriculture, Forestry, Horticulture and Veterinary Medicine</i>	129
<i>Gastroenterology, Metabolism</i>	7	<i>Agriculture, Forestry, Horticulture and Veterinary Medicine</i>	91
<i>Endocrinology, Diabetology</i>	2	<i>Soil Sciences</i>	21
<i>Rheumatology, Clinical Immunology, Allergology</i>	2	<i>Plant Cultivation</i>	3
<i>Dermatology</i>	1	<i>Plant Nutrition</i>	1
<i>Pediatric and Adolescent Medicine</i>	7	<i>Ecology of Agricultural Landscapes</i>	25
<i>Gynaecology and Obstetrics</i>	4	<i>Plant Breeding</i>	4
<i>Urology</i>	1	<i>Agricultural and Food Process Engineering</i>	5
<i>Gerontology and Geriatric Medicine</i>	3	<i>Agricultural Economics and Sociology</i>	12
<i>Radiology and Nuclear Medicine</i>	6	<i>Inventory Control and Use of Forest Resources</i>	13
<i>Radiation Oncology and Radiobiology</i>	3	<i>Basic Forest Research</i>	20

<i>Biomedical Technology and Medical Physics</i>	15	<i>Animal Husbandry, Breeding and Hygiene</i>	3
<i>Neurosciences</i>	58	<i>Cellular Neuroscience</i>	1
<i>Molecular Neuroscience and Neurogenetics</i>	6	<i>Developmental Neurobiology</i>	1
<i>Basic Veterinary Medical Science</i>	3	<i>Preparatory and Physical Chemistry of Polymers</i>	2
<i>Basic Research on Pathogenesis, Diagnostics and Therapy and Clinical Veterinary Medicine</i>	3	<i>Experimental and Theoretical Physics of Polymers</i>	1
<i>Natural Sciences</i>	958	<i>Polymer Materials</i>	1
<i>Chemistry</i>	175	<i>Physics</i>	250
<i>Molecular Chemistry</i>	39	<i>Condensed Matter Physics</i>	19
<i>Inorganic Molecular Chemistry</i>	5	<i>Experimental Condensed Matter Physics</i>	1
<i>Organic Molecular Chemistry</i>	8	<i>Optics, Quantum Optics and Physics of Atoms, Molecules and Plasmas</i>	66
<i>Chemical Solid State and Surface Research</i>	11	<i>Optics, Quantum Optics, Atoms, Molecules, Plasmas</i>	18
<i>Solid State and Surface Chemistry, Material Synthesis</i>	3	<i>Particles, Nuclei and Fields</i>	45
<i>Physical Chemistry of Solids and Surfaces, Material Characterisation</i>	3	<i>Particles, Nuclei and Fields</i>	8
<i>Theory and Modelling</i>	2	<i>Statistical Physics, Soft Matter, Biological Physics, Nonlinear Dynamics</i>	3
<i>Physical and Theoretical Chemistry</i>	28	<i>Astrophysics and Astronomy</i>	147
<i>Physical Chemistry of Molecules, Interfaces and Liquids - Spectroscopy, Kinetics</i>	19	<i>Astrophysics and Astronomy</i>	22
<i>General Theoretical Chemistry</i>	2	<i>Mathematics</i>	22
<i>Analytical Chemistry, Method Development (Chemistry)</i>	17	<i>Mathematics</i>	5
<i>Analytical Chemistry, Method Development (Chemistry)</i>	3	<i>Geosciences (including Geography)</i>	604
<i>Biological Chemistry and Food Chemistry</i>	18	<i>Food Chemistry</i>	
<i>Biological and Biomimetic Chemistry</i>	5	<i>Polymer Research</i>	
<i>Geochemistry, Mineralogy</i>	70	<i>Biomaterials</i>	1

<i>and Crystallography</i>			
<i>Geochemistry, Mineralogy and Crystallography</i>	16	<i>Computer Science, Electrical and System Engineering</i>	86
<i>Geography</i>	129	<i>Systems Engineering</i>	12
<i>Physical Geography</i>	14	<i>Automation, Control Systems, Robotics, Mechatronics</i>	1
<i>Human Geography</i>	17	<i>Traffic and Transport Systems, Logistics</i>	6
<i>Water Research</i>	123	<i>Human Factors, Ergonomics, Human-Machine Systems</i>	2
<i>Hydrogeology, Hydrology, Limnology, Urban Water Management, Water Chemistry, Integrated Water Resources Management</i>	34	<i>Electrical Engineering</i>	10
<i>Engineering Sciences</i>	256	<i>Electronic Semiconductors, Components, Circuits, Systems</i>	1
<i>Mechanical and industrial Engineering</i>	7	<i>Communication, High-Frequency and Network Technology, Theoretical Electrical Engineering</i>	3
<i>Mechanics and Constructive Mechanical Engineering</i>	4	<i>Electrical Energy Generation, Distribution, Application</i>	3
<i>Acoustics</i>	4	<i>Computer Science</i>	53
<i>Thermal Engineering/Process Engineering</i>	15	<i>Software Technology</i>	4
<i>Process Engineering, Technical Chemistry</i>	10	<i>Operating, Communication and Information Systems</i>	2
<i>Chemical and Thermal Process Engineering</i>	1	<i>Artificial Intelligence, Image and Language Processing</i>	14
<i>Technical Chemistry</i>	1	<i>Construction Engineering and Architecture</i>	28
<i>Biological Process Engineering</i>	2	<i>Construction Engineering and Architecture</i>	27
<i>Heat Energy Technology, Thermal Machines, Fluid Mechanics</i>	7	<i>Architecture, Building and Construction History, Sustainable Building Technology, Building Design</i>	12
<i>Energy Process Engineering</i>	1	<i>Urbanism, Spatial Planning, Transportation and Infrastructure Planning, Landscape Planning</i>	11
<i>Technical Thermodynamics</i>	1	<i>Structural Engineering, Building Informatics, Construction Operation</i>	1
<i>Materials Science and</i>	23	<i>Biomaterials</i>	1



<i>Engineering</i>			
<i>Materials Engineering</i>	5	<i>Computer Science, Electrical and System Engineering</i>	86
<i>Metallurgical and Thermal Processes, Thermomechanical Treatment of Materials</i>	1	<i>Systems Engineering</i>	12
<i>Composite Materials</i>	1	<i>Automation, Control Systems, Robotics, Mechatronics</i>	1
<i>Materials Science</i>	16	<i>Traffic and Transport Systems, Logistics</i>	6
<i>Thermodynamics and Kinetics of Materials</i>	2	<i>Human Factors, Ergonomics, Human-Machine Systems</i>	2
<i>Geotechnics, Hydraulic Engineering</i>	3		
<b>Total de repositórios de dados de pesquisa indexados no diretório</b>	<b>11.510</b>		

**APÊNDICE H – Formulário de identificação de recursos constatados nos repositórios com publicação ampliada**

	<b>Características</b>	<b>Opção</b>
<b>Título:</b>		
<b>Assunto (Área):</b>		
<b>Tipos de conteúdo (formas):</b>		
<b>País:</b>		
<b>Identificador digital</b> (AuthorClaim; ISNI; <i>ORCID</i> ; <i>ResearcherID</i> ; nenhum ou outro):		
<b>API</b> – Padrões de programação (FTP; NetCDF; OAI-PMH; <i>OpenDAP</i> ; <i>REST</i> ; <i>SOAP</i> ; <i>SPARQL</i> ; <i>SWORD</i> ; outros):		
<b>Restrições de acesso a dados</b> ( <i>feeRequired</i> – taxa requisitada; <i>institutional membership</i> – associação institucional; <i>registration</i> – cadastro; outros):		
<b>Carregamento de dados</b> (fechado, aberto ou restrito):		
<b>Tipo de instituição</b> (sem fins		

lucrativos ou comercial):		
<b>Padrões de metadados</b> (ABCD; AVM; CF; CIF; CIM; DCAT; DDI; DIF; <i>Darwin Core</i> ; <i>DataCite Metadata Schema</i> ; <i>Dublin Core</i> ; EML; FGDC/CSDGM; FITS; <i>Genome Metadata</i> ; ISA-Tab; ISO 19115; <i>International Virtual Observatory Alliance Technical Specification</i> ; MIBBI; OAI-ORE; PROV; <i>RDF Data Cube Vocabulary</i> ; <i>Repository-Developed Metadata Schemas</i> ; <i>SDMX</i> ; outros):		
<b>Sistemas de identificação de objetos digitais</b> (ARK; DOI ; PURL; URN; Hdl; nenhum; outros):		
<b>Idioma:</b>		
<b>Software utilizado</b> (CKAN; <i>DSpace</i> ; <i>DataVerse</i> ; <i>DigitalCommons</i> ; <i>EPrints</i> ; <i>Fedora</i> ; <i>MySQL</i> ; <i>Nesstar</i> ; <i>Opus</i> ; <i>dLibra</i> ; <i>eSciDoc</i> ; desconhecido; outros):		
<b>Tipo de repositório:</b> (Institucional; temático; outro):		