

Dissertação de Mestrado

**pRRoject: Um Ambiente de Suporte ao  
Desenvolvimento e Acompanhamento de  
Pesquisas Quantitativas Reprodutíveis**

Ana Maria Aquino dos Santos  
ana\_maria@lccv.ufal.br

Orientador:  
Dr. Alejandro C. Frery

Maceió, Novembro de 2017

Ana Maria Aquino dos Santos

**pRRoject: Um Ambiente de Suporte ao  
Desenvolvimento e Acompanhamento de  
Pesquisas Quantitativas Reprodutíveis**

Dissertação apresentada como requisito parcial para obtenção do grau de Mestre pelo Programa de Pós-Graduação em Modelagem Computacional de Conhecimento do Instituto de Computação da Universidade Federal de Alagoas.

Orientador: Dr. Alejandro C. Frery

**Catálogo na fonte**  
**Universidade Federal de Alagoas**  
**Biblioteca Central**  
**Divisão de Tratamento Técnico**

Bibliotecária Responsável: Helena Cristina Pimentel do Vale – CRB4 - 661

S237p

Santos, Ana Maria Aquino dos.

Project: um ambiente de suporte ao desenvolvimento e acompanhamento de pesquisas quantitativas reprodutíveis / Ana Maria Aquino dos Santos. – 2018.  
55 f. : il.

Orientador: Alejandro C. Frery.

Dissertação (mestrado em Modelagem Computacional de Conhecimento) –  
Universidade Federal de Alagoas. Instituto de Computação. Maceió, 2017.

Bibliografia: f. 50-55.

1. Reproducible research. 2. Ambiente computacional. 3. Projetos – Gerência.  
4. Pesquisa computacional quantitativa. I. Título.

CDU: 004.4:001.891



Ata da defesa de dissertação da aluna  
Ana Maria Aquino dos Santos

Realizou-se no dia 17 de novembro de dois mil e dezessete, a partir das 9h30min, na sala de aula do Laboratório de Computação Científica e Visualização da Universidade Federal de Alagoas, a defesa de dissertação de Mestrado em Modelagem Computacional de Conhecimento, intitulada "pRProject: Um Ambiente de Suporte ao Desenvolvimento e Acompanhamento de Pesquisas Quantitativas Reprodutíveis", apresentada por Ana Maria Aquino dos Santos, graduada em Matemática, como requisito parcial para a obtenção do grau de Mestre em Modelagem Computacional de Conhecimento, perante a seguinte comissão examinadora:

Professor e orientador Alejandro Cesar Frery  
Instituto de Computação – Ufal


Professor Heitor Soares Ramos Filho  
Instituto de Computação – Ufal

Professor Leonardo Melo de Medeiros  
Instituto Federal de Alagoas – Ifal

Professor Paolo Ettore Gamba  
Departamento de Engenharia Elétrica, Computacional e Biomédica – UNIPV


A dissertação foi considerada aprovada pela Comissão. Finalizados os trabalhos, às 12h00min, lavrou-se a presente ata, que vai assinada pelos membros da Comissão.

Maceió, 17 de novembro de 2017.

  
Alejandro Cesar Frery  
Doutor em Computação Aplicada – Instituto Nacional de Pesquisas Espaciais/Brasil

  
Heitor Soares Ramos Filho  
Doutor em Ciência da Computação – Universidade Federal de Minas Gerais/Brasil

  
Leonardo Melo de Medeiros  
Doutor em Ciência da Computação – Universidade Federal de Campina Grande/Brasil

  
Paolo Ettore Gamba  
Doutor em Engenharia Eletrônica – Universidade de Pavia/Itália

## RESUMO

Pesquisadores da área computacional enfrentam dificuldades para publicar resultados precisos dos seus artigos e isso tem sido um dos motivos cruciais para se pensar em *Reproducible Research* (RR). RR é uma sistematização de boas práticas em metodologia científica voltada especialmente para a ciência computacional quantitativa. Boas práticas de RR se relacionam com boas práticas de gerência de projetos colaborativos. Diante disso, apresentamos o pRRoject, um ambiente computacional comprometido em auxiliar o pesquisador a adotar boas práticas de RR do início ao fim do projeto de pesquisa, através do uso de ferramentas para gerenciamento de projetos.

**Palavras-chave:** *Reproducible Research*. Ambiente Computacional. Pesquisa Quantitativa. Gerência de Projetos.

## ABSTRACT

Computational researchers are confronted with difficulties to publish accurate and reproducible results, and this is been one of the crucial reasons for thinking about Reproducible Research (RR). RR is a systematization of good practices in scientific methodology focused especially to quantitative computational science. Good RR practices relate to good collaborative project management practices. Therefore, we present pRRoject, a computational environment committed to help researchers adopting good RR practices from the start until the end of the research project, through the use of tools for project management.

**Keywords:** Reproducible Research. Computing Environment. Quantitative Research. Management Project. Support

## **AGRADECIMENTOS**

Durante a produção desse trabalho científico, passei por um processo de aprendizagem que contribuiu significativamente para o meu crescimento pessoal e profissional. Nesse tempo, foi essencial o apoio de familiares, amigos, professores e, principalmente, do grande mestre Alejandro C. Frery. A todos esses, que gentilmente estiveram comigo e me deram um pouco de si, meus sinceros votos de agradecimento.

Além disso, agradeço aos membros do Laboratório de Computação Científica e Visualização (LCCV) da UFAL por disponibilizarem a infraestrutura técnica necessária para a hospedagem da ferramenta pRProject, produzida no decorrer desse trabalho. Em especial, agradeço ao Rodrigo de Lima Pinheiro, analista de sistema deste laboratório, por toda a paciência, atenção e sugestões disponibilizadas.

Também agradeço aos membros do Laboratório de Computação Científica e Análise Numérica (LaCCAN) da UFAL por promoverem seminários que permitiram apresentar o desenvolvimento deste trabalho e a participação de alguns dos membros no processo de validação da ferramenta.

Por fim, e não menos importante, agradeço à banca examinadora, pela leitura atenta, questionamentos e sugestões.

Ana Maria Aquino dos Santos

## LISTA DE FIGURAS

2.1	Ciclo de vida do projeto de pesquisa. . . . .	15
2.2	Uma visão geral dos componentes do conceito de compêndio. . . . .	17
2.3	Síntese das diretrizes de boas práticas de <i>Reproducible Research</i> . . . . .	18
3.1	Ambiente Reproducible Research pRroject. . . . .	31
4.1	Página com a lista de projetos cadastrados no pRroject. . . . .	33
4.2	Página para cadastro de um novo projeto, com destaque para o campo de seleção de módulos. . . . .	34
4.3	Página para adicionar artigos, com marcação nos campos de seleção e preenchimento . . . . .	35
4.4	Página para associar membros ao projeto com destaque nos campos seleção de usuários e papéis.) . . . . .	35
4.5	Página de usuário, com foto, nome, email, com destaque para as abas (atividade, arquivos e projetos) e para o espaço de informações pessoais e profissionais. . . . .	36
4.6	Projeto Ambiente <i>Reproducible Research</i> criado no pRroject . . . . .	36
4.7	Página de cadastro de um dos membros do projeto Ambiente <i>Reproducible Research</i> . . . . .	37
4.8	Página para associação de membros ao projeto Ambiente <i>Reproducible Research</i> , com destaque para os membros já associados ao projeto. . . . .	37
4.9	Página de configuração do repositório referente ao projeto Ambiente <i>Reproducible Research</i> . . . . .	38
4.10	Estrutura padrão de diretórios. . . . .	38
4.11	Página correspondente ao menu repositório, com destaque para a idade dos arquivos enviados e os comentários dos membros. . . . .	39
4.12	Diferenças do arquivo “fundamentacao.tex” antes e depois da revisão de membros. . . . .	40
4.13	A página <i>Global statistics</i> mostra o total de <i>commits</i> , o total de membros, a data do primeiro e do último <i>commit</i> realizado até o momento de captura dessa imagem, a quantidade de dias em que o projeto está em atividade até então, a média de <i>commits</i> por dia e a média de <i>commits</i> por membro. . . . .	40
4.14	A página <i>Commits statistics</i> mostra os gráficos relativos ao número de <i>commits</i> categorizados por mês, dia, hora e semana. . . . .	41
4.15	A página <i>Contributors statistics</i> mostra os gráficos dos <i>commits</i> referentes a cada membro do projeto e um gráfico detalhando a quantidade de <i>commits</i> e alterações realizadas por cada autor. . . . .	41
4.16	Página inicial da “Base de Conhecimento” com a lista dos artigos adicionados. . . . .	42
4.17	Página de um dos artigos adicionados à base de conhecimento na categoria “Lidos” e classificado, por um dos membros do projeto, com quatro estrelas. . . . .	43
4.18	Comentário emitido por um dos membros do projeto com relação ao conteúdo do artigo. . . . .	43
4.19	Subpágina (Arquivos) do artigo <i>Liberating field science samples and data</i> adicionado à categoria “Lidos” da base de conhecimento. . . . .	44
4.20	Subpágina (Histórico) do artigo <i>Liberating field science samples and data</i> adicionado à categoria “Lidos” da base de conhecimento. . . . .	44
4.21	Página pública do projeto Ambiente <i>Reproducible Research</i> com menu (1), informações gerais (2) e descrição dos membros com <i>links</i> para suas respectivas páginas (3). . . . .	45



4.22	Telas de execução da <i>box</i> pRProject no <i>Windows</i> . . . . .	46
4.23	Página do pRProject acessada localmente em máquina com sistema operacional <i>Windows</i> . . . . .	46
4.24	Página Wiki do projeto “Ambiente <i>Reproducible Research</i> ”. . . . .	47

## LISTA DE TABELAS

2.1	Quadro comparativo de algumas soluções . . . . .	23
2.2	Alinhamento da pesquisa com categorias de gerenciamento de projetos. . . . .	24

# SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO</b>	<b>9</b>
1.1	<b>Definição do Problema</b> . . . . .	9
1.2	<b>Contribuições</b> . . . . .	10
1.2.1	Contribuições Tecnológicas . . . . .	10
1.2.2	Contribuições Científicas . . . . .	11
1.3	<b>Objetivos</b> . . . . .	11
1.4	<b>Organização da Dissertação</b> . . . . .	11
<b>2</b>	<b>FUNDAMENTAÇÃO TEÓRICA</b>	<b>13</b>
2.1	<b>A lógica da pesquisa científica</b> . . . . .	13
2.2	<b>Reproducible Research</b> . . . . .	15
2.3	<b>Soluções computacionais para Reproducible Research</b> . . . . .	19
2.4	<b>Gerência de Projetos de Pesquisa</b> . . . . .	24
<b>3</b>	<b>AMBIENTE REPRODUCIBLE RESEARCH</b>	<b>26</b>
3.1	<b>Metodologia</b> . . . . .	26
3.2	<b>Ambiente computacional <i>open source</i> para suporte a Reproducible Research</b> . . . . .	28
3.3	<b>Do Redmine ao pRRoject</b> . . . . .	29
3.4	<b>Virtualização do pRRoject</b> . . . . .	32
3.5	<b>Validação do pRRoject</b> . . . . .	32
<b>4</b>	<b>RESULTADOS E ANÁLISES</b>	<b>33</b>
4.1	<b>O pRRoject</b> . . . . .	33
4.2	<b>Desenvolvimento e Acompanhamento da Pesquisa</b> . . . . .	36
4.3	<b>Ambiente virtual do pRRoject</b> . . . . .	45
4.4	<b>Impacto do pRRoject na pesquisa</b> . . . . .	47
4.5	<b>Análises e Discussão</b> . . . . .	48
<b>5</b>	<b>CONCLUSÕES</b>	<b>50</b>
5.1	<b>Considerações Finais</b> . . . . .	50
5.2	<b>Trabalhos futuros</b> . . . . .	50
	<b>Referências Bibliográficas</b>	<b>52</b>

# 1

## INTRODUÇÃO

### 1.1 Definição do Problema

As publicações científicas possuem ao menos dois propósitos: proclamar um resultado e provar aos leitores de que o resultado é correto. Esses objetivos são um desafio para quem faz pesquisa em computação científica. A necessidade da divulgação detalhada das pesquisas computacionais é um assunto discutido internacionalmente nos últimos anos e um dos tópicos enfatizados é o incentivo ao desenvolvimento de ferramentas apropriadas à *Reproducible Research* (Mesirov, 2010; Stodden et al., 2013b).

*Reproducible Research* (RR) consiste numa sistematização de boas práticas em metodologia da pesquisa científica, especialmente organizadas para a pesquisa computacional quantitativa. RR é tanto sobre os hábitos que garantam investigação reproduzível como as tecnologias que possam tornar esses processos eficientes e realistas (Fomel & Claerbout, 2009; Sandve et al., 2013).

Recentes mudanças culturais, em algumas áreas de conhecimento, tiveram um impacto positivo sobre a disponibilidade de dados e códigos. E para isso, novas ferramentas têm sido desenvolvidas: *Research Compendia*<sup>1</sup>, *The ReScience journal*<sup>2</sup>, *Open Science Framework*<sup>3</sup>. Algumas dessas ferramentas são associadas a revistas, que estão começando a exigir a divulgação de dados como condição para publicação. Os periódicos *Science* e *Nature*, por exemplo, adotaram novas diretrizes para aumentar a confiança nos estudos publicados e disponibilizaram editoriais em favor de políticas de financiamento e publicação acadêmica, promovendo a ciência aberta e a reprodutibilidade<sup>4</sup> (Leek & Peng, 2015; McNutt, 2014; Skaggs et al., 2015).

---

<sup>1</sup>Disponível em: <<http://146.20.110.164/>>. Acesso em: 18 out. 2017.

<sup>2</sup>Disponível em: <<http://rescience.github.io/>>. Acesso em: 20 set. 2017.

<sup>3</sup>Disponível em: <<https://osf.io/>>. Acesso em: 09 nov. 2017.

<sup>4</sup>Disponível em: <<http://www.nature.com/news/reproducibility-1.17552>>. Acesso em: 20 jun. 2017.

Isso mostra que há uma tendência na comunidade científica das Ciências Computacionais de que trabalhos de impacto devem ser apresentados na forma de *RR*. Todavia, a reprodutibilidade computacional não deve ser implantada apenas durante a publicação da pesquisa, mas desde o início do projeto de pesquisa. Algumas das vantagens disso são: tornar o trabalho mais transparente para outros pesquisadores; melhorar o trabalho em equipe; aumentar o impacto do artigo; e favorecer a continuidade do trabalho (Donoho, 2010).

De acordo com Sandve et al. (2013), o requisito mínimo para um pesquisador é reproduzir os resultados da sua própria pesquisa. A partir disso, deve ir para um patamar onde “rotineiramente” possa reutilizar o seu trabalho anterior, aumentar a sua produtividade e, num segundo momento, assegurar que outros pesquisadores tenham uma possibilidade prática de reproduzir os seus resultados.

Apesar dos ganhos decorrentes da prática de *RR*, ainda há resistência por parte de pesquisadores. Algumas das suas justificativas para não aderir a *RR* são: não há benefícios diretos ou incentivos; é necessário muito esforço para limpar códigos e dados; se perdem vantagens competitivas sobre outros pesquisadores; não há disponibilidade de tempo para responder a perguntas de usuários; há dificuldade em corresponder aos requisitos de propriedade intelectual (Skaggs et al., 2015).

Diante disso, há a necessidade de mostrar, detalhadamente, uma nova maneira de se trabalhar com *RR*, respeitando as diretrizes de boas práticas disponíveis na literatura e causando o menor impacto possível nas práticas de pesquisas consolidadas.

Como forma de auxiliar esse processo, recomenda-se a implementação e o uso de um ambiente computacional que facilite a aplicação das práticas de *RR* durante todo o desenvolvimento de um projeto de pesquisa computacional quantitativa.

Esse ambiente é um sistema de gerenciamento de projetos *open source* que auxilia o processo do ciclo de vida da pesquisa computacional quantitativa e oferece ao pesquisador suporte para executar boas práticas de *RR*.

## 1.2 Contribuições

Disponibilizamos um ambiente computacional *Open Source* especialmente customizado para auxiliar laboratórios, grupos e pesquisadores interessados em realizar pesquisa computacional quantitativa comprometida com as práticas de *Reproducible Research* do início ao fim do projeto de pesquisa.

### 1.2.1 Contribuições Tecnológicas

- Disponibilizamos o modelo de um ambiente computacional *Open Source* e customizável voltado para projetos de pesquisas quantitativas reprodutíveis;

- Oferecemos um ambiente virtual e reproduzível do modelo;
- Colocamos à disposição do usuário o manual de utilização do ambiente computacional<sup>5</sup> e um fórum de discussão<sup>6</sup>.

### 1.2.2 Contribuições Científicas

- Relacionamos projetos de pesquisa com gerência de projetos e boas práticas de *Reproducible Research*.
- Realizamos uma pesquisa participante para averiguar o impacto do uso do ambiente computacional no desenvolvimento de pesquisas quantitativas.
- Desenvolvemos o presente trabalho respeitando as práticas de *RR*, disponíveis na literatura.

## 1.3 Objetivos

Os principais objetivos relacionados ao trabalho são:

1. Estender um sistema *open source*, com suporte a *Reproducible Research*, para que o mesmo atenda de forma mais completa as necessidades dos laboratórios de estudos científicos, comprometidos com pesquisa computacional quantitativa.
2. Propor, implementar e avaliar novas funcionalidades para o ambiente computacional.
3. Disponibilizar o ambiente computacional para indivíduos, laboratórios e grupos interessados em pesquisa computacional quantitativa.
4. Obter uma avaliação qualitativa das vantagens e desvantagens do uso do ambiente computacional.

## 1.4 Organização da Dissertação

As próximas partes deste trabalho abordarão: no capítulo 2, os aspectos históricos e científicos sobre *Reproducible Research* e uma breve análise dos softwares, desenvolvidos nos últimos dois anos, que se propõem a publicar artigos científicos reproduzíveis. No capítulo 3 veremos detalhes do processo de customização de um ambiente computacional *open source*, voltado a auxiliar pesquisadores da ciência computacional quantitativa a adotarem boas

---

<sup>5</sup>Disponível em: <<https://rerisip.lccv.ufal.br/projects/ambiente-reproducible-research/wiki/Wiki>>. Acesso em: 15 nov. 2017.

<sup>6</sup>Disponível em: <<https://rerisip.lccv.ufal.br/projects/ambiente-reproducible-research/boards/1>>. Acesso em: 15 nov. 2017.

práticas de pesquisa reprodutível. No capítulo 4 apresentaremos as análises e os resultados do uso do ambiente computacional, durante o desenvolvimento deste trabalho. No capítulo 5 descreveremos a conclusão e sugestões para trabalhos futuros.

O capítulo apresentou a abordagem do problema foco nesta dissertação. Em seguida, foram apresentadas as contribuições e objetivos desta pesquisa. Dando sequência, o segundo capítulo apresentará a fundamentação base e aspectos cruciais ao desenvolvimento desta dissertação.

# 2

## FUNDAMENTAÇÃO TEÓRICA

**I**NICIALMENTE, este capítulo apresentará um breve histórico sobre a pesquisa científica, os desafios e progresso. Em seguida, como surgiu a necessidade da reprodutibilidade na pesquisa, destacando a pesquisa computacional quantitativa e seus desafios. Por fim, abordará as soluções computacionais desenvolvidas nos últimos dois anos para publicar artigos científicos reprodutíveis.

### 2.1 A lógica da pesquisa científica

A palavra ciência deriva do latim *scientia* e significa o ato de conhecer o universo por meio de base empírica, método e discurso lógico. Muitas contribuições filosóficas, ao longo da história da humanidade, auxiliaram na construção dessa definição.

Francis Bacon (1561 – 1626) ao analisar e criticar abordagens empiristas e racionalistas, cria a primeira forma sistemática de ciência empírica, alegando ser uma forma para aquisição de conhecimento útil e confiável. Esse método é um dos responsáveis por caracterizar a distinção entre ciência e filosofia (Volpato, 2013).

No entanto, Galileu Galilei (1564 – 1642) mostrou, por meio de experimentos, o que ninguém fez nos dois mil anos que antecederam 1582 no Ocidente, ano em que Galileu lançou dois objetos de massas diferentes e observou que ambos caíam ao mesmo tempo, caso não ocorresse alguma intervenção. Isso contrariou as ideias de Aristóteles com relação ao movimento dos objetos e revolucionou a forma de pensar e agir da sociedade europeia. Dessa forma, Galileu é considerado o pioneiro do método científico (Kuhn, 1998; Volpato, 2013).

Johannes Kepler (1571 – 1630), contemporâneo de Galileu e considerado um revolucionário da ciência moderna, descobriu que as órbitas do sistema planetário não são circulares como supôs Copérnico (1473 – 1543), e sim, elípticas. Tal descoberta deu origem a uma nova



astronomia que contou com contribuições do dinamarquês Tycho Brahe (1546 – 1601), a primeira explicação física para as órbitas celestes.

Centenas de anos depois, Karl Popper (1902 – 1994) considerou que o sucesso de Kepler decorreu do fato de sua hipótese inicial (círculo) ser relativamente fácil de falsear (Popper, 2004).

Todo enunciado científico deve ser passível de verificação e falsificação: “recuso-me a aceitar a concepção de que, em ciência, existam enunciados que devemos resignadamente aceitar como verdadeiros, simplesmente pela circunstância de não parecer possível, devido a razões lógicas, submetê-los a teste” (Popper, 2004, p.50).

Thomas Kuhn (1922 – 1996) não considerou que o critério de falseabilidade seja suficiente para definir o que é científico, ou não. Não há teoria que possa resolver todos os problemas, os “paradigmas<sup>1</sup>” determinam a evolução das ciências (Kuhn, 1998).

Um bom exemplo de “paradigma” na ciência contemporânea são os feitos de Alan Turing (1912 – 1954), durante a Segunda Guerra Mundial, que se tornaram um marco para a ciência computacional (Hodges & Hofstadter, 2014). Devemos a Turing o fato de utilizarmos computadores para avançarmos na pesquisa científica.

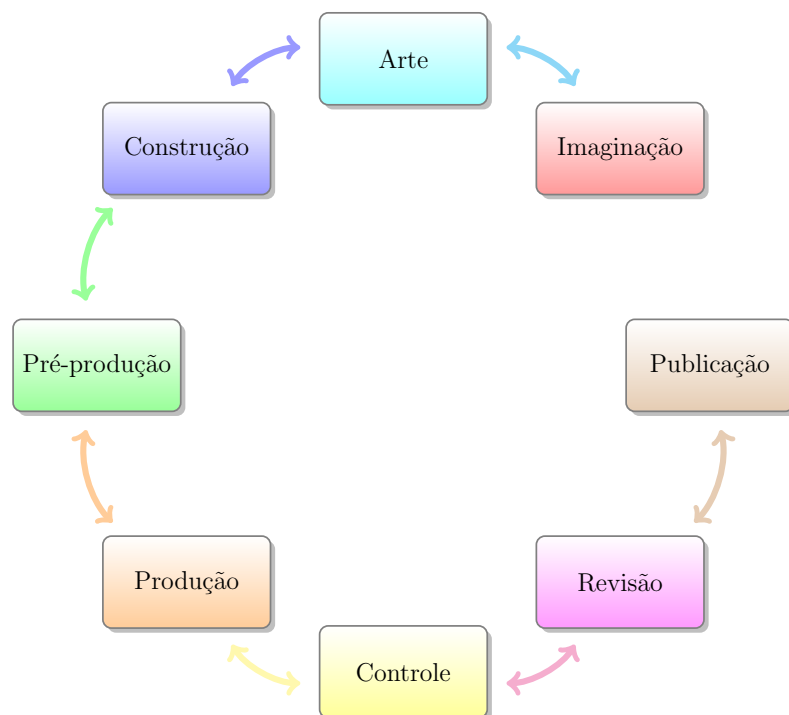
Apesar das divergências entre os pensamentos de Popper e Kuhn, suas contribuições são inerentes ao objetivo da ciência atual, pois o processo científico inicia com uma boa ideia promissora (Kuhn) que precisa ser testada (Popper) (Volpato, 2013).

Atualmente, no processo científico, uma ideia promissora dá origem a um projeto de pesquisa que segue um ciclo de vida com início, meio e fim (Figura 2.1), caracterizado, respectivamente, pela revisão bibliográfica, experimentos ou atividades desenvolvidas e publicação dos resultados em artigos de jornais ou revistas (Volpato, 2013; Wintage, 2015).

---

<sup>1</sup>“Considero “paradigmas” as realizações científicas universalmente reconhecidas que, durante algum tempo, fornecem problemas e soluções modelares para uma comunidade de praticantes de uma ciência” (Kuhn, 1998, p.13).

Figura 2.1: Ciclo de vida do projeto de pesquisa.



Fonte: Diagrama fundamentado em Wintage, 2015.

## 2.2 Reproducible Research

A reprodutibilidade é um critério fundamental para o método científico. De acordo com Popper (2004), “não tomamos muito seriamente nem mesmo nossas próprias observações e não as vemos como observações científicas, até as havermos repetido e submetido a prova.” O físico e matemático árabe, Ibn Al-Haitham, também defendeu a ideia de que uma hipótese deveria ser provada por experimentos e foi o primeiro a utilizar o critério de reprodutibilidade de forma sistemática, para tentar provar a propagação retilínea da luz através de uma câmera obscura, no século XI (Atmanspacher & Maasen, 2016).

Mais tarde, no século XVII, a reprodutibilidade é incluída no método científico e tem como um dos seus principais protagonistas, Robert Boyle, um dos fundadores da *Royal Society* e considerado o promotor da “filosofia experimental”. A *Royal Society*, uma das primeiras instituições de pesquisa, desde sua formação considerou a replicabilidade uma atividade central (Stodden, 2010; Steinle, 2016).

Em 1672, Isaac Newton envia para essa instituição um de seus trabalhos mais polêmicos, no qual descreveu uma “nova teoria” sobre a luz e as cores, contestado por muitos cientistas devido à dificuldade de replicabilidade e aos argumentos metodológicos insuficientes (Steinle, 2016; Fara, 2015). Verifica-se, dessa forma, indícios da necessidade de obrigatoriedade de transparência na pesquisa.

Nos séculos XVIII e XIX, respectivamente, o fisiologista Felice Fontana e o físico Hans Christian Ørsted relatam seus experimentos e, com isso, os tornam passíveis de serem reproduzidos. Mas é no século XX quando Jon Claerbout, geofísico e professor da universidade de Stanford, percebe a dificuldade dos pesquisadores para reproduzir os resultados expressos nas suas próprias publicações, que se iniciam os primeiros movimentos para a criação do termo *Reproducible Research* no âmbito da pesquisa computacional (Steinle, 2016; Fomel & Claerbout, 2009; Mesirov, 2010; Schwab et al., 2000).

*Reproducible Research*, segundo Fomel & Claerbout (2009), define-se como uma ideia de replicação na pesquisa computacional quantitativa. Essa expressão propõe uma nova cultura de publicação no meio acadêmico e desde sua criação tem atraído a atenção de alguns pesquisadores que passaram a adotar métodos reprodutíveis ou os introduziram nos seus artigos (Stodden et al., 2015; Mesirov, 2010).

David Donoho, matemático, estatístico e professor da universidade de Stanford, é um dos primeiros pesquisadores exortado por Claerbout à se envolver com “pesquisa verdadeiramente reprodutível”. Ele é o pioneiro em revelar livremente códigos e dados de pesquisas utilizando *wavelets*<sup>2</sup> (Buckheit & Donoho, 1995; Stodden, 2010) e um dos mentores de *RR*.

De acordo com Donoho (2010), há a necessidade dos pesquisadores abandonarem seus costumes arcaicos e se disporem a aprender novos hábitos para fazer pesquisa computacional quantitativa. O apego aos velhos costumes gera resistência à adoção de *RR* e isso é um dos empecilhos para o desenvolvimento científico (Peng, 2011).

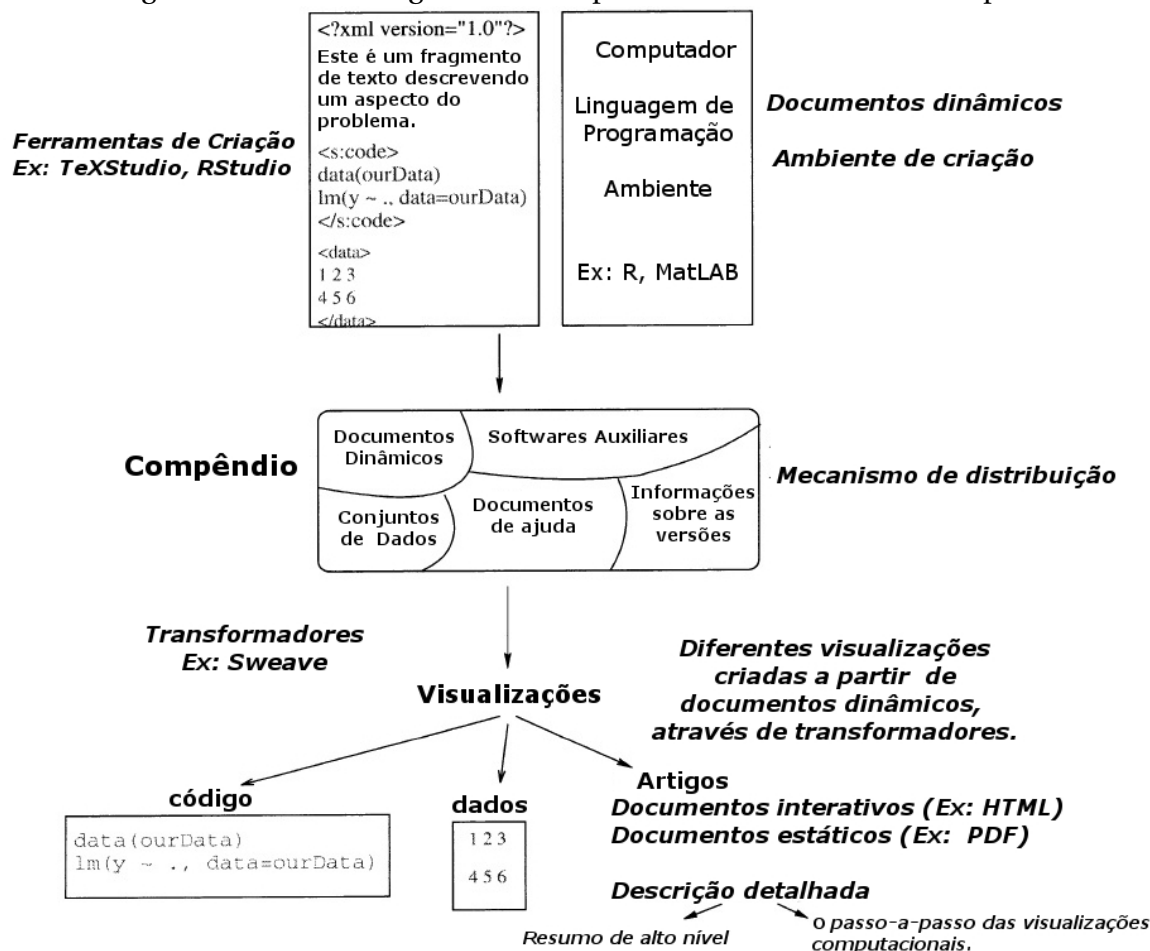
Buckheit & Donoho (1995) propuseram que ao publicar artigos que contenham figuras ou gráficos, gerados computacionalmente, também deve-se publicar os softwares necessários para gerá-los. Fundamentados nessa ideia, Gentleman & Temple Lang (2004) apresentam o conceito de “compêndio” (Figura 2.2), uma coleção de artefatos da pesquisa composta por arquivos, documentos dinâmicos, conjunto de dados e softwares. O objetivo do “compêndio” é fornecer aos autores e leitores de artigos a possibilidade de gerar seus resultados de forma independente (Stodden et al., 2015; Gentleman & Lang, 2007).

Dessa forma, o “compêndio” pode auxiliar a publicação de artigos conforme a proposta de Buckheit & Donoho (1995), pois funciona como um recipiente de armazenamento que contém um ou mais documentos dinâmicos (texto, código, dados), com informações detalhadas sobre os procedimentos necessários para reprodução dos resultados em diferentes contextos.

---

<sup>2</sup>Segundo Goswami & Chan (2011), *wavelet* pode ser considerada uma função gerada de uma função única, através de operações de dilatações e translações.

Figura 2.2: Uma visão geral dos componentes do conceito de compêndio.



Fonte: Imagem adaptada de Gentleman & Lang, 2007, p. 3.

A partir dessas contribuições e do avanço tecnológico, muitas ferramentas foram desenvolvidas e percebeu-se um grande potencial para o compartilhamento e replicação dos resultados nas pesquisas científicas computacionais. Diante disso, em novembro de 2009, um grupo formado por cientistas, advogados e editores de revistas se reuniram para discutir quais ações os cientistas computacionais deveriam executar para conseguir credibilidade em suas pesquisas. Essa reunião indicou o que mais tarde ficou conhecido como “boas práticas de *RR*” (Stodden et al., 2010; Stodden & Miguez, 2014).

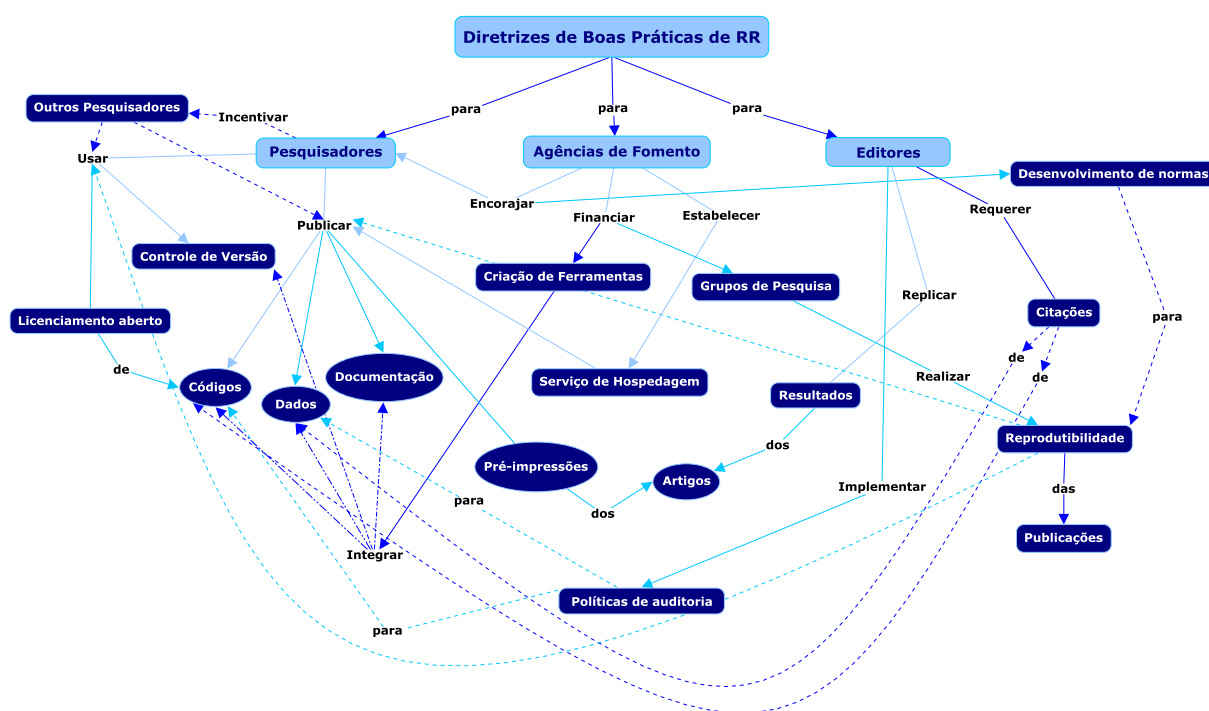
As boas práticas de *Reproducible Research* propõem atitudes proativas à publicação, regulamentadas através de diretrizes e tendo em vista, principalmente, a “universalidade” e a “transparência”, características inerentes à ciência (Buckheit & Donoho, 1995; Donoho, 2010; Sandve et al., 2013). Segundo Stodden et al. (2010), cientistas computacionais, instituições de fomento à pesquisa e editores de jornais ou revistas científicas, precisam obedecer requisitos, conforme sua categoria, para “reintroduzir” a reprodutibilidade na pesquisa científica (Figura 2.3).

A Figura 2.3 mostra, de forma resumida, as principais diretrizes de boas práticas de

pesquisa reprodutível para pesquisadores, agências de fomento e editores. Nessa ordem temos:

- Pesquisadores: usar controle de versão e licenciamento aberto de códigos; publicar códigos, dados, documentação, pré-impressões de artigos e incentivar outros pesquisadores a fazerem o mesmo.
- Agências de fomento: encorajar pesquisadores; financiar criação de ferramentas para integrar códigos, dados e documentação; financiar grupos de pesquisa para realizar reprodutibilidade das publicações de forma plena; estabelecer serviço de hospedagem para publicação; fomentar o desenvolvimento de normas de reprodutibilidade.
- Editores: implementar políticas de auditoria para dados e códigos; replicar resultados dos artigos e requerer citações de dados e códigos.

Figura 2.3: Síntese das diretrizes de boas práticas de *Reproducible Research*



Fonte: Mapa Conceitual fundamentado em Stodden et al., 2010.

Conforme os relatos de Stodden & Míguez (2014), nada referente à pesquisa deve ficar restrito, tudo deve ser revelado nos mínimos detalhes e segundo os padrões estabelecidos

nas diretrizes. Não há espaço para “caixa-preta” quando o assunto é *Reproducible Research*. Restringir código, dados, documentação, detalhes do ambiente computacional e artefatos, essenciais para a reprodutibilidade dos resultados da pesquisa, é inaceitável (Morin et al., 2012).

Com base nessas orientações, iniciativas foram tomadas e algumas diretrizes estão sendo obedecidas por pesquisadores, revistas e instituições de fomento, principalmente, no que se refere à divulgação de códigos e dados, mas a credibilidade dos resultados apresentados nos artigos ainda é questionada, não há coerência, variáveis são omitidas, faltam dados. Novas ações precisam ser tomadas, principalmente, no processo de formação de cientistas computacionais que trabalham com análise de dados (Peng, 2015; Munafò et al., 2017).

## 2.3 Soluções computacionais para Reproducible Research

A necessidade de solucionar o problema da pouca confiabilidade dos resultados quantitativos expressos nos artigos científicos computacionais é inerente ao termo *Reproducible Research* (Schwab et al., 2000). Ferramentas são desenvolvidas desde a década de 90 para tentar resolver esse problema e promover a pesquisa reprodutível.

No geral, esses softwares são categorizados nos seguintes tipos: programação literária, gerenciamento de fluxo de trabalho, controle de versão, captura do ambiente computacional (Stodden et al., 2013a). Diante disso, mencionamos algumas das ferramentas que se destacaram na comunidade científica, devido atenderem às necessidades de determinados grupos de pesquisa.

O *ReDoc* é uma das ferramentas pioneiras, desenvolvida em 1990 por membros da Universidade de Stanford, que oferece requisitos de software referentes à documentação, ao detalhamento do fluxo de trabalho e, atualmente, ainda é utilizada por membros do *Stanford Exploration Project* (SEP)<sup>3</sup>.

Em 1999, uma outra ferramenta ganha destaque na área quantitativa, a *WaveLab*, biblioteca do *MATLAB* desenvolvida por David Donoho e sua equipe, com o propósito de atender às necessidades de transparência, universalidade, reprodutibilidade e proporcionar aos pesquisadores uma análise detalhada dos trabalhos com análises *wavelet*.

Mais tarde, novas funcionalidades são implementadas e disponibilizadas para o *MATLAB* através dos pacotes *BeamLab*<sup>4</sup>, *SymmLab*<sup>5</sup> e *SparseLab*<sup>6</sup>. Todos esses pacotes seguem o conceito de pesquisa reprodutível e permitem o trabalho colaborativo e a reprodução dos resultados publicados em artigos.

<sup>3</sup>Uma associação acadêmica, financiada pela indústria, com o objetivo de melhorar o processo de construção de imagens sísmicas em três e quatro dimensões.

<sup>4</sup>Disponível em: <<http://statweb.stanford.edu/~beamlab/>>. Acesso em: 10 jun. 2017.

<sup>5</sup>Disponível em: <<http://statweb.stanford.edu/~symmlab/>>. Acesso em: 10 jun. 2017.

<sup>6</sup>Disponível em: <<https://sparselab.stanford.edu/>>. Acesso em: 11 jun. 2017.

A linguagem *R* também disponibiliza funções para facilitar o processo de adoção de hábitos reprodutíveis, como integração de dados e relatórios, através do *Sweave*<sup>7</sup>. Os pacotes *Knitr* e *RMarkdown*, também permitem a integração da análise estatística e apresentação dos resultados em um único documento, trabalham com linguagens de marcação  $\LaTeX$ , *Markdown* e *HTML* e outras linguagens de programação, além do *R*, como *Bash*, *Python* e *Ruby* (Gandrud, 2016).

Apesar dessas ferramentas auxiliarem o processo reprodutível, ainda há muito a ser feito para amenizar a dificuldade em reunir todos os artefatos necessários para reproduzir as experiências computacionais e como elas foram construídas ao longo do processo de pesquisa. Em tal caso, membros da Universidade de Nova York (NYU) desenvolveram a ferramenta *ReproZip*<sup>8</sup>. Essa ferramenta permite reproduzir um experimento de forma menos complicada, fornecendo ao usuário, que utiliza o sistema operacional *Linux*, a possibilidade de capturar com facilidade todos os componentes imprescindíveis para a reprodutibilidade dos resultados do seu projeto, em um único pacote compartilhável. (Chirigati et al., 2013).

Existem sistemas de gerenciamento de fluxos de trabalhos (*Workflow Management Systems* – WMS) desenvolvidos para prestar suporte ao processo da pesquisa científica e fornecer infraestrutura para configurar, executar e monitorar fluxos de trabalhos científicos.

Freire et al. (2014), por exemplo, apresentam um *framework* que auxilia o ciclo de vida dos experimentos computacionais e é utilizado como parte do sistema *open-source* denominado *VisTrails*, este é baseado em exploração de dados, fluxos de trabalho e sistema de visualização. Foi desenvolvido com a linguagem de programação *Python* e funciona com o auxílio das ferramentas *open-source*: *VTK*<sup>9</sup>, *Matplotlib*<sup>10</sup> e *ImageMagick*<sup>11</sup> (Freire et al., 2012).

O *Sumatra*<sup>12</sup> também é um WMS, pacote da linguagem de programação *Python* que permite a captura sistemática do contexto de simulações, ou análises numéricas (Davison et al., 2014). Similar ao *VisTrails*, ambos se preocupam com a reprodutibilidade. De acordo com Davison (2012), a principal diferença entre essas ferramentas é que a primeira limita-se ao gerenciamento de fluxo de trabalho, e a segunda à captura do ambiente.

Em 2013, o *Center for Open Science (COS)*<sup>13</sup>, um centro de pesquisa aberta comprometido em aumentar a abertura, integralidade e reprodutibilidade da pesquisa científica, independente da área de conhecimento, disponibiliza o *Open Science Framework (OSF)*<sup>14</sup>, a ferramenta que auxilia o trabalho colaborativo e transparente durante o ciclo de vida da pesquisa. Entretanto, estudos de caso realizados com o *OSF* mostram que a ferramenta apresentou mais benefícios para estudos reprodutíveis na área de psicologia quantitativa

<sup>7</sup>Sincroniza a linguagem de programação *R*, ou *S-Plus* com a linguagem de marcação  $\LaTeX$  (Leisch, 2002).

<sup>8</sup>Disponível em: <<http://vida-nyu.github.io/reprozip/>>. Acesso em: 15 jun. 2017.

<sup>9</sup>Disponível em: <<http://www.vtk.org/>>. Acesso em: 05 jun. 2017.

<sup>10</sup>Disponível em: <<http://matplotlib.org/>>. Acesso em: 09 jun. 2017.

<sup>11</sup>Disponível em: <<http://www.imagemagick.org/>>. Acesso em: 09 jun. 2017.

<sup>12</sup>Disponível em: <<http://neuralensemble.org/sumatra/>>. Acesso em: 20 jun. 2017.

<sup>13</sup>Disponível em: <<https://cos.io/>>. Acesso em: 09 out. 2017.

<sup>14</sup>Disponível em: <<https://osf.io/>>. Acesso em: 15 ago. 2017.



(Foster, 2017).

Em 2015, pesquisadoras da Universidade da Columbia, motivadas pelo conceito de “compendio” (Gentleman & Temple Lang, 2004), desenvolvem o *ResearchCompendia*<sup>15</sup>, um serviço web sem fins lucrativos que permite compartilhar os softwares e dados associados a uma publicação científica e, além disso, possibilita que jornais e revistas utilizem a ferramenta para a publicação de seus artigos, ou simplesmente para promover boas práticas de reprodutibilidade (Stodden et al., 2015).

Editores de periódicos indexados recomendam fortemente a divulgação de dados e códigos vinculados às boas práticas de *RR*, pois isso agrega valor e confiança aos resultados publicados (Frery, 2015; McNutt, 2014). À vista disso, revistas e jornais científicos estão agregando ferramentas para compartilhamento de artefatos da pesquisa em seus repositórios eletrônicos.

O *Image Processing On Line (IPOL)*<sup>16</sup> é considerado um dos jornais científicos pioneiros a se comprometer com reprodutibilidade na área de processamento de imagens. Foi idealizado em 2011 como uma plataforma experimental, atualmente é indexado e tem mais de 80 artigos publicados. Todos os algoritmos publicados são pseudo códigos nas linguagens C, C++, ou *MATLAB* (Limare & Morel, 2011; Colom et al., 2015).

O *ReScience Journal*<sup>17</sup> é uma revista hospedada no *GitHub*, desenvolvida por pesquisadores comprometidos com a comunidade de código aberto, revisada por pares e voltada para a pesquisa computacional reprodutível. O objetivo central do *ReScience* é encorajar a replicação explícita de pesquisas publicadas na forma tradicional, promovendo implementações novas para garantir que a pesquisa original seja reprodutível. Dessa forma, para cada novo estudo computacional são disponibilizados comentários, explicações e testes.

O *Code Ocean*<sup>18</sup> é uma plataforma de reprodutibilidade computacional baseada em nuvem, lançado em fevereiro de 2017. Tem funcionalidades similares ao *IPOL*, mas com alguns acréscimos, permite que pesquisadores disponibilizem e reproduzam códigos e dados publicados em revistas ou conferências acadêmicas, sem precisar realizar qualquer tipo de instalação na máquina local. Essa ferramenta permite que códigos escritos, em até 11 linguagens de programação diferentes, sejam executados por meio de um navegador web (Silver, 2017).

A maioria dessas ferramentas, disponíveis durante o fim do século XX e início do século XXI, estão atreladas às necessidades particulares de laboratórios de pesquisa e instituições preocupados apenas com a publicação e a reprodução dos resultados, porém não oferecem ao pesquisador o suporte necessário para atender as diretrizes de boas práticas de *RR* durante todo o desenvolvimento do projeto de pesquisa.

<sup>15</sup>Disponível em: <<http://146.20.110.164/>>. Acesso em: 20 ago. 2017.

<sup>16</sup>Disponível em: <<http://www.ipol.im/>>. Acesso em: 18 jul. 2017.

<sup>17</sup>Disponível em: <<http://rescience.github.io/>>. Acesso em: 05 ago. 2017.

<sup>18</sup>Disponível em: <<https://codeocean.com>>. Acesso em: 10 out. 2017.



Isso abre espaço para a implementação da ferramenta proposta neste trabalho, que disponibiliza aos pesquisadores um software livre e customizável com diversas funcionalidades para trabalhar de forma colaborativa e sistemática, aplicando boas práticas de *RR* e gerência de projetos durante todo o ciclo de vida do projeto de pesquisa quantitativa.

Diante disso, a tabela 2.1 mostra uma comparação entre algumas das ferramentas supracitadas e a ferramenta proposta neste trabalho, a qual denominamos pRRoject. A comparação tem como parâmetro os seguintes critérios:

- Oferece funcionalidades que atendem aos requisitos de boas práticas de *RR*:
  1. Um repositório com acesso público pela web.
  2. O relatório da pesquisa, o(s) conjunto(s) de dados, scripts, comandos, instruções e qualquer artefato relevante.
  3. Um fórum com recursos de discussão entre autores e leitores.
  4. Informações dos autores com links às respectivas páginas web.
  5. Artigo em PDF e informações sobre seu estado (em preparação, submetido, aceito).
  6. Uma página com o resumo do projeto de pesquisa.
  7. Referências bibliográficas da pesquisa com seus respectivos abstracts.
  8. Informações da plataforma na qual o resultado da sua pesquisa foi testado.
- Disponibiliza o código fonte da ferramenta de forma livre;
- Permite adicionar e retirar funcionalidades da ferramenta;
- Suporta funcionalidades primordiais ao gerenciamento e construção do projeto científico;
- Mostra todo o processo de desenvolvimento da pesquisa científica do início até a publicação.

Tabela 2.1: Quadro comparativo de algumas soluções

Software	Práticas de RR <sup>a</sup>	Código livre	Customizável	Gerência de projetos	Ciclo de vida
IPOL	2, 4, 6, 7, 8	—	—	—	—
ReScience Journal	1, 2, 4, 6, 7	—	—	—	—
ResearchCompendia	2, 3, 4, 5, 6	—	—	—	—
Sumatra	8	X	—	—	—
ReproZip	8	X	—	—	—
VisTrails	2, 8	X	—	—	—
OSF	1, 2, 4, 5, 6, 8	—	—	—	X
Code Ocean	1, 2, 4, 6, 8	—	—	—	—
pRRoject	1, 2, 3, 4, 5, 6, 7, 8	X	X	X	X

Fonte: Autor, 2017.

<sup>a</sup>Requisitos de boas práticas enumerados no primeiro item da página anterior.

## 2.4 Gerência de Projetos de Pesquisa

A gerência de projetos surgiu na década de 50 como uma metodologia eficaz para controlar atividades e exige aplicação de conhecimentos, habilidades, ferramentas e técnicas na projeção de atividades para atender aos objetivos do projeto (Wintage, 2015).

Os projetos de pesquisa e desenvolvimento, geralmente, iniciam com um levantamento bibliográfico, em seguida são realizadas algumas experiências ou atividades e, no fim, ocorre uma publicação. No entanto, o progresso na pesquisa não segue uma linearidade, há processos criativos e muitas incógnitas.

Wintage (2015) afirma que as características inerentes ao progresso da pesquisa não impedem a gerência desse tipo de projeto, ao contrário, o gerenciamento dos projetos de pesquisa e desenvolvimento deve auxiliar a sanar empecilhos, oferecer informações, aumentar a visibilidade, garantir credibilidade e facilitar a reprodução dos resultados.

As atividades dos projetos de pesquisa e desenvolvimento são equiparadas com as categorias de gerenciamento de projetos, determinadas por Wintage (2015) como: liderança, definição, planejamento e controle (Tabela 2.2).

Tabela 2.2: Alinhamento da pesquisa com categorias de gerenciamento de projetos.

Categoria de Gerência de Projetos	Atividades de Pesquisa
Liderança	Gerar ideias Ganhar apoio Obter recursos
Definição	Documentar a ideia Projetar estudo Projetar experimento
Planejamento	Executar testes Executar experimentos Verificar e validar Examinar a qualidade Avaliar riscos
Controle	Interpretar resultados Documentar resultados Fechar atividades

Fonte: Tabela adaptada de Wintage, 2015.

Dessa forma, percebe-se a flexibilidade das “boas práticas” de gerência de projetos que são a aplicação de conhecimentos, habilidades e técnicas em projetos diversos, com o intuito de alcançar resultados bem sucedidos.

O conceito de “boas práticas” (Project Management Institute, 2013) na gerência de projetos assemelha-se à proposta de “melhores práticas” (Stodden et al., 2010) para a pesquisa

reprodutível, pois ambos propõem uma abordagem sistemática que agrega valor ao projeto de pesquisa.

Segundo [Lippe & vom Brocke \(2016\)](#), uma boa gerência de projetos beneficia os projetos de pesquisa colaborativa em todas as suas dimensões. Esse fato dá suporte ao trabalho aqui apresentado.

O presente capítulo elencou alguns fatores históricos e científicos sobre a pesquisa científica e *Reproducible Research*. Além disso, apresentou os principais softwares utilizados para publicar artigos científicos reprodutíveis e relacionou boas práticas de *RR* com gerência de projetos. No próximo capítulo, apresentaremos um ambiente computacional *open source* para suporte a *Reproducible Research*.

# 3

## AMBIENTE REPRODUCIBLE RESEARCH

ESTE capítulo apresentará um ambiente computacional *open source* com funcionalidades que obedecem alguns dos requisitos de boas práticas para a pesquisa reprodutível. Além disso, mostrará módulos de extensões acrescentados ao ambiente para estender suas funcionalidades, visando obedecer todos os requisitos de boas práticas de *RR* desde o início da pesquisa.

### 3.1 Metodologia

A proposta deste trabalho é fundamentada na necessidade apresentada por [Donoho et al. \(2009\)](#): reprodutibilidade computacional deve ser realizada desde o início da pesquisa, atrelada às recomendações de boas práticas de *Reproducible Research*.

Para atender a essa demanda verificamos, com base na literatura, a necessidade da implementação de um ambiente computacional com funcionalidades que auxiliem os pesquisadores da área computacional quantitativa a adotarem boas práticas de *Reproducible Research* durante todo o ciclo de vida do projeto de pesquisa colaborativa.

À vista disso e do tempo para a execução deste trabalho (2 anos) consideramos viável a seleção de um *software open source* multiplataforma<sup>1</sup>, conforme os seguintes parâmetros: percentual de utilização do software no mundo, quantidade de plugins desenvolvidos, presença de comunidade ativa e tempo médio de atualização.

Em seguida, realizamos um estudo minucioso do *software* Redmine, estendemos as suas funcionalidades, o utilizamos no processo de escrita deste trabalho para realização de um estudo de caso, o disponibilizamos através de um ambiente virtual e realizamos uma validação.

---

<sup>1</sup>Sistema que pode ser configurado e executado em diferentes sistemas operacionais como, por exemplo: *Linux, Windows, macOS* e etc.

Para realizar o processo de pesquisa e desenvolvimento deste trabalho não foi necessário levar em conta custos, como mostra o detalhamento das atividades executadas:

- Estudo aprofundado sobre as funcionalidades da ferramenta Redmine;
- Estudos sobre o *framework Ruby on Rails* e o desenvolvimento de plugins para Redmine;
- Montagem de um ambiente de desenvolvimento para customizar o Redmine;
- Criação e configuração de banco de dados *Mysql*;
- Pesquisas sobre templates compatíveis com o Redmine;
- Criação de um template filho;
- Criação de novos campos na página cadastro de usuário;
- Pesquisas e testes de plugins para Redmine;
- Seleção, instalação, configuração e utilização de plugins;
- Configuração dos requisitos de segurança e acessibilidade da ferramenta;
- Hospedagem da ferramenta em um dos servidores do Laboratório de Computação Científica e Visualização (LCCV/UFAL);
- Criação de projetos e repositórios;
- Cadastro de usuários;
- Criação de manual de utilização da ferramenta;
- Utilização das funcionalidades da ferramenta, após hospedagem no servidor;
- Solução de erros técnicos apresentados durante o processo de configuração da ferramenta e instalação de novos plugins.
- Instalação e configuração de um ambiente virtual para disponibilizar a versão da ferramenta hospedada no servidor;
- Criação de um formulário de validação da ferramenta.

## 3.2 Ambiente computacional *open source* para suporte a Reproducible Research

Um ambiente para suporte à pesquisa reprodutível é um sistema que agrega ferramentas *Free/Libre and Open Source Software* (FLOSS)<sup>2</sup> e facilita a adoção das práticas de *RR* por parte do pesquisador.

As diretrizes de boas práticas de *RR*, como supracitado, recomendam ao pesquisador: uso de controle de versão<sup>3</sup>; repositório com acesso público na *web*; disponibilidade do relatório, dados, *scripts*, comandos, instruções e qualquer artefato relevante para o processo de desenvolvimento da pesquisa; fórum ou algo similar que possua recursos de discussão entre autores e leitores; um relatório com estrutura IMRAD (Introdução, Metodologia, Resultados e Discussão).

De acordo com isso e considerando o ciclo de vida de um projeto de pesquisa, averiguamos a necessidade de um sistema *web* multiplataforma que disponibilize no mínimo: controle de versão, gerência de usuários, escrita colaborativa e fóruns.

No entanto, existem centenas de ferramentas *open source* que atendem essa demanda, por exemplo, os sistemas classificados como gestores de projetos. Dessa forma, escolhemos um sistema com licença *GNU General Public License*<sup>4</sup>, flexível e extensível, denominado *Redmine*.

O *Redmine* é um sistema de gerenciamento de projetos desenvolvido pelo engenheiro Jean Phillippe Lang, com o auxílio de um grupo de desenvolvedores integrantes da comunidade *Redmine Lesyuk* (2013).

As funcionalidades do *Redmine* permitem ao pesquisador criar projetos colaborativos, com *wiki*<sup>5</sup>, gerência de tempo, controle de versão, fórum, *workflow*<sup>6</sup> do processo de execução das atividades.

Conforme as diretrizes de boas práticas de *RR*, o pesquisador deve disponibilizar a pesquisa em uma página *web* com as seguintes informações:

- autores, com links às respectivas páginas *web*, por exemplo, link para o currículo Lattes;
- resumo da pesquisa;
- o artigo em PDF e informações sobre seu estado (em preparação, submetido, aceito);
- os códigos que geram os dados, as figuras e as tabelas documentados e com informações sobre como executá-los;

---

<sup>2</sup>Softwares licenciados que disponibilizam o código fonte de forma livre e permitem utilização, estudo e modificação.

<sup>3</sup>É uma ferramenta que salva todas as modificações realizadas em um documento e permite recuperá-las.

<sup>4</sup>Uma licença gratuita com livre direito de cópia para softwares e outros trabalhos. *GNU GPL* garante a liberdade para compartilhar e alterar todas as versões de um programa. Disponível em: <<http://www.gnu.org/licenses/gpl.html>>. Acesso em: 15 jul. 2017.

<sup>5</sup>Uma ferramenta para criação de páginas *web* de forma colaborativa.

<sup>6</sup>Automação das etapas dos processos realizados no decorrer do projeto.

- os dados e informações sobre como lê-los;
- informações da plataforma onde o software foi testado;
- email ou outras formas de contato;
- referências com os seus *abstracts*.

Para auxiliar o pesquisador a disponibilizar essas informações e obedecer as normas de boas práticas de *RR*, durante todo o processo de desenvolvimento da pesquisa, customizamos o *Redmine*, através da instalação de alguns *plugins*<sup>7</sup>. Essa customização mostra um ambiente *reproducible research* associado a algumas práticas de gerência de projetos e considera essa agregação um critério vantajoso para dar suporte aos pesquisadores no decorrer do desenvolvimento do projeto de pesquisa.

### 3.3 Do Redmine ao pRRoject

Estendemos o *Redmine* através da instalação e configuração dos seguintes *plugins*:

#### *Knowledgebase Plugin*

Bevilacqua (2014) percebeu o número de funcionalidades oferecidas pelo sistema *open source Redmine* e a facilidade para sua extensão, criou o *plugin redmine\_knowledgebase*.

O *redmine\_knowledgebase*<sup>8</sup> é uma base de conhecimento que permite armazenar artigos e, além disso, categorizar, avaliar, comentar, etiquetar e gerenciar permissões e versões. Esse *plugin* é *open source* e o código fonte está disponível através da licença *Massachusetts Institute of Technology (MIT)*<sup>9</sup>.

Esse *plugin* permite que o *pRRoject* disponibilize ao pesquisador a possibilidade de atender uma das diretrizes de boas práticas de *RR*: publicar códigos, dados, documentação, pré-impressões de artigos.

Portanto, o *redmine\_knowledgebase* possibilita a publicação, na web, das seguintes informações dos artigos:

- resumo da pesquisa;
- o artigo em PDF e informações sobre seu estado (em preparação, submetido, aceito);
- os códigos que geram os dados, as figuras e as tabelas, documentados e com informações sobre como executá-los;
- referências com os seus *abstracts*.

<sup>7</sup>Uma extensão da ferramenta para agregar novas funcionalidades.

<sup>8</sup>Disponível em: <[https://github.com/alexbevi/redmine\\_knowledgebase](https://github.com/alexbevi/redmine_knowledgebase)>. Acesso em: 20 ago. 2017.

<sup>9</sup>Uma licença com livre direito de cópia utilizada em *software* livre.



### ***Git Hosting Plugin***

O *plugin redmine\_git\_hosting* tem como objetivo facilitar a integração da ferramenta de controle de versão *Git* com o *Redmine* e a configuração do *Git* no servidor de hospedagem. Esse *plugin* utiliza a ferramenta *gitolite*<sup>10</sup>, responsável pelo gerenciamento das funcionalidades do *Git* dentro do *Redmine*.

Com o auxílio desse *plugin*, o *pRRproject* oferece ao pesquisador a possibilidade de atender a mais uma das diretrizes de boas práticas de *RR*: usar controle de versão e licenciamento aberto de códigos.

### ***Redmine People Plugin***

Uma ferramenta utilizada para gerenciar recursos humanos, suas funcionalidades disponibilizam ao *pRRproject* a possibilidade de cadastrar mais informações sobre o usuário, gerenciá-los e categorizá-los.

Dessa forma, o *pRRproject* auxilia o pesquisador a publicar na web as informações dos autores e dos demais participantes do projeto de pesquisa, com links às respectivas páginas web e etc.

### ***Redmine Questions and Answers Plugin***

O *plugin* permite que os usuários possam publicar perguntas e respostas e classificá-las, através de uma seção de “perguntas mais frequentes”, denominada *Frequently Asked Questions (FAQ)*.

Além disso, é possível criar um espaço para fornecer informações gerais sobre o projeto e também uma seção para sugestões de ideias.

O usuário tem autonomia para votar nas melhores respostas e o administrador do projeto tem como categorizar o conteúdo disponibilizado e gerenciá-lo.

Dessa maneira, o *plugin Questions and Answers* torna a ferramenta fórum, disponibilizada pelo *Redmine*, mais interativa e oferece ao pesquisador a possibilidade de atender a mais um requisito de boas práticas de *RR*: disponibilizar o projeto de pesquisa em página web com variadas fontes de contato, que permitam discussão entre autores e leitores.

---

<sup>10</sup>Disponível em: <<http://gitolite.com/gitolite/>>. Acesso em: 25 jul. 2017.

Os plugins mencionados anteriormente incorporam funcionalidades ao Redmine e dão origem ao ambiente *reproducible research*, denominado *pRRoject* (Figura 3.1).

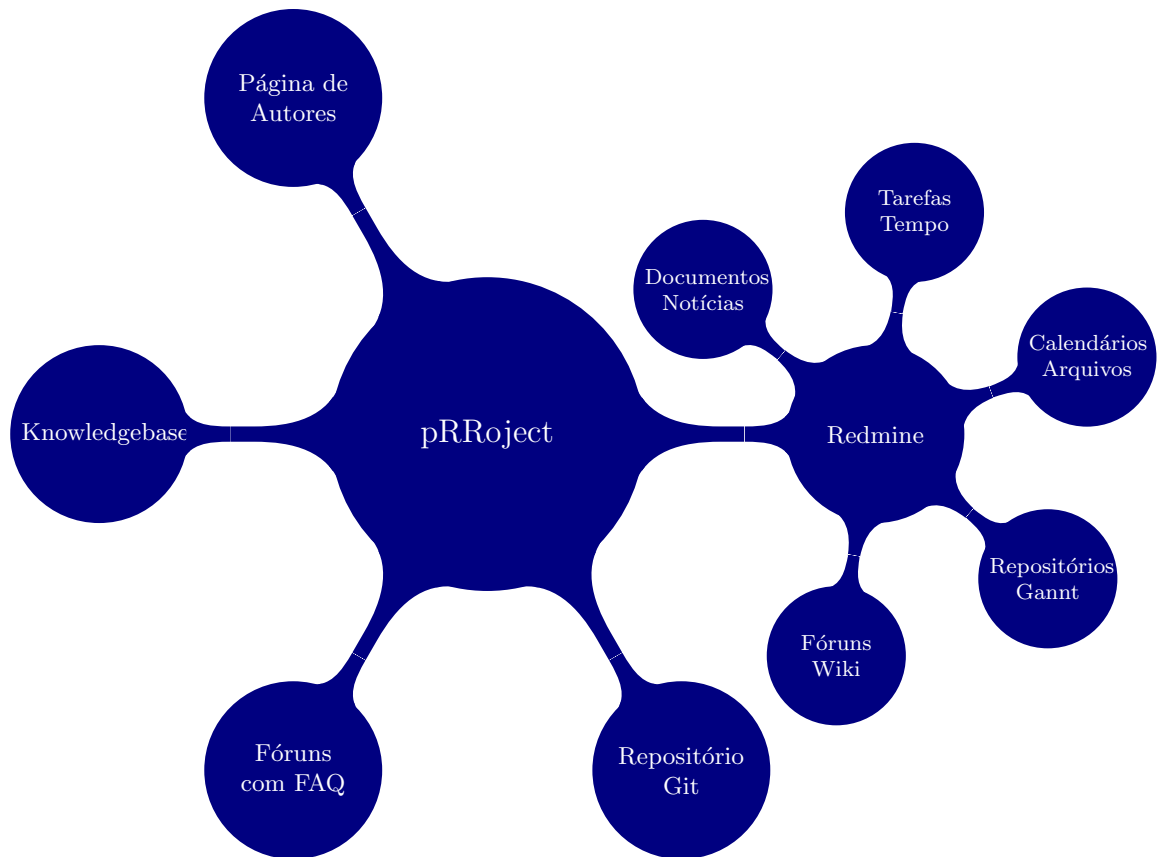


Figura 3.1: Ambiente Reproducible Research pRRoject.

As melhorias disponíveis nas funcionalidades de cada plugin oferecem ao usuário a possibilidade de iniciar projetos científicos conforme os requisitos de boas práticas de *RR*:

1. Um repositório com acesso público pela web.
2. O relatório da pesquisa, o(s) conjunto(s) de dados, scripts, comandos, instruções e qualquer artefato relevante.
3. Um fórum com recursos de discussão entre autores e leitores.
4. Informações dos autores com links às respectivas páginas web.
5. Artigo em PDF e informações sobre seu estado (em preparação, submetido, aceito).
6. Uma página com o resumo do projeto de pesquisa.
7. Referências bibliográficas da pesquisa com seus respectivos abstracts.
8. Informações da plataforma na qual o resultado da sua pesquisa foi testado.

### 3.4 Virtualização do pRProject

O ambiente virtual permite a realização de testes de uma determinada ferramenta por parte do usuário, através da replicação das configurações da máquina onde a ferramenta está hospedada.

Dessa maneira, resolvemos criar um ambiente virtual para o pRProject utilizando o *Vagrant*<sup>11</sup>, uma ferramenta que facilita o processo de virtualização.

O *Vagrant* trabalha com *boxes*, que são máquinas virtuais personalizadas. Assim, selecionamos uma *box* compatível com os requisitos de um ambiente de desenvolvimento *Ruby on Rails*, para facilitar a instalação do *Redmine*.

A *box* denomina-se *fnando/hellobits-trusty64*<sup>12</sup> e está configurada com o sistema operacional *Ubuntu Linux Server 14.04 LTS 64-bits* e os programas *Vim*, *MySQL*, *PostgreSQL*, *Ruby*, *Bundler*, *Git*, *Curl* e *Node.js*.

A partir dessa *box* instalamos o *framework Rails* e a versão 2.6 do *Redmine*, com as devidas configurações. Esses procedimentos deram origem a uma nova *box*, denominada *pRProject.box* que permite a fácil replicação do ambiente *Reproducible Research pRProject* em diferentes máquinas, independente do sistema operacional.

### 3.5 Validação do pRProject

O intuito de validar o pRProject é coletar informações para averiguar o impacto da ferramenta em referência aos requisitos de boas práticas de *Reproducible Research*, como um paradigma de pesquisa qualitativa através da técnica de observação participante. Para isso, criamos um formulário de pesquisa por meio da ferramenta *Google Forms*<sup>13</sup> e convidamos três membros do Laboratório de Computação Científica e Análise Numérica (LaCCAN), da Universidade Federal de Alagoas para participarem. A escolha de apenas três membros, decorre da necessidade da realização de uma análise detalhada do uso da ferramenta (pRProject), por parte dos usuários, durante o desenvolvimento de uma pesquisa científica quantitativa. Os convidados já estão cadastrados no pRProject, conhecem algumas funcionalidades da ferramenta e a utilizaram no processo de desenvolvimento dos seus respectivos projetos.

O presente capítulo detalhou o processo de extensão do ambiente computacional *Reproducible Research*. A seguir, no quarto capítulo, apresentaremos os resultados e análises com base no uso do ambiente.

<sup>11</sup>Disponível em: <<https://www.vagrantup.com/>>. Acesso em: 10 out. 2017

<sup>12</sup>Disponível em: <<http://files.hellobits.com/vagrant/hellobits-trusty64-virtualbox.box>>. Acesso em: 10 out. 2017.

<sup>13</sup>Disponível em: <<https://www.google.com/forms/about/>> Acesso em: 20 set. 2017.

# 4

## RESULTADOS E ANÁLISES

As seções deste capítulo mostram, respectivamente, as funcionalidades do pRRoject, sua utilidade frente a adoção de boas práticas de *Reproducible Research* durante o processo de escrita deste trabalho e uma breve discussão com relação ao seu uso. Dessa forma, disponibilizamos um ambiente computacional com suporte a *RR* e estudo de caso do uso desse ambiente.

### 4.1 O pRRoject

O pRRoject<sup>1</sup> permite o cadastro de múltiplos projetos de pesquisa. Cada projeto é especificado pelo título e descrição (Figura 4.1).

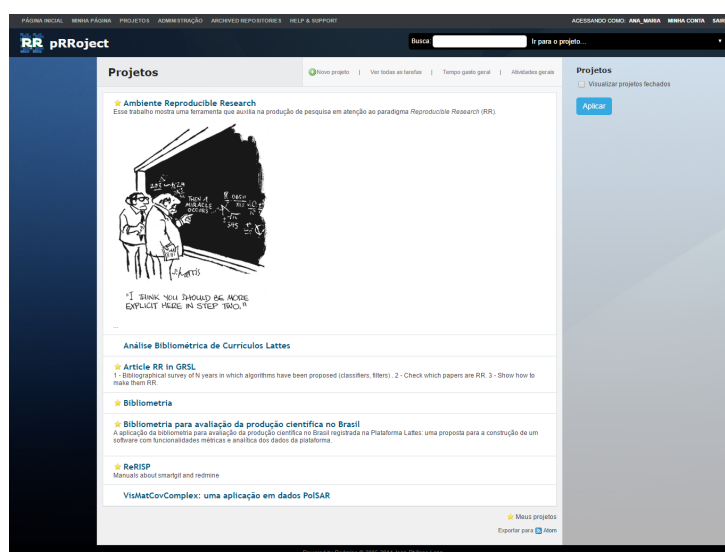


Figura 4.1: Página com a lista de projetos cadastrados no pRRoject.

<sup>1</sup>Disponível em: <<https://rerisip.lccv.ufal.br>>. Acesso em: 19 ago. 2017.

Para cada novo projeto criado no pRROject há opções para preencher título, descrever informações e selecionar módulos (Figura 4.2). Os módulos disponibilizam uma variedade de funcionalidades.

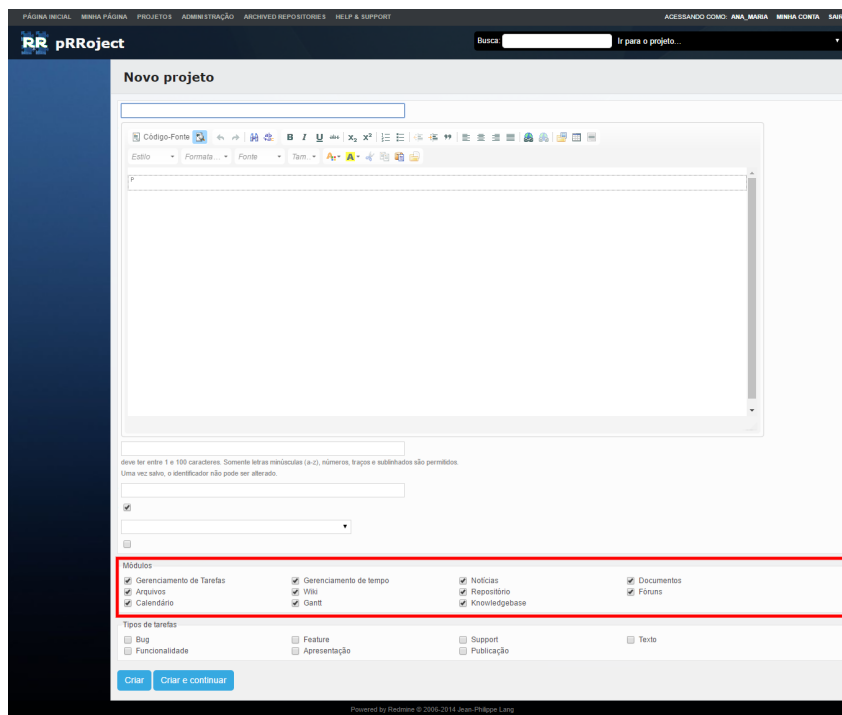


Figura 4.2: Página para cadastro de um novo projeto, com destaque para o campo de seleção de módulos.

As funcionalidades disponíveis através dos módulos são:

- gráfico de *Gantt* e calendário de tarefas;
- notícias, documentos e arquivos;
- fóruns interativos com *FAQ*;
- integração com repositório *Git*;
- base de conhecimento para adicionar artigos e categorizá-los como: em preparação, aceitos, publicados, submetidos, lidos;
- *Wiki* para criação de página web colaborativa.

Dentre essas funcionalidades destacamos a “base de conhecimento”, que permite a adição e categorização de artigos. Os membros do projeto têm autonomia para publicar pré-impresões dos seus artigos, escrever artigos de forma colaborativa e até mesmo, montar uma base bibliográfica. Ao adicionar os artigos é possível selecionar categoria, acrescentar título, resumo, conteúdo, palavras-chave e arquivos com breve descrição (Figura 4.3).

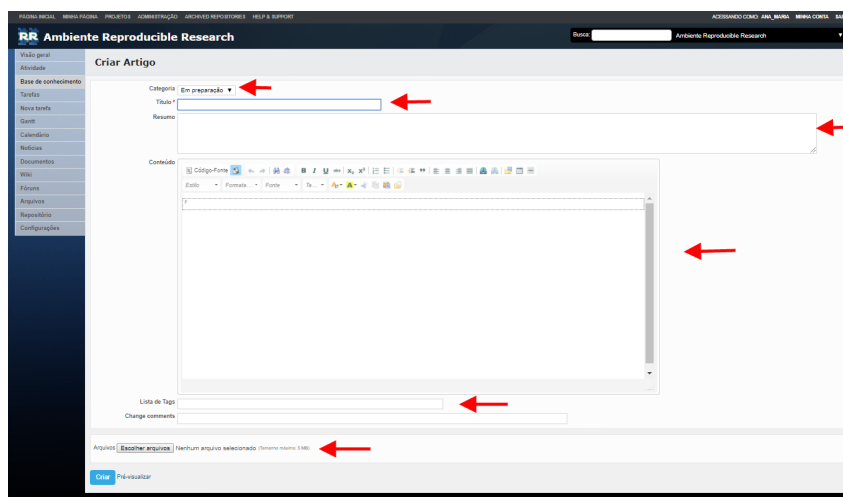


Figura 4.3: Página para adicionar artigos, com marcação nos campos de seleção e preenchimento

Além disso, é possível associar membros ao projeto e definir a função (papéis) de cada indivíduo (Figura 4.4). O papel determina as funcionalidades e informações às quais o usuário tem acesso.

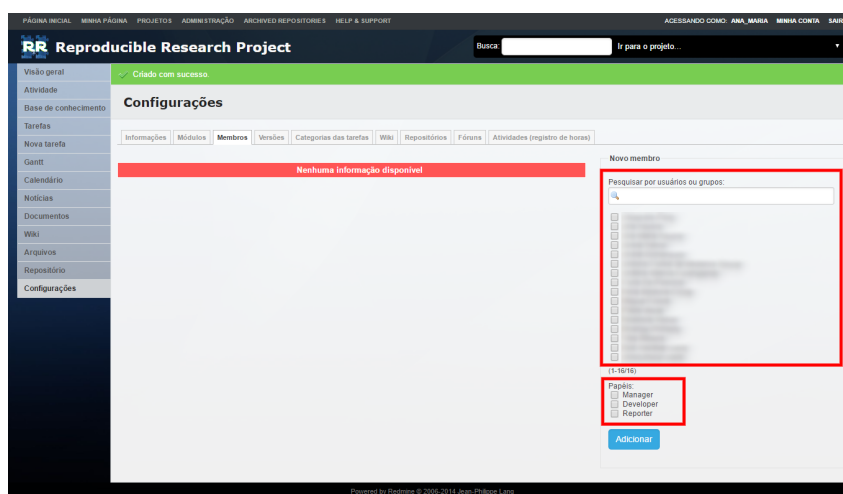


Figura 4.4: Página para associar membros ao projeto com destaque nos campos seleção de usuários e papéis.)

Cada usuário cadastrado no pRRoject tem uma página individual com descrição automática das atividades realizadas, arquivos enviados, campo para acrescentar suas informações pessoais e profissionais e lista dos projetos dos quais faz parte (Figura 4.5).

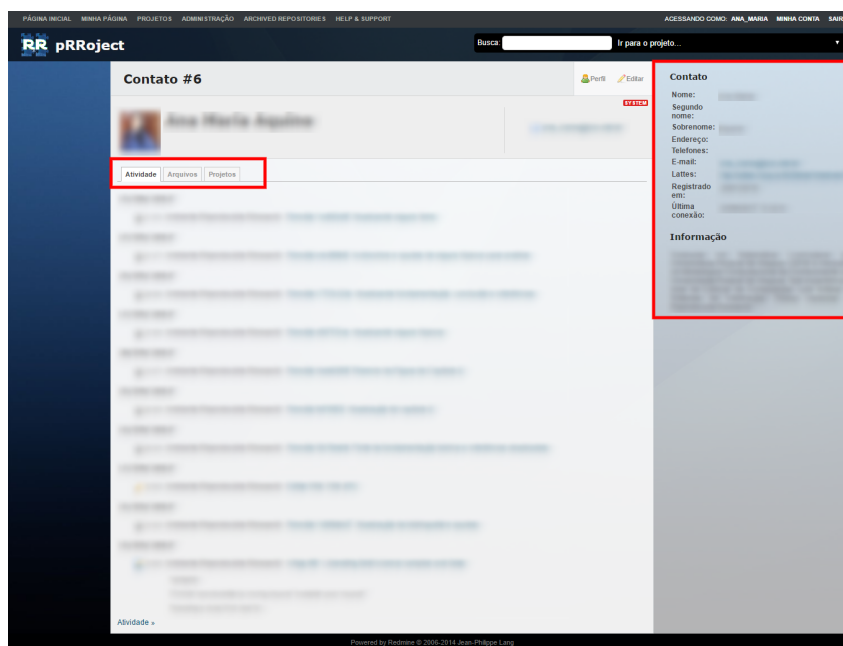


Figura 4.5: Página de usuário, com foto, nome, email, com destaque para as abas (atividade, arquivos e projetos) e para o espaço de informações pessoais e profissionais.

Todas informações adicionadas ao pRROject têm a opção de acesso público, ou privado (apenas para membros do projeto) e o administrador do sistema tem flexibilidade e facilidade para definir essas configurações no pRROject.

## 4.2 Desenvolvimento e Acompanhamento da Pesquisa

O pRROject foi utilizado no processo de escrita deste trabalho. Inicialmente, criamos um projeto intitulado *Ambiente Reproducible Research* e adicionamos uma breve descrição (Figura 4.6).

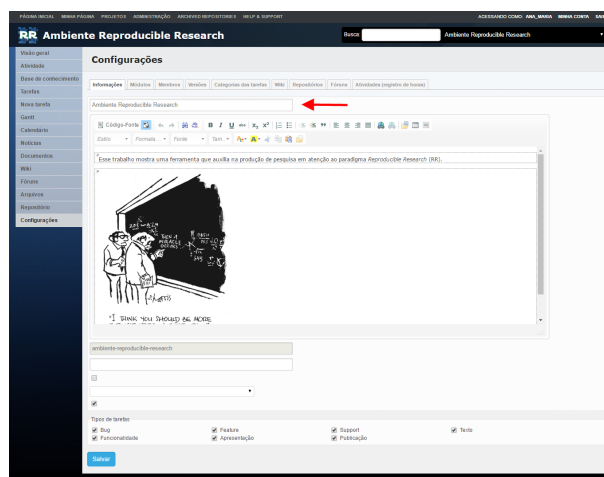


Figura 4.6: Projeto *Ambiente Reproducible Research* criado no pRROject

Cadastramos os membros participantes do projeto com algumas informações pessoais e profissionais (nome, sobrenome, email, endereço do currículo Lattes) e dados para autenticação no sistema (nome de usuário e senha) (Figura 4.7).

Figura 4.7: Página de cadastro de um dos membros do projeto Ambiente *Reproducible Research*.

Associamos os membros cadastrados ao projeto (Figura 4.8)

Usuário / Grupo	Papéis
Alejandro Frey	Manager
Ana Maria Aquiño	Manager

Figura 4.8: Página para associação de membros ao projeto Ambiente *Reproducible Research*, com destaque para os membros já associados ao projeto.

Criamos um repositório *Git*, para garantir o versionamento do projeto, com endereço: `ssh://git@rerisip.lccv.ufal.br:2221/ambiente-reproducible-research.git`<sup>2</sup> (Figura 4.9).

<sup>2</sup>Disponível em: <<https://rerisip.lccv.ufal.br/projects/ambiente-reproducible-research/repository>>. Acesso em: 10 set. 2017.



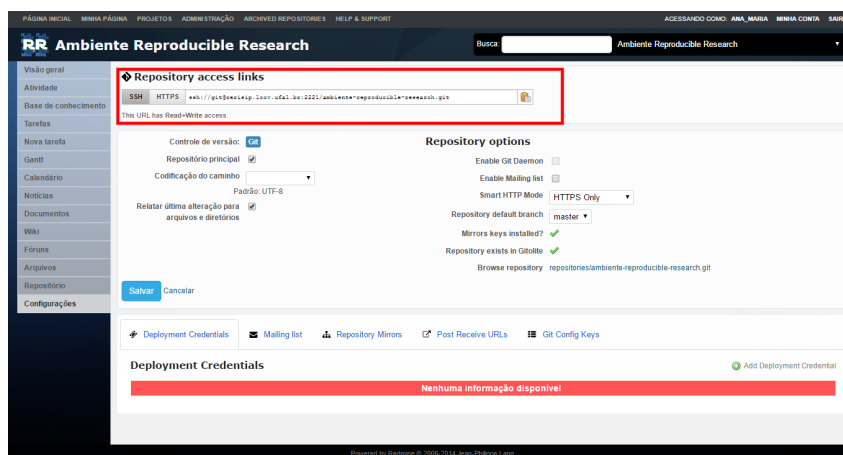


Figura 4.9: Página de configuração do repositório referente ao projeto *Ambiente Reproducible Research*.

Além disso, definimos uma estrutura de diretórios para organizar os arquivos enviados ao servidor através do *Git* (Figura 4.10). Essa estrutura pode ser alterada pelos autores segundo a forma que acharem oportuna para o projeto.

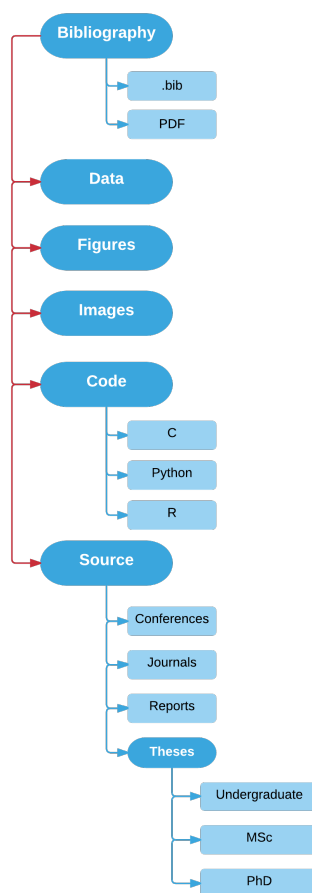


Figura 4.10: Estrutura padrão de diretórios.

No menu repositório é possível visualizar todos os comentários realizados pelos membros em referência aos arquivos enviados, com registro de data e hora. Assim, temos a informação do período de tempo em que o arquivo foi enviado e alterado (Figura 4.11).

The screenshot displays the RR repository interface. The main content area shows a file listing table with columns for Name, Size, Revision, Age, Author, and Comment. A red box highlights the 'Idade' (Age) column and the 'Comentário' (Comment) column. Below the file listing is a table titled 'Últimas revisões' (Latest revisions) with columns for #, Date, Author, and Comment, also highlighted with a red box. The interface includes a sidebar with navigation options and a right sidebar with repository access links and download buttons.

Nome	Tamanho	Revisão	Idade	Autor	Comentário
Bibliography		e4c86b5	30 dias	Ana Maria Aquino	Acréscimos e ajustes de alguns tópicos para aná...
Code		80afc719	10 meses	Ana Maria Aquino	Estrutura de diretórios
Data		80afc719	10 meses	Ana Maria Aquino	Estrutura de diretórios
Figures		e4c86b5	30 dias	Ana Maria Aquino	Acréscimos e ajustes de alguns tópicos para aná...
Images		80afc719	10 meses	Ana Maria Aquino	Estrutura de diretórios
Source		1ad92e9f	10 dias	Ana Maria Aquino	Atualizando alguns itens.
.Rhistory	0 Bytes	ea80ee52	7 meses	Ana Maria Aquino	Alteração no arquivo references bib
README.md	18 Bytes	3322e38e	7 meses	Ana Maria Aquino	Testando repositório.

#	Data	Autor	Comentário
1ad92e9f	15/06/2017 11:03 h	Ana Maria Aquino	Atualizando alguns itens.
8f485d75	30/05/2017 16:00 h	Alejandro Frery	Capítulo 3 revisado e com comentários
50da3c37	29/05/2017 18:38 h	Alejandro Frery	Capítulos 1 e 2 corrigidos e revisados. Implementar a sugestões.
e4c86b5	26/05/2017 21:27 h	Ana Maria Aquino	Acréscimos e ajustes de alguns tópicos para análise.
1753c32e	25/05/2017 09:09 h	Ana Maria Aquino	Atualizando fundamentação, conclusão e referências.
df3753ce	17/05/2017 15:46 h	Ana Maria Aquino	Atualizando alguns tópicos.
f4339d1d	28/04/2017 15:45 h	Alejandro Frery	Limpeando trailhas
4609d9e8	28/04/2017 15:44 h	Alejandro Frery	Um detalhe, e vamos em frente!
4add395f	28/04/2017 13:27 h	Ana Maria Aquino	Reenvio da Figura do Capítulo 2.
fef1823	25/04/2017 00:26 h	Ana Maria Aquino	Atualização do capítulo 2.

Figura 4.11: Página correspondente ao menu repositório, com destaque para a idade dos arquivos enviados e os comentários dos membros.

Ainda na página repositório, visualizamos as diferenças entre os arquivos após a revisão de algum dos membros (Figura 4.12). A página mostra as diferenças (lado a lado) do arquivo “fundamentacao.tex”, a cor vermelha representa a versão anterior do arquivo e a cor verde a versão mais recente com modificações.



Figura 4.12: Diferenças do arquivo “fundamentacao.tex” antes e depois da revisão de membros.

Também averiguamos algumas estatísticas referentes à participação dos membros do projeto conforme o número de *commits*<sup>3</sup> realizados e o tempo de atividade do projeto. Dessa forma, temos um resumo quantitativo das informações do repositório do projeto (Figura 4.13), a quantidade dos *commits* realizados categorizados por mês, dia, hora e semana (Figura 4.14) e o número de *commits* por membro (Figura 4.15).

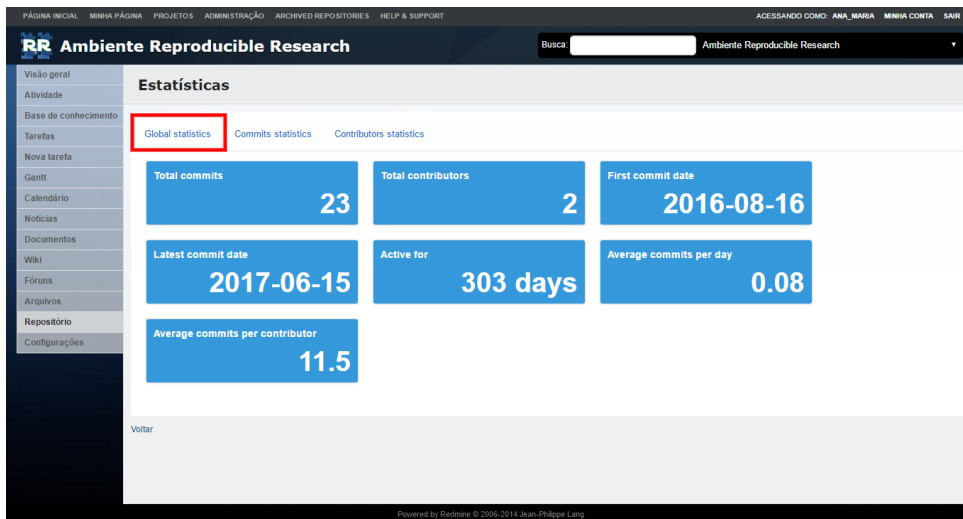


Figura 4.13: A página *Global statistics* mostra o total de *commits*, o total de membros, a data do primeiro e do último *commit* realizado até o momento de captura dessa imagem, a quantidade de dias em que o projeto está em atividade até então, a média de *commits* por dia e a média de *commits* por membro.

<sup>3</sup>Envio de arquivo para o servidor com uma breve descrição referente à realização de modificações nesse arquivo.

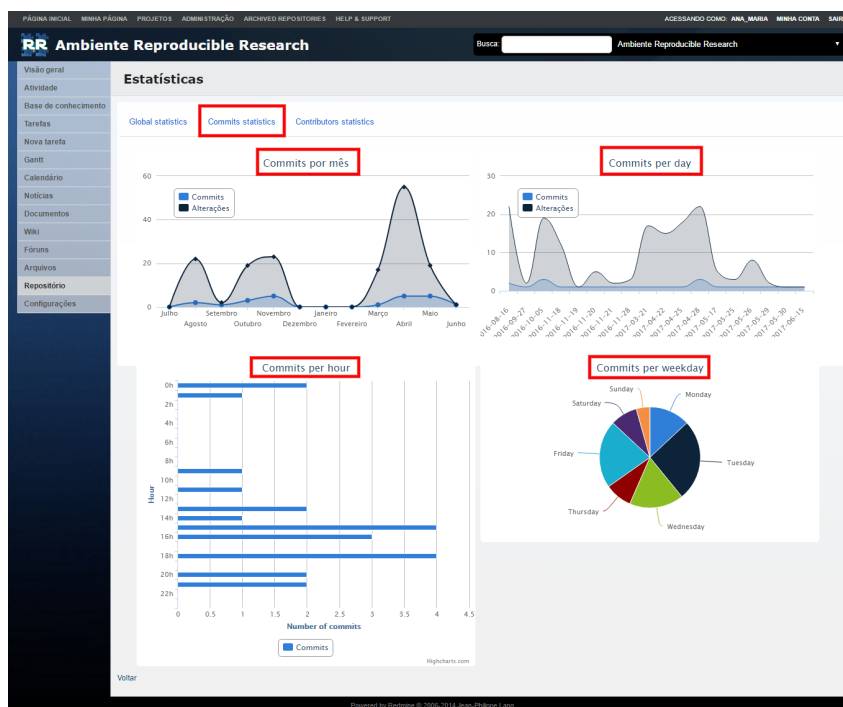


Figura 4.14: A página *Commits statistics* mostra os gráficos relativos ao número de *commits* categorizados por mês, dia, hora e semana.

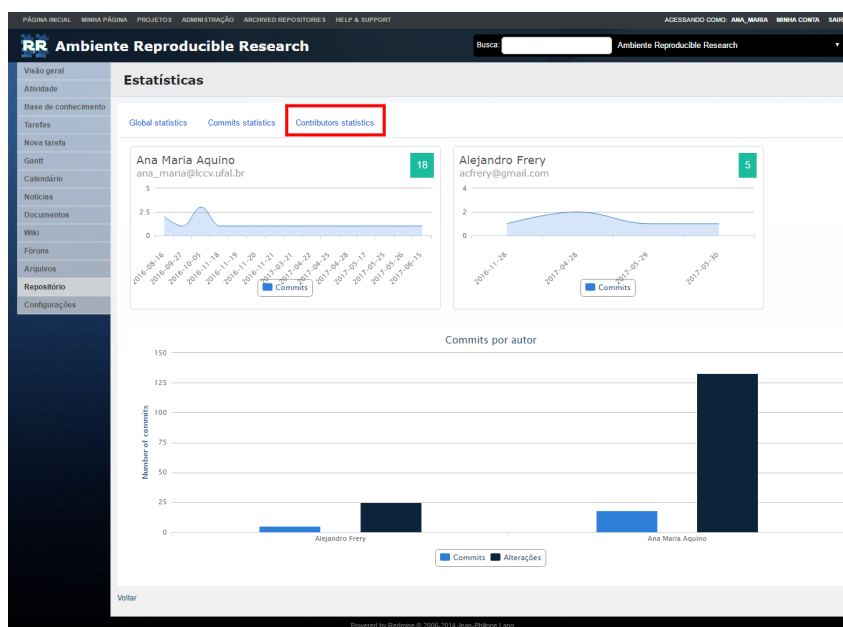


Figura 4.15: A página *Contributors statistics* mostra os gráficos dos *commits* referentes a cada membro do projeto e um gráfico detalhando a quantidade de *commits* e alterações realizadas por cada autor.

Após utilizar o repositório e verificar a possibilidade de ter o controle de todas as versões do trabalho, averiguamos a necessidade de atestar as funcionalidades da base de conheci-

mento. Para isso, adicionamos alguns dos artigos lidos durante a elaboração deste trabalho (Figura 4.16).

Figura 4.16: Página inicial da “Base de Conhecimento” com a lista dos artigos adicionados.

Definimos cinco categorias para os artigos adicionados: em preparação, submetido, aceito, publicado e lido. As quatro primeiras, respectivamente, definem o estado do artigo e a última funciona como uma categoria para criar base de dados bibliográfica vinculada ao projeto de pesquisa.

A Figura 4.16 mostra um sumário com o total de categorias cadastradas, o total de artigos inseridos e o total de artigos escrito pelo membro que está acessando o projeto, nesse caso, a ana\_maria. Todavia, é importante ressaltar que o total de artigos classificados como “Articles written by me” não significa de autoria do usuário.

À vista disso, apenas os artigos das categorias **em preparação, submetido, aceito e publicado**, são de autoria do usuário.

A página inicial (Figura 4.16) mostra a lista dos artigos recentemente atualizados, artigos mais populares e artigos mais bem classificados. No momento, os cinco artigos adicionados estão relacionados em todas as listas, pois não ocorreu classificação dos artigos por parte de outros membros do projeto.

As funcionalidades da base de conhecimento permitem que os membros do projeto votem nos artigos, os quais podem receber até cinco estrelas (Figura 4.17).

The screenshot shows the 'Ambiente Reproducible Research' website. The main article is titled 'Liberating field science samples and data' and is categorized as 'Lidos'. The article text discusses the importance of transparency and reproducibility in research. A red box highlights the 'Classificação do Artigo (Votos)' section, which shows a 4-star rating. Another red box highlights the 'New Comment' link in the right sidebar. The page also features a 'Highlights' section and a 'Full text' link.

Figura 4.17: Página de um dos artigos adicionados à base de conhecimento na categoria “Lidos” e classificado, por um dos membros do projeto, com quatro estrelas.

Ainda na página do artigo (Figura 4.17), é possível adicionar comentários, clicando no link “New comment” listado acima do espaço destinado à classificação (Figura 4.18).

The screenshot shows the same article page as Figure 4.17, but with a new comment added. The comment is highlighted with a red box and reads: 'Adicionado por Ana Maria Aquino menos de um minuto atrás. Uma pesquisa científica transparente aumenta a confiabilidade nos resultados.' The 'New Comment' link in the right sidebar is also highlighted with a red box. The article title and 4-star rating are visible at the top of the main content area.

Figura 4.18: Comentário emitido por um dos membros do projeto com relação ao conteúdo do artigo.

Ademais, associamos arquivos aos artigos adicionados na base de conhecimento (Figura 4.19).

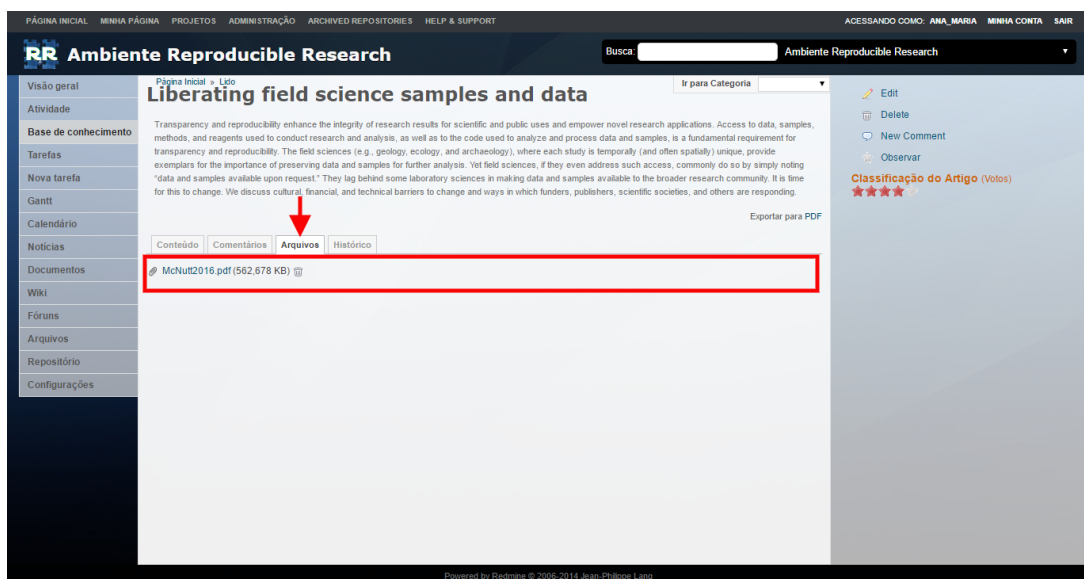


Figura 4.19: Subpágina (Arquivos) do artigo *Liberating field science samples and data* adicionado à categoria “Lidos” da base de conhecimento.

Assim como no repositório, na aba histórico temos registros de todas as versões do artigo e a opção de observar as diferenças (Figura 4.20).

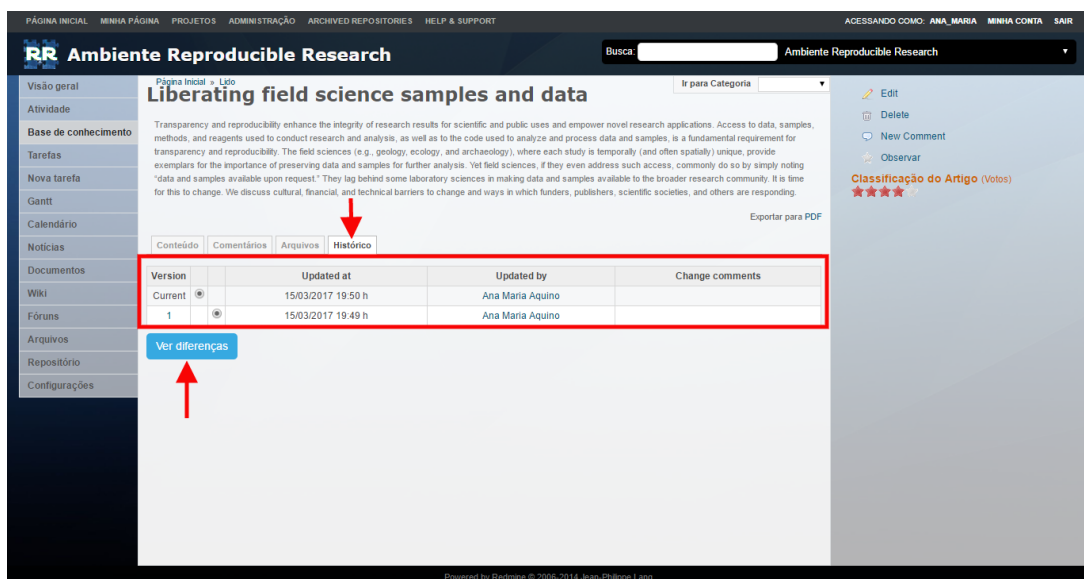


Figura 4.20: Subpágina (Histórico) do artigo *Liberating field science samples and data* adicionado à categoria “Lidos” da base de conhecimento.

A visualização das diferenças entre as versões dos artigos está disponível apenas para os membros do projeto, mas há a possibilidade de acesso público.

O acesso público ao projeto deste trabalho (*Ambiente Reproducible Research*) disponibiliza um site com as informações do ciclo de vida deste projeto acessível através do ende-

reço: <https://rerisip.lccv.ufal.br/projects/ambiente-reproducibile-research><sup>4</sup> (Figura 4.21).

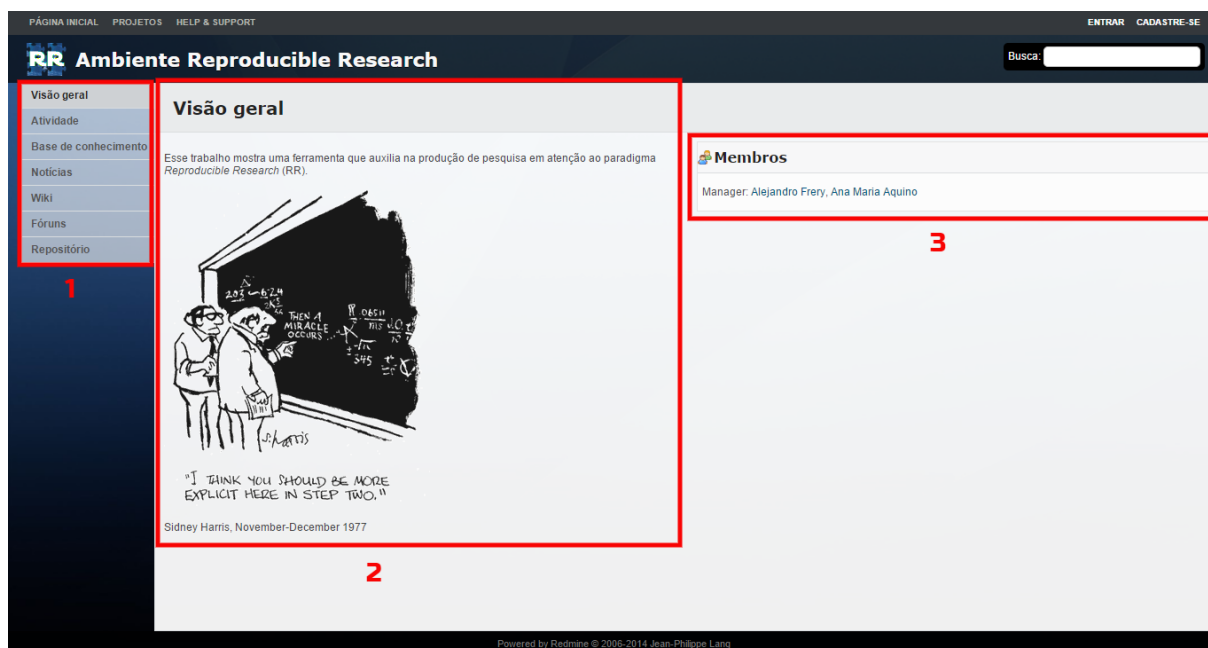


Figura 4.21: Página pública do projeto *Ambiente Reproducible Research* com menu (1), informações gerais (2) e descrição dos membros com *links* para suas respectivas páginas (3).

### 4.3 Ambiente virtual do pRProject

O ambiente virtual do pRProject está disponível através da *pRProject.box*<sup>5</sup>. Adicionamos e executamos essa *box* numa máquina com sistema operacional *Windows 10 64-bits* (Figura 4.22).

<sup>4</sup>Disponível em: <<https://rerisip.lccv.ufal.br/projects/ambiente-reproducibile-research>>. Acesso em: 10 ago. 2017.

<sup>5</sup>Disponível em: <<https://goo.gl/F5hbSz>>. Acesso em: 25 out. 2017.



The image shows two terminal windows. The left window is a Windows PowerShell terminal running Vagrant commands to start a VM named 'default'. The right window is a PuTTY terminal showing the output of a bundle install command, listing various gems and their versions, and ending with a message that the bundle is complete and 45 gems are installed.

```

S C:\> cd Teste
S C:\teste> vagrant up
Bringing machine 'default' up with 'virtualbox' provider...
=> default: Clearing any previously set forwarded ports...
=> default: Clearing any previously set network interfaces...
=> default: Preparing network interfaces based on configuration...
  default: Adapter 1: nat
=> default: Forwarding ports...
  default: 3000 (guest) => 3000 (host) (adapter 1)
  default: 22 (guest) => 2222 (host) (adapter 1)
=> default: Booting VM...
=> default: Waiting for machine to boot. This may take a few minutes...
  default: SSH address: 127.0.0.1:2222
  default: SSH username: vagrant
  default: SSH auth method: private key
  default: Warning: Connection reset. Retrying...
  default: Warning: Connection reset. Retrying...
  default: Warning: Connection reset. Retrying...
  default: Warning: Connection reset. Retrying...
  default: Warning: Connection reset. Retrying...
  default: Warning: Connection reset. Retrying...
=> default: Machine booted and ready!
[default] Guest additions %s.1.22 running --- OK.
=> default: Checking for guest additions in VM...
=> default: Mounting shared folders...
  default: /vagrant => C:/Teste
=> default: Machine already provisioned. Run 'vagrant provision' or use the '--provision'
=> default: flag to force provisioning. Provisioners marked to run always will still run.
S C:\teste> vagrant ssh
Please configure /home/vagrant/.bash/user.sh with your git info.
vagrant@127.0.0.1:~$ vagrant
[default] Using gitlab-grit 2.8.1
[default] Using sass-listen 4.0.0
[default] Using activerecord 3.2.22.2
[default] Using activerecord 3.2.22.2
[default] Using paperclip 3.5.4
[default] Using actionpack 3.2.22.2
[default] Using capybara 2.15.1
[default] Using jbox-gitolite 1.2.6
[default] Using shoulda 3.3.2
[default] Using sass 3.5.1
[default] Using railties 3.2.22.2
[default] Using actionmailer 3.2.22.2
[default] Using kaminari 0.17.0
[default] Using jquery-rails 3.1.4
[default] Using sass-rails 3.2.6
[default] Using rails 3.2.22.2
[default] Using acts-as-taggable-on 2.4.1
[default] Using rich 1.4.6
[default] Using redmine_acts_as_taggable_on 1.0.0
Bundle complete! 45 Gemfile dependencies, 93 gems now installed.
Use 'bundle info [gemname]' to see where a bundled gem is installed.

[vagrant@127.0.0.1:~] /vagrant/redmine2.6
$ rails server
/home/vagrant/.rbenv/versions/2.2.2/lib/ruby/gems/2.2.0/gems/railties-5.1.3/lib/
rails/app_loader.rb:40: warning: Insecure world writable dir /vagrant/redmine2.6
/. in PATH, mode 040777
/home/vagrant/.rbenv/versions/2.2.2/lib/ruby/gems/2.2.0/gems/htmlentities-4.3.1/
lib/htmlentities/mappings/expanded.rb:465: warning: duplicated key at line 466 i
gnored: "inodot"
=> Booting WEBrick
=> Rails 3.2.22.2 application starting in development on http://0.0.0.0:3000
=> Call with -d to detach
=> Ctrl-C to shutdown server
/vagrant/redmine2.6/lib/redmine/scm/adapters/git_adapter.rb:26: warning: already
initialized constant Redmine::Scm::Adapters::GitAdapter; GIT BIN
/vagrant/redmine2.6/lib/redmine/scm/adapters/git_adapter.rb:26: warning: previou
s definition of GIT BIN was here

```

Figura 4.22: Telas de execução da *box* pRRoject no *Windows*.

O resultado das telas de execução da *box* é a ferramenta pRRoject<sup>6</sup> replicada na máquina local (Figura 4.23).



Figura 4.23: Página do pRRoject acessada localmente em máquina com sistema operacional *Windows*.

Detalhamos os procedimentos para a execução dos comandos mostrados na Figura 4.22, na *wiki*<sup>7</sup> do projeto “Ambiente *Reproducible Research*” (Figura 4.24).

<sup>6</sup>Disponível em: <<https://rerisip.lccv.ufal.br>>. Acesso em: 11 out. 2017.

<sup>7</sup>Disponível em: <<https://rerisip.lccv.ufal.br/projects/ambiente-reproducible-research/wiki>>. Acesso em: 22 ago. 2017.

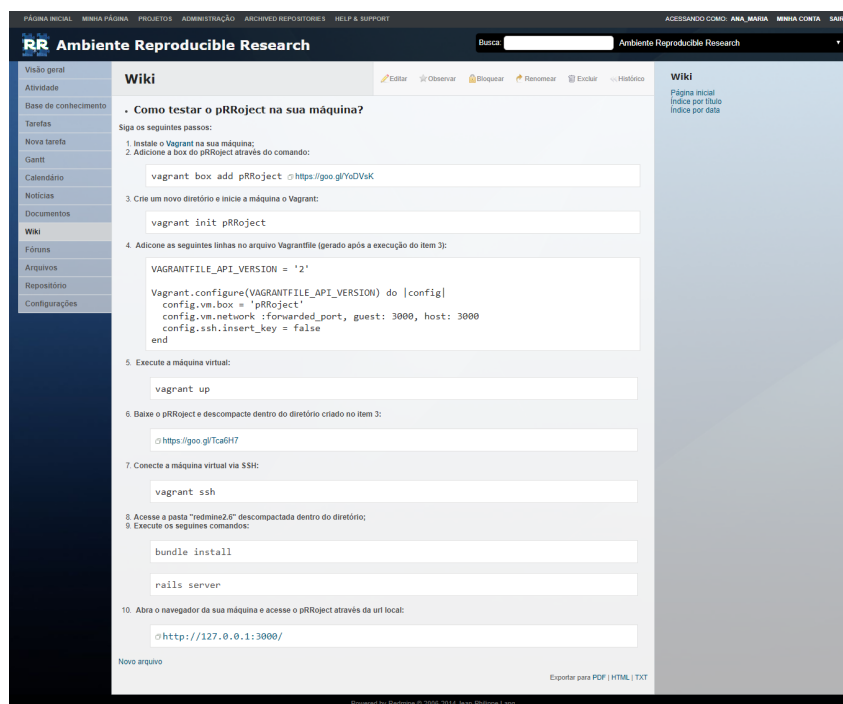


Figura 4.24: Página Wiki do projeto “Ambiente *Reproducible Research*”.

Os comandos descritos na *wiki*, referente ao *Vagrant*, são válidos para qualquer sistema operacional, o que pode apresentar alteração é o processo de instalação da ferramenta *Vagrant*.

## 4.4 Impacto do pRroject na pesquisa

Os membros convidados para responder o formulário<sup>8</sup> identificaram que o pRroject:

- Auxilia no processo de organização, arquivamento e documentação dos artefatos da pesquisa;
- Permite o desenvolvimento da pesquisa científica de forma transparente;
- Melhora a comunicação durante o processo de desenvolvimento do projeto de pesquisa;
- Auxilia no processo de replicação dos resultados durante o desenvolvimento do projeto de pesquisa;
- Permite que os pesquisadores divulguem os artefatos da pesquisa e os resultados;
- Auxilia o pesquisador a compreender e adotar boas práticas de RR.

<sup>8</sup>Disponível em: <<https://goo.gl/QvRMys>>. Acesso em: 11 out. 2017.

Quando questionados sobre quais dos oito requisitos de boas práticas de *RR* conseguiram obedecer com o uso da ferramenta, apenas dois requisitos foram atendidos por todos (Um repositório com acesso público pela web; O relatório da pesquisa, o(s) conjunto(s) de dados, scripts, comandos, instruções e qualquer artefato relevante) outros dois foram atendidos por 66,7% (Informações dos autores com links às respectivas páginas web; Uma página com o resumo do projeto de pesquisa), outros três requisitos foram obedecidos por 33,3% (Referências bibliográficas da pesquisa com seus respectivos *abstracts*; Informações da plataforma na qual o resultado da sua pesquisa foi testado) e um deles não foi obedecido (Artigo em PDF e informações sobre seu estado). Segundo usuários, o espaço disponibilizado para divulgação do artigo em PDF não é perceptível.

## 4.5 Análises e Discussão

Averiguamos o pRRoject funciona como um compêndio, mas não só isso, mostra detalhes de todo o ciclo de vida do projeto de pesquisa. O pRRoject não é a ferramenta pioneira em considerar o ciclo de vida da pesquisa científica e auxiliar na organização de múltiplos projetos científicos, mas permite um gerenciamento mais completo do projeto de forma flexível e, além disso, é um software livre.

Ao utilizar o pRRoject o pesquisador consegue usar controle de versão, publicar códigos, dados, documentação, pré-impressões de artigos e disponibilizar um produto final transparente, com detalhes necessários a replicação dos resultados. Entretanto, essa ferramenta ainda não oferece funcionalidade para replicar os resultados de cada projeto.

O pRRoject suporta diversas funcionalidades essenciais para a construção e gerência de projetos: gestão de versões, documentos, arquivos, repositórios, planejamento, acompanhamento e *workflow* de execução das tarefas.

Essas e outras funcionalidades agregadas ao pRRoject, através dos *plugins: knowledgebase, git hosting, people e questions and answers*, são suficientes para auxiliar pesquisadores e colaboradores de projetos de pesquisa computacional quantitativa a executarem boas práticas *RR*.

Todavia, no processo de validação, os usuários mostram que não conseguem perceber 100% de suficiência em todas funcionalidades, devido a ausência de informações sobre o uso da ferramenta. Dessa forma, averiguamos há a necessidade de criação de um assistente de utilização de *software* sobre como obedecer boas práticas de *RR* utilizando o pRRoject.

O pRRoject precisa de uma boa infraestrutura técnica de hospedagem para evitar imprevistos no decorrer do projeto de pesquisa. Também há a necessidade do pRRoject ser gerenciado por um indivíduo que conheça todas as funcionalidades do sistema e preste suporte aos pesquisadores para resolver os problemas técnicos que podem ocorrer.

Percebemos que a originalidade deste trabalho está no fato de utilizarmos um gerenciador

de projetos *open source* para auxiliar pesquisadores a adotarem boas práticas de *RR* durante o ciclo de vida da pesquisa científica.

Nós evidenciamos essa associação no processo de desenvolvimento deste trabalho ao mostrar que é possível adotar boas práticas de *RR* a partir do uso de um sistema de gerenciamento de projetos.

Não há estudos na literatura científica que mostrem essa associação e a proponham como um suporte ao pesquisador para que ele possa adotar boas práticas de *RR*.

Diante disso, nós disponibilizamos um ambiente computacional com funcionalidades que atendem aos requisitos de boas práticas de *Reproducible Research*, baseado em um sistema de gerenciamento de projetos. Todavia, necessitamos de uma avaliação com referência as vantagens e desvantagens do uso desse ambiente.

Atualmente o pRRoject está hospedado em um dos servidores do Laboratório de Computação Científica e Visualização (LCCV), da Universidade Federal de Alagoas (UFAL), e acessível através do endereço *web*: <https://rerisip.lccv.ufal.br><sup>9</sup>.

Ademais, é possível acessar uma versão teste do pRRoject, em sua máquina local, por meio da ferramenta de virtualização *Vagrant*. Entretanto, há a necessidade de validar o processo de instalação e configuração do pRRoject.

Esse capítulo apresentou as vantagens e desvantagens do ambiente *Reproducible Research* durante todo o processo de desenvolvimento e acompanhamento deste trabalho. Finalizamos com o próximo capítulo, no qual serão relatados os principais impactos e contribuições deste trabalho e os passos futuros.

<sup>9</sup>Disponível em: <<https://rerisip.lccv.ufal.br>>. Acesso em: 10 set. 2017.

# 5

## CONCLUSÕES

**E**STE capítulo abordará o avanço que o presente trabalho traz para o cenário científico e porquê isso é interessante. Além disso, apresentará sugestões para trabalhos futuros.

### 5.1 Considerações Finais

Disponibilizamos um modelo de ambiente computacional com suporte a *Reproducible Research*, baseado em um sistema de gerenciamento de projetos, denominado pRRoject.

O *pRRoject* oferece condições a pesquisadores e laboratórios de pesquisa para gerenciarem seus projetos, adotando boas práticas de *RR* desde o início das atividades inerentes ao ciclo de vida da pesquisa. Isso garante um projeto de pesquisa que atende às diretrizes de *RR* propostas na literatura.

Essa ferramenta, *open source*, está acessível na versão web e disponível para testes, através de uma máquina virtual personalizada. Tal acessibilidade visa facilitar o processo de avaliação qualitativa do pRRoject, não realizado no decorrer deste trabalho.

Recomendamos um novo olhar para a pesquisa reprodutível ao associarmos gerência de projetos com as diretrizes de boas práticas de *RR*. O gerenciamento de projetos de pesquisa e desenvolvimento interage com as diretrizes de boas práticas para a pesquisa reprodutível, com atenção ao ciclo de vida do projeto de pesquisa.

### 5.2 Trabalhos futuros

Pretendemos expandir o uso da ferramenta proposta neste trabalho e agregar novas funcionalidades. Para tanto, pretendemos alcançar os seguintes objetivos:

- Divulgar a proposta deste trabalho e a ferramenta pRRoject, em eventos internacionais.

- 
- Instalar a ferramenta *pRProject* em laboratórios de pesquisa computacional quantitativa, através de máquinas virtuais ou plataformas de contêiner de *softwares*.
  - Obter uma avaliação qualitativa dos pesquisadores da área computacional quantitativa, referente ao uso da ferramenta.
  - Desenvolver plugins que agreguem funcionalidades ao ambiente *RR* como, por exemplo: integração com ferramentas para gerência de referências bibliográfica, integração com editores de texto  $\text{\LaTeX}$ , integração com ferramentas *Google* e execução de algoritmos escritos em diferentes linguagens.
  - Averiguar a possibilidade de utilizar o *pRProject* em projetos categorizados como propriedade intelectual.

## REFERÊNCIAS BIBLIOGRÁFICAS

- Atmanspacher, H. & Maasen, S. (2016), *Reproducibility: Principles, Problems, Practices, and Prospects*, Wiley.
- Bevilacqua, A. (2014), *Redmine Plugin Extension and Development*, Community experience distilled, Packt Publishing.
- Buckheit, J. B. & Donoho, D. L. (1995), *WaveLab and Reproducible Research*, Springer New York, pp. 55–81.
- Chirigati, F. S., Shasha, D. & Freire, J. (2013), Rezip: Using provenance to support computational reproducibility, *in* ‘Presented as part of the 5th USENIX Workshop on the Theory and Practice of Provenance’, USENIX.
- Colom, M., Kerautret, B., Limare, N., Monasse, P. & Morel, J.-M. (2015), Ipol: a new journal for fully reproducible research; analysis of four years development, *in* ‘2015 7th International Conference on New Technologies, Mobility and Security (NTMS)’, pp. 1–5.
- Davison, A. P. (2012), ‘Automated capture of experiment context for easier reproducibility in computational research’, *Computing in Science and Engineering* **14**, 48–56.
- Davison, A. P., Mattioni, M., Samarkanov, D. & B. Teleńczuk (2014), Sumatra: A toolkit for reproducible research, *in* V. Stodden, F. Leisch & R. Peng, eds, ‘Implementing Reproducible Research’, Chapman & Hall/CRC The R Series, CRC Press, pp. 57–79.
- Donoho, D. L. (2010), ‘An invitation to reproducible computational research’, *Biostatistics* pp. 385–388.
- Donoho, D., Maleki, A., Rahman, I., Shahram, M. & Stodden, V. (2009), ‘Reproducible research in computational harmonic analysis’, *Computing in Science & Engineering* **11**(1), 8–18.
- Fara, P. (2015), ‘Newton shows the light: a commentary on newton (1672) ‘a letter . . . containing his new theory about light and colours. . .’’, *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*.
- Fomel, S. & Claerbout, J. E. (2009), ‘Reproducible research introduction’, *Computing in Science & Engineering* **11**(1), 5–7.
- Foster, MSLS, E. D. . D. M. A. (2017), ‘Open science framework (osf)’, *Journal of the Medical Library Association*,.

- Freire, J., Bonnet, P. & Shasha, D. (2012), Computational reproducibility: State-of-the-art, challenges, and database research opportunities, in 'Proceedings of the 2012 ACM SIGMOD International Conference on Management of Data', ACM, pp. 593–596.
- Freire, J., Koop, D., Chirigati, F. & Silva, C. (2014), *Implementing Reproducible Research*, Taylor & Francis, chapter Reproducibility Using VisTrails, pp. 33–56.
- Frery, A. C. (2015), 'How to successfully make a scientific contribution through ieee geoscience and remote sensing letters', *IEEE Geoscience and Remote Sensing Letters* **12**(6), 1167–1169.
- Gandrud, C. (2016), *Reproducible Research with R and R Studio, Second Edition*, Chapman & Hall/CRC The R Series, CRC Press.
- Gentleman, R. & Lang, D. T. (2007), 'Statistical analyses and reproducible research', *Journal of Computational and Graphical Statistics* **16**(1), 1–23.
- Gentleman, R. & Temple Lang, D. (2004), 'Statistical analyses and reproducible research', *Bioconductor Project Working Papers*.
- Goswami, J. & Chan, A. (2011), *Fundamentals of Wavelets: Theory, Algorithms, and Applications*, Wiley Series in Microwave and Optical Engineering, Wiley.
- Hodges, A. & Hofstadter, D. (2014), *Alan Turing: The Enigma*, Princeton University Press.
- Kuhn, T. S. (1998), *A Estrutura das revoluções científicas*, Perspectiva.
- Leek, J. T. & Peng, R. D. (2015), 'Opinion: Reproducible research can still be wrong: Adopting a prevention approach', *Proceedings of the National Academy of Sciences* **112**(6), 1645–1646.
- Leisch, F. (2002), Sweave: Dynamic generation of statistical reports using literate data analysis, in 'Compstat: Proceedings in Computational Statistics', Physica-Verlag HD, pp. 575–580.
- Lesyuk, A. (2013), *Mastering Redmine*, Packt Publishing.
- Limare, N. & Morel, J.-M. (2011), 'The ipol initiative: Publishing and testing algorithms on line for reproducible research in image processing', *Procedia Computer Science* **4**, 716 – 725.
- Lippe, S. & vom Brocke, J. (2016), 'Situational project management for collaborative research projects', *Project Management Journal* **47**(1), 76–96.
- McNutt, M. (2014), 'Reproducibility', *Science* **343**, 229–229.
- Mesirov, J. P. (2010), 'Computer science. accessible reproducible research.', *Science (New York, N.Y.)*.



- Morin, A., Urban, J., Adams, P. D., Foster, I., Baker, D. & Sliz, P. (2012), 'Shining light in black boxes', *Science* **336**, 159–160.
- Munafò, M. R., Nosek, B. A., Bishop, D. V. M., Button, K. S., Chambers, C. D., Percie du Sert, N., Simonsohn, U., Wagenmakers, E.-J., Ware, J. J. & Ioannidis, J. P. A. (2017), 'A manifesto for reproducible science', *Nature Human Behaviour*.
- Peng, R. (2015), 'The reproducibility crisis in science: A statistical counterattack', *Significance* **12**(3), 30–32.
- Peng, R. D. (2011), 'Reproducible research in computational science', *Science* **334**(6060), 1226–1227.
- Popper, K. (2004), *A lógica da pesquisa científica*, Cultrix.
- Project Management Institute, I. (2013), *A Guide to the Project Management Body of Knowledge*, Foundational Standards.
- Sandve, G. K., Nekrutenko, A., Taylor, J. & Hovig, E. (2013), 'Ten simple rules for reproducible computational research', *PLoS Computational Biology*.
- Schwab, M., Karrenbach, M. & Claerbout, J. (2000), 'Making scientific computations reproducible', *Computing in Science Engineering* **2**(6), 61–67.
- Silver, A. (2017), 'Software simplified', *Nature News* (546), 173–174.
- Skaggs, T. H., Young, M. H. & Vrugt, J. A. (2015), 'Reproducible research in vadose zone sciences', *Vadose Zone Journal*.
- Steinle, F. (2016), *Stability and Replication of Experimental Results: A Historical Perspective*, John Wiley & Sons, Inc., pp. 39–63.
- Stodden, V. (2010), 'The scientific method in practice: Reproducibility in the computational sciences', *MIT Sloan Research Paper*.
- Stodden, V. & Miguez, S. (2014), 'Best practices for computational science: Software infrastructure and environments for reproducible and extensible research', *Journal of Open Research Software*.
- Stodden, V., Bailey, D. H., Borwein, J., LeVeque, R. J., Rider, W. & Stein, W. (2013a), Setting the default to reproducible, in 'ICERM Workshop on Reproducibility in Computational and Experimental Mathematics', pp. 1–19. URL [https://icerm.brown.edu/tw12-5-rcem/icerm\\_report.pdf](https://icerm.brown.edu/tw12-5-rcem/icerm_report.pdf).

- Stodden, V., Borwein, J., Bailey, D. H., LeVeque, R. J., Rider, W. & Stein, W. (2013b), 'Setting the default to reproducible in computational science research', *Society for Industrial and Applied Mathematics (SIAM)*.
- Stodden, V., Donoho, D., Fomel, S., Friedlander, M. P., Gerstein, M., LeVeque, R., Mitchell, I., Ouellette, L. L. & Wiggins, C. (2010), 'Reproducible research', *Computing in Science Engineering* **12**(5), 8–13.
- Stodden, V., Miguez, S. & Seiler, J. (2015), 'Researchcompendia.org: Cyberinfrastructure for reproducibility and collaboration in computational science', *Computing in Science Engineering* **17**(1), 12–19.
- Volpato, G. (2013), *Ciência: da filosofia à publicação*, Cultura Acadêmica.
- Wintage, L. M. (2015), *Project Management for Research and Development: Guiding Innovation for Positive R&D Outcomes*, Best Practices and Advances in Program Management (Book 10), Auerbach Publications.

Este trabalho foi redigido em  $\text{\LaTeX}$  utilizando uma modificação do estilo IC-UFAL. As referências bibliográficas foram preparadas no JabRef e administradas pelo  $\text{\BIBTeX}$  com o estilo LaCCAN. O texto utiliza fonte Fourier-GUTenberg e os elementos matemáticos a família tipográfica Euler Virtual Math, ambas em corpo de 12 pontos.