

Ambiente Computacional em Nuvem Utilizando Docker para Suporte a *Reproducible Research* em Computação

**Sandro M. de Carvalho^{1,2}, Fernando da F. de Souza¹, Ricardo J. R. Amorim^{3,4},
Dinani G. Amorim^{3,4}**

¹Centro de Informática – Universidade Federal de Pernambuco (UFPE)
Caixa Postal 7851 – 50.732-970 – Recife – PE – Brasil

²Instituto Federal do Sertão Pernambucano (IFSertão-PE) – Reitoria
Rua Coronel Amorim, 76 – Centro – 56.302-320 – Petrolina – PE – Brasil

³Universidade do Estado da Bahia (UNEB)
Av. Edgard Chastinet, S/N – São Geraldo – 48.905-680 – Juazeiro – BA – Brasil

⁴Faculdade de Ciências Aplicadas e Sociais de Petrolina (FACAPE)
Campus Universitário, S/N – Vila Eduardo – 56.328-903 – Petrolina – PE – Brasil
{smc2, fdfd}@cin.ufpe.br, {amorim.ricardo, dinaniamorim}@gmail.com

***Abstract.** In general, replicate research results is an arduous task, which is especially perceived by post graduate students. To mitigate this, Reproducible Research is a technique that systematize best practices in scientific research, focused on quantitative methods in computer science. This article describes the scientific methodology, analyzes the main computer found solutions and proposes a computing environment using cloud Docker to support this technique. The aim is to make it available for public access.*

***Resumo.** No geral, reproduzir resultados de pesquisas é uma tarefa árdua, especialmente percebida por estudantes de pós-graduação. Na mitigação disso, a Reproducible Research é uma técnica que sistematiza boas práticas na pesquisa científica, voltada à métodos quantitativos em computação. Este artigo descreve a metodologia científica utilizada, analisa as principais soluções computacionais encontradas e propõe um ambiente computacional em nuvem utilizando Docker para suporte a esta técnica. O objetivo é disponibilizá-lo para acesso público.*

1. Introdução

A reprodutibilidade dos resultados é considerada um fator primordial na ciência. O sucesso e a credibilidade de uma pesquisa estão ancorados na vontade dos cientistas de expor suas ideias e resultados a testes independentes e replicação por outros cientistas. Isso requer uma completa abertura dos dados, procedimentos e materiais [Fomel and Claerbout 2009].

Em um contexto educacional, de formação para a pesquisa, as dificuldades e os problemas enfrentados pelos estudantes e pesquisadores para reproduzir os resultados gerados por pesquisas científicas estão relacionados, principalmente, à falta de

publicação do trabalho científico junto aos seus artefatos. Segundo Donoho (2010), tais resultados são apresentados de forma pouco convincente em periódicos, conferências e livros, devido principalmente as limitações impostas pelo formato dessas mídias.

Como resultado de sua pesquisa, Pinheiro (2013) desenvolveu um ambiente computacional para suporte a práticas de *Reproducible Research* (RR) e propôs alguns trabalhos futuros relacionados ao ambiente desenvolvido, dentre eles, disponibilizar o ambiente computacional por meio de uma máquina virtual (VM). Referenciou que segundo Howe (2012), uma VM pode conter todo o ambiente de trabalho de um pesquisador, incluindo dados, software, dependências, anotações, *logs*, *scripts* e muito mais. A grande vantagem da utilização da máquina virtual reside no fato de que ela pode ser instalada de maneira simples, por meio de um software de virtualização, no computador pessoal do pesquisador.

Neste contexto, este trabalho propõe o desenvolvimento e implantação de um ambiente computacional em nuvem utilizando *containers* Docker para suporte às práticas recomendadas por *Reproducible Research* em Computação. Tal ambiente deverá garantir que as premissas de RR sejam aplicadas de forma transparente, e que a sua adoção impacte minimamente no modo de trabalho do pesquisador. Partindo do princípio de que a ciência é universal, ou seja, deve ser acessível de forma livre a todos, esse mesmo princípio será aplicado na concepção do ambiente proposto, incentivando a prática da pesquisa computacional e contribuindo com a educação no contexto da formação de pesquisadores.

2. Metodologia

A pesquisa, no contexto científico, também pode ser classificada de acordo com diferentes critérios. Entre eles, é possível diferenciar tipos de pesquisa de acordo com sua natureza, objetivos ou procedimentos técnicos. Nem sempre um trabalho de pesquisa limita-se a um único tipo. Além disso, alguns tipos de pesquisa podem ser a base para outros [Wazlawick 2014].

Neste sentido, esta é uma pesquisa aplicada, com abordagem qualitativa, que em relação aos procedimentos foi feita uma revisão bibliográfica inicial, com base em livros conceituados relacionados com o tema em questão e trabalhos publicados em edições da Revista Brasileira de Informática na Educação (RBIE)¹, anais do Simpósio Brasileiro de Informática na Educação (SBIE)² e artigos disponíveis em bibliotecas científicas, tais como: *Association for Computing Machinery* (ACM)³, *Scientific Electronic Library Online* (SciELO)⁴, *Institute of Electrical and Electronic Engineers* (IEEE Xplore)⁵, dentre outras, explorando publicações dos últimos 5 (cinco) anos.

Juntamente à revisão de literatura, foi realizado um estudo a fim de analisar trabalhos relacionados à pesquisa. A revisão de literatura somada a esse estudo permitiu conhecer mais profundamente os temas em questão, possibilitando a compreensão de

1 www.br-ie.org/pub/index.php/rbie

2 www.br-ie.org/pub/index.php/sbie

3 dl.acm.org

4 www.scielo.org

5 ieeexplore.ieee.org

seus conceitos e da sua atual situação. De posse desses conhecimentos, foram levantadas as funcionalidades do ambiente computacional em nuvem, que deverão nortear o objetivo principal do trabalho.

Por fim, o ambiente proposto está sendo desenvolvido e implantado a partir da combinação do resultado das seguintes etapas:

- Levantamento dos conceitos de RR, por meio de investigação da literatura;
- Identificação e classificação das principais práticas recomendadas por RR;
- Análise das principais soluções computacionais que dão suporte a RR; e
- Análise do Docker quanto ao suporte a RR visando construir o ambiente computacional em nuvem.

3. Ambientes Computacionais

O estudo de ferramentas e soluções tem importância para a proposta do ambiente computacional em nuvem, já que proporciona uma análise de estrutura, interface e funcionalidades que podem ser aproveitadas no novo ambiente.

3.1. Apresentação e Análise de Ambientes Computacionais

No que se refere à ambiente computacional para suporte a *Reproducible Research*, o tema tem sido abordado discretamente na literatura científica. Apesar disso, é possível elencar diferentes trabalhos que tratam do referido tema, como mostrado a seguir.

Um desses trabalhos é o de Pinheiro (2013), o qual desenvolveu um ambiente computacional para suporte a práticas de *Reproducible Research* a partir da integração de ferramentas computacionais *Free/Libre and Open-Source Software* (FLOSS). Segundo o autor, assim como qualquer produto de software, o ambiente desenvolvido deverá sofrer modificações (correções e novas funcionalidades) ao longo do tempo.

Um trabalho bem citado sobre o tema é o *An introduction to Docker for reproducible research*, de Boettiger (2015). Segundo o autor, Docker⁶ tem o potencial para resolver as deficiências de certas abordagens existentes para os desafios da pesquisa reprodutível que decorrem de recriar ambientes computacionais complexos.

Outro trabalho relevante nesse contexto é o de Junqueira *and* Lóscio (2014), o qual realiza uma análise comparativa entre diversos repositórios de Objetos de Aprendizagem (OA). A análise buscou identificar os requisitos necessários para novas propostas de organização e compartilhamento de objetos de aprendizagem na Web que promovam e facilitem o reuso de conteúdos educacionais.

Outro trabalho nesse contexto é o RECOBA, apresentado em outro artigo de Junqueira *and* Lóscio (2015). O RECOBA tem por objetivo estimular a reutilização de Objetos de Aprendizagem, contribuindo para o fortalecimento da aprendizagem, disseminação e aquisição de novos conhecimentos.

Um outro trabalho relacionado a este é o de Otsuka et. al (2015), que apresenta o Livre Saber (LiSa), um repositório digital de recursos educacionais abertos criado com

6 www.docker.com

o intuito de organizar, preservar e compartilhar um acervo de materiais didáticos em diferentes formatos.

Os três últimos trabalhos destacam soluções que indiretamente estão relacionadas ao termo *Reproducible Research*, pois abordam uma área que também está em evidência: Recursos Educacionais Abertos (REA), que será a base para os materiais produzidos no ambiente proposto, descrito a seguir.

3.2. Ambiente Computacional em Nuvem Utilizando Docker

O Docker está sendo analisado a partir do mapeamento das práticas de RR com funcionalidades oferecidas pelo software, o qual visa diminuir os custos de implementação. Encontra-se em estágio pleno de utilização, emprega formatos de arquivos não proprietários e possui grande comunidade de usuários e desenvolvedores que interagem garantindo atualizações, testes e melhorias em suas funcionalidades. Além de contar com investimentos e o apoio de grandes corporações da área de Tecnologia da Informação.

A Figura 1 apresenta o Espectro da Reprodutibilidade, foi adaptada para representar o ambiente computacional proposto.

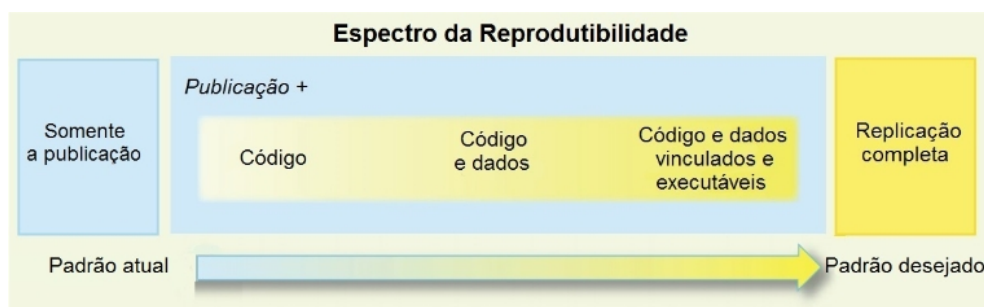


Figura 1. Espectro da Reprodutibilidade. Fonte: Adaptada de [Pinheiro 2013]

Um fator relevante na proposta de RR diz respeito à sua organização e estrutura. Para tanto, é necessário que o ambiente possua as seguintes funcionalidades:

- Armazenamento de projetos em servidores com disponibilidade em nuvem;
- Alocação de cada projeto em um *container* Docker;
- Acesso ao projeto pelo aluno e pesquisador por meio de um navegador web;
- Versionamento da pesquisa. De maneira geral, à medida que um projeto de pesquisa evolui, os seus resultados vão se modificando; e
- Armazenamento dos artefatos da pesquisa (conjunto de dados, *scripts*, relatórios, imagens, instruções, dentre outros) de acordo com o que preconiza RR e seguindo um formato de REA.

4. Resultados Iniciais

O Quadro 1 apresenta os resultados iniciais obtidos a partir da análise comparativa dos trabalhos apresentados no item 3.1.

Quadro 1. Comparativo entre os trabalhos relacionados. Fonte: [Os Autores]

Autor(es)	Ano	Resultado
Pinheiro	2013	Necessita aprimorar a gestão dos artefatos, finalizar a geração de página web e também a possibilidade de disponibilização do ambiente por meio de uma VM.
Boettiger	2015	Utiliza o Docker e ilustra apenas com exemplos para o ambiente estatístico R.
Junqueira and Lóscio	2014	Realiza uma análise e constata que algumas características dos OAs, dos Dados Abertos e dos REAs, ainda não são atendidas pelos repositórios existentes.
	2015	Evolui o trabalho anterior e propõe a criação de um ambiente para publicação e compartilhamento de OAs fragmentados, utilizando formatos de dados abertos.
Otsuka et. al	2015	Proposta interessante, porém não faz relação entre uma pesquisa e seus artefatos.

5. Considerações Finais

A pesquisa desenvolvida neste trabalho visa incentivar as práticas de RR, pois propõe o desenvolvimento e implantação de um ambiente computacional em nuvem utilizando Docker para suporte a *Reproducible Research* em Computação. Dessa forma, com a disponibilização do ambiente para acesso público, espera-se contribuir para auxiliar laboratórios, grupos, estudantes e pesquisadores interessados em realizar pesquisa computacional comprometida com as práticas de *Reproducible Research*.

Referências

- Boettiger, C. (2015) An introduction to Docker for reproducible research. In *ACM SIGOPS Operating Systems Review*, v. 49, n. 1, p. 71-79.
- Donoho, D. L. (2010) An invitation to reproducible computational research. *Biostatistics*, v. 11, n. 3, p. 385-388.
- Fomel, S. and Claerbout, J. F. (2009) Guest Editors' Introduction: Reproducible Research. In *Computing in Science & Engineering*, v. 11, n. 1, p. 5-7.
- Howe, B. (2012) Virtual Appliances, Cloud Computing, and Reproducible Research. In *Computing in Science & Engineering*, v. 14, n. 4, p. 36-41.
- Junqueira, R. de P. and Lóscio, B. F. (2014) Repositórios de Objetos de Aprendizagem: uma análise comparativa com ênfase no reúso de conteúdos. In *Simpósio Brasileiro de Informática na Educação, SBIE*.
- Junqueira, R. de P. and Lóscio, B. F. (2015) RECOBA: Um ambiente para publicação e compartilhamento de Objetos de Aprendizagem no formato aberto e fragmentado. In *Anais dos Workshops do Congresso Brasileiro de Informática na Educação, CBIE*.
- Otsuka, J. L. et al. (2015) Livre Saber (LiSa): Um Repositório de Recursos Educacionais Abertos de Cursos a Distância. *Revista Brasileira de Informática na Educação*, [S.l.], v. 23, n. 1.
- Pinheiro, R. (2013) Ambiente computacional para suporte a “*Reproducible Research*”. Dissertação de Mestrado em Modelagem Computacional de Conhecimento. Universidade Federal de Alagoas, Maceió.
- Wazlawick, R. (2014) *Metodologia de Pesquisa para Ciência da Computação*. 2ª edição. Elsevier Brasil.