

UMA ANÁLISE COMPARATIVA DE REPOSITÓRIOS PARA COMPARTILHAMENTO DE OPEN DATA DE EXPERIMENTOS DE ENGENHARIA DE SOFTWARE

Pedro Henrique de Melo Costa (PIBIC/CNPq/FA/Uem), Edson A. Oliveira Junior (Orientador), e-mail: ra112653@uem.br, edson@din.uem.br

Universidade Estadual de Maringá / Departamento de Informática,
Maringá/PR.

Ciência da computação - Engenharia de software

Palavras-chave: Repositórios, digitais, confiáveis.

Resumo:

Seguindo os princípios do Open Science, os repositórios digitais confiáveis seguem com afinco a disponibilidade de acesso, onde os dados devem estar apresentados integralmente e com custo nulo ou mínimo para sua reprodução, de preferência com a possibilidade do download dos dados via internet. Os dados fornecidos devem apresentar a possibilidade de serem reutilizados e redistribuídos. Os dados oferecidos não devem discriminar quanto à sua disponibilidade de uso, seja para uma instituição acadêmica, científica, comercial ou não comercial, todos devem possuir o direito de uso integral dos dados para quaisquer que sejam seus fins.

Introdução

Com o avanço tecnológico, a informação está se tornando cada vez mais digital, a facilidade de acesso à informação no clique de um botão tem sido cada vez mais otimizada. Contudo, essa digitalização da informação ainda apresenta inúmeras dificuldades, tais como URL's inativas. Tendo em vista esse déficit na informação digital, os repositórios digitais confiáveis surgem como uma solução para sanar estes problemas.

Ao ser feita uma análise de uma gama de experimentos sendo desenvolvidos em engenharia de software (FURTADO, 2018), foi constatado que, num total de 174 experimentos analisados, apenas 64

desses informavam explicitamente a disponibilidade de seus dados. E ainda, desses 64 experimentos, 25% das fontes não possuíam mais os seus dados disponíveis (URL's inativas). A adoção de um repositório digital confiável poderia sanar esse problema.

O conceito de Openness, quando aplicado à ciência, é, sem dúvidas, um dos pilares para a construção do conhecimento. Afinal, Openness nada mais é que a acessibilidade ao conhecimento, tecnologias e outros recursos. A chave para a construção do conhecimento envolve a transparência do mesmo, permitindo assim sua reproducibilidade e replicabilidade. E é com base nisso que os repositórios digitais confiáveis moldam seus alicerces.

A importância desta pesquisa está em mostrar as semelhanças e diferenças entre os repositórios selecionados, para que a construção de novos repositórios de dados abertos seja cada vez mais otimizada e tenha um maior impacto educacional na sociedade.

A Figura 1 apresenta uma taxonomia de Open Science, na qual podemos observar a importância do conceito de *Open Repositories* diretamente relacionada ao conceito mais abrangente de *Open Science Tools*.

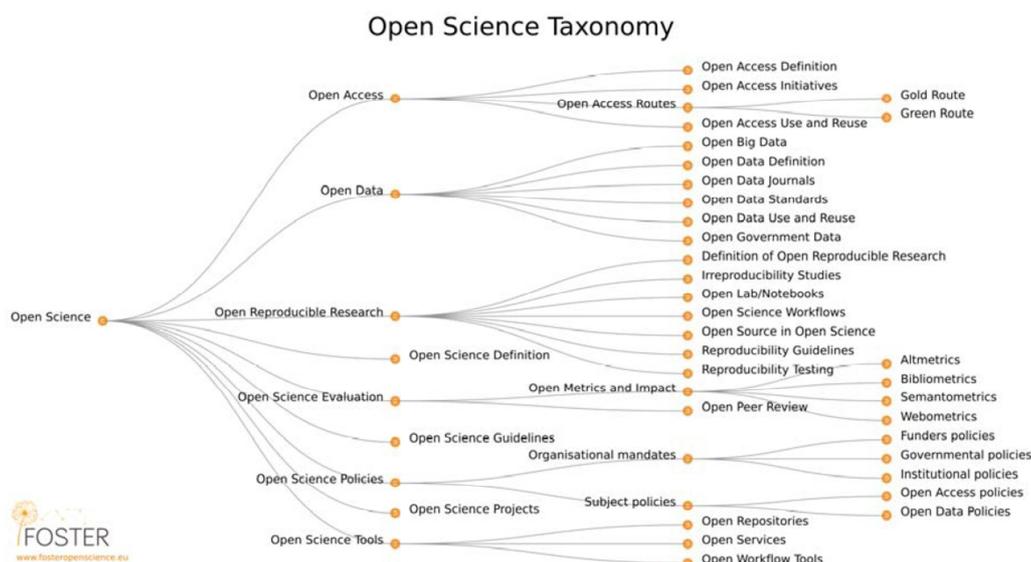


Figura 1 - Taxonomia do Open Science de acordo com <https://www.fosteropenscience.eu/>

Materiais e métodos

Este trabalho foi realizado utilizando-se de busca e análise nas seguintes fontes:

- <https://zenodo.org/>

- <https://arxiv.org/>
- <https://citeseerx.ist.psu.edu/index>
- <https://www.data-archive.ac.uk/>
- <https://figshare.com/>

Foi utilizado o método comparativo neste trabalho. Em cada site foram analisadas suas diversas estruturas, tais como API's, métodos de compartilhamento de dados, custo, segurança armazenamento dos dados, tipagem dos dados aceita pelos repositórios e tempo de vida dos repositórios.

Resultados e Discussão

A seguir apresentamos nossos resultados em forma de tabela.

Tabela 1: Tabela representativa dos resultados obtidos na comparação dos repositórios digitais confiáveis.

Repositórios	Zenodo	arXiv	CiteSeerX	UK Data Archive	Figshare
Tempo de vida	7 anos	29 anos	23 anos	18 anos	8 anos
Restrição de dados (tipos de arquivos)	nenhuma	LaTeX, AMSLaTeX, PDFLaTeX, PDF, HTML	PDF, ZIP, GZIP, UNIX, PostScript	TXT, DOC, XTHML, PDF, RTF, SAV, DTA, HTML, outros	nenhuma
Identificador digital	DOI ¹	DOI	nenhum	ORCID	DOI
API ²	uploads, busca e download	upload, acesso aos dados e metadados	acesso e extração de metadados, tabelas e figuras.	acesso aos dados	acesso aos dados
Segurança e armazenamento de dados	CERN's EOS service	Cornell University	Pennsylvania State University	UK Data Service	Amazon Web Services

¹ DOI: Digital Object Identifier

² API: Interface de programação de aplicações

Conclusões

Após análise dos repositórios em questão, foi possível concluir a eficácia dos repositórios digitais confiáveis, principalmente quanto a sua responsabilidade quanto a manutenção e zelo dos dados. Contudo, talvez por ser um serviço relativamente novo, ainda há uma heterogeneidade quanto aos tipos de dados aceitos e também quanto ao custo pelo uso. É evidente que é necessário criar um padrão para homogeneizar as restrições de dados dos repositórios. Devido a não restrição de dados e também à segurança do armazenamento dos dados, dos repositórios comparados, os melhores repositórios para uso hoje seriam o Zenodo e o Figshare.

Agradecimentos

Os autores agradecem à Fundação Araucária e ao CNPq pelo suporte financeiro.

Referências

- WOHLIN, C, et al. **Experimentation in Software Engineering**. Springer ed. 2012
- RICHARDSON, L; AMUNDSEN, M; RUBY, S. **RESTful Web APIs: Services for a changing World**. O'REILLY ed. 2013